Evaluating the (over)use of diagnostics in rheumatoid arthritis care

Evy Ulijn

Sint Maartenskliniek

Radboudumc university medical center

Radboud University

Radboudumc

Evaluating the (over)use of diagnostics in rheumatoid arthritis care

Evy Ulijn

Evaluating the (over)use of diagnostics in rheumatoid arthritis care

Evy Ulijn

Colofon

Design/lay-out and print
Promotie In Zicht | www.promotie-inzicht.nl

Publishing of this thesis was financially supported by the Sint Maartenskliniek, Nijmegen, The Netherlands.

Evaluating the (over)use of diagnostics in rheumatoid arthritis care

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 8 april 2025
om 14.30 uur precies

door

Evy Ulijn
geboren op 21 oktober 1996
te Oss

Promotoren
Dr. A.A. den Broeder
Prof. dr. R.B.M. Landewé (Amsterdam UMC)

Copromotoren
Dr. N. van Herwaarden
Dr. N. den Broeder

Manuscriptcommissie
Prof. dr. S. Middeldorp
Prof. dr. J.W.L. Cals (Maastricht University)
Prof. dr. R. B. Kool

Table of contents

Chapter 1

General introduction

When investigating the value of diagnostics, I encountered the phrase stating that 'laboratory medicine influences 70% of clinical decisions' and that '70% of the electronic patient record is composed of laboratory data' (1–3). These statements have been used since 1996 to promote the importance of laboratory testing (4). Without making claims about their validity or scientific foundation, these statements are intriguing as they underscore the prominent role of diagnostic tests in modern medicine. Diagnostic tests, such as laboratory testing, are embedded in modern medicine. Yet, the question arises: do we truly need all these tests, do we need them in in as many patients, and what is their actual impact on clinical decision-making? This PhD thesis aims to evaluate the clinical value of various diagnostic and prognostic tests, focusing specifically on the care of patients with Rheumatoid Arthritis (RA), and to assess their appropriateness using different approaches.

Diagnostic and prognostic tests cover a wide range of uses and appear in various forms. Examples are history taking and physical examinations, Patient Reported Outcome Measures (PROMs), blood tests, imaging modalities (radiographs or PET/CT scans), tissue biopsies, function tests, and so on. As the medical field evolved, numerous tests were developed and integrated into diagnostic and treatment guidelines and protocols for routine patient care. *Figure 1* shows the temporal trend in use of diagnostic tests in UK primary care (5). Notably, the total test use per person year has increased markedly over time. This study by O'Sullivan et al. illustrates that this trend was found for both sexes, across all age groups, and across various test types (laboratory, imaging, and miscellaneous).

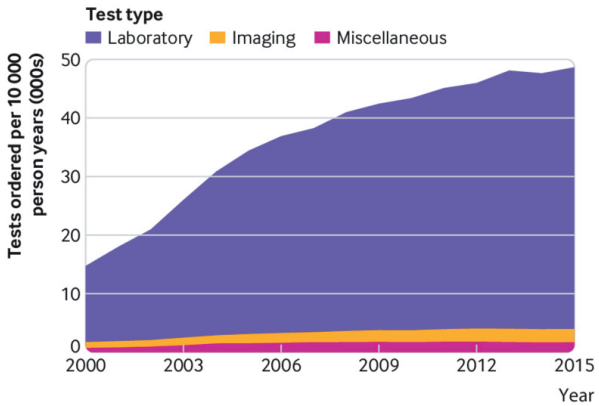

Figure 1. *Temporal trends in use of tests in UK primary care, 2000-15 (O'Sullivan JW, et al., 2018)* (5).

This growth in the development and usage of diagnostic and prognostic tests raises the question whether usage of all these tests improves patient outcomes.

Appropriateness of testing, general points to consider

Diagnostic tests play an important role in modern healthcare as they help identifying diseases, monitor patient conditions, and guide treatment decisions. Diagnostic testing can improve patient outcomes by reducing uncertainty about diagnoses and tailoring treatments to individual needs. Effective diagnostic testing can lead to timely and accurate diagnoses, enabling early intervention and more effective management of health conditions. For a test to be used appropriately and improve healthcare, it must meet several conditions: it must have sufficient sensitivity and specificity, it should relate to a significant difference (untreated or treated) in outcomes, it must provide added value beyond existing knowledge, its results should have clear consequences, its use (followed by appropriate treatment) should lead to better health outcomes, and its use must be cost-effective. By understanding the benefits and conditions for appropriate use of testing, we can better appreciate the importance of balancing its use to avoid the pitfalls of overuse of tests.

The inappropriate use of diagnostics, predictive testing and monitoring can lead to an array of pitfalls. Starting with the burden for the patient, consisting of possible procedural (physical and psychological) harms and procedural morbidity, (travel) time and -costs. Every test that is conducted could result in side-effects and complications which could enhance patient burden due to testing. Furthermore, ineffective allocation of limited resources should be considered: every test that is performed inappropriately produces unjustified incremental costs, deranges healthcare efficiency and widens the gap between demand and supply of healthcare workers (2,6). With each test that is performed, the likelihood for an irrelevant, false-positive or false-negative finding increases. A false-negative result could lead to a missed diagnosis and/or delay in effective treatment. A false-positive result or incidental finding could lead to follow-up testing and treatment, increasing all aforementioned patient- and healthcare burden, and generating indirect costs (2,6). False-positive results can cause considerable patient anxiety that may persist months after additional testing has shown negative findings that cancel out the initial test results (7). Additionally, despite the widespread yet unsubstantiated belief that diagnostic testing reassures patients, evidence shows that diagnostic tests for symptoms with a low risk of serious illness do little to reassure patients in the short or long term (8).

On the other hand, the use of too few diagnostics, predictive testing and monitoring could again result in a missed diagnosis and delaying treatment, misclassifying prognosis, or missing serious adverse events due to prescribed treatments that require starting, altering or discontinuation of currently used treatment(s) and/or treatment for an occurred adverse event itself. All of these potential problems could be prevented by finding the right balance between the effective use of our extensive range of diagnostics and limiting inappropriate use testing in scope and frequency to what was shown to be of added value to make sure patients receive the care they need, while minimizing the risks associated with excessive testing.

By continuously adding new tests the problem of inappropriate use of testing has emerged, with its accompanying risks. In this thesis we will evaluate several existing tests, specifically within the rheumatological care for patients with RA, in which diagnostic and prognostic testing have a role in diagnosing RA, assessment of disease prognosis and detection of treatment (side-)effects. Throughout this thesis I will evaluate these aspects of rheumatological care, as well as investigating topics that are currently not (yet) implemented in treatment guidelines but are investigated or used in research, i.e. predicting the effectiveness of new treatment options and other imaging modalities for visualizing arthritis.

(In)Appropriate testing in RA

RA is an inflammatory rheumatic disease characterized by symmetric arthritis of hands and feet. The disease onset is usually between 35 and 60 and it is more prevalent in females. The typical course of RA involves periods of remissions and exacerbations (9). High disease activity in RA is strongly associated with the progression of joint damage and a greater disease burden. Therefore, treatment goals focus on early inflammation reduction to alleviate symptoms and minimize joint damage. Timely initiation of a treatment plan asks for a fast diagnosis and prediction of the progression of the disease (prognosis). Thereafter, ongoing disease management is needed to guide treatments and minimize disease flares (10). Therefore, the appropriate use of diagnostic and prognostic testing at diagnosis and during the disease course is be of importance.

The burden of untreated or suboptimally treated RA manifests as active disease and eventual joint damage, carrying significant consequences for both patients and their loved ones. However, the prognosis of modern RA patients, who are diagnosed early and treated according to the latest standards, is less dominated by inflammation and joint destruction or extra-articular disease (e.g., affecting the lungs and heart). Instead, it is shaped more by elements common to chronic diseases, such as fatigue, widespread pain, sensitization, and consequent loss of participation in daily activities. This shift implies that for many patients, the need for frequent monitoring of late-

stage disease complications might be reduced. Nevertheless, the broader society is still affected by the ongoing need for disease management, work disability, reduced productivity, and early retirements (11).

Given this context, it is crucial to evaluate the appropriateness of various diagnostic and prognostic tests used throughout the patient journey of a typical RA patient. In this thesis, we focus on several tests used in different contexts, starting with routine radiographs of hands and feet.

Routine radiographs of hands and feet

For clinically diagnosing RA, a range of procedures can be used. Firstly, history taking and physical examination are of importance to establish the history and presence of arthritis of the joints of hands and feet in a typical symmetric pattern, the hallmark for an RA diagnosis. These factors are also pivotal in the American College of Rheumatism (ACR)/ European Alliance of Associations for Rheumatology (EULAR) 2010 criteria, along with presence of elevated acute phase reactants (C-reactive protein and erythrocyte sedimentation rate), Rheumatoid Factor (RF), Anti-Citrullinated Protein Antibodies (ACPA) and the duration of symptoms (<6 or ≥6 weeks) (12). These criteria have a sensitivity of 97% and a specificity of 55% for classification of RA (13).
Routine radiographic imaging of both hands and feet is also recommended in the workup for patients with arthritis that is clinically suspected of RA. This recommendation was added to clinical guidelines as presence of RA-associated erosions may be of diagnostic value, or of prognostic value for a more severe disease course requiring more intensive treatment (14). However, the additional value of routine radiographic imaging of both hands and feet for RA diagnosis and prognosis has not yet been established. Results of existing studies suggest that additional value of routine radiographs of hands and feet is limited (15–17). Therefore we have set up a large cohort study with patients suspected for RA to assess the value of these radiographs. This cohort study will be addressed within the **second chapter** of this thesis.

(Bio)Markers in guiding treatments and predicting treatment effects.

Biomarkers are measurable and specific biological indicators that are able to objectively assess various aspects of the disease, such as inflammation, joint damage, or disease activity. These markers may include specific proteins, antibodies, genetic factors, or imaging features that provide valuable information for diagnosing, monitoring, and managing RA. The aforementioned RF, ACPA and acute phase reactants are examples of biomarkers used in the diagnosis of RA (18). Biomarkers might also be helpful further on in the disease course as the array of available Disease-Modifying Anti-Rheumatic Drugs (DMARDs) has significantly expanded over the years. Current treatment strategies for RA follow a trial-and-error treat-to-target

(T2T) approach, setting a target and adjusting treatment accordingly until the goal is reached. After failure of (a combination of) conventional synthetic (cs) DMARDs, current treatment guidelines advise to start a biological (b) or targeted synthetic (ts) DMARD (19,20). Available b/tsDMARDs generally show response rates between 50 to 60% (21). Potential biomarkers have been studied for their possible value in improving this T2T strategy by guiding treatment decisions to offer each patient the treatment to which they have the best probability of responding and thereby optimizing and individualizing treatment plans (22). An example of such a biomarker is Therapeutic Drug Monitoring (TDM) of biologics by means of serum drug levels and levels or Anti-Drug Antibodies (ADAbs) that may be present in Tumor Necrosis Factor inhibitor (TNFi) users. Reportedly, these ADAbs could predict response to a second DMARD with the same mode of action (23). The theory is that patients with ADAbs are more likely to respond to another TNFi than patients who failed TNFi therapy without ADAbs, as the antibodies act against the drug and cause non-response (24–26). We have evaluated the test characteristics of adalimumab serum drug levels and ADAbs as predictors for response to a next bDMARD in a cohort of RA patients who failed a TNFi (adalimumab). This study will be addressed in the **third chapter** of this thesis.

Considering the amount of research that is done into predictive biomarkers, it is of interest to investigate what characteristics (e.g. costs, sensitivity/specificity) a biomarker should have in order to have a meaningful impact on treatment outcomes. In other words, would a biomarker that accurately predicts response to a subsequent DMARD treatment significantly improve clinical outcomes? Markov modelling allows us to investigate the effects of a putative biomarker on the course of RA treatment, even though such a biomarker is currently not available (i.e. a hypothetical biomarker) (27,28). Therefore, we have examined this by building a Markov model that compared a T2T strategy without a biomarker with a biomarker-steered strategy in RA patients starting biologic therapy, in terms of time spent in remission of low disease activity and costs. This study will be addressed in the **fourth chapter** of this thesis.

FDG-PET/CT scanning

In addition to serum drug levels and ADAbs, the use of FDG-PET/CT scanning is being explored as a potential biomarker in the context of RA. Although whole-body ^{18}FDG-PET combined with CT scanning (FDG-PET/CT) is not recommended routinely for establishing and quantifying arthritis in the context of RA, it is occasionally used by physicians. Reasons to use FDG-PET/CT scans are to diagnose arthritis or to guide treatment decisions, as FDG uptake in affected joints can reflect disease activity (29,30). Recommendations for the use of imaging techniques for the joints in the clinical management of RA state that imaging may be used to predict response to treatment better than clinical features of disease activity (31). There are, however,

limited studies investigating the rate of incidental extra-articular findings that are associated with use of whole-body FDG-PET/CT scans for assessment of arthritis. Whole-body FDG-PET/CT scanning could also be used as a cancer screening tool in asymptomatic adults. This idea has been conceptually challenged, as suboptimal test characteristics might increase false-positive, false-negative or irrelevant abnormal findings (32–34). False positive findings result in inappropriate follow-up testing and sometimes even treatment whereas false negative results lead to false reassurance. In the DRESS trial, baseline and follow-up whole-body FDG-PET/CT scans were performed to assess arthritis activity in RA patients treated with TNF inhibitors, with close clinical monitoring of the patients during a 3-year period (35,36). This provided an opportunity to study the cancer screening performance of whole-body FDG-PET/CT scanning in an RA population. This study will be addressed in the **fifth chapter** of this thesis.

Laboratory monitoring for medication toxicity

As (chronic) treatment with DMARDs may induce unintended side effects, routine Laboratory Toxicity Monitoring RLTM) is recommended. During the initial 6-month period of drug use more frequent monitoring is recommended as most drug toxicity events are seen within this period. The laboratory parameters measured are considered surrogate markers for clinical morbidity related to DMARD use and include measures of liver- (Alanine Transferase (ALT), renal- (estimated Glomerular Filtration Rate (eGFR)), and hematologic toxicity (hemoglobin (Hb), White Blood Cells (WBC), and Platelets).

During long term DMARD use, laboratory monitoring is recommended every 3-6 months to monitor for drug toxicity (long term routine laboratory toxicity monitoring (lt-RLTM). Despite its widespread usage, the added clinical value of routine lt-RLTM has not been established, leaving uncertainty whether the benefits of lt-RTLM outweigh potential drawbacks. This uncertainty is also reflected in varying, cautiously formulated guidelines, which are often based on initial safety protocols from registration trials that follow a defensive strategy (19,37–42). Less frequent lt-RLTM has the potential to reduce patient- and environmental burden, and save costs. Therefore, we have examined the value of the currently used and recommended testing strategy by addressing three research questions: 1) what is the prevalence of (very) abnormal long-term RLT tests? 2) is there a difference in the incidence between patients using a DMARD for which monitoring is and is not advised?, and 3) what are characteristics of very abnormal laboratory tests? This study will be addressed in the **sixth chapter** of this thesis.

Aim and outline of this thesis

As outlined above, diagnostic and predictive testing occurs at several stages of the diagnostic and treatment process of RA, and there is a limited evidence base for the appropriateness of use of various test types in these contexts. As both the use of too many and too few tests has their own accompanying drawbacks, optimization of the use of diagnostic tests in RA is crucial. This PhD project will assess the value of several types of testing that are currently used and/or proposed to be used in for diagnosis and during treatment of RA.

Research questions that will be addressed:

1. What is the prevalence, diagnostic and prognostic value of RA-associated erosions seen on routine X-hands and feet in patients with newly presenting arthritis suspected for RA? (chapter 2)
2. What is the diagnostic test accuracy of adalimumab serum drug levels and anti-adalimumab antibodies in predicting response to a subsequent bDMARD in patients with RA? (chapter 3)
3. What is the additional value of a hypothetical biomarker in predicting response to treatment for RA when considering treatment efficacy and costs? (chapter 4)
4. What are unexpected findings of whole body FDG-PET/CT scans to measure arthritis in patients with RA, and what is their association with clinically relevant disease? (chapter 5)
5. What is the value of the current strategy of long-term routine laboratory toxicity monitoring in RA patients using DMARDs? (chapter 6)

References

1. Hicks AJ, Carwardine ZL, Hallworth MJ, Kilpatrick ES. Using clinical guidelines to assess the potential value of laboratory medicine in clinical decision-making. Biochem medica. 2021;31(1):74–84.
2. Lippi G, Bovo C, Ciaccio M. Inappropriateness in laboratory medicine: an elephant in the room? Ann Transl Med. 2017;5(4).
3. Hallworth MJ. The '70% claim': what is the evidence base? Vol. 48, Annals of clinical biochemistry. SAGE Publications Sage UK: London, England; 2011. p. 487–8.
4. Forsman RW. Why is the laboratory an afterthought for managed care organizations? Clin Chem. 1996;42(5):813–6.
5. O'Sullivan JW, Stevens S, Hobbs FDR, Salisbury C, Little P, Goldacre B, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. bmj. 2018;363.
6. Mendelson R, Bairstow PJ. Inappropriate imaging: why it matters, why it happens, what can be done. 2010;
7. Kroenke K. Diagnostic testing and the illusory reassurance of normal results: comment on "Reassurance after diagnostic testing with a low pretest probability of serious disease." JAMA Intern Med. 2013;173(6):416–7.
8. Rolfe A, Burton C. Reassurance after diagnostic testing with a low pretest probability of serious disease: systematic review and meta-analysis. JAMA Intern Med. 2013;173(6):407–16.
9. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. Jama. 2018; 320(13):1360–72.
10. Singh JA. Treatment guidelines in rheumatoid arthritis. Rheum Dis Clin. 2022;48(3):679–89.
11. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics. 2004;22:1–12.
12. Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatol (United Kingdom). 2012;51(SUPPL. 6):5–9.
13. Kennish L, Labitigan M, Budoff S, Filopoulos MT, McCracken WA, Swearingen CJ, et al. Utility of the new rheumatoid arthritis 2010 ACR/EULAR classification criteria in routine clinical care. BMJ Open. 2012;2(5):e001117.
14. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis. 2023;82(1):3–18.
15. Le Loët X, Nicolau J, Boumier P, Daragon A, Mejjad O, Pouplin S, et al. Validation of the 2010-ACR/EULAR classification criteria using newly EULAR-defined erosion for rheumatoid arthritis on the very early arthritis community-based (VErA) cohort. Jt Bone Spine. 2015;82(1):38–41.
16. Brinkmann GH, Norli ES, Bøyesen P, van der Heijde D, Grøvle L, Haugen AJ, et al. Role of erosions typical of rheumatoid arthritis in the 2010 ACR/EULAR rheumatoid arthritis classification criteria: results from a very early arthritis cohort. Ann Rheum Dis. 2017;76(11):1911–4.
17. den Hollander NK, Verstappen M, Huizinga TWJ, van der Helm-van Mil A. Management of contemporary early undifferentiated arthritis: data on EULAR's recommendation on the risk of persistent disease. Ann Rheum Dis. 2022;81(5):740–1.
18. Taylor PC, Deleuran B. Biologic markers in the assessment of rheumatoid arthritis. UpToDate. 2022.
19. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020;79(6):685–99.
20. Fraenkel L, Bathon JM, England BR, St. Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol. 2021; 73(7):1108–23.
21. Kerschbaumer A, Sepriano A, Smolen JS, van der Heijde D, Dougados M, van Vollenhoven R, et al. Efficacy of pharmacological treatment in rheumatoid arthritis: a systematic literature research informing the 2019 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis. 2020;79(6):744–59.
22. Wei K, Jiang P, Zhao J, Jin Y, Zhang R, Chang C, et al. Biomarkers to predict DMARDs efficacy and adverse effect in rheumatoid arthritis. Front Immunol. 2022;13:865267.

23. L'Ami MJ, Ruwaard J, Krieckaert CLM, Nurmohamed MT, van Vollenhoven RF, Rispens T, et al. Serum drug concentrations to optimize switching from adalimumab to etanercept in rheumatoid arthritis. Scand J Rheumatol. 2019;1–5.
24. Bartelds GM, Wijbrandts CA, Nurmohamed MT, Stapel S, Lems WF, Aarden L, et al. Anti-infliximab and anti-adalimumab antibodies in relation to response to adalimumab in infliximab switchers and anti-tumour necrosis factor naive patients: A cohort study. Ann Rheum Dis. 2010;69(5):817–21.
25. Jamnitski A, Bartelds GM, Nurmohamed MT, Van Schouwenburg PA, Van Schaardenburg D, Stapel SO, et al. The presence or absence of antibodies to infliximab or adalimumab determines the outcome of switching to etanercept. Ann Rheum Dis. 2011;70(2):284–8.
26. Plasencia C, Pascual-Salcedo D, García-Carazo S, Lojo L, Nuño L, Villalba A, et al. The immunogenicity to the first anti-TNF therapy determines the outcome of switching to a second anti-TNF therapy in spondyloarthritis patients. Arthritis Res Ther. 2013;15(4):12–6.
27. Welsing PMJ, Severens JL, Hartman M, van Gestel AM, van Riel PLCM, Laan RFJM. The initial validation of a Markov model for the economic evaluation of (new) treatments for rheumatoid arthritis. Pharmacoeconomics. 2006;24:1011–20.
28. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. J Health Serv Res Policy. 2004;9(2):110–8.
29. Mandl P, Ciechomska A, Terslev L, Baraliakos X, Conaghan PG, D'Agostino MA, et al. Implementation and role of modern musculoskeletal imaging in rheumatological practice in member countries of EULAR. RMD open. 2019;5(2):e000950.
30. Kubota K, Yamashita H, Mimori A. Clinical Value of FDG-PET/CT for the Evaluation of Rheumatic Diseases: Rheumatoid Arthritis, Polymyalgia Rheumatica, and Relapsing Polychondritis. Semin Nucl Med [Internet]. 2017;47(4):408–24. Available from: http://dx.doi.org/10.1053/j.semnuclmed.2017.02.005
31. Colebatch AN, Edwards CJ, Østergaard M, Van Der Heijde D, Balint P V., D'Agostino MA, et al. EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. Ann Rheum Dis. 2013;72(6):804–14.
32. Schöder H, Gönen M. Screening for cancer with PET and PET/CT: potential and limitations. J Nucl Med. 2007;48(1 suppl):4S-18S.
33. Saquib N, Saquib J, Ioannidis JPA. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. Int J Epidemiol. 2015;44(1):264–77.
34. Nishizawa S, Kojima S, Okada H, Shinke T, Torizuka T, Teramukai S, et al. Ten-year prospective evaluation of whole-body cancer screening with multiple modalities including [18F] fluorodeoxyglucose positron emission tomography in a healthy population. Ann Nucl Med. 2020;34(5):358–68.
35. Van Herwaarden N, Van Maas A Der, Minten MJM, Van Den Hoogen FHJ, Kievit W, Van Vollenhoven RF, et al. Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis:Open label, randomised controlled, non-inferiority trial. BMJ. 2015;350:1–8.
36. Bouman CAM, Van Herwaarden N, Van Den Hoogen FHJ, Fransen J, Van Vollenhoven RF, Bijlsma JWJ, et al. Long-term outcomes after disease activity-guided dose reduction of TNF inhibition in rheumatoid arthritis: 3-year data of the DRESS study - A randomised controlled pragmatic non-inferiority strategy trial. Ann Rheum Dis. 2017;76(10):1716–22.
37. Ledingham J, Gullick N, Irving K, Gorodkin R, Aris M, Burke J, et al. BSR and BHPR guideline for the prescription and monitoring of non-biologic disease-modifying anti-rheumatic drugs. Rheumatology. 2017;56(6):865–8.
38. Holroyd CR, Seth R, Bukhari M, Malaviya A, Holmes C, Curtis E, et al. The British Society for Rheumatology biologic DMARD safety guidelines in inflammatory arthritis. Rheumatology. 2019;58(2):e3–42.
39. Rigby WFC, Lampl K, Low JM, Furst DE. Review of Routine Laboratory Monitoring for Patients with Rheumatoid Arthritis Receiving Biologic or Nonbiologic DMARDs. Int J Rheumatol. 2017;2017:1–15.
40. Usdadiya J, Mehra S, Jain V, Singh B, Emmanuel D, Jain A, et al. Comparison of combination disease modifying antirheumatic drugs (DMARDs) with methotrexate monotherapy in early rheumatoid arthritis: an open label randomized trial. Indian J Rheumatol. 2016;11(5).

41. Tsakas JJ, Liew DFL, Adams CL, Hill CL, Proudman S, Whittle S, et al. Attitudes and practices in the laboratory monitoring of conventional synthetic disease modifying anti-rheumatic drugs by rheumatologists and rheumatology trainees. BMC Rheumatol. 2022;6(1):59.
42. European Medicines Agency. The European regulatory system for medicines and the European Medicines Agency. 2014; Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Brochure/2014/08/WC500171674.pdf

1

Chapter 2

Limited Diagnostic and Prognostic Value of Routine Radiographs in Newly Presenting Arthritis Suspected of Rheumatoid Arthritis: A Retrospective Study

Evy Ulijn
Nathan den Broeder
David Ten Cate
Kim van Overdijk
Hatice Demirel
Robert Landewé
Noortje van Herwaarden
Alfons A den Broeder

Published in Arthritis Care & Research, March 2024

Abstract

Objective: Current recommendations suggest that patients with newly presenting arthritis suspected of Rheumatoid Arthritis (RA) should undergo routine radiographs of hands and feet (X-HF) as the presence of RA-associated erosions might be of diagnostic and prognostic value. Our objective was investigate the prevalence, diagnostic and prognostic value of RA-associated erosions seen on routine X-HF in a large, recent cohort of newly presenting arthritis patients.

Methods: A retrospective cohort study was performed between 2016-2019 in patients with newly presenting arthritis suspected of RA. Patients were included if arthritis was present at diagnosis, Rheumatoid Factor and Anti-Citrullinated Protein Antibodies were measured, RA was noted in the differential diagnosis, and routine X-HF were conducted. Outcomes were the prevalence of ≥1 RA-associated erosion, and whether diagnostic or prognostic classification were changed by erosivity. Seronegative patients, patients without Acute Phase Reactants (APR) and patients with longer symptom duration were analyzed as subgroups.

Results: RA-associated erosions were found in 32/724 patients (4.4%, CI 3.1-6.2%). Erosions led to a change of diagnostic classification in two patients (0.3%, 95%CI 0.01-1.1%) and changed prognostic classification in three patients (0.4%, 95%CI 0.1-1.3%). Seronegative patients and patients without elevated APR had significantly lower prevalence of erosions (χ^2_1 9.4, P=0.002, χ^2_1 6.5, P=0.01). Longer symptom duration was not associated with a different prevalence of erosions (χ^2_1 0.4, P=0.81).

Conclusion: The recommendation of conducting routine X-HF in patients with newly presenting arthritis suspected of RA might be reconsidered, due to low prevalence of early erosive disease and lack of diagnostic and prognostic value.

Keywords: RA, radiographs, hands, feet, diagnostics, prognostics, early arthritis, undifferentiated arthritis.

Significance and innovations

- Routine radiographs of hands and feet are recommended in the diagnostic and prognostic work-up for patients with newly presenting arthritis suspected of RA. Despite the limited evidence supporting the use of these routine radiographs in yielding significant new findings, it is still considered standard practice and endorsed in the 2010 classification criteria.
- Our study results depict that prevalence of RA-associated erosions on routine radiographs of hands and feet in patients with newly presenting arthritis suspected of RA was low, and performing routine radiographs rarely leads to a change in diagnosis or prognosis (even in relevant subgroups).
- Although radiographs of hands and feet in patients with newly presenting arthritis suspected of RA are of value in specific cases and on indication, the recommendation to perform these radiographs routinely might be reconsidered as prevalence, diagnostic and prognostic value seems limited.

Introduction

Routine radiographs of hands and feet (routine X-HF) are recommended in the workup for patients with newly presenting arthritis suspected of Rheumatoid Arthritis (RA) (1,2). The rationale for this is threefold. Firstly, RA-associated erosions detected through X-HF can aid in diagnosing atypical cases, such as seronegative RA or patients without elevated Acute Phase Reactants (APR). Secondly, the presence of RA-associated erosions may be of prognostic value for a more severe disease course, requiring more intensive treatment. Lastly, routine X-HF can serve as a reference for future radiographic progression.

The diagnostic role of routine X-HF in RA is reflected in the 2010 ACR/EULAR criteria, with erosions typical for RA allowing classification of RA even without fulfillment of the scoring system (3,4). The 2010 ACR/EULAR criteria defined erosions typical for RA as; 'when an erosion (defined as a cortical break) is seen in at least three separate joints at any of the following sites: the proximal interphalangeal, the metacarpophalangeal, the wrist and the metatarsophalangeal joints'(4). Of note, this addition to the other criteria was made due to high specificity, and despite low sensitivity for new RA.

Prognostically, early erosive disease, along with factors like Rheumatoid Factor (RF) or Anti-Citrullinated Protein Antibodies (ACPA), and high disease activity, indicates a poor RA prognosis (5). However, the additional value of specifically RA-associated erosions at baseline without positivity of the other predictors remains uncertain.

Several somewhat older and smaller studies have investigated the prevalence and/or diagnostic value of routine X-HF in patients with newly presenting arthritis. A community based Early-Arthritis Cohort (EAC) classified 170/269 (63.2%) as 2010 ACR/EULAR criteria positive. Among them, 28/170 (16.5%) showed at least one erosion and only 6/170 (3.5%) patients had ≥3 RA-associated erosions on routine X-HF. 99/269 (12.1%) patients had a 2010 ACR/EULAR score of <6, 12/99 (21.1%) had ≥1 erosion and none of the 99 patients had ≥3 erosive joints (6). In another EAC 120/289 (41.5%) patients fulfilled the 2010 ACR/EULAR criteria, of whom 49 (40.8%) had ≥1 erosive joint at baseline, and 17 (14.2%) had at ≥3 erosive joints at baseline. Of the remaining 169 patients not fulfilling the 2010 ACR/EULAR criteria 55 patients had ≥1 erosive joint, and 15 patients had ≥3 erosive joints (7). It has been suggested that RA is evolving into a milder disease at presentation in the last decades, which is probably partly due to patients presenting themselves earlier in the disease course (8,9). This trend, would decrease the prevalence of erosions at diagnosis, and limit diagnostic and/or prognostic value of routine X-HF (10,11). A more recently published letter reported prevalence of erosions of only 1.8% in a cohort of 710 patients (12). Results of existing studies suggest that additional diagnostic value of routine X-HF may be limited. Considering the limited diagnostic value and lack of evidence regarding prognostic value, routine X-HF are still recommended and widely used in the diagnostic and prognostic work-up of RA.

To address this, we set out to explore the prevalence and diagnostic/prognostic value of RA-associated erosions seen on routine X-HF in a large, recent cohort of patients with newly presenting arthritis suspected of RA.

Patients and methods

Design and participants

A retrospective cohort study was conducted at the department of rheumatology of the Sint Maartenskliniek, a large tertiary referral center specialized in rheumatology in The Netherlands. All new patients over 16 years of age who visited the outpatient clinic between January 2016 and January 2019 were considered for inclusion. We aimed to include all patients with newly presenting arthritis who were suspected of RA. This was operationalized by including patients who met four entry criteria. Firstly, arthritis should have been present at the first consultation (judged by a rheumatologist). Secondly, RF and ACPA were measured within 6 months before or within four days after the first consultation. Thirdly, RA should have been noted in the differential diagnosis and/or in the order of the radiograph. Lastly, only a set radiographs of both hands and feet were considered for inclusion, as a radiograph of

hands or feet alone is more likely to be performed on indication rather than routinely. Patients were excluded if they had a preexisting RA diagnosis, another diagnosis of inflammatory arthritis, or if a different indication was noted for the radiograph.

Patient data were collected on demographics, diagnostics, disease characteristics and final clinical diagnosis. The final clinical diagnosis by a rheumatologist was cross-referenced with the reimbursement code (DBC 101), or ICD codes for RA (ICD 9 714.x, ICD 10 M06.9) to ensure the clinical diagnosis that was used was accurate and up-to date. Reports of routine X-HF (assessed by specialized musculoskeletal radiologists in usual care) were also examined by the treating rheumatologist. To increase specificity, all routine X-HF with the term 'erosion' mentioned in the report were retrospectively reassessed by a rheumatologist (AB, DTC) for presence of at least one typical RA-associated erosion. Reports of routine X-HF in which 'erosions' were not mentioned by the radiologist were not retrospectively re-assessed.

Outcomes and subgroup analyses

Our primary outcome was the prevalence of at least one typical RA-associated erosion found on routine X-HF. Secondary outcomes were firstly the additional value of routine X-HF for diagnostic classification, defined as prevalence of patients being reclassified as positive for 2010 ACR/EULAR criteria based only on RA-associated erosions on routine X-HF. The other secondary outcome was relevance of routine X-HF for prognostic classification, defined as the prevalence of patients being reclassified as prognostically unfavorable based only on early-onset joint damage (erosions), and therefore without RF/ACPA positivity or high disease activity (Disease Activity Score based on C-Reactive Protein (DAS28CRP) >5.1), conform the 2022 EULAR recommendations for RA (5). Additionally, the number needed to screen (NNS) of both secondary outcomes was calculated.

Subgroup analyses were performed for three groups. Firstly, the seronegative patients who did not have presence of RF nor ACPA. Secondly, patients without elevated APR, including Erythrocyte Sedimentation Rate (ESR) and CRP. The third subgroup consisted of patients with a longer symptom duration of ≥6 weeks at presentation (as described in the 2010 ACR/EULAR criteria). Additionally the association between symptom duration and erosive disease with a logistic regression analysis. The groups of seronegative patients and patients with non-elevated APR were analyzed separately as they might show a different prevalence of erosions found on routine X-HF, and they were expected to gain most diagnostic and prognostic value out of making routine X-HF. Patients with a longer symptom duration were analyzed separately as these patients might be expected to have a higher rate of erosive disease.

As the 2010 ACR/EULAR criteria are classification criteria and not diagnostic criteria, the association between the criteria and final clinical diagnosis of RA was assessed to test study integrity and generalizability.

Statistical analysis

Baseline characteristics (age, gender, 2010 ACR/EULAR criteria positivity, duration of symptoms, RF-/ACPA positivity and DAS28CRP) were described using descriptive statistics and provided with mean (+/- Standard Deviation (SD)), median (interquartile ranges (IQR (p25-p75)), or n (%) depending on data distribution. To describe the association between 2010 ACR/EULAR criteria and clinical diagnosis sensitivity and specificity were calculated.

Outcomes were described by providing percentages and 95%-Confidence Intervals (95%-CI). Missing data was mentioned respective tables. We did not make use of statistical techniques to handle missing data in the dataset as there was no missing data on the variables supporting the primary analyses. To examine differences between relevant subgroups Pearson chi-square tests for categorical data were performed. A p-value <0.05 was considered statistically significant. No correction for multiple testing was performed. Castor EDC was used to enter and store the data. All statistical analyses were performed in STATA/IC version 13.1.

Results

Between 01/2016 and 01/2019, 5,836 new consultations took place in the rheumatology department of the Sint Maartenskliniek that included patients without history of a rheumatic disease in whom RF and ACPA was tested within 6 months before or within four days after the first consultation, and routine X-HF were made within 6 months before or after the first consultation. Arthritis was observed and RA was clinically suspected in 724 patients, of whom 299 (41.3%) were eventually clinically diagnosed with RA (*table 1).* Sensitivity and specificity of the 2010 ACR/EULAR classification at baseline for final clinical diagnosis of RA were 80.3% (95%-CI 75.3-84.6) and 96.4% (95%-CI 94.2-98.0), respectively.

Prevalence of RA-associated erosions on routine X-HF

The radiologist mentioned the term 'erosion(s)' in 107/724 (14.8%) patients' radiographs. After review by a rheumatologist, at least one RA-associated erosion was found in 32 of the 724 patients with newly presenting arthritis (4.4%, CI 3.1-6.2%). The other 75 patients' radiographs contained other types of erosions such as gouty- or osteo-arthritis-associated erosions. Of the 32 patients with at least one erosion, 11/32

(34.4%, CI 18.6-53.2%) patients were seronegative and 10/32 (31%, CI 16.1-50%) (partly overlapping) patients had non-elevated APR. Prevalence of RA-associated erosions was significantly lower for both seronegative patients (χ^2_1 9.4, *P* 0.002) and patients without elevated APR (χ^2_1 6.5, *P* 0.010).

Table 1. Baseline characteristics of all included patients, and patients eventually clinically diagnosed with RA.

	Clinically diagnosed with RA (N=299)	Not clinically diagnosed with RA (N=425)	All patients with newly presenting arthritis (N=724)
Age [mean ± SD]	58.8 ± 14.7 *N = 299*	55.7 ± 15.2 *N = 425*	57 ± 15.0 *N = 724*
Female [n (%)]	167 (55.8%) *N = 299*	282 (66.3%) *N = 425*	449 (62%) *N = 724*
Past or present smoker [n (%)]	139 (46.5%) *N = 245*	166 (39.1%) *N = 333*	305 (42.1%) *N = 573*
Disease status			
2010 ACR/EULAR criteria positive [n (%)]	240 (80.3%) *N = 299*	15 (3.5%) *N = 425*	266 (36.7%) *N = 724*
Symptom duration at first visit in months [median (IQR)]	5 (12) *N = 292*	9.6 (36) *N = 404*	6 (28) *N = 696*
Rheumatoid factor positive [n (%)]	177 (60%) *N = 299*	59 (14.0%) *N = 425*	236 (33%) *N = 724*
ACPA positive [n (%)]	177 (59.8%) *N = 299*	33 (7.8%) *N = 425*	210 (29.3%) *N = 724*
DAS28CRP [mean ± SD]	3.8 ± 1.2 *N = 287*	3.0 ± 1.1 *N = 412*	3.4 ± 1.3 *N = 699*
BSE [mean ± SD]	31.3 ± 24.6 *N = 224*	18.6 ± 16.5 *N = 296*	24.1 ± 21.3 *N = 520*
CRP [mean ± SD]	22.5 ± 37.0 *N = 291*	9.7 ± 21.2 *N = 418*	14.9 ± 29.4 *N = 709*
Tender joint count [mean ± SD]	6.0 ± 6.1 *N = 289*	5.2 ± 6.1 *N = 397*	5.5 ± 6.1 *N = 686*
Swollen joint count [mean ± SD]	5.6 ± 5.2 *N = 288*	2.4 ± 3.6 *N = 396*	3.8 ± 4.6 *N = 684*

N=sample size with available data, n=number of patients in referenced group,
SD= standard deviation, %= population proportion of n, IQR=inter quartile range.

Relevance for diagnostic and prognostic classification

The presence of at least one RA-associated erosion seen on the routine X-HF caused a change to positive RA-classification in two out of 724 (0.3%, 0.01-1.1%) patients leading to a NNS of 362. Subgroup analyses showed that two out of 437 (0.5%, 0.1-1.8%) seronegative patients had RA-associated erosions changing the RA-classification to positive (NNS 219). Concerning patients without elevated APR classification changed to positive in two out of 385 (0.5%, 0.1-2.1%) patients (NNS 193). The prevalence of RA-associated erosions changing diagnostic classification was not significantly higher for seronegative patients (χ^2_1 1.3, *P* 0.25) nor for patients with non-elevated APR (χ^2_1 1.7, *P* 0.18).

RA-associated erosions seen on routine X-HF changed prognostic classification in three out of 724 patients (0.4%, 0.1-1.3%), leading to a NNS of 241. Subgroup analysis showed that RA-associated erosions seen on routine X-HF changed prognostic classification in three out of 437 (0.7%, 0.2-2.1%) seronegative patients (NNS 146), and two out of 385 (0.5%, 0.1-2.1%) patients with non-elevated APR (NNS 193). Prevalence of RA-associated erosions changing prognostic classification was not significantly higher for seronegative patients (χ^2_1 2.0, *P* 0.16) or for patients with non-elevated APR (χ^2_1 0.2, *P* 0.64).

Erosions were seen in five out of 96 (5.2%, CI 2.1-12.9%) patients with short symptom duration (<6 weeks), and in 27/615 (3.4%, CI 3-6.3%) patients with a longer symptom duration (≥ 6 weeks). Erosive disease was not significantly more prevalent in patients with longer symptom duration (χ^2_2, 0.4 *P*=0.81), *table 2*. There was a weak and non-significant association between erosive disease and a shorter symptom duration (odds ratio 0.99, 95%-confidence interval 0.97-1.01).

Table 2. Prevalence of typical RA erosive disease, and diagnostic/prognostic relevance of radiographs. Subgroup analyses for seronegative patients, non-elevated APR and short/long symptom duration.

	Seronegative (N=437)	Non-elevated APR (N=385)	All patients (N=724)
Erosive disease [n (%, 95%-CI)]	11 (2.5%, CI 1.4-4.5%) *N = 437*	10 (2.6%, CI 1.4-4.8%) *N = 385*	32 (4.4%, CI 3.1-6.2%) *N = 724*
Symptom duration < 6 weeks	2 (4.1%, CI 1.0-15.6%) *N = 49*	1 (3.6%, CI 0.4-23.7%) *N = 28*	5 (5.2%, CI 2.1-12.9%) *N = 96*
Symptom duration ≥ 6 weeks	9 (2.4%, CI 1.2-4.5%) *N = 379*	9 (2.6%, CI 1.3-4.9%) *N = 350*	27 (3.4%, CI 3-6.3%) *N = 615*
Erosions resulting in changed 2010 ACR/EULAR RA-classification [n (%, 95%-CI)]	2 (0.5%, 0.1-1.8%) *N = 437*	2 (0.5%, 0.1-2.1%) *N = 385*	2 (0.3%, 0.01-1.1%) *N = 724*
Erosions resulting in reclassification to prognostically unfavorable [n (%, 95%-CI)]	3 (0.7%, 0.2-2.1%) *N = 437*	2 (0.5%, 0.1-2.1%) *N = 385*	3 (0.4%, 0.1-1.3%) *N = 724*

N=sample size with available data, n=number of patients in referenced group,
%= population proportion of n, 95%, CI= 95% confidence interval of population proportion

Discussion

The prevalence of RA-associated erosions in patients with newly presenting arthritis suspected of RA was low, and rarely led to a change in diagnosis or prognosis. Subgroups of seronegative patients and patients without elevated APR had a significantly lower prevalence of RA-associated erosions found on routine X-HF. The prevalence of erosions altering diagnostic or prognostic classification in these subgroups were not significantly higher. Therefore, there seems no need to recommend routine X-HF in patients with newly presenting arthritis suspected of RA. Of note, we only evaluate the use of *routine* radiographs of hands and feet in this group of patients. Radiographs of hands and/or feet that are made for any specific indication can be of value, and our study results do not generalize to routine X-HF for a specific indication. Additionally, in some contexts, an alternative diagnosis may be more probable. Our study did not specifically investigate alternative diagnoses, so further research is required in this area to determine when radiographs should be considered in such cases.

This study's strengths lie in its relatively recent data, ample sample size, and low risk of ascertainment bias due to its retrospective nature. Rheumatologists scored erosions as typical for RA if at least one erosion was seen on X-HF as would probably better reflect how interpretation would be done in clinical care. This is a more sensitive cut-off than the 2010 ACR/EULAR criteria, which require erosions in at least three separate joints. Therefore, using the 2010 ACR/EULAR definition would yield an even lower prevalence of RA-associated erosions on routine X-HF.

A potential limitation of the study is the method of identifying patients with newly presenting arthritis, as there is no specific registration code for this patient group in our center. We therefore had to operationalize how to select patients with newly presenting arthritis suspected of RA from the electronic health records. This was done by requiring the four entry criteria mentioned in the method section, which seem sensible and face valid. Selection bias might have occurred as the entry criteria included mandatory testing of RF/ACPA. However, patients in whom these tests had not been performed but still had arthritis were likely to have a lower chance for RA. Therefore, the magnitude of this bias is unknown, but the direction of the potential bias is in the conservative direction. Also, there could have been patients with newly presenting arthritis who were suspected of RA but did not undergo routine X-HF. Presumably these patients had a milder disease presentation, and therefore any bias resulting from leaving these patients out would have led to over-, not underestimation of the prevalence of erosions. We also did not reassess all radiographs that were negative for initial erosivity. However, it is unlikely that a relevant number of erosions would have been missed by both the specialized musculoskeletal radiologist and the treating rheumatologist, when the indication of the X-HF was presence of erosions.

Although we were able to show the lack of value of routine X-HF for RA diagnosis or prognosis, one might argue that this radiograph could be valuable as baseline assessment to assess future change. However, given the low prevalence of RA-associated erosions, one can assume that erosions seen on subsequent routine X-HF could be interpreted as new erosions. Another reason for routine X-HF might be to find alternative diagnoses. However, only a handful of exceedingly uncommon alternative diagnoses in early oligo/polyarthritis are associated with these specific radiographical abnormalities which does not make it rational to routinely perform X-HF (13). Additionally, there are no recommendations supporting the use of routine X-HF for identifying alternative diagnoses in this context.

Another argument may be that use, or assessment of routine X-HF might be suboptimal in our clinic. However, the prevalence of erosions we observe is consistent with the observed prevalence in other studies conducted in a similar setting (12,14,15). It is

possible that another test, such as CT scans, would be more sensitive for capturing erosive disease. Investigating those alternative modalities is beyond the scope of our current study as we aimed to evaluate the existing recommendations, and not to explore new optimizations.

Our study shows high generalizability as baseline characteristics of our population were very comparable with the Leiden EAC and the Argentinian EAC, *table S1* (14,15). Additionally, the Leiden EAC reports a similar low prevalence of erosions on routine X-HF (1.8%) (12). The distribution of final diagnoses also indicates that our cohort consisted of patients with newly presenting arthritis suspected of RA, as the two largest categories of final diagnoses were RA (40.5%) and undifferentiated arthritis (18.1%). Additionally, sensitivity and specificity of the 2010 ACR/EULAR criteria for clinical RA diagnosis were high, validating our clinical endpoint. The generalizability of our study remains limited to comparable health care systems where patients have timely access to high quality rheumatological care. In contexts with more delay or other barriers the prevalence of erosive disease at presentation may be substantially higher, as would be the associated diagnostic and prognostic value of routine X-HF.

In conclusion, the recommendation of conducting routine radiographs of hands and feet in patients with newly presenting arthritis suspected of RA might be reconsidered, due to low prevalence of early erosive disease, and lack of diagnostic and prognostic value of RA-associated erosions.

Acknowledgements: We would like to acknowledge Stijn van der Plassche for his valuable research assistance.

Authors' contributors: EU, NdB, AdB, NvH, HD, DTC and KvO were involved in the conception and design of the study. EU, KvO and NdB were involved in the data collection. EU, NdB, NvH and AdB contributed to the data analysis. EU, KvO, NdB, NvH and AdB drafted the manuscript, and all co-authors have critically reviewed the manuscript and gave final approval for its submission.

Funding: This study did not receive external funding.

Conflict of interests: None of the authors have any conflict of interests for the submitted work.

Patient involvement statement: Patients and the public were not involved in the design and conduct of the study, choice of outcome measures, or recruitment to the study due to the (retrospective) nature of data collection.

Patient consent: Potentially eligible participants were identified through the electronic health records of the Sint Maartenskliniek. Pseudonymized data were extracted. Based on WGBO (law on medical treatment agreement) 458 it was judged not feasible to acquire informed consent from all patients. Therefore, following WGBO 458 sec 2, pseudonymized data were used without opt in consent, but after informing all patients and providing opt-out option.

Ethics committee approval: This study has been judged to fall outside the scope of the WMO (law on medical research). Therefore, Medical Ethical Review Board approval was not required (CMO: 2020-6806).

Data sharing statement: Additional unpublished data can be obtained from the corresponding author upon reasonable request.

Table S1. Comparison of baseline variables of all patients with newly presenting arthritis in the current cohort with existing early arthritis cohorts (21,22).

	All patients with newly presenting arthritis in current cohort (N=724)	Leiden EAC (2010) (N=570)	Argentinian EAC (2010) (N=413)
Age [mean ± SD]	57 ± 15.0	60 ± 16.8	47 ± 14
Female [n(%)]	449 (62%)	329 (57.7%)	137 (82%)
Past or present smoker [n (%)]	305 (42.1%)	271 (48%)	35 (23%)
Disease status			
Symptom duration at first visit in months [median (IQR)]	6 (28)	5 (5)*	8 (6)*
Rheumatoid factor positive [n(%)]	236 (33%)	140 (24.6%)	33 (27.7%)
ACPA positive [n(%)]	210 (29.3%)	121 (21.2%)	..
DAS28crp [mean ± SD]	3.4 ± 1.3	..	4.6 ± 1
ESR [mean ± SD]	24.1 ± 21.3	29.5 ± 24.8	..
CRP [mean ± SD]	14.9 ± 29.4	21.4	2.34 ± 6
TJC [mean ± SD]	5.5 ± 6.1	..	6.48 ± 6
SJC [mean ± SD]	2.5 ± 3.6	3.8 ± 4	5.31 ± 5

*based on mean (SD), median (IQR) unknown.
N=total sample size, n=number of patients in referenced group,
SD=standard deviation, IQR=interquartile range, %= population proportion of n, 95%-CI= 95% confidence interval of population proportion.

Table S2. Comparison of baseline variables of all patients clinically diagnosed with RA in the current cohort with existing early arthritis cohorts (21,22).

	Patients diagnosed with RA (N=293)	RA patients Leiden EA cohort (N= 676)	RA patients Argentina EAC (N=183)
Age [mean ± SD]	59 ± 14.6	56.4 ± 15.7	47 ± 14
Female [n(%)]	162 (55.3%)	459 (67.9%)	148 (81%)
Past or present smoker [n(%)]	137 (57.1%)	..	62 (38%)
Disease status			
Symptom duration at first visit in months [median (IQR)]	5 (12)	6 (5)*	8 (6)*
Rheumatoid factor positive [n(%)]	173 (59.9%)	378 (58%)	146 (85%)
ACPA positive [n(%)]	174 (60%)	217 (32%)	..
DAS28crp [mean ± SD]	3.9 ± 1.2	..	5.4 ± 1
ESR [mean ± SD]	30.9 ± 24.1	39.7 ± 27.4	35.8 ± 24
CRP [mean ± SD]	22.4 ± 37.1	30.4 ± 34.7	7.6 ± 19
TJC [mean ± SD]	6.0 ± 6.1	..	9.5 ± 7
SJC [mean ± SD]	5.5 ± 5.2	9.5 ± 7.4	8.5 ± 6

*based on mean (SD), as median (IQR) are unknown.
N=total sample size, n=number of patients in referenced group,
SD=standard deviation, IQR=interquartile range, %= population proportion of n, 95%-CI= 95% confidence interval of population proportion.

References

1. Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatol (United Kingdom). 2012;51(SUPPL. 6):5–9.
2. Mackenzie AH. Differential diagnosis of rheumatoid arthritis. Am J Med. 1988;85(4 SUPPL. 1):2–11.
3. American College of Rheumatology, European League Against Rheumatism. 2010 Classification Criteria for Rheumatoid Arthritis. Prescriber. 2011;22(13–14):39–39.
4. Van Der Heijde D, Van Der Helm-Van AHM, Aletaha D, Bingham CO, Burmester GR, Dougados M, et al. EULAR definition of erosive disease in light of the 2010 ACR/EULAR rheumatoid arthritis classification criteria. Ann Rheum Dis. 2013;72(4):479–81.
5. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis. 2023;82(1):3–18.
6. Le Loët X, Nicolau J, Boumier P, Daragon A, Mejjad O, Pouplin S, et al. Validation of the 2010-ACR/EULAR-classification criteria using newly EULAR-defined erosion for rheumatoid arthritis on the very early arthritis community-based (VErA) cohort. Jt Bone Spine. 2015;82(1):38–41.
7. Brinkmann GH, Norli ES, Bøyesen P, van der Heijde D, Grøvle L, Haugen AJ, et al. Role of erosions typical of rheumatoid arthritis in the 2010 ACR/EULAR rheumatoid arthritis classification criteria: results from a very early arthritis cohort. Ann Rheum Dis. 2017;76(11):1911–4.
8. Diffin JG, Lunt M, Marshall T, Chipping JR, Symmons DPM, Verstappen SMM. Has the severity of rheumatoid arthritis at presentation diminished over time? J Rheumatol. 2014;41(8):1590–9.
9. Alcorn N, Meng MC, Murdoch R, Madhok R. Rheumatoid arthritis in recession. J Rheumatol. 2009; 36(7):1353–4.
10. Welsing PMJ, Fransen J, van Riel PLCM. Is the disease course of rheumatoid arthritis becoming milder?: Time trends since 1985 in an inception cohort of early rheumatoid arthritis. Arthritis Rheum. 2005;52(9):2616–24.
11. Finckh A, Choi HK, Wolfe F. Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment. Ann Rheum Dis. 2006;65(9):1192–7.
12. den Hollander NK, Verstappen M, Huizinga TWJ, van der Helm-van Mil A. Management of contemporary early undifferentiated arthritis: data on EULAR's recommendation on the risk of persistent disease. Ann Rheum Dis. 2022;81(5):740–1.
13. Ezzati F, Pezeshk P. Radiographic Findings of Inflammatory Arthritis and Mimics in the Hands. Diagnostics. 2022;12(9):2134.
14. de Rooy DPC, van der Linden MPM, Knevel R, Huizinga TWJ, van der Helm-van Mil AHM. Predicting arthritis outcomes—what can be learned from the Leiden Early Arthritis Clinic? Rheumatology. 2011;50(1):93–100.
15. Marcos J, Waimann C, Dal Pra F, Hogrefe J, Retamozo S, Caeiro F, et al. General characteristics of an early arthritis cohort in Argentina. Rheumatology. 2011;50(1):110–6.

Chapter 3

Therapeutic drug monitoring of adalimumab in RA: no predictive value of adalimumab serum levels and anti-adalimumab antibodies at time of adalimumab failure for prediction of response to the next bDMARD

Evy Ulijn
Nathan den Broeder
Maike HM Wientjes
Noortje van Herwaarden
Inger Meek
Lieke Tweehuysen
Aatke van der Maas
Bart JF van den Bemt
Alfons A den Broeder

Published in Annals of Rheumatic Diseases, April 2020

Abstract

Background: After adalimumab treatment failure, TNFi and non-TNFi bDMARDs are equally viable options on a group level as subsequent treatment in RA based on the current best evidence synthesis. However, preliminary data suggest that anti-adalimumab antibodies (ADA) and adalimumab serum levels (ADL) during treatment predict response to a TNFi as subsequent treatment.

Objective: To validate the association of presence of ADA and/or low ADL with response to a subsequent TNFi bDMARD or non-TNFi bDMARD. Sub-analyses were performed for primary and secondary non-responders.

Methods: a diagnostic test accuracy retrospective cohort study was done in consenting RA patients who discontinued adalimumab after >3 months of treatment due to inefficacy and started another bDMARD. Inclusion criteria included the availability of (random timed) serum samples between ≥8 weeks after start and ≤2 weeks after discontinuation of adalimumab, and clinical outcome measurements (DAS28-CRP) between 3-6 months after treatment switch. Test characteristics for EULAR good response (DAS28-CRP based) after treatment with the next (non-) TNFi bDMARD were assessed using ROC AUC and sensitivity/specificity.

Results: 137 patients were included. ADA presence was not predictive for response in switchers to a TNFi (sensitivity/specificity 18%/75%) or a non-TNFi (sensitivity/specificity 33%/70%). The same was true for ADL levels in TNFi switchers (sensitivity/specificity 50%/52%) and non-TNFi switchers (sensitivity/specificity 32%/69%). Predictive value of ADA and ADL were similar for both primary and secondary non-responders to adalimumab.

Conclusions: In contrast to earlier research, we could not find predictive value for response to a second TNFi or non-TNFi for either ADA or random timed ADL.

Keywords: RA, adalimumab, TNFi, prediction, therapeutic drug monitoring

Key messages

What is already known about this subject?

- Anti-adalimumab antibody (ADA) presence has been suggested to correlate with response to a second bDMARD after discontinuation of adalimumab use.

What does this study add?

- We investigated the predictive value of ADA and adalimumab serum levels (ADL) for EULAR clinical response to subsequent treatment with a second bDMARD (TNFi or non-TNFi) after discontinuing adalimumab because of treatment failure.

How might this impact on clinical practice?

- ADA presence nor ADL had predictive value for clinical response to a subsequent TNFi or non-TNFi treatment after failure of adalimumab treatment.
- Combining these data with 4 earlier studies that did find some predictive value of adalimumab and etanercept (anti)drug levels, the next research step might be doing a well-dimensioned prospective trial

Introduction

Biologic disease-modifying anti-rheumatic drugs (bDMARDs) are important in the treatment of rheumatoid arthritis (RA). bDMARDs with several modes of action are available, such as TNF inhibition (TNFi: adalimumab, etanercept, golimumab, infliximab, certolizumab) and non-TNFi (e.g. rituximab, tocilizumab, abatacept). Adalimumab – a human monoclonal antibody TNFi – is one of the most frequently used bDMARDs, and is a safe and effective treatment for RA.

However, approximately 41% of RA patients do not achieve good response after 6 months of treatment with adalimumab.[1] After non-response to adalimumab treatment (or any bDMARD treatment), the current guidelines state that another TNFi or a non-TNFi bDMARD could be prescribed as a subsequent treatment with equal chance of response.[2] This is supported by current available evidence from three randomized controlled trials,[3–5] and a systematic review on predictive factors for response to a bDMARD in RA.[5] Based on this, no preference should be given to starting either another TNFi, or a non-TNFi bDMARD after primary or secondary non-response to adalimumab.

However, it has been suggested that measurement of adalimumab serum levels and/or anti-adalimumab antibodies (therapeutic drug monitoring, TDM) might be helpful for channeling the right patients to a TNFi or a non-TNFi thus increasing overall

response chances.[6] The rationale for this is that approximately 20% of the RA patients treated with adalimumab develop antibodies against this drug (anti-drug antibodies, ADA) and this can result in primary or secondary non-response. Another possible reason for non-response, however, is innate insensitivity to TNFi in a proportion of patients. It can be hypothesized that the first group of non-responders will have adequate response chances to a second TNFi, whereas in the second group of patients, TNFi response will be much lower. A recent systematic review indeed supports this notion based on three small studies in RA and axial spondylarthritis, for adalimumab and infliximab.[3,7,8]

Following this rationale, the optimal strategy after adalimumab non-response might be a second TNFi in patients with low adalimumab levels/ADA presence, and a non-TNFi in patients with adequate levels and no ADA presence. One could argue that just giving a non-TNFi in all adalimumab non-responders would negate the need for testing. However, as many adalimumab non-responding patients experience secondary non-response rather than primary non-response, patients in which secondary non-response occurred were indeed TNFi responding patients. Therefore, response rates to a second TNFi in these patients might be *higher* than response rates to a non-TNFi, resulting a better outcome for all patients after TDM.

The abovementioned hypothesis has – in part – been tested in two studies with infliximab and adalimumab.[7,9] These studies showed that presence of ADA against either infliximab or adalimumab was associated with a larger decrease in disease activity after the next TNFi. Additionally, the same mechanism has been replicated using infliximab in RA, and adalimumab in axial spondyloarthritis (3,8). However, these studies have some limitations. Firstly, the number of patients was somewhat limited, and no differentiation was made between primary and secondary non-responders, a distinction that might be important for response chances to a second TNFi as argued earlier. Also, these studies did not mention test characteristics (sensitivity, specificity), only difference in mean improvement, thus hampering judgement of test characteristics. In addition, the studies did not assess the predictive value of adalimumab TDM for response to non-TNFi after adalimumab, which is relevant to determine whether ADA presence is simply a marker of more refractory disease or able to differentially predict response to a second TNFi compared to a non-TNFi . Finally, testing with a newer competitive enzyme-linked immunosorbent assay (ELISA) is now possible in order to quantify anti-drug antibodies even in the presence of large amounts of TNF inhibitor. As this is a drug-tolerant assay, it is a more precise measure of ADA than conventional testing methods where ADA cannot be detected in the presence of large amounts of the drug.

Therefore, we set out to investigate this predictive value in a larger study population, estimating sensitivity and specificity of both presence of ADA and random timed adalimumab levels (ADL), and validate currently proposed thresholds, in both TNFi and non-TNFi switchers.

Methods

Design

A retrospective diagnostic test accuracy cohort study to assess the predictive value of ADA and ADL for response to a subsequent TNFi or non-TNFi bDMARD in RA patients.

Patients

All RA patients who received adalimumab and subsequently another TNFi (etanercept, golimumab, infliximab, certolizumab) or a non-TNFi bDMARD (Rituximab, Tocilizumab, Abatacept) in the Sint Maartenskliniek or Radboud University Medical Centre between January 2012 and January 2018 were considered for inclusion in the current study. Potentially eligible participants were identified through the electronic patient records of the Sint Maartenskliniek and the Radboudumc. Patients included in this study had a diagnosis of RA according to ACR 1987/2010 criteria, EULAR criteria or clinical diagnosis,[10] and were ≥16 years of age. They had received adalimumab for at least 3 months (+/- 2 weeks) in standard dosing (40mg subcutaneously every other week). Acceptable reasons for stopping adalimumab were either inefficacy (primary or secondary, no formal disease activity cut-off) or toxicity, but not tapering because of remission. The next bDMARD should also have been administered in standard dosing (registered dose, exception being rituximab 1×1000/2×500 mg instead of 2×1000mg) for at least 3 months (+/- 2 weeks). Furthermore, a serum sample that is suitable for analysis should be available, being samples taken ≥8 weeks after start adalimumab and within 2 weeks after discontinuing adalimumab (for ADL) or within 12 weeks after discontinuation (for ADA),[11] Finally, DAS28 scores had to be available to assess EULAR clinical response to subsequent bDMARDs, a baseline DAS at start and a follow-up DAS after 3-6 months of treatment (+/- 8 weeks).

Ethical approval, consent and funding

Approval from the local ethics committee (Commissie Mensgebonden Onderzoek (CMO) region Arnhem-Nijmegen) was obtained (CMO: 2019-5443). Patients had either previously consented to inclusion in several biobanking studies, including the Nijmegen RA protocollaire follow-up[12] (CMO-number: 2016-2281) and the BIOTOP study[13] (CMO region Arnhem-Nijmegen, ***NL47946.091.14)*** or were sent opt-out

informed consent letters with information about the aims and methods of the study. Patients were given 4 weeks to read the information and respond in case they are not willing to participate (according to Dutch law: WGBO art 458 sub 2). This study received no external funding. The laboratory analyses of adalimumab and ADA levels and personnel costs were funded by the Sint Maartenskliniek.

The study was conducted according to the principles of the Declaration of Helsinki and in accordance to Dutch law: WMO, AVG, WGBO, code Goed Gedrag and NFU 'richtlijn kwaliteitsborging mensgebonden onderzoek'.

Testing of serum adalimumab levels and anti-adalimumab antibodies

After collection, the serum samples were stored at -80°C until analysis. Blood samples were pseudonymised and stored in the Sint Maartenskliniek or the Radboudumc biobank for collection. A drug-tolerant competitive enzyme-linked immunosorbent assay (Sanquin, the Netherlands) was used to quantify ADA, enabling measurement of ADA in the presence of large amounts of TNF-inhibitor. In short, a high affinity adalimumab mutant (variant cb1-3, murine origin[14]) was used, which can efficiently remove the TNFi from TNF due to increased affinity.

Thereafter, the adalimumab concentration was determined via an ELISA. Concentrations <0.004 µg/ml were deemed not detectable. Concentrations <5 µg/ml were considered as not effective. ADA were quantified with the antigen binding test (RIA). The reference value for this test was >12 AU/ml.

Testing was performed by Sanquin, The Netherlands. The treating physician (who was responsible for the choice of subsequent bDMARD) was blinded to test results as the sample analysis has been done retrospectively.

Assessment of clinical outcome

The primary outcome of this study was the association between ADA or ADL and EULAR good response to the bDMARD after adalimumab failure ('EULAR response'). Response was operationalized as EULAR good response to the subsequent bDMARD after adalimumab failure, measured between 3-6 months (+/- 8 weeks) after start of the next and subsequent bDMARDs based on the Disease Activity Score in 28 joints (DAS28-CRP/DAS28-ESR), which is a valid, reliable and broadly accepted indicator of the clinical activity of rheumatoid arthritis.[15,16] When DAS28 response was unavailable / if glucocorticoid injection could have influenced the DAS28 score outcome, clinical assessment by a rheumatologist was used to assess response ('clinical response'). When DAS28 was already low at baseline and remained low in follow-up, clinical response assessment by a rheumatologist was also used. Of note, both

DAS28-ESR-scores and DAS28-CRP scores were used during the study period, and slightly different cut-offs for response were used to consistently assess response.[17]

Finally, a sub-analysis was performed for primary and secondary failure on adalimumab. Non-response is classified as primary non-response if adalimumab is used for less than 6 months, and as secondary non-response if adalimumab is used for longer than 6 months.

Statistical analyses

Data management systems Castor EDC and Microsoft powerBI database were used to enter and store the data. Data was extracted to a STATA database and analysed (version 13.1).

Descriptive statistics are provided with mean (+/– standard deviation), median (interquartile ranges (p25-p75)) or n (%) depending on data distribution. Baseline characteristics of the TNFi vs non-TNFi as second treatment groups were compared using a Student's t-test (or, if not normally distributed, Wilcoxon rank sum) and χ^2 test for continuous and categorical data, respectively.

Correlations between ADA presence and clinical variables (i.e. age, gender, smoking, disease duration, rheumatoid factor (RF), anti-citrullinated protein antibodies (ACPA), DAS28-CRP/ESR and its' components, C-Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR)) were first cross-sectionally explored by Spearman correlation analysis.

Area Under the Receiver Operating Characteristic (AUROC) curves were generated to evaluate the predictive value of ADA presence and adalimumab concentrations for EULAR clinical response in respectively TNFi and non-TNFi as consecutive treatment. Sensitivity and specificity were calculated using the cut-offs suggested by earlier studies (ADL<5mg/L and ADA12AU/ml[18]), and precision is shown with a 95% confidence interval. A p-value <0.05 was considered statistically significant.

Results

Participants

137 patients were included (*figure 1*), 47 of whom switched to a second TNFi and 90 to a bDMARD with another mode of action. ADA were measured in all patients and ADL) were measured in 95 patients due to timing of serum samples.

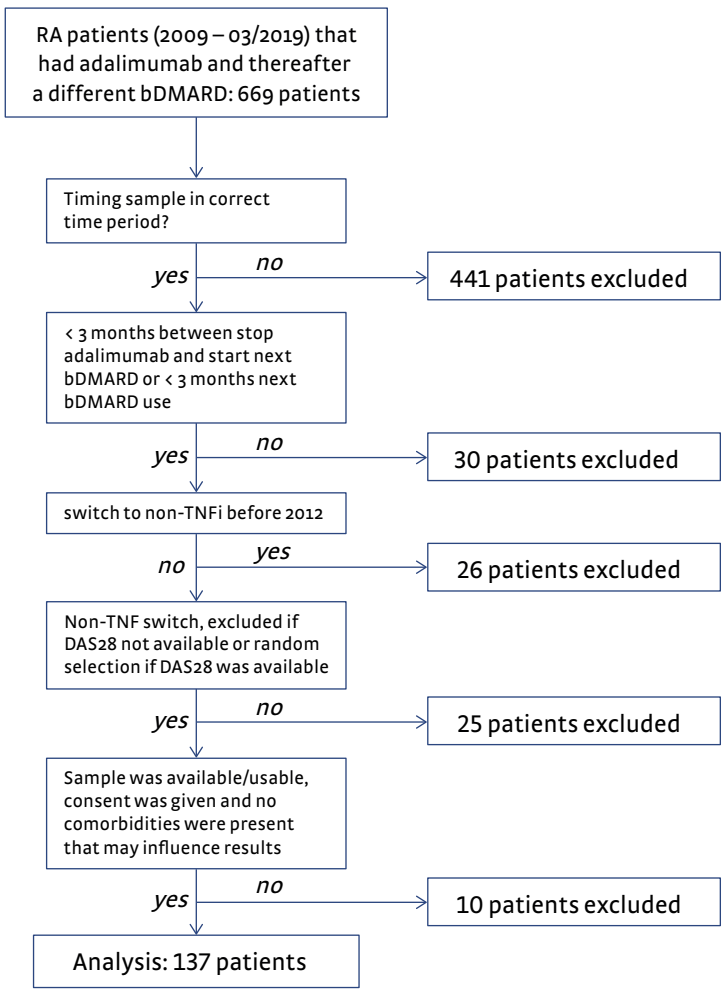

Figure 1. *Flow of participants.*

Baseline characteristics and group differences are shown in *table 1*. In patients receiving a second TNFi, 36% achieved good EULAR clinical response, while 23.4% achieved good EULAR clinical response in the non-TNFi group.

Table 1. baseline values and differences between groups.

	All patients (137)	TNFi switchers (47)	Non-TNFi switchers (90)	Difference between groups
Demographics				
age (mean ± SD)	64.4 ± 13.2	64.7 ± 12.9	64.2 ± 13.4	p= 0.83
female (n%)	94 (68.6)	30 (63.8)	64 (71.1)	p= 0.38
adalimumab levels measured (n %)	95 (67.4)	38 (27)	57 (40.4)	p= 0.01
Concomitant treatments at baseline (n%)				
csDMARDS (any)	105 (76.6)	33 (70.2)	72 (80)	χ^2= 1.65, p= 0.20
none	32 (23.4)	14 (29.8)	18 (20)	
csDMARD (azathioprine)	20 (14.6)	4 (8.5)	16 (17.8)	
csDMARD (methotrexate)	60 (43.8)	21 (44.7)	39 (43.3)	
csDMARD (leflunomide)	23 (16.8)	5 (10.7)	18 (20)	
csDMARD (hydroxychloroquine)	13 (9.5)	3 (6.4)	10 (11.1)	
Glucocorticoid oral (prednisone/ prednisolone)	24 (17.5)	10 (21.3)	14 (15.6)	χ^2= 0.70 p= 0.40
none	113 (82.5)	37 (78.7)	76 (84.4)	
bDMARD treatments				
nr. bDMARD previous to adalimumab (mean ± SD)	0.8 (0 - 4)	0.32 (0 - 2)	1,1 (0 - 4)	
Time until start bDMARD after adalimumab (days) (mean (IQR))	26.71 (24)	26.1 (24)	26.9 (25)	p=0.36
Duration of adalimumab use (years) (mean IQR)	2.2 (3.2)	3.6 (4.7)	1.3 (0.85	
Difference (days) stop adalimumab and date serum sample (mean IQR)	23.08 (27)	40.7 (81)	13.9 (24)	
Disease status				
disease duration (years until sample) (median (IQR))	11.4 (12.7)	11.5 (11.4)	11.6 (14.1)	p=0.76
Rheumatoid factor positive (n%)	96 (70.1)	35 (74.5)	61 (67.7)	p=0.36
anti-CCP positive (n%)	83 (60.6)	27 (57.4)	56 (62.2)	p=0.79

Correlations between ADA/ADL and patient characteristics

ADL showed a negative correlation with baseline DAS28 (Spearman's ρ = -0.68, *p* = 0.00). However, ADA presence did not correlate significantly with baseline DAS28 (ρ = 0.23, *p* = 0.28) and both ADA and ADL did not correlate with follow-up DAS28 (ρ = -0.29, *p* = 0.17, and ρ = 0.10, p = 0.65 respectively).

ADA correlates with baseline ESR (ρ = 0.49, *p* = 0.01) and ADL with baseline CRP (ρ = – 0.67, *p* = 0.00) and ESR (ρ = – 0.546, *p* = 0.006).

Predictive value of ADA and ADL

No clear predictive value of ADA could be found in either TNFi or non-TNFi groups (*figure 2*). In the TNFi switchers, a sensitivity of 18% and specificity of 75% were found for presence of ADA predicting EULAR good response, with an AUROC value 0.46 (95% confidence interval (CI) = 0.32 to 0.59). For non-TNFi switchers, a sensitivity of 33% and specificity of 70% were found and the AUROC value was 0.52 (95% CI = 0.42 to 0.63).

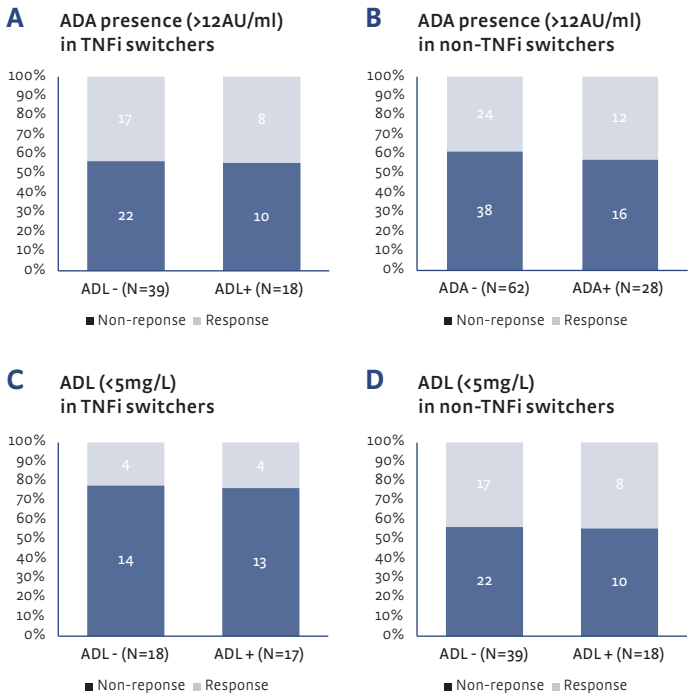

Figure 2. *Response and ADA presence in TNFi switchers (A) and non-TNFi switchers (B). Adalimumab levels <5mg/L in TNFi switchers (C) and non-TNFi switchers (D). AUROC of ADA in TNFi switchers (E) and non-TNFi switchers (F). AUROC of ADL in TNFi switchers (G) and non-TNFi switchers (H).*

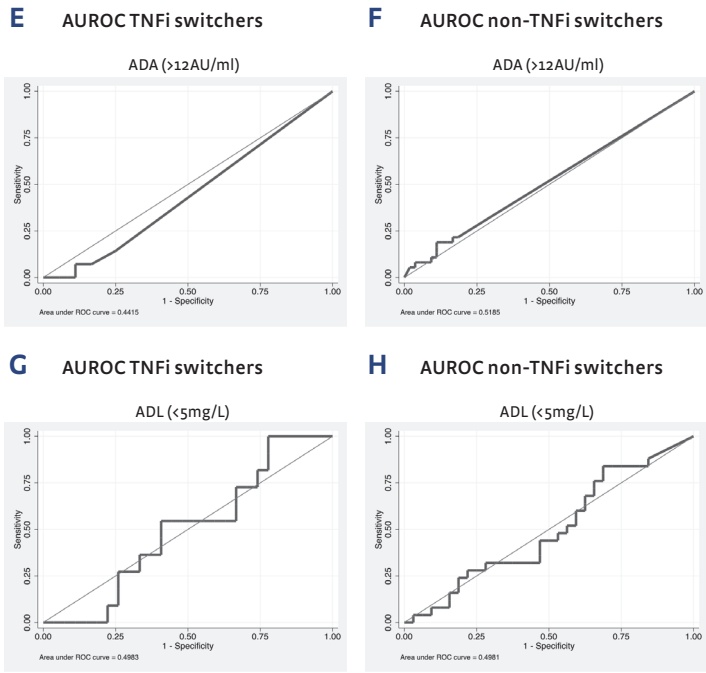

Figure 2. *Continued.*

Additionally, in respect to ADL levels no predictive value was observed in the TNFi or non-TNFi group. In the TNFi switchers a sensitivity of 32% and specificity of 69% were found for ADL predicting EULAR clinical response, with an AUROC value of 0.50 (95% CI = 0.29 to 0.71), whereas in the non-TNFi switchers a sensitivity of 50% and specificity of 52% were found, with an AUROC value of 0.50 (95% CI = 0.34 to 0.65).

Secondary outcomes

ROC analysis was conducted for patients with primary and secondary non-response as a mechanistic difference was expected between these groups. There were 74 patients with primary failure of which 10 had switched to a TNFi and 64 had switched to a non-TNFi. There were 63 patients with secondary failure of which 37 had switched to a non-TNFi and 26 had switched to a TNFi. Clinical response was significantly lower in the secondary failures than in the primary failures (23,8% vs 43,2% respectively, P=0.02). This lower rate in response is mostly seen in the non-TNFi switchers of the secondary failures (19.2% non-TNFI vs 58.7% TNFi). Additionally, there was no significant difference in ADA presence (29.7% vs 27.0%, p=0.850) or drug levels (36.0% vs 40.5% p=0.673) between the primary and secondary non-response groups.

ADA and ADL also did not show predictive value for response to either a second TNFi or a non-TNFi in sub-analyses restricted to primary or secondary non-responders specifically (*table 2*).

Table 2. Predictive values of ADA and ADL for primary and secondary non-responders in TNFi and non-TNFi switchers.

	sensitivity (%)	specificity (%)	AUC	CI
primary non-responders		***TNFi switchers***		
ADA presence (>12AU/mL)	0	44	0.28	0.11-0.45
low ADL (<5mg/L)	0	50	0.56	0.14-0.97
		non-TNFi switchers		
ADA presence (>12AU/mL)	29	27	0.51	0.40-0.63
low ADL (<5mg/L)	54	47	0.49	0.31-0.67
secondary non-responders		***TNFi switchers***		
ADA presence (>12AU/mL)	20	18.5	0.49	0.35-0.63
low ADL (<5mg/L)	57.2	47.6	0.47	0.24-0.70
		non-TNFi switchers		
ADA presence (>12AU/mL)	60	33.3	0.61	0.34-0.89
low ADL (<5mg/L)	0	30	0.48	0.13-0.82

Discussion

In this diagnostic test accuracy study no predictive value for response to a second (non) TNFi was found for either ADA or random timed ADL. Secondary failure had a lower clinical response rate. This difference was mostly seen in the non-TNFi switchers. These results were expected as the secondary failures did have an effect of the TNFi treatment at the start of the treatment.

In contrast to other studies, results of this study showed no predictive values. This is due to the fact that sensitivity and specificity was assessed instead of mean DAS values. There were, however, some significant correlations found as previously reported in other studies. Not only did the results of this study show no predictive values, in some subjects a prediction is found in the opposite direction of what was expected. The AUROC-values were all ±0,5 which shows that this is not due to lack of power.

This study has several strengths: Firstly, the choice of treatment and outcome assessment were blinded for ADA/ADL as these had not been determined at time of treatment. Secondly, a larger patient sample was achieved than in previous studies, except for the 2019 L'Ami study.[19] Thirdly, there was solely focused on adalimumab. Fourthly, selection bias is unlikely because the inclusion criteria for the several studies in which patients were included when their sample was drawn were very broad. Finally, a control group was included with patients that switched to a non-TNFi treatment, to assess whether any predictive value of ADA/ADL was different for a second TNFi vs a non-TNFi bDMARD, as otherwise ADA/ADL might simply predict a more severe disease phenotype instead of differing chances of response to TNFi vs non-TNFi.

However, several limitations of this study should also be addressed. Firstly, the samples were not taken at through level but rather timed at random related to adalimumab injection. This might have reduced the association between (anti)drug levels and response. However, it should be noted that random timed drug levels, and moreover ADA, are strongly correlated with trough level sampling.[20] So this should not have resulted in absence of any predictive value. In addition, random timed drug sampling is more feasible in clinical practice, thus increasing generalizability.

Secondly, as this was a retrospective study, both serum samples and clinical outcomes were not always available, and this might have resulted in selection bias. However, it seems hard to conceive how this selection of patients with measurements would be biased with regards to the predictive value of ADA/ADL and could have resulted in a false negative study.

Third, misclassification of the outcome can occur, both by incorrectly classifying patients as responders (e.g. by glucocorticoid injections resulting in spuriously low DAS28) or incorrectly classifying as non-responder (e.g. a patient starting with a low baseline DAS28 that remains low during treatment). To correct this misclassification, the physician judgement of response was also assessed in a sensitivity analysis, and this did not lead to a different result.

Further research should be done to confirm if ADA presence and ADL are indeed not predictive for disease activity. This could be done in a prospective study with a large sample size in which DAS28 measurements and sample collection are done on the correct timepoints in all patients. A randomized trial to address these is currently evaluating whether a switching strategy based on ADL is superior to usual care switching in RA patients failing adalimumab treatment.

Acknowledgements: The authors would like to thank all the patients who were willing to participate and Thea van Gaalen for helping with the collection of samples.

Contributors: EU, NdB and AdB were involved in the conception and design of the study. EU, NdB and AB were involved in the data collection. EU, NdB and AdB contributed to the data analysis. EU drafted the manuscript and all co-authors reviewed the manuscript critically and gave final approval for its submission.

Funding: This study was not supported by any external funding. The laboratory analyses of adalimumab and antibodies and serum levels (ADA and ADL) were performed in the laboratory of Sanquin, Amsterdam.

Competing interests: None of the authors have any competing interests for the submitted work.

Patient consent: Obtained.

Ethics approval: This study has been judged to be not a WMO scope study. Therefore, Medical Ethical Review Board approval was not required (CMO: 2019-5443).

Provenance and peer review: Non-commissioned

Data sharing statement: Additional unpublished data can be obtained from the corresponding author upon reasonable request.

Supplementary table

Table S1. Absolute values of the DAS28, baseline CRP and ESR of patients with ADA and ADL.

	Patients with ADA (n=39)	Patients without ADA (n=98)
Baseline DAS (mean, SD)*	4.6 (1.2)	4.3 (1.1)
Baseline CRP (median, p25-p27)	18 (5-34)	7 (1-27)
Baseline ESR (median, p25-p27)	41 (25-63)	25 (9-38)
	Patients with ADL >5μg/ml (n=35)	**Patients without ADL >5μg/ml (n=57)**
Baseline DAS (mean, SD)*	4.2 (1.0)	4.7 (0.86)
Baseline CRP (median, p25-p27)	12 (1-19)	24 (2-40)
Baseline ESR (median, p25-p27)	41 (11-25)	45 (35-69)

*consists of both DAS28(CRP) and DAS28(ESR), which were combined due to the low number of DAS28(ESR) (n=13/137)

References

1 Hetland ML, Christensen IJ, Tarp U, *et al.* Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: Results from eight years of surveillance of clinical practice in the nationwide Danish . *Arthritis Rheum* 2010;**62**:22–32. doi:10.1002/art.27227

2 Smolen JS, Landewé R, Bijlsma J, *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;**76**:960–77. doi:10.1136/annrheumdis-2016-210715

3 Bartelds GM, Wijbrandts CA, Nurmohamed MT, *et al.* Anti-infliximab and anti-adalimumab antibodies in relation to response to adalimumab in infliximab switchers and anti-tumour necrosis factor naive patients: A cohort study. *Ann Rheum Dis* 2010;**69**:817–21. doi:10.1136/ard.2009.112847

4 Nam JL, Takase-Minegishi K, Ramiro S, *et al.* Efficacy of biological disease-modifying antirheumatic drugs: A systematic literature review informing the 2016 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis* 2017;**76**:1108–13. doi:10.1136/annrheumdis-2016-210713

5 Cuppen BVJ, Welsing PMJ, Sprengers JJ, *et al.* Personalized biological treatment for rheumatoid arthritis: A systematic review with a focus on clinical applicability. *Rheumatol (United Kingdom)* 2016;**55**:826–39. doi:10.1093/rheumatology/kev421

6 Van Herwaarden N, Van Den Bemt BJF, Wientjes MHM, *et al.* Clinical utility of therapeutic drug monitoring in biological disease modifying anti-rheumatic drug treatment of rheumatic disorders: a systematic narrative review. *Expert Opin Drug Metab Toxicol* 2017;**13**:843–57. doi:10.1080/17425255.2017.1353602

7 Jamnitski A, Bartelds GM, Nurmohamed MT, *et al.* The presence or absence of antibodies to infliximab or adalimumab determines the outcome of switching to etanercept. *Ann Rheum Dis* 2011;**70**:284–8. doi:10.1136/ard.2010.135111

8 Plasencia C, Pascual-Salcedo D, García-Carazo S, *et al.* The immunogenicity to the first anti-TNF therapy determines the outcome of switching to a second anti-TNF therapy in spondyloarthritis patients. *Arthritis Res Ther* 2013;**15**:12–6. doi:10.1186/ar4258

9 l'Ami MJ, Ruwaard J, Krieckaert CLM, *et al.* Serum drug concentrations to optimize switching from adalimumab to etanercept in rheumatoid arthritis. *Scand J Rheumatol* 2019;:1–5. doi:10.1080/03009742.2019.1577915

10 Aletaha D, Neogi T, Silman AJ, *et al.* 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;**62**:2569–81. doi:10.1136/ard.2010.138461corr1

11 van den Bemt BJF, Vos K, Broeder AA d., *et al.* A single course of rituximab does not abrogate anti-infliximab antibodies in patients with rheumatoid arthritis. *Ann Rheum Dis* 2009;**68**:1368–9. doi:10.1136/ard.2008.095448

12 Kievit W, Fransen J, de Waal Malefijt MC, *et al.* Treatment changes and improved outcomes in RA: an overview of a large inception cohort from 1989 to 2009. *Rheumatology* 2013;**52**:1500–8. doi:10.1093/rheumatology/ket166

13 Tweehuysen L, den Broeder AA, Schraa K, *et al.* Predictive value of ex-vivo drug-inhibited cytokine production for clinical response to biologic DMARD therapy in rheumatoid arthritis. *Clin Exp Rheumatol* 2019;**37**:367–72.

14 Votsmeier C, Scheidig A, Strerath M, *et al.* Femtomolar Fab binding affinities to a protein target by alternative CDR residue co-optimization strategies without phage or cell surface display. *MAbs* 2012;**4**:341–8.

15 Prevoo MLL, Van'T Hof M, Kuper HH, *et al.* Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum Off J Am Coll Rheumatol* 1995;**38**:44–8.

16 Van Gestel AM, Prevoo MLL, Van't Hof MA, *et al.* Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism cri. *Arthritis Rheum Off J Am Coll Rheumatol* 1996;**39**:34–40.

17 Fleischmann RM, Van Der Heijde D, Gardiner P V., *et al.* DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. *RMD Open* 2017;**3**:e000382. doi:10.1136/rmdopen-2016-000382

18 Vogelzang EH, Kneepkens EL, Nurmohamed MT, *et al.* Anti-adalimumab antibodies and adalimumab concentrations in psoriatic arthritis; an association with disease activity at 28 and 52 weeks of follow-up. *Ann Rheum Dis* 2014;**73**:2178–82. doi:10.1136/annrheumdis-2014-205554

19 Berkhout LC, L'Ami MJ, Ruwaard J, *et al.* SAT0189 Dynamics of circulating tnf during adalimumab treatment of rheumatoid arthritis using a novel drug-tolerant tnf assay. *Sci Transl Med* | 2019;**11**:eaat3356. doi:10.1136/annrheumdis-2018-eular.3749

20 Hooijberg F, L'ami, Merel J. LC, Berkhout SA, *et al.* TROUGH VERSUS NON-TROUGH ADALIMUMAB DRUG LEVEL MEASUREMENTS. *Dutch Soc Rheumatol*

Chapter 4

The added value of predictive biomarkers in treat-to-target strategies for rheumatoid arthritis patients: a conceptual modelling study

Maike H. M. Wientjes
Evy Ulijn
Wietske Kievit
Robert B. M. Landewé
Inger Meek
Nathan den Broeder
Noortje van Herwaarden
Bart J. F. van den Bemt
Lise M. Verhoef
Alfons A. den Broeder

Published in Rheumatology, July 2020

Abstract

Objectives: To quantify the additional value of a hypothetical biomarker predicting response to treatment for RA regarding efficacy and costs by using a modelling design.

Methods: A Markov model was built comparing a usual care T2T strategy with a biomarker-steered strategy for RA patients starting biologic therapy. Outcome measures include time spent in remission or low disease activity (LDA) and costs. Four additional scenario analyses were performed by varying biomarker or clinical care characteristics: (i) costs of the biomarker; (ii) sensitivity and specificity of the biomarker; (iii) proportion of eligible patients tapering; and (iv) medication costs.

Results: In the base model, patients spent 2.9 months extra in LDA or remission in the biomarker strategy compared with usual care T2T over 48 months. Total costs were €43 301 and €42 568 for, respectively, the usual care and biomarker strategy, and treatment costs accounted for 91% of total costs in both scenarios. Cost savings were driven due to patients in the biomarker strategy experiencing remission or LDA earlier, and starting tapering sooner. Cost-effectiveness was not so much driven by costs or test characteristics of the biomarker (scenario 1/2), but rather by the level of early and proactive tapering and drug costs (scenarios 3/4).

Conclusions: The use of a biomarker for prediction of response to b/tsDMARD treatment in RA can be of added value to current treat-to-target clinical care. However, gains in efficacy are modest and cost gains are depending on a combination of early proactive tapering and high medication costs.

Keywords: RA, modelling, biomarker, DMARD, prediction, treat-to-target.

Key messages

- Biomarkers for prediction of response to treatment can be cost-effective in RA T2T clinical care.
- Decrease in medication costs or less/later tapering reduces the benefits of the biomarker
- Efficacy of the biomarker remains depending on rather optimistic assumptions about the test

Introduction

Current treatment strategies for RA follow a trial-and-error treat-to-target (T2T) approach. T2T includes measuring disease activity, setting a target and adjusting treatment accordingly until the goal is reached. After failure of (a combination of) conventional synthetic DMARDs (csDMARDs), current guidelines advise to give a biological (b) or a targeted synthetic (ts) DMARD [1, 2]. Efficacy of available b/tsDMARDs is approximately similar on a group level, with response rates of between 50 and 60% [3]. Medication is adjusted until the patient responds and the target [often low disease activity (LDA) or remission] is achieved.

Prediction of response could conceptually improve this T2T treatment strategy, as this would enable us to offer the more effective treatment options sooner in the treatment process. Timely initiation of more effective treatment should lead to better short- and long-term outcomes as the treatment target is reached earlier, flaring of the disease is prevented and longterm joint damage is minimized [4]. Additionally, fewer patients with active disease and fewer disease flares can lead to cost savings, as fewer (extra) visits to a rheumatologist, additional diagnostic tests and escape medication (e.g. corticosteroids or NSAIDs) are needed. From a societal point of view, this is also relevant, as improved disease control might lead to higher work participation.

Potential biomarkers that predict treatment response in RA have been extensively studied [5, 6]. Unfortunately, currently no clinically useful (bio)markers are available that are able to guide tailored treatment. RF positivity and ACPA confer a slight added positive predictive value to treatment success with rituximab (between 1.9 and 8.9%, and 1.1 and 7.5%, respectively), while a polymorphism in the TNF-a promotor adds between 1.3 and 8.9% to the predictive value of response to TNF-inhibitor therapy [5]. However, their limited predictive value does not support use in clinical practice. Therefore, current research into genetic variants, differential gene expression, proteomics and clinical and demographic factors as predictors of response to RA treatment is ongoing [7].

Considering the amount of research that is done into predictive biomarkers, it is of interest to investigate what characteristics (e.g. costs, sensitivity/specificity) a biomarker should have in order to have a meaningful impact on treatment and outcomes. In the current paper we aim to approach prediction-based T2T strategies in RA by modelling. Modelling allows us to investigate the effects of a biomarker on the course of RA treatment, even if such a biomarker is currently not available (i.e. a hypothetical biomarker) [8, 9]. Data on T2T strategies are widely available, and by

using computer simulations based on these data it is possible to provide insight into several 'what if' scenarios. It can be investigated to which extent characteristics of a hypothetical biomarker influence clinical outcomes like time to/in remission or low disease activity, but also costs. For example, sensitivity and specificity of the biomarker and the costs related to measuring the biomarker can be varied. Also, the influence of using biomarker response prediction on several treatment possibilities in clinical care, including tapering of b/ts DMARDs, can be investigated. These scenarios can inform policy makers, but also inform researchers for future research.

We aim to quantify the additional value of a hypothetical biomarker predicting response to b/tsDMARD treatment to current clinical T2T care for RA in terms of time spent in remission/LDA and costs. Additionally, we investigate the influence of biomarker characteristics (sensitivity/specificity and costs), adherence to tapering strategies and medication costs.

Methods

Model structure

We used a Markov model to assess the differences in terms of time in remission/LDA and costs between a usual care T2T strategy and a biomarker enhanced T2T strategy for RA patients failing csDMARDs and starting biologic therapy. In a Markov model, hypothetical individuals reside in predefined health states and can move from one health state to another at fixed time intervals (cycles). We defined the health states based on the 28-joint disease activity score based on erythrocyte sedimentation rate (DAS28-ESR) as 'high disease activity (HDA, DAS28-ESR>5.1)', 'moderate disease activity (MDA, DAS28-ESR>3.2 and 5.1)', 'low disease activity (LDA, DAS28-ESR >2.6 and 3.2)' and 'remission (DAS28ESR 2.6)'. A graphical representation of the model is shown in Fig. 1. Each cycle, hypothetical RA patients transitioned to any other state or stayed in the same state. Patients that moved towards HDA or MDA switched treatment, while patients moving to remission stayed on the same treatment. For patients transitioning to LDA, it depends on their previous state whether or not treatment is switched. If they moved from remission to LDA, their treatment was switched, but if they moved from HDA or MDA to LDA they remained on the same treatment. Additionally, if patients experienced remission or low disease activity for two cycles in a row (equalling a time period of 8months), they moved to 'sustained LDA' or 'sustained remission' and remained in that state for the rest of the time horizon. For patients in the 'sustained remission' and 'sustained LDA' states, we assumed that disease activity guided tapering (defined as reaching a lower than standard dose or discontinuation with maintenance of low disease activity) was

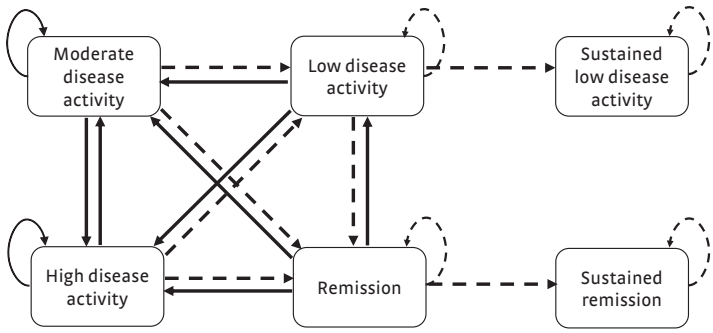

Figure 1. *Graphical representation of the Markov model, in which hypothetical RA patients reside in pre-defined health states and can move from one health state to another, or remain in the same health state, at fixed time intervals. Arrows indicate a patient's transition from one health state to another. For patients transitioning according to a dashed line, it was assumed that the patient would, as a result of the transition, continue current treatment. Likewise, for solid lines it was assumed that treatment was escalated. Curved lines indicate that the patient remains in the same health state*

performed in 60% of the eligible patients, leading to a reduction in treatment dose and thereby a reduction of average treatment costs of 50% for these patients. The time horizon of the model was four years (12 cycles) in order to include long-term effects like tapering of treatment. Cycle duration was set at four months from a clinical point of view, as patients visit the clinic with a mean of three visits per year. Death was not included in the model, as mortality differences between usual care and the biomarker strategy over four years were not expected.

Input parameters

In the model, a hypothetical cohort of 1000 RA patients was followed that started biological treatment. These patients had active disease, with 500 patients starting in the HDA state and 500 patients in the MDA state. Probabilities to respond to DMARD treatment (transition probabilities) are based on existing cohort data based on routine clinical practice, the BIO-TOP (n¼400) and the Nijmegen inception cohort of early rheumatoid arthritis (n¼1100) (baseline characteristics can be found in the original publications, only patients were selected in whom bDMARDs were started) [10-12]. Transition probabilities were calculated by evaluating the number of patients moving to a specific health state, divided by the total number of patients in the health state they were moving from. Both cohorts are based on T2T routine clinical practice data, which has the benefit of capturing effects of concomitant csDMARD use and bridging with glucocorticoids. Treatment options are hypothetical (e.g. treatment 1,

2, 3, etc.) as we are not primarily interested in comparing specific treatments and because available bDMARD and tsDMARD options are considered more or less equally effective [1]. Therefore, transition probabilities were similar for each treatment option. After failure to six treatment options, patients could be classified as refractory RA and their chances of response are probably relevantly lower compared with patients starting their 1–6 treatment. The first six treatment options in RA do not include multiple within-class options with lower response chance such as a third or later TNF blocker [13], or inherently lower efficacy drugs such as anakinra [14] (a classical order would be, for example, using TNFi 1, RTX, anti-IL6, abatacept, TNFi 2, JAKi in any particular order). Therefore, we think that these transition probabilities would suit the response changes for the first six treatment. Thereafter, the treatment options will include more within the mode of action switching, and some inherently less effective treatment options will be used. This implies that from treatment option seven onwards, the chances to respond to treatments were 50% of response chances for the first six treatments (based on expert opinion). All parameters that were used are reported in Supplementary Table S1, available at Rheumatology online.

Costs

We assigned costs to each health state in the Markov model based on a previously validated model for RA [8]. These costs are divided into direct healthcare costs related to RA (i.e. rheumatologist consultations, X-rays, laboratory tests, hospital admissions and other RA medication besides the biologic) and costs for treatment (i.e. the biologic drug). After correcting for inflation over time and difference in cycle length, direct healthcare costs per cycle were e226 for remission, e435 for LDA, e658 for MDA and e886 for HDA. The price for each treatment option was estimated at e12000/ year and thus e4000 per cycle, based on currently estimated b/tsDMARD prices in Europe [15]. For patients in 'sustained remission' or 'sustained LDA', tapering of treatment was included. It is assumed that in these states, for 60% of patients', treatment is tapered and that treatment costs for these patients are reduced by 50%. Total costs have been discounted at 4% per year, respectively, as per the Dutch pharmaco-economic guidelines [16].

Biomarker strategy

The biomarker-steered strategy is an enhancement of the usual care RA T2T treatment strategy in which a hypothetical biomarker is introduced. Several assumptions were made regarding the characteristics of the biomarker and strategy, which will be explained below. It is of note that these assumptions were all chosen to maximize the benefit of the biomarker scenario while remaining realistic, thus providing the biomarker with a context with optimal effectiveness.

- Dichotomous test characteristics, with predictive value for multiple treatments
  It was assumed that the biomarker had dichotomous test characteristics, i.e. each patient will have either Profile A or Profile B. Both Profile A and B refer to specific efficacy of a set of multiple treatments, and were assumed to be present in 50% of the population, as a more skewed profile prevalence would result in a lower number of cases in which treatment decision can be meaningfully altered post-test, resulting in a higher number needed to diagnose.

- Differential prediction is present
  Differential prediction entails that a positive biomarker test (for example, Profile A) results in a higher chance of response for three specific treatments, but also unchanged or a lower chance of response for the three other treatment options of Profile B. Differential predictive value is needed, because when a predictor predicts equally higher or lower response chances for all treatment options, it does not have clinical relevance for clinical decision making. Therefore, we assumed that patients identified as Profile A would have higher response rates to three of the first six available treatments (treatments 1-2-3) and decreased response rates to the other half of available treatments (treatments 4-5-6), to maintain an equal total response rate. Patients identified as having Profile B would have the inverse chances of Profile A, meaning an increased chance of response to treatment (4-5-6) and a decreased chance of response to treatment (1-2-3). Logic also dictates that prediction can never increase overall response chances in a T2T setting in which all treatment options can be tried within the timeframe.

- Sensitivity and specificity are ambitious but reasonable. Base sensitivity and specificity were set at 70% each, as an ambitious but realistic estimate based on what would be achievable with currently available biomarkers [5]. Because pre-test chances were set at 50% (see Dichotomous test characteristics, with predictive value for multiple treatments), positive and negative predictive values (PPV and NPV) are both 70% as well. In the scenario analyses, we added a sensitivity analysis of a decrease and increase in both sensitivity and specificity.

- The biomarker is measurable irrespective of cotreatment, and results are available without delay. It was assumed that the biomarker is measurable in all patients, regardless of background treatment with, for example, glucocorticoids or a current (ineffective) DMARD. Additionally, test results were available without delay, as any delay in test results directly reduces the added value of the biomarker measurement by inducing the same amount of delay in time to LDA/remission.

- Perfect test-retest profile consistency
  Testing was assumed to be only needed once before the start of treatment and the result would remain the same (value as well as predictive value) within a patient over time, regardless of received treatments and course of the RA (e.g. perfect test-retest consistency). Suboptimal testretest or patient profile consistencies would induce more costs, and—when test delay is present—multiply these delays in treatment change and response.

- Reasonable pricing
  The price of the biomarker, including personnel and lab costs, was estimated at e100 per patient, this was based on prices of several available biomarkers (explanation in Supplementary Table S1, available at Rheumatology online). In the scenario analyses, we assessed the influence of a decrease and increase in biomarker costs.

Analyses

A probabilistic sensitivity analysis was performed by drawing at random from the assigned beta (transition probabilities) and gamma (costs) distributions, using Monte Carlo simulation with 10000 iterations. Based on these Monte Carlo simulations, mean values and 95% credible intervals were calculated. Validation was performed to evaluate whether the model could depict a realistic number of patients in remission at different time points. For this, the modelled clinical outcomes were compared with outcomes from the IMPROVED study as external validation set, which consisted of a different but comparable cohort of early RA patients [17]. In the base case analyses, the primary outcomes of the biomarker enhanced T2T strategy were compared with the usual care T2T strategy; these outcomes include the time spent in remission/LDA and the associated costs. Additionally, several scenario analyses were performed using the biomarker strategy. Characteristics of the biomarker as well as clinical care were varied to assess their influence on the outcomes. Included scenarios are: (i) decrease and increase in costs of the biomarker from €100 to €10 or €500; (ii) decrease and increase of the sensitivity and specificity of the biomarker from 70% to 60% or 80%; (iii) decrease and increase in the percentage of tapering of eligible patients in 'sustained remission' or 'sustained LDA' from 60% to 40% and 80%; and (iv) decrease and increase in treatment costs from 100% (€12000 per year) to 50% (€6000 per year) and 150% (€18000 per year).

Incremental costs (difference between biomarker enhanced T2T strategy and usual care T2T) for each scenario were compared with the incremental costs of the base case.

Results

Validation

First, the clinical outcomes of the observed data from the IMPROVED study were compared with our modelled data from the usual care T2T strategy (Table 1) [17]. Absolute differences between the modelled and observed proportions after 4months (1cycle) are 16% remission and 0% LDA, and after 12months (three cycles) 1% remission and 2% LDA. Therefore, the model approaches values from the IMPROVED study after 12months.

Table 1. Comparison between modelled and observed clinical outcomes.

	Observed (IMPROVED)	Modelled
	Outcomes at 4months	
Remission (%)	53%	37%
LDA (%)	19%	19%
	Outcomes at 12months	
Remission (%)	51%	52%
LDA (%)	22%	20%

The modelled outcomes reflect outcomes generated by our Markov model in which hypothetical RA patients reside in, and move between, pre-defined health states. The observed outcomes reflect data from the IMPROVED study, a clinical trial in which early RA patients were steered using T2T [14].

Base case

The results of the probabilistic base case analysis are displayed in Table 2. Both time in remission and time in LDA are slightly higher for the biomarker enhanced T2T strategy compared with usual care T2T, although these effects are non-significant. Patients have spent 2.0months extra in remission and 0.9months extra in LDA in the biomarker strategy compared with usual care T2T over the total time horizon of the model (48months). Fig. 2 shows that this difference is mainly seen at the beginning of the time horizon, as patients in the biomarker strategy experienced remission or LDA earlier in the treatment process compared with the usual care T2T strategy. After approximately two years, the outcomes of the strategies converged and the curves flattened. Earlier remission also led to an increase in the number of patients in whom treatment was tapered, as at 12months 15% of patients in the usual care T2T strategy tapered medication compared with 23% in the biomarker strategy. At 24months this difference was still present, with 39% of patients tapering treatment in the usual care

T2T strategy and 49% in the biomarker strategy. The biomarker strategy was dominant over the usual care T2T strategy with a certainty of 94%, as—next to the results for time in remission—it was also cost-saving compared with usual care T2T (Fig. 3). Costs related to treatment (i.e. the biologic drug) were the main contributor to the total costs. The mean costs related to treatment in the usual care T2T and biomarker enhanced T2T strategies over the four-year time horizon were e39334 and e38780, respectively, while total costs were e43301 and e42568, so treatment costs accounted for 91% of total costs in both scenarios.

Scenario analyses

Using a biomarker with a lower sensitivity and specificity of both 60% leads to an incremental time in remission of 1month for the biomarker enhanced T2T strategy over usual care T2T. This is a loss of time in remission of 1month compared with the base case analysis. In contrast, an increase towards 80% sensitivity and specificity gains approximately half a month in remission over the base case analyses, with an incremental time in remission of 2.4 months for the biomarker strategy over usual care T2T. The other strategies do not influence time in remission. The largest incremental cost savings are seen for scenarios 3 and 4, in which, respectively, the percentage of eligible patients initiating dose tapering and b/tsDMARD costs are increased (Fig. 4). This indicates that the influence of (the combination of) early and widespread tapering and high treatment costs on the total costs is large.

Table 2. Results probabilistic base case analysis

	Usual care T2T	Biomarker enhanced T2T strategy	Differences
Months in remission [a]	30.3 (27.4–33.2)	32.3 (29.1–35.2)	2.0 (-0.5–4.3)
Months in low disease activity [a]	7.6 (5.7–6.5)	8.5 (6.5–10.8)	0.9 (-1.3–3.1)
Total costs	€43301 (€41864–€44847)	€42568 (€41118–€44122)	€733 (-€138 – -€1371)
Costs related to treatment	€39334 (€38529–€40145)	€38780 (€37924–€39661)	€554 (-€144 – -€970)

Data are mean values (95% credible interval) of 10000 simulations per patient accrued over the complete time horizon of the model (4 years).

[a] Of 48months of follow-up.

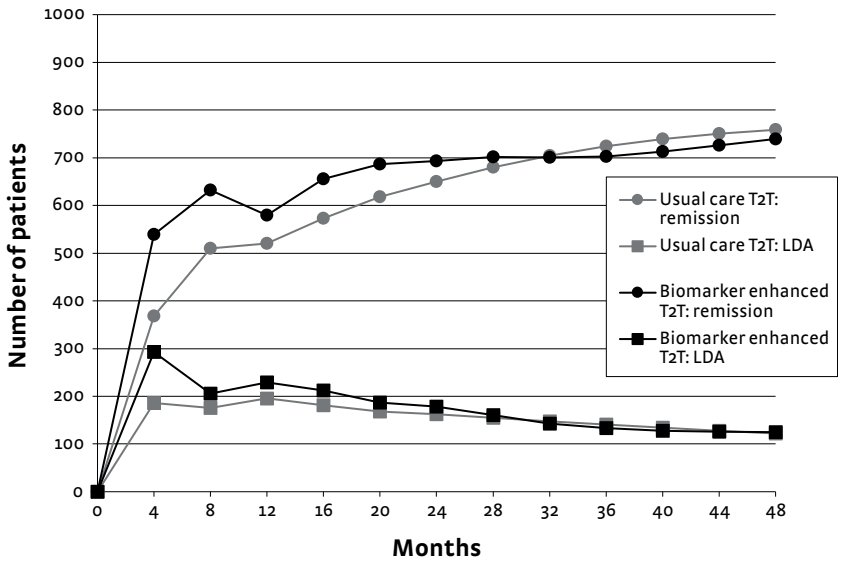

Figure 2. *Mean number of patients (over 10 000 iterations) in remission or low disease activity in the usual care T2T strategy and in the biomarker enhanced T2T strategy at each visit in the model*

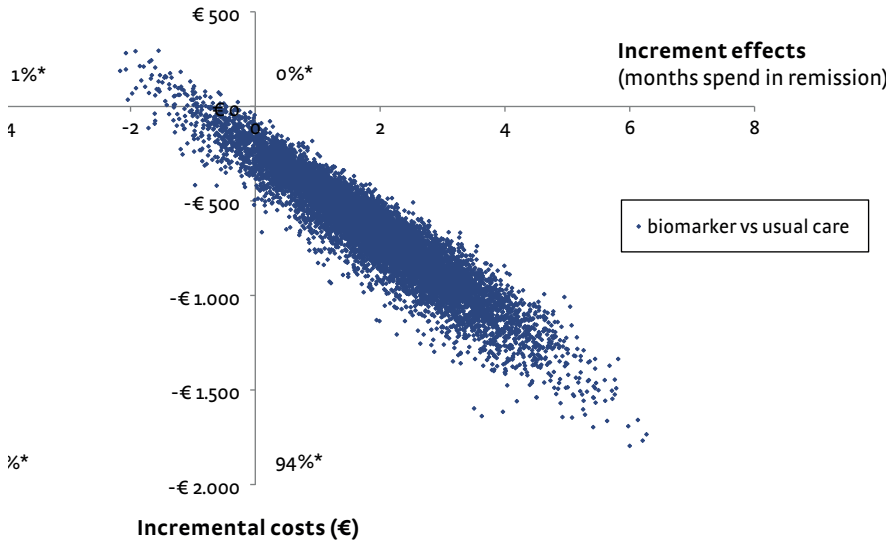

Figure 3. *Incremental cost-effectiveness plane for the biomarker enhanced T2T strategy with respect to the usual care T2T strategy. *Percentages of simulations in each quadrant of the figure*

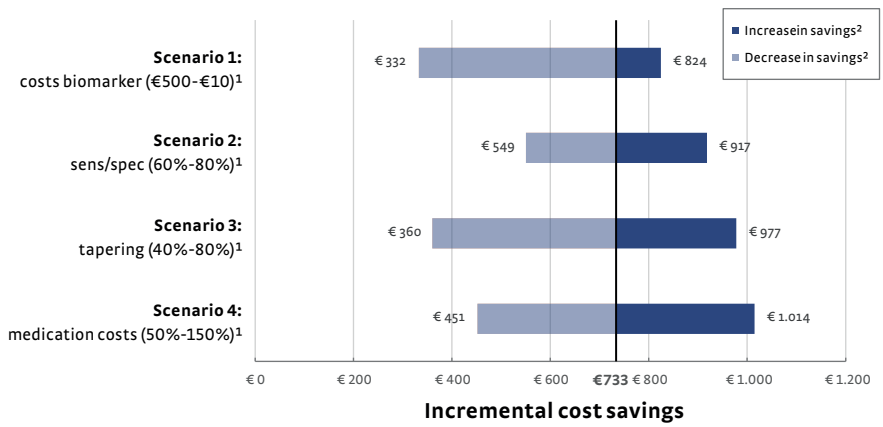

Figure 4. *Incremental cost savings for each scenario relative to the incremental costs of the base case analyses on the x-axis. [a]Scenario (value with lowest increase in savings value with highest increase in savings). Example for scenario 1: biomarker costs of e500 reduces the incremental cost savings of the biomarker enhanced T2T strategy compared to the usual care T2T strategy from e733 (base case) to e332. [b]Compared to base case (y-axis)*

Discussion

The use of a biomarker for prediction of response to b/tsDMARD treatment in RA can be of added value to a T2T RA strategy. The base biomarker enhanced T2T strategy of our model was dominant compared with usual care T2T regarding both effectiveness and costs. The dominance is mainly caused by earlier reach of the remission or LDA target for patients in the biomarker enhanced T2T strategy, causing tapering to be initiated earlier. A decrease in the percentage of patients tapered, or a decrease in medication costs reduces the benefits of the biomarker enhanced T2T strategy. However, the biomarker must meet many assumptions that otherwise directly limit their cost-effectiveness.

The maximum number of patients that is able to experience remission or low disease activity is similar for both strategies, as addition of a biomarker does not influence effectiveness of medication overall. The main difference between the biomarker enhanced T2T strategy and the usual care T2T strategy is that the treatment target is reached earlier in the treatment process for the biomarker strategy. These differences are mainly present in the first two years after start of b/tsDMARD treatment, which is comparable to previous strategy trials in RA [17, 18]. Achieving the treatment target earlier is beneficial, as a shorter period of active inflammation is associated with

increased functional performance and quality of life and a reduced risk of irreversible joint damage [18–20]. Validation analyses showed that after one cycle, differences between the model and clinical care are substantial, as patient distribution over the health states is still largely depending on initial probabilities. However, after 12months the differences are minimal and the biomarker enhanced model approaches the T2T-only model.

The biomarker strategy was based on a number of essential assumptions. Therefore, results of the biomarker strategy only apply in the rather positive scenario that was outlined. However, the question remains if this is achievable in practice, and it has been acknowledged before that the cost-effectiveness of biomarkers also depends on the context in which they are used [21]. For example, results of laboratory biomarkers are often not immediately available, leading to a delay in starting the next treatment. Our results show that a biomarker strategy leads to a maximum of only 3months extra in remission or LDA, so every week of waiting for results before starting the next treatment is directly limiting effectiveness of the biomarker. In addition, no biomarkers are currently available that show differential prediction between several RA treatment options, as most studies show predictive value for response to a specific bDMARD and unclear or similar predictive values for another bDMARD. Differential prediction entails that a result of the biomarker results in a higher chance of response for a certain treatment, but also unchanged or lower chance of response for another treatment. The 50/50 split chosen here represents the maximal uncertainty, and the number needed to test and the value per test would deteriorate with every deviation. Therefore, this results in the highest possible added value for the biomarker. If the biomarker is not able to show differential prediction to different treatment options, it does not alter clinical decision making, as prediction of a patient's response chances to most or all treatments does not provide added information for clinical practice. Additionally, we assumed that the biomarker was measurable before starting the new treatment. There are also biomarkers available that can only be measured after start of treatment; for example, drug serum levels [22, 23]. Despite the potential of these biomarkers, the most optimal scenario includes measurement before start of treatment for preventing the start of ineffective treatment. Lastly, we assumed that after failure to six treatment options the chance of response to future treatment would be lower, and we classified these patients as refractory RA. Results of our study would not change relevantly if these response chances would drop slightly more or less, as this will only impact a very limited number of patients.

Our scenario analyses showed that costs of the biomarker as well as the sensitivity and specificity of the biomarker did not seem to affect the outcomes as much. The costs of the biomarker are small compared with medication costs, and therefore do not have much influence on the outcomes. Also, an increase in the sensitivity and specificity of the biomarker from 60% to 80% adds only one and a half months to the total time in remission, illustrating the difficulty to substantially improve efficacy of the current trial-and-error strategy. On the contrary, costs of medication and the amount of tapering have a substantial impact on the total costs. The development towards lower drug prices (through biosimilars, etc.) will lower the effects of a biomarker, and will mainly be in the area of reaching treatment targets earlier.

In conclusion, there seems to be modest potential for the use of a biomarker enhanced T2T strategy compared with usual T2T in RA. Patients experience remission and LDA earlier in the treatment process, and associated costs are lower compared with usual care. However, effects are very dependent on high drug costs combined with tapering, and rather optimistic assumptions about the test. The current extensive amount of research regarding biomarkers may not be justifiable in light of these modest benefits.

Data availability
All data relevant to the study are included in the article or uploaded as supplementary information.

Contribution statement
M.H.M.W., E.U., W.K., L.MV. and A.A.dB. contributed to model design and analyses. M.H.M.W., E.U., W.K., L.MV. and A.A.dB. were responsible for drafting this manuscript. All other authors contributed to data interpretation and writing of the manuscript. Furthermore, all authors approved the final version of the manuscript.

Funding
No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary tables

Table S1. Parameters used in the model.

	Deterministic value	Standard error	Distribution
Discount rates			
Discount rate for costs	0,04	-	Fixed
Costs			
Mean treatment costs	€ 4.000	-	Fixed
Direct healthcare costs for low disease activity	€ 435	€ 61	gamma
Direct healthcare costs for moderate disease activity	€ 658	€ 161	gamma
Direct healthcare costs for high disease activity	€ 886	€ 219	gamma
Direct healthcare costs for remission	€ 226	€ 58	gamma
Costs of Biomarker	€ 100	-	Fixed
Transition probabilities usual care strategy			
Low to remission	0,574	0,056	Dirichlet
Low to low disease 2	0,191	0,049	Dirichlet
Low to moderate	0,213	0,050	Dirichlet
Low to high	0,021	0,020	Dirichlet
Moderate to remission	0,449	0,024	Dirichlet
Moderate to low disease	0,180	0,021	Dirichlet
Moderate to moderate	0,320	0,023	Dirichlet
Moderate to high	0,051	0,013	Dirichlet
High to remission	0,333	0,044	Dirichlet
High to low disease	0,153	0,037	Dirichlet
High to moderate	0,389	0,045	Dirichlet
High to high	0,125	0,035	Dirichlet
Remission to remission2	0,746	0,047	Dirichlet
Remission to low disease	0,095	0,034	Dirichlet
Remission to moderate	0,143	0,039	Dirichlet
Remission to high	0,016	0,015	Dirichlet
Low2 to remission	0,333	0,049	Dirichlet
Low2 to low sustained	0,333	0,049	Dirichlet
Low2 to moderate	0,333	0,049	Dirichlet
Low2 to high	0,000	0,000	Dirichlet

	Alpha	Beta	Source
	-	-	Dutch pharmacoeconomic guideline
	-	-	Assumed average of current accepted b/tsDMARD prices in Europe. Based on The National Health Care Institute (Zorginstituut Nederland). Price information, in Dutch (www.medicijnkosten.nl).
	51,36	8,46	Welsing 2006, inflation of 4% for 15 years
	16,68	39,45	Welsing 2006, inflation of 4% for 15 years
	16,33	54,28	Welsing 2006, inflation of 4% for 15 years
	15,37	14,69	Welsing 2006, inflation of 4% for 15 years
	-	-	Expert opinion*
	27	47	BIO-TOP and the Nijmegen inception cohort of early rheumatoid arthritis
	9	47	"
	10	47	"
	1	47	"
	115	256	"
	46	256	"
	82	256	"
	13	256	"
	24	72	"
	11	72	"
	28	72	"
	9	72	"
	47	63	"
	6	63	"
	9	63	"
	1	63	"
	19	57	"
	19	57	"
	19	57	"
	0	57	"

Table S1. Continued.

	Deterministic value	Standard error	Distribution
Remission2 to sustained remission	0,651	0,029	Dirichlet
Remission2 to low disease	0,189	0,025	Dirichlet
Remission2 to moderate	0,149	0,024	Dirichlet
Remission2 to high	0,011	0,008	Dirichlet
Transition probabilities – difficult to treat RA			
Low to remission	0,391	0,055	Dirichlet
Low to low disease 2	0,130	0,043	Dirichlet
Low to moderate	0,435	0,056	Dirichlet
Low to high	0,043	0,028	Dirichlet
Moderate to remission	0,258	0,022	Dirichlet
Moderate to low disease	0,103	0,017	Dirichlet
Moderate to moderate	0,552	0,024	Dirichlet
Moderate to high	0,087	0,016	Dirichlet
High to remission	0,164	0,038	Dirichlet
High to low disease	0,075	0,029	Dirichlet
High to moderate	0,575	0,045	Dirichlet
High to high	0,185	0,039	Dirichlet
Remission to remission2	0,566	0,048	Dirichlet
Remission to low disease	0,072	0,030	Dirichlet
Remission to moderate	0,325	0,047	Dirichlet
Remission to high	0,036	0,023	Dirichlet
Low2 to remission	0,200	0,045	Dirichlet
Low2 to low sustained	0,200	0,045	Dirichlet
Low2 to moderate	0,600	0,050	Dirichlet
Low2 to high	0,000	0,000	Dirichlet
Remission2 to sustained remission	0,494	0,029	Dirichlet
Remission2 to low disease	0,143	0,023	Dirichlet
Remission2 to moderate	0,338	0,028	Dirichlet
Remission2 to high	0,026	0,012	Dirichlet
Transition probabilities biomarker strategy – high chances			
Low to remission	0,710	0,055	Dirichlet
Low to low disease 2	0,237	0,051	Dirichlet
Low to moderate	0,048	0,030	Dirichlet
Low to high	0,005	0,010	Dirichlet

	Alpha	Beta	Source
	114	175	“
	33	175	“
	26	175	“
	2	175	“
			“
	18,39	47	“
	6,13	47	“
	20,43	47	“
	2,04	47	“
	66,01	256	“
	26,40	256	“
	141,20	256	“
	22,39	256	“
	11,84	72	“
	5,42	72	“
	41,42	72	“
	13,32	72	“
	35,67	63	“
	4,55	63	“
	20,49	63	“
	2,28	63	“
	11,40	57	“
	11,40	57	“
	34,20	57	“
	0,00	57	“
	86,36	175	“
	25,00	175	“
	59,09	175	“
	4,55	175	“
			“
	33,38	47	“
	11,13	47	“
	2,27	47	“
	0,23	47	“

Table S1. Continued.

	Deterministic value	Standard error	Distribution
Moderate to remission	0,644	0,024	Dirichlet
Moderate to low disease	0,258	0,022	Dirichlet
Moderate to moderate	0,084	0,016	Dirichlet
Moderate to high	0,013	0,007	Dirichlet
High to remission	0,574	0,045	Dirichlet
High to low disease	0,263	0,042	Dirichlet
High to moderate	0,123	0,035	Dirichlet
High to high	0,040	0,022	Dirichlet
Transition probabilities biomarker strategy – low chances			
Low to remission	0,282	0,053	Dirichlet
Low to low disease 2	0,094	0,039	Dirichlet
Low to moderate	0,568	0,056	Dirichlet
Low to high	0,057	0,032	Dirichlet
Moderate to remission	0,170	0,020	Dirichlet
Moderate to low disease	0,068	0,015	Dirichlet
Moderate to moderate	0,658	0,024	Dirichlet
Moderate to high	0,104	0,017	Dirichlet
High to remission	0,102	0,032	Dirichlet
High to low disease	0,047	0,024	Dirichlet
High to moderate	0,645	0,045	Dirichlet
High to high	0,207	0,040	Dirichlet

*Price of the biomarker was based on expert opinion and known pricing of different biomarkers. E.g. CRP €4,- (based on https://www.cz.nl/service-en-contact/vind-en-vergelijk-de-kosten-van-uw-behandeling), MBDA Vectra marker €1000,- (https://www.labcorp.com/tests/504965/vectra), drug levels / anti-drug antibody testing €90 (based on https://www.sanquin.org/products-and-services/diagnostics/index), PRISM RA €75 out of pocket, but more expensive for health insurers (https://www.prismra.com/).

	Alpha	Beta	Source
	164,98	256	“
	65,99	256	“
	21,61	256	“
	3,43	256	“
	41,34	72	“
	18,95	72	“
	8,86	72	“
	2,85	72	“
			“
	13,23	47	“
	4,41	47	“
	26,69	47	“
	2,67	47	“
	43,41	256	“
	17,36	256	“
	168,51	256	“
	26,72	256	“
	7,31	72	“
	3,35	72	“
	46,42	72	“
	14,92	72	“

References

1. Smolen JS, Landewe´ RBM, Bijlsma JWJ et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis 2020;79:685–99.
2. Fraenkel L, Bathon JM, England BR et al. 2021 American college of rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res 2021;73:924–39.
3. Kerschbaumer A, Sepriano A, Smolen JS et al. Efficacy of pharmacological treatment in rheumatoid arthritis: a systematic literature research informing the 2019 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis 2020;79:744–59.
4. Combe B, Landewe R, Daien CI et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017;76:948–59.
5. Cuppen BVJ, Welsing PMJ, Sprengers JJ et al. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology 2016;55: 826–39.
6. Tweehuysen L, van den Ende CH, Beeren FM et al. Little evidence for usefulness of biomarkers for predicting successful dose reduction or discontinuation of a biologic agent in rheumatoid arthritis: a systematic review. Arthritis Rheumatol 2017;69:301–8.
7. Wu X, Sheng X, Sheng R, Lu H, Xu H. Genetic and clinical markers for predicting treatment responsiveness in rheumatoid arthritis. Front Med 2019;13:411–9.
8. Welsing PM, Severens JL, Hartman M et al. The initial validation of a Markov model for the economic evaluation of (new) treatments for rheumatoid arthritis. Pharmacoeconomics 2006;24: 1011–20.
9. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. J Health Serv Res Pol 2004;9:110–8.
10. Tweehuysen L, den Broeder N, van Herwaarden N et al. Predictive value of serum calprotectin (S100A8/A9) for clinical response after starting or tapering anti-TNF treatment in patients with rheumatoid arthritis. RMD Open 2018;4:e000654.
11. Welsing PM, van Riel PL. The Nijmegen inception cohort of early rheumatoid arthritis. J Rheumatol Suppl 2004;69:14–21.
12. Welsing PM, Fransen J, van Riel PL. Is the disease course of rheumatoid arthritis becoming milder? Time trends since 1985 in an inception cohort of early rheumatoid arthritis. Arthritis Rheum 2005;52:2616–24.
13. Karlsson JA, Kristensen LE, Kapetanovic MC et al. Treatment response to a second or third TNF-inhibitor in RA: results from the South Swedish Arthritis Treatment Group Register. Rheumatology 2008;47:507–13.
14. Mertens M, Singh JA. Anakinra for rheumatoid arthritis: a systematic review. J Rheumatol 2009;36:1118–25.
15. The National Health Care Institute (Zorginstituut Nederland). Price information, in Dutch. www.medicijnkosten.nl (21 January 2022, date last accessed).
16. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2016. https://www.zorgin stituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg (30 June 2022, date last accessed).
17. Akdemir G, Markusse IM, Bergstra SA et al. Comparison between low disease activity or DAS remission as treatment target in patients with early active rheumatoid arthritis. RMD Open 2018;4:e000649.
18. Goekoop-Ruiterman YPM, De Vries-Bouwstra JK, Allaart CF et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum 2005;52:3381–90.
19. Conaghan PG, O'Connor P, McGonagle D et al. 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.
20. Taylor PC, Steuer A, Gruber J et al. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. Arthritis Rheum 2004;50:1107–16.
21. Sluiter RL, Kievit W, van der Wilt G. Can health care be mademore affordable by the use of biomarkers? In: C Carini, M Fidock, A van Gool, eds. Handbook of biomarkers and precision medicine. Chapter 9.6, 1st edn. New York: Chapman and Hall/CRC, 2019: 596–608.

22. Verweij N, Zwezerijnen G, Ter Wee M et al. Early prediction of treatment response in rheumatoid arthritis by quantitative macrophage PET. RMD Open 2022;8:e002108.
23. Van Herwaarden N, Van Den Bemt BJF, Wientjes MHM, KramersC, Den Broeder AA. Clinical utility of therapeutic drug monitoring in biological disease modifying anti-rheumatic drug treatment of rheumatic disorders: a systematic narrative review. Expert Opin Drug Metabol Toxicol 2017;13:843–57.

Chapter 5

Extra-articular findings with FDG-PET/CT in Rheumatoid Arthritis patients: more harm than benefit

Evy Ulijn
Alfons A den Broeder
Nadine Boers
Martin Gotthardt
Chantal AM Bouman
Robert Landewé
Nathan den Broeder
Noortje van Herwaarden

Rheumatology Advances in Practice, February 2022

Abstract

Background: Whole-body Positron Emission Tomography with CT-scanning using fluorine-18 fluorodeoxyglucose (18F-FDG) is occasionally used in Rheumatoid Arthritis (RA) patients to detect arthritis. FDG-PET/CT might also detect malignancies, but the amount of incidental findings and the number of relevant malignant disease that could be missed are currently unknown.

Objective: To study the malignancy screening performance of whole-body FDG-PET/CT in longstanding RA patients with low disease activity.

Methods: FDG-PET/CT-scanning was done in the intervention arm of the Dose REduction Strategy of Subcutaneous TNF-inhibitors (DRESS) study, a randomized controlled trial on dose-tapering of biological Disease Modifying Anti-Rheumatic Drugs (bDMARDs). The reference standard was clinical diagnosis of malignancy during the 3 year follow-up of the study. Prevalence of extra-articular abnormalities, follow-up, and treatments were summarized post-hoc.

Results: 121 scans were made in 79 patients. Extra-articular abnormalities were found in 59/121 (49%) scans, resulting in additional diagnostic procedures in 21/79 (26.6%) patients. Nine patients (7.4%) were suspected of malignancy, none turned out to be malignant. Six clinical malignancies that developed during follow-up were all negative on baseline FDG-PET/CT.

Conclusion: Whole-body FDG-PET/CT-scanning used in RA patients for imaging of arthritis results in frequent incidental extra-articular findings, while some who apparently had normal scans also developed malignancies.

Keywords: Rheumatoid Arthritis, Tumor Necrosis Factor inhibitors, Biological therapy

Key messages

- Using FDG-PET/CT for assessing arthritis in longstanding RA patients with low disease activity results in a substantial number of incidental findings, while some who apparently had normal scans also developed malignancies.
- Whole-body FDG-PET/CT-scanning with musculoskeletal indication requires properly informing patients of risks and benefits.

BACKGROUND

Whole-body fluorine 18 Fluorodeoxyglucose (18F-FDG) Positron Emission Tomography scanning – often combined with low-dose CT-scanning (FDG-PET/CT) – has the ability to noninvasively detect various malignancies at potentially curable stages and is used as diagnostic tool as well as for follow-up [1–3]. Other clinical indications for FDG-PET/CT include cardiac conditions (myocardial functioning), workup of infectious and inflammatory diseases (fever of unknown origin) and neurologic conditions (epilepsy, dementia) [4].

Although FDG-PET/CT is not routinely recommended for establishing and quantifying arthritis in the context of Rheumatoid Arthritis (RA), it is occasionally used by physicians. Reasons to use FDG-PET/CT-scans are to diagnose arthritis or guide decisions on systemic therapy, as FDG uptake in affected joints may reflect disease activity [5,6]. Elzinga et al. [7] found that FDG-PET/CT of hands and wrists might be used as a predictor of therapeutic response. Partly based on these findings, the EULAR recommendations for the use of imaging techniques for the joints in the clinical management of rheumatoid arthritis state: "Inflammation seen on imaging may be more predictive of a therapeutic response than clinical features of disease activity; imaging may be used to predict response to treatment" [8].

However, no data are available on extra-articular incidental findings associated with the use of whole-body FDG-PET/CT-scans for assessment of arthritis. Although whole-body FDG-PET/CT could be used as a cancer screening tool in asymptomatic adults, there are few data on this subject. This idea has been conceptually challenged for PET and other screening modalities [9–12]. Suboptimal test characteristics, especially low specificity, might increase the likelihood of false positive or irrelevant abnormal findings, resulting in additional follow-up diagnostics/treatment and generating patient burden and costs. Likewise, suboptimal sensitivity of a test in the incorrect setting could lead to false reassurance in case of a false negative result. Procedural drawbacks of whole-body FDG-PET/CT-scanning are exposure to radiation, patient burden, use of scarce resources and costs. Nevertheless, whole-body FDG-PET/CT is often perceived as a valuable whole-body screening tool by both patients and physicians.

In the DRESS trial [13–15], we performed baseline and follow-up whole-body FDG-PET/CT to assess arthritis activity in longstanding RA patients treated with TNF-inhibitors (a class of biological DMARDs), with close clinical monitoring of the patients during a three year period. This provided an opportunity to study the cancer screening performance of whole-body FDG-PET/CT in this specific population.

Design, participants and methods

Longstanding RA patients with stable disease activity treated with subcutaneous TNFi were randomized to either stepwise tapering or continuation of their TNFi [13–15]. Baseline whole-body FDG-PET/CT-scans were performed in consenting patients in the tapering arm, to assess predictive value of subclinical PET-arthritis for risk of flaring [13]. Scanning was done with a Siemens Biograph mCT FDG-PET/CT-scanner according to the European Association of Nuclear Medicine (EANM) procedure guidelines as described by Boellaard et al. (2015) [16]. The scanning protocol, arthritis scoring system, and results on arthritis activity were reported elsewhere [17]. The scans were also read by experienced nuclear medicine specialists at the academic hospital immediately after they were performed for any unexpected extra-articular findings. At this time, a report was made of all incidental extra-articular findings conform routine clinical care. The reader was not blinded for the clinical information of the patient, and reported results on a probability scale. When necessary, the treating physician could consult the nuclear medicine specialist for further advice. One patient could show multiple abnormalities on one scan.

Patients were followed for three years, and all clinical outcomes and FDG-PET/CT related follow-up diagnostics and treatments were noted. The reader that retrospectively summarized the clinical outcomes was not blinded for the FDG-PET/CT-test result. Similar abnormalities found on both scans (in case of repeated scans) were counted as one. The DRESS study was performed at the Sint Maartenskliniek, from December 2011 to May 2014. Patients gave written informed consent. This manuscript and the clinical study did not receive any external funding.

Results

Baseline FDG-PET/CT-scans were performed in 79 patients, and in 42 patients a follow-up scan was performed at time of maximal tapering/discontinuation (between 3-18 months after baseline, depending on whether and when a flare occurred). This led to a total number of 121 FDG-PET/CT-scans.

Incidental findings

One or more abnormal results were found in 45/79 (57%) patients and on 59/121 (48.8%) scans. Extra-articular abnormal results are specified in *table 1*. Of these 59 abnormal scan results, the research physician (consulting with a nuclear medicine specialist) categorized 36 (61%) scan results as clinically insignificant and no further action was undertaken.

Table 1. Abnormalities found on FDG-PET/CT scans.

	# abnormal results found on scans (%)
No PET/CT result obtained	**3 (2.5)**
Claustrophobia	2
Moved during scan	1
No abnormalities found on any scan	**59 (48.8)**
One or more abnormalities found per scan*	**59 (48.8)**
Total	**121**
Inflammatory	**7 (5.7)**
Upper respiratory tract infection	3
Mediastinal lymphadenopathy	3
Pneumonia (known)	1
Suspected malignancy	**9 (7.4)**
breast, caecum, uterus, lymphoma, adrenal, larynx, sigmoid, pulmonary, prostate	9
Cardiovascular	**2 (1.6)**
Aneurism	2
Pulmonary	**7 (5.8)**
Nodules	6
Pleural thickening	1
Gastrointestinal	**10 (8.3)**
Gallstones	1
Esophagitis/gastritis	5
Intestinal/rectal focal lesions (non-specific)	4
Muscles/tendons	**3 (2.5)**
Bone-related	**3 (2.5)**
Fractures (known)	1
osteoarthritis / osteoporosis**	2
Hypermetabolic lymph nodes (non-specific)	**16 (13.2)**
Thyroid	**4 (3.3)**
Enlarged	1
High uptake / Metabolism (diffused)	3

* Fifteen of these abnormalities were found on the second PET/CT, the rest was found on the first scan. 11 abnormalities on the second PET/CT were the same as the one seen on the first scan, and 7 abnormalities resolved after the first scan. One scan can show multiple abnormalities, from different categories.
** suggestive image on CT

Follow-up action was undertaken for 23 (39%) abnormalities in 21 patients which could consist of referral to a specialist or reassessing and/or scheduling diagnostics directly by the treating rheumatologist (*table 2*). In 5 (6.3%) patients, the rheumatologist followed-up. In three cases (*table 2, patients 6, 12, 14)* this follow-up took place without referral to another specialist. In the first patient physical examination of the thorax and lungs took place and an X-ray was conducted. In the second patient a skin lesion was examined and in a third patient the thyroid was clinically examined and thyroid stimulating hormone (TSH) lab was performed. These tests did not result in clinically relevant abnormal findings. In two cases (*table 2, patients 5 and 21)* the rheumatologist referred the patient to another specialist. In the first patient clinical evaluation of the tonsils took place, after which the patient was reffered to an ENT-specialist. In the other patient the thyroid was clinically examined after which the patient was referred to an endocrinologist. The ENT-specialist clinically examined the thyroid after 6 and 12 months to follow-up but did not find any clinically relevant abnormalities.

For 19 (32.2%) abnormalities in 17 (21.5%) patients a consultation with a different specialist was scheduled. One patient *(table 2, patient 20)* consulted two; an Ear Nose and Throat (ENT) specialist and a urologist, but no additional diagnostics or treatments were performed.

Non-invasive treatment

In one patient *(table 2, patient 17)* an ultrasound, which was conducted after the abnormalities found on the FDG-PET/CT, found an aneurysm of 43mm just above the aortic bifurcation. Referral to a vascular surgeon resulted in advice regarding lifestyle interventions and statins. In another patient *(table 2, patient 5)*, that was referred by the rheumatologist, the throat was diffusely tender with palpation. As a tonsillectomy previously took place and due to globus sensation in the throat combined with a productive cough, the ENT-specialist prescribed antibiotics. Thereafter, the follow-up was expectant and no standard follow-up consultation was planned.

Surgical interventions

One patient *(table 2, patient 13)* with an enlarged thyroid gland was referred to internal medicine after which an ultrasound showed a non-homogenous hypervascular nodule that took up most of the enlarged left thyroid gland, with focal calcifications and cystic components. Thyroid fine needle aspiration biopsy was performed three times and were all inconclusive. A hemi-thyroidectomy was performed where the mass turned out to be a follicular adenoma. This resection was complicated by persistent recurrent laryngeal nerve paresis and hoarseness.

In another patient *(table 2, patient 15)*, irregularities were found in the cervical smear test. A subsequent ultrasound showed a myoma located in the uterus. This was followed by a myomectomy. In a third patient *(table 2, patient 9),* a colonoscopy was performed based on the FDG-PET/CT-scan results, and found two polyps in the rectosigmoid colon which were both resected. Histopathology of the two colonic polyps showed two low-grade adenoma's. A follow-up colonoscopy was planned 5 years after the polyp resection.

In a fourth patient *(table 2, patient 18)*, an emergency consultation with an ENT-specialist was planned based on the FDG-PET/CT-scan which showed a hypermetabolic process in the larynx without clinical symptoms, initially suspected to be activity of the vocal cords. After review, however, the abnormality was diagnosed as a thyroglossal cyst. Eventually a sistrunk procedure was performed to extract the cyst in the neck. Lastly, a patient *(table 2, patient 10)* with the FDG-PET/CT-scan showing a paraspinal muscle mass (level L3/L4) underwent marginal myotomy after multidisciplinary consultation. The abnormality turned out to be a benign schwannoma.

Development of malignancy

None of the 9/79 (7.4%) on PET/CT-scan suspected malignant lesions were confirmed to be or developed into a malignancy. During the three year follow-up, six clinical malignancies (bladder, penile, lymphoma, 2x melanoma and prostate) were found in six patients. None of these malignancies had been identified by the study-related whole-body FDG-PET/CT-scans (*table S1).* The malignancies were diagnosed after an interval of between 5 and 34 months, with a mean of 13 months (*table S2*).

Table 2. Follow-up diagnostics and treatment after abnormal FDG-PET/CT scan

Patient	Consultation Rheumatologist	Consultation other specialist	Follow-up diagnostics
1	-	Pulmonologist	-
2	-	Internal medicine Internal medicine#	CT-thorax CT-colon + colonoscopy
3	-	General practitioner	Mammogram + ultrasound of breast
4	-	Dermatologist	-
5	Clinical evaluation tonsils at next planned consultation	ENT specialist	-
6	Physical examination of thorax and lungs		X-thorax
7	-	ENT specialist	
8	-	Internal medicine##	-
9	-	Internal medicine	Colonoscopy with polyp resection
10	-	Internal medicine	-
11	-	Pulmonologist	-
12	Clinical evaluation of skin lesion at next planned consultation		-
13	-	Internal medicine	Fine needle aspiration biopsy (3x) + laboratory testing
14	Evaluation TSH and palpation thyroid at next planned consultation		-
15	-	Gynecologist	Cervical smear test + Ultrasound uterus
16	-		CT thorax + abdomen
17	-	Vascular surgeon	Ultrasound abdominal aorta
18	-	ENT specialist##	-

	Non-invasive and Surgical intervention	Conclusion and diagnosis
	-	Increased FDG uptake in the right inferior lobe of the lungs combined with several nodules. However, no malignancy/other clinically relevant diagnosis. No further action.
	-	Increased FDG uptake in hilar/mediastinal lymph nodes and intestines. No malignancy/other clinically relevant diagnosis. No further action.
	-	Increased FDG uptake in breast tissue. No malignancy/other clinically relevant diagnosis. No further action.
	-	Increased FDG uptake in tissue on the right upper leg. No malignancy/other clinically relevant diagnosis. No further action.
	Antibiotics	Increased FDG uptake due to previously performed tonsillectomy, no further action.
	-	Increased FDG uptake dorsally around the tenth rib. No malignancy/other clinically relevant diagnosis. No further action.
	-	Increased FDG uptake due to speaking during scan, no further action.
	-	Increased FDG uptake in hilar/mediastinal lymph nodes and intestines. No malignancy/other clinically relevant diagnosis. No further action.
	-	2 low grade adenomas. A follow-up colonoscopy was planned after 5 years post-resection.
	Marginal myotomy paraspinal muscle mass	Schwannoma
	-	Increased nodular FDG uptake in the basal segment of the left lung. Turned out to be a stable rheumatoid nodule
	-	Increased FDG uptake at cutaneous lesion in axilla/upper arm. No malignancy/other clinically relevant diagnosis. No further action.
	Hemi-thyroidectomy	Benign follicular adenoma.
	-	Increased FDG uptake in the thyroid. No malignancy/other clinically relevant diagnosis. No further action.
s	Myomectomy	Myoma in the uterus
	-	Increased FDG uptake in the lungs and adrenal glands. No malignancy/other clinically relevant diagnosis. No further action.
	Statins and advise for lifestyle interventions	Aneurysm of the abdominal aorta (43mm) + atherosclerosis
	Cyst extraction in the right medial side of the neck via sistrunk procedure	Thyroglossal cyst.

Table 2. Continued

Patient	Consultation Rheumatologist	Consultation other specialist	Follow-up diagnostics
19	-	Pulmonologist	CT-thorax (2x, after 6 and 18 months)
20	-	ENT specialist Urologist	-
21	Clinical evaluation thyroid next planned consultation	Endocrinologist	Clinical evaluation thyroid after 6 and 12 months

[#]Telephone consultation
[##]Emergency consultation

Discussion

This study is the first to highlight the incidental extra-articular findings and test characteristics of whole-body FDG-PET/CT-scans for malignancy in a population of longstanding RA patients with low disease activity treated with bDMARDs. A large number of extra-articular abnormalities was found, leading to follow-up consultations, additional (invasive) diagnostic testing, referral to other specialists and in some cases treatment. These were associated with anxiety, patient burden, costs and adverse effects. Conversely, the diagnostic value for malignancies was low.

Our study has some limitations concerning the conclusion regarding test characteristics for malignancy. It should be taken into consideration that test characteristics of whole-body FDG-PET/CT-scans depend on the type and stage of a malignancy. For example, bladder and penile carcinoma are not visualized very well on FDG-PET/CT due to urine contamination. For prostate carcinoma, there is no indication for FDG-PET/CT, as Prostate-Specific Membrane Antigen PET-scan is better suited for this type of tumor. In case of lymphomas, imaging on whole-body FDG-PET/CT-scans is strongly dependent of the tumor subtype.

The major difference of this study compared to a standard diagnostic test accuracy study is the lack of blinding for the FDG-PET/CT assessment, with subsequent diagnostic analyses in patients. This could, however, only have led to overestimation of test accuracy, not to underestimation.

This study paints a sobering picture of the risks and benefits of whole-body FDG-PET/CT-scanning in longstanding RA patients with low disease activity treated with bDMARDs. Using FDG-PET/CT-scanning for assessing arthritis results in a substantial

	Non-invasive and Surgical intervention	Conclusion and diagnosis
	-	Increased nodular FDG uptake in the inferior lobe of the right lung. Turned out to be a stable nodular lesion.
	-	Increased FDG uptake in larynx and prostate. No malignancy/other clinically relevant diagnosis. No further action.
12	-	Increased FDG uptake in thyroid. No malignancy/other clinically relevant diagnosis. No further action.

number of incidental findings, while some who had apparently normal scans also developed malignancies. Based on our findings, the use of whole-body FDG-PET/CT-scanning for a musculoskeletal indication – either in case of research or as a clinical tool – requires properly informing patients of risk and benefits of this type of imaging.

Supplementary tables

Supplementary Table S1. FDG-PET/CT results for suspected malignancy compared to development of malignancy in the following 3 years

	Malignancy present	No malignancy present	Total
FDG-PET; malignancy suspected	0	9	9
FDG-PET; no malignancy suspected	6	64	70
Total	**6**	**73**	**79**

Supplementary Table S2. Intervals between the most recent FDG-PET/CT and diagnosis of malignancies

Type of malignancy	Date of scan	Date of diagnosis malignancy	Time lag between scan and diagnosis (months)
Lymphoma	jun-13	apr-16	34
Penile	jan-13	sep-13	8
Bladder	apr-12	sep-12	5
Prostate	oct-12	aug-13	10
Melanoma	dec-11	jan-13	13
Melanoma 2	feb-12	nov-12	9

RERENCES

1 Yasuda S, Ide M, Fujii H, *et al.* Application of positron emission tomography imaging to cancer screening. *Br J Cancer* 2000;**83**:1607–11.

2 Terauchi T, Murano T, Daisaki H, *et al.* Evaluation of whole-body cancer screening using 18 F-2-deoxy-2-fluoro-d-glucose positron emission tomography: a preliminary report. *Ann Nucl Med* 2008;**22**:379–85.

3 Kojima S, Zhou B, Teramukai S, *et al.* Cancer screening of healthy volunteers using whole-body 18F-FDG-PET scans: The Nishidai clinic study. *Eur J Cancer* 2007;**43**:1842–8.

4 Edinburgh RC of P of, Committee A of RSA. Evidence-based indications for the use of PET-CT in the United Kingdom 2016. *Clin Radiol* 2016;**71**:e171–88.

5 Mandl P, Ciechomska A, Terslev L, *et al.* Implementation and role of modern musculoskeletal imaging in rheumatological practice in member countries of EULAR. *RMD open* 2019;**5**:e000950.

6 Kubota K, Yamashita H, Mimori A. Clinical Value of FDG-PET/CT for the Evaluation of Rheumatic Diseases: Rheumatoid Arthritis, Polymyalgia Rheumatica, and Relapsing Polychondritis. *Semin Nucl Med* 2017;**47**:408–24. doi:10.1053/j.semnuclmed.2017.02.005

7 Elzinga EH, van der Laken CJ, Comans EFI, *et al.* 18F-FDG PET as a tool to predict the clinical outcome of infliximab treatment of rheumatoid arthritis: an explorative study. *J Nucl Med* 2011;**52**:77–80.

8 Colebatch AN, Edwards CJ, Østergaard M, *et al.* EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. *Ann Rheum Dis* 2013;**72**:804–14. doi:10.1136/annrheumdis-2012-203158

9 Schöder H, Gönen M. Screening for cancer with PET and PET/CT: potential and limitations. *J Nucl Med* 2007;**48**:4S-18S.

10 Saquib N, Saquib J, Ioannidis JPA. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol* 2015;**44**:264–77.

11 Kwee RM, Kwee TC. Whole-body MRI for preventive health screening: A systematic review of the literature. *J Magn Reson Imaging* 2019;**50**:1489–503.

12 Nishizawa S, Kojima S, Okada H, *et al.* Ten-year prospective evaluation of whole-body cancer screening with multiple modalities including [18F] fluorodeoxyglucose positron emission tomography in a healthy population. *Ann Nucl Med* 2020;**34**:358–68.

13 Bouman CAM, Van Herwaarden N, Van Den Hoogen FHJ, *et al.* Long-term outcomes after disease activity-guided dose reduction of TNF inhibition in rheumatoid arthritis: 3-year data of the DRESS study - A randomised controlled pragmatic non-inferiority strategy trial. *Ann Rheum Dis* 2017;**76**:1716–22. doi:10.1136/annrheumdis-2017-211169

14 Van Herwaarden N, Van Maas A Der, Minten MJM, *et al.* Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis:Open label, randomised controlled, non-inferiority trial. *BMJ* 2015;**350**:1–8. doi:10.1136/bmj.h1389

15 den Broeder AA, van Herwaarden N, van der Maas A, *et al.* Dose REduction strategy of subcutaneous TNF inhibitors in rheumatoid arthritis: design of a pragmatic randomised non inferiority trial, the DRESS study. *BMC Musculoskelet Disord* 2013;**14**:1–9.

16 Boellaard R, Delgado-Bolton R, Oyen WJG, *et al.* FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015;**42**:328–54. doi:10.1007/s00259-014-2961-x

17 Bouman CAM, van Herwaarden N, Blanken AB, *et al.* 18F-FDG PET-CT scanning in rheumatoid arthritis patients tapering TNFi: reliability, validity and predictive value. *Rheumatology* 2021.

Chapter 6

Long-term routine laboratory monitoring of immunomodulatory drugs in rheumatoid arthritis, a retrospective cohort study

Evy Ulijn
Nathan den Broeder
Karen Bevers
Rosalie Pruijs
Bram van Es
Tjebbe Tauber
Robert Landewé
Noortje van Herwaarden
Alfons A den Broeder

Under review in Annals of Internal Medicine

Abstract

Background: There are limited data on the value of long-term routine laboratory toxicity monitoring (lt-RLTM) during DMARD use which results in frequent monitoring in clinical practice.

Objectives: 1) determining the cumulative incidence of abnormal and very abnormal results, 2) comparing new very abnormal result rates between periods with and without recommended monitoring, 3) assessing indicators of (in) appropriate testing for all new very abnormal results.

Design and setting: Monocenter retrospective cohort study from July 2008 to April 2020 in the Netherlands.

Participants: Rheumatoid arthritis patients undergoing lt-RLTM after ≥6 months of DMARD use.

Measurements: Cumulative incidences of patient-DMARD exposure periods progressing to abnormal or very abnormal results were calculated for ALT, eGFR, Hb, white blood cell-, and platelet counts. Incidence densities (ID) and incidence rate ratios (IRR) were calculated for periods with and without recommended monitoring. New very abnormal results were chart-reviewed for indicators of (in) appropriate testing.

Results: 4,819 patients underwent 330,435 lt-RLTM tests over 30,505 patient years. Progression to very abnormal results occurred in 1.3% (95%-CI 1.2 to 1.4) of 41,585 patient-DMARD exposure periods. The ID of very abnormal results was similar between periods with and without recommended monitoring (IRR 0.94, 95%-CI 0.75 to 1.17). New very abnormal results (n=487) mostly occurred after dose increase, were often suspected, considered unrelated to DMARD use, or did not lead to action.

Limitation: We cannot conclude that the lack of serious adverse outcomes indicates inappropriate testing, as regular monitoring was conducted.

Conclusion: Monitoring seems valuable in the first six months of treatment, after dose escalation, based on patient characteristics and symptoms. Remaining lt-RLTM strategies warrant critical revision.

Primary funding source: None.

Introduction

Long-term routine laboratory toxicity monitoring (lt-RLTM) is used for management of rheumatoid arthritis (RA) patients using disease-modifying antirheumatic drugs (DMARDs). Intensive monitoring is recommended during the first six months due to higher adverse reaction risks, followed by less frequent monitoring for the duration of DMARD use, which can last decades (1–3). Laboratory parameters used for lt-RLTM include liver, renal, and hematologic toxicity markers (3). Despite high consistency in the types of tests ordered, the frequency and responses to abnormal results vary widely across practices (4).

While short-term DMARD toxicity is well-documented, less is known about long-term toxicity and abnormal lt-RLTM results. Nakafero et al. found methotrexate and leflunomide discontinuation rates due to abnormal results were low, but comparatively higher in the first year (6.16 and 9.42/1000 person-years) than in subsequent years (2.84 and 4.4/1000 person-years) (5). Fraser et al. observed mostly normal results over two years for methotrexate in primary care. (6). Notably, current research often focusses on individual DMARDs, while often combinations are used in clinical practice. Additionally, there are no test-treatment trials that compare different lt-RLTM strategies in terms of their impact on relevant clinical outcomes (3).

The limited data on the optimal frequency of lt-RLTM is reflected in existing guidelines, which vary in their recommendations and often tend to be cautious in their approach (1–4,7–11). Summaries of product characteristics also differ in their recommendations per parameter, per DMARD and even between different brands of the same DMARD (12). Typically, lt-RLTM occurs every three months for csDMARD mono- and combination therapies, and for rituximab and tocilizumab monotherapy. Other biologic DMARDs and hydroxychloroquine are mostly judged not to require lt-RLTM. The necessity of frequent lt-RLTM for patients on stable DMARD doses is debated and reduced monitoring during the COVID-19 pandemic did not show evident harm (13). Reducing lt-RLTM frequency could lessen patient- and healthcare burden and save costs (14).

We set out to investigate the value of current lt-RLTM strategies in DMARD using RA patients by addressing three research questions: 1) what is the cumulative incidence of abnormal and very abnormal lt-RLTM tests? 2) is there a difference in incidence density (ID) between patient-DMARD exposure periods during which monitoring is and is not recommended? and 3) are new very abnormal laboratory results linked to indicators of (in)appropriate testing?

Methods

This retrospective cohort study was performed at the rheumatology department of the Sint Maartenskliniek, a specialized multicenter rheumatology practice in the Netherlands. Eligible participants were identified through the electronic health records, and relevant data were extracted anonymously. All new very abnormal laboratory results were further analyzed by manual chart review.

Patients with a clinical diagnosis of RA in outpatient care between July 2008 and April 2020, that have used a DMARD for at least 6 months were included. The censoring in 2020 was chosen to exclude data collected during the COVID-19 pandemic as it caused (rheumatological) care to be organized differently. Diagnosis of RA was operationalized by use of international classification of disease 9 code 714.x and 10 code M06.9. Follow-up visits took place every 3-6 months based on disease activity. Patients received protocol-based treatment with conventional-, biologic-, and targeted synthetic- DMARDs in authorized dosing. Oral, intra-articular or intramuscular glucocorticoids were administered on indication. The treat-to-target strategy, aiming for disease activity score using C-reactive protein (DAS28-CRP) to achieve low disease activity or remission, was consistently applied, and treatment was tapered where possible. Overall, care adhered to current European Alliance of Associations for Rheumatology RA treatment recommendations (10).

Routine laboratory toxicity monitoring was conducted in all patients after start of a DMARD, once a month for the first 6 months, followed by lt-RLTM measurements every 3-6 months. Additional monitoring was recommended during dose escalation.

Laboratory parameters

The test suite of laboratory parameters within this cohort included five laboratory parameters, divided in abnormal- and very abnormal results: ALT (>100U/L, >300U/L), eGFR (<60ml/min/1.73m^2, <45ml/min/1.73m^2), Hb (<7.5mmol/L for females, <8mmol/L for males, <6mmol/L), white blood cells (WBC) (<3.5mmol/L, <2.0mmol/L) and platelet counts (<140*10^9/L, <100*10^9/L). Cut-offs for abnormal and very abnormal results were based on guidelines (2), and derived from several clinical treatment recommendations on when clinical intervention is deemed necessary, or based on expert opinion (3,17–19).

We excluded laboratory tests that were not collected in the test suite of 5 parameters used for lt-RLTM, aligning with our focus solely on routine monitoring, as we presumed such tests were ordered for specific indications rather than for routine monitoring.

DMARDs and patient-DMARD exposure periods

DMARDs included in this study consisted of regularly used conventional synthetic (cs) DMARDS: methotrexate, leflunomide, sulfasalazine, hydroxychloroquine and azathioprine, biological (b)DMARDs: adalimumab, certolizumab, etanercept, golimumab, infliximab, abatacept, rituximab, sarilumab, tocilizumab and anakinra, and targeted synthetic (ts)DMARDs: baricitinib and tofacitinib. As we focus on long-term RLTM, the laboratory test results retrieved during the first six months of a DMARD after starting/switching/adding (to) a new DMARD were not analyzed.

Patient-DMARD exposure periods were generated based on three levels: 1/ the patient, 2/ the DMARD used, and 3/ the laboratory parameter. Due to this multilevel structure, one patient can be represented several times in different patient-DMARD exposure periods throughout the follow-up of the study and could have several abnormal and very abnormal laboratory results per testing moment (*Figure S1).*

Monitoring recommended and monitoring not recommended periods

To assess differences between patient-DMARD exposure periods for which monitoring is and is not recommended, they were clustered into two period categories. The first consisted of patient-DMARD exposure periods where a DMARD (combination) was used for which lt-RLTM is generally recommended and performed in clinical practice (methotrexate, leflunomide, sulfasalazine, azathioprine, baricitinib, tofacitinib, rituximab, tocilizumab and any combination therapy including these DMARDs). The second consisted of exposure periods in which a DMARD was used during which routine lt-RLTM is generally not recommended and performed in clinical practice but was still performed within our centre for logistical reasons (bDMARD monotherapy, excluding rituximab and tocilizumab and/or hydroxychloroquine monotherapy).
As sensitivity analysis, a more stringent definition was used, defining only hydroxychloroquine monotherapy as 'monitoring not recommended', and all patient-DMARD exposure periods (mono- and combination therapy) as 'monitoring recommended' (1,2).

Statistical analysis

Patient demographics, disease-, test- and treatment characteristics were described by means of descriptive statistics provided with mean (+/- standard deviation (SD)), median (25^{th} percentile (p25),75^{th} percentile (p75), interquartile ranges (IQR)), or n (%) depending on data distribution, with corresponding 95%-confidence intervals. Missing data were mentioned in respective tables. Castor EDC was used to enter and store data on characteristics of new very abnormal laboratory results. Statistical analyses were performed using STATA version 13.1 and R version 4.2.2.

IRB approval

Exemption from ethical review was obtained from the Medical Ethical Committee of the eastern region of the Netherlands (file number 2022-15833).

Patient and public involvement

No patients were directly involved in setting the research question or outcome measures for this study. However, 17 practicing rheumatologists from 10 hospitals were consulted regarding their insights on the acceptability of a less intensive monitoring strategy. Rheumatologists indicated a general willingness to accept a less intensive monitoring approach if it was proven to be safe. Prior to implementation of the study results input from patients and the public will be required to ensure effective communication tailored to their needs and understanding.

Role of the Funding source

This study did not receive any external funding

Objective 1: Cumulative incidence of abnormal and very abnormal lt-RLTM results

Firstly, the absolute event rates of abnormal and very abnormal laboratory results were given per laboratory parameter and DMARD type. The primary outcome of this study, however, was the cumulative incidence (number of events/patient-DMARD exposure periods) of abnormal and very abnormal laboratory results occurring during lt-RLTM. This was done by identifying three patterns in results during patient-DMARD exposure periods. 1/ patient-DMARD exposure periods with only normal laboratory results. 2/ patient-DMARD exposure periods with only abnormal or very abnormal laboratory results (both crossed in *Figure S2* as our study aims to identify new abnormalities occurring during DMARD use), and 3/ patient-DMARD exposure periods that consisted of lt-RLTM results that varied from normal to abnormal or very abnormal at least once during the period (circled in *Figure S2*, these periods are of interest for this study. They depict new abnormal results occurring during DMARD use).

The cumulative incidence was given for each type of patient-DMARD exposure period. Additionally, the cumulative incidence of variable patient-DMARD exposure periods were stratified for each DMARD (combination) and laboratory parameter and were provided with 95%-confidence intervals (95%-CI).

Finally, the proportion of patient-DMARD exposure periods during which the DMARD was stopped or switched after an abnormal or very abnormal laboratory result was calculated.

Objective 2: Difference in the incidence density (ID) of very abnormal laboratory results between DMARDs where monitoring is and is not recommended

The ID (new events / patient years) of very abnormal laboratory results were compared between patient-DMARD exposure periods for which lt-RLTM is and is not recommended. The ID was calculated for all very abnormal laboratory results in the monitoring recommended, and the monitoring not recommended periods. Subsequently, the relative risk reduction was determined by calculating the incidence rate ratio (IRR) assessing the difference between the groups of periods. Thereafter the ID and IRR were calculated within the sensitivity analysis.

Objective 3: Characteristics of new very abnormal laboratory results

All new very abnormal laboratory results were further analyzed by manual chart review, conducted by a treating physician (resident) who held consensus meetings with two rheumatologists, one of whom was also qualified as a clinical pharmacologist and the other a clinical epidemiologist. We chose not to analyze laboratory results that were abnormal, not very abnormal, as we judged the clinical relevance being very low. Data extraction focused on pre-specified characteristics consisting of patient- and test characteristics, judgement of pretest probability for a very abnormal laboratory result, judgement of the likelihood of causality between DMARD use and the laboratory result, the actions that were undertaken after the new very abnormal laboratory result and the clinical outcome severity of the laboratory result. Underlying disease pathology and clinical outcome severity were assessed following Common Terminology Criteria for Adverse Events, version 5.0 (CTCAE) (20). The topic 'investigations' was removed as its inclusion would have resulted in circularity (all very abnormal laboratory results would have been classified as high-grade adverse events). The proportion of tests fulfilling each characteristic were provided. One new very abnormal laboratory result may have met multiple characteristics of (in) appropriate use of tests.

Finally, case vignettes were included to offer insight in the clinical context surrounding new very abnormal laboratory results.

Results

Of the 5,341 DMARD-using RA patients, 476 patients never underwent lt-RLTM as they never used a DMARD for longer than 6 months. Therefore, 4,819 RA patients were included, with a follow-up of 30,505 patient years (24,726 patient years in monitoring recommended periods and 5,779 patient years in monitoring not recommended periods) (*Table 1).* Mean DAS28-CRP scores during total follow-up of the cohort was 2.1 (SD 0.83).

Table 1. Patient demographics, disease-, test- and treatment characteristics

Patient demographics (N=4,819)	
Sex, female [n (%)]	3,244 (67.3%)
Age in years* [mean (SD)]	59.6 (13.9)
Disease characteristics	
Anti citrullinated peptide antibodies positive [n (%)]	2,920 (60.6%)
Rheumatoid factor positive [n (%)]	3,150 (65.4%)
Test- and treatment characteristics	
Mean number of DMARDs used per patient [mean (min-max)]**	1.7 (1-8)
Number of lt-RLTM test suites per patient [mean (SD)]**	13.7 (11.5)

Missing cases: **Anti citrullinated peptide antibodies** (156), **Rheumatoid factor** (39)
* As documented at first lt-RLTM lab.
** within total follow-up of the cohort

Table S1 shows the number of patients per exposure period, the mean duration of DMARD use per patient and the testing frequency within each patient-DMARD exposure period. The most prevalent combinations of DMARDs were analyzed separately (if ≥2,500 tests were retrieved during that exposure period).

The total number of tests was 330,435, containing 66,087 test sets (containing five tests on the same day), with a mean of 13.7 test sets per patient, and 2.17 test sets per patient year. 275,475 lt-RLTM tests were retrieved during monitoring recommended periods, of which 16,178 (5.9%, 95%-CI 5.8%-6.0%) results were abnormal and 1,678 (0.6%, 95%-CI 0.58%-0.64%) results were very abnormal. 54,960 lt-RLTM tests were retrieved during monitoring not recommended periods, of which 3,397 (6.2%, 95%-CI 6.0%-6.4%) results were abnormal and 659 (1.2%, 95%-CI 1.1%-1.3%) results were very abnormal (*Figure 1*). The proportion of abnormal and very abnormal lt-RLTM test results split per DMARD are shown in table S2 and S3.

Objective 1

The cohort consisted of 41,585 patient-DMARD exposure periods, of which 33,625 were clustered as monitoring recommended periods and 7,960 were clustered as monitoring not recommended periods.

In the monitoring recommended patient-DMARD exposure periods, the cumulative incidence of lt-RLTM results changing from normal to abnormal was 9.2% (95%-CI 8.9%-9.5%). Furthermore, 88% (95%-CI 87.6%-88.3%) of patient-DMARD exposure periods showed only normal results and 2.7% (95%-CI 2.5%-2.9%) showed consistent abnormal results.

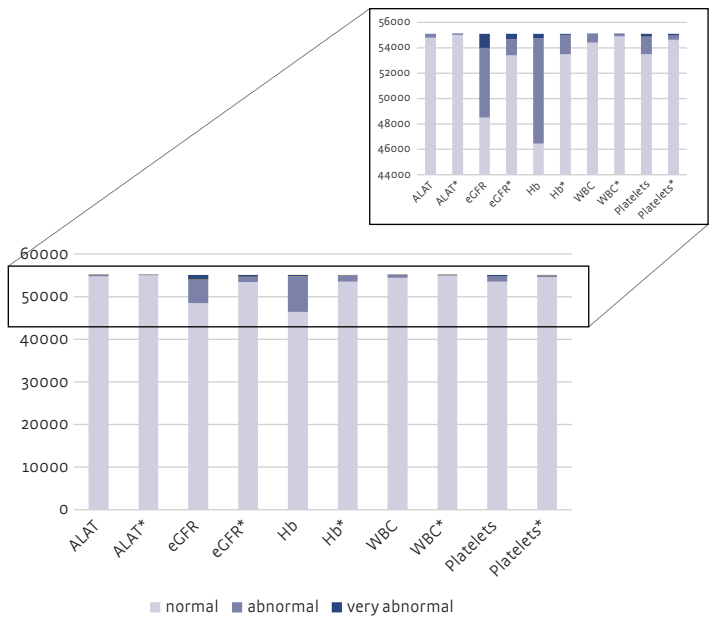

Figure 1. *Proportion of abnormal/very abnormal lt-RLTM results during monitoring recommended and monitoring not recommended periods.*

Of note, mentioned abnormalities are overall event rates and include the abnormal results retrieved during patient-DMARD exposure periods in which results were always abnormal.

Regarding very abnormal lt-RLTM results, the cumulative incidence of changes from (ab)normal to very abnormal was 1.3% (95%-CI 1.2%-1.4%) of the patient-DMARD exposure periods. 98.5% (95%-CI 98.4%-98.6%) of the patient-DMARD exposure periods never showed very abnormal results and 0.3% (95%-CI 0.2%-0.4%) consistently showed very abnormal results.

In the monitoring not recommended patient-DMARD exposure periods, the cumulative incidence of lt-RLTM results changing from normal to abnormal was 7.9% (95%-CI 7.3%-8.5%), while 88.9% (95%-CI 88.2%-89.6%) of the patient-DMARD exposure periods showed only normal results and 3.2% (95%-CI 2.8%-3.6%) showed consistently abnormal results.

The cumulative incidence of changes from (ab)normal to very abnormal was 1.3% (95%-CI 1.0%-1.5%) of the patient-DMARD exposure periods for which monitoring was not recommended. Furthermore, 98.1% (95%-CI 97.8%-98.4%) of the patient-

DMARD exposure periods never showed very abnormal results, 0.6% (95%-CI 0.4%-0.8%) of the patient-DMARD exposure periods consistently showed very abnormal results.

Patient-DMARD exposure periods with results varying from normal to abnormal or very abnormal at least once, split for all DMARD subgroups and for the different laboratory parameters are shown in *Table S4, S5, S6 and S7.* This analysis showed that the highest proportions of variable patient-DMARD exposure periods were seen in the parameters Hb and eGFR, and during infliximab monotherapy and methotrexate & tocilizumab combination therapy.

To assess whether abnormal results led to changes in treatment, we analyzed the final result of each patient-DMARD exposure period. Of the 41,585 patient-DMARD exposure periods, the final lt-RLTM result was abnormal in 2,658 (6.8%), of which 719 (1.7%) progressed from normal to abnormal in the last laboratory measurement within the patient-DMARD exposure period suggesting that the progression to abnormal result could have been a reason for switching or discontinuing the DMARD. In 369 (0.89%) of these, there had been no previous abnormal results in that patient-DMARD exposure period.

Objective 2

During follow-up, 486 new very abnormal lt-RLTM results were found, of which 389 were observed in monitoring recommended periods, with an ID of 15.7 per 1000 patient-years (py) (95%-CI 14.2-17.3), and 97 new very abnormal results occurred in the monitoring not recommended periods, with an ID of 16.8 per 1000py (95%-CI 13.5-20.1), resulting in an IRR of 0.94 (95%-CI 0.75-1.17).

For the sensitivity analysis, defining only hydroxychloroquine monotherapy as 'monitoring not recommended', and all other patient-DMARD exposure periods as 'monitoring recommended', 456 new very abnormal laboratory results occurred during monitoring recommended periods with an ID 16.1/1000py (95%-CI 14.6-17.6), and 30 new very abnormal laboratory results occurred during the monitoring not recommended periods with an ID of 13.3/1000py (95%-CI 8.59-18.1, resulting in an IRR of 1.2 (95%-CI 0.83-1.82).

Objective 3

The characteristics of the new very abnormal laboratory results are shown in *Table 2*. Notable are the high median age of patients and the large proportion of patients with elevated pre-test probabilities for abnormal laboratory results. A considerable proportion of abnormalities were judged as unrelated to DMARD use, and for a sizeable proportion of very abnormal results no further action was undertaken.

Table 2. Characteristics of new* very abnormal laboratory results (n=486)

Patient characteristics	
Age *[median (p25, p75, IQR)]*	73 (65.5, 79.7, 14.3)
Sex *[female N (%)]*	339 (69%)
Test Characteristics	
Type of test *[N, median test result (p25, p75, IQR), % change (median)]*	
ALT:	9, 456 (35, 490, 155), +1633%
eGFR	246, 42 (39, 43, 4), -20%
Hb	157, 5.7 (5.5, 5.8, 0.3), -14%
WBC count	10, 1.65 (1.1, 1.8, 0.7), -51%
Platelet count	64, 90 (79.5, 95, 15.5), -28%
Laboratory error *[N (%)]*	10 (2.1%)
Measurement error:	8 (1.6%)
Test result was from another patient:	2 (0.4%)
Pretest probability for abnormal laboratory result	
Occurred <6months after DMARD dose increase *[N (%)]*	31 (6.4%), m*issing: 59/486*
Current lab result was already known from external source (e.g. previous hospitalization or tested by GP) *[N (%)]*	236 (48.5%), m*issing: 10/486*
A previous result was very abnormal *[N (%)]*	
In the first 6 months of this DMARD use:	95 (19.5%), m*issing: 137/486*
During previously used DMARDs:	111 (22.8%), m*issing: 87/486*
Testing would have been performed on indication** *[N (%)]*	399 (82.1%), m*issing: 16/486*
No increased pretest probability (none of the above) *[N (%)]*	16 (3.3%)
Causality	
A different cause than the DMARD was (eventually) deemed a more probable cause of the abnormal laboratory result *[N (%]*	117 (24.1%), m*issing: 3/486*
Actions	
No action was undertaken*** *[N (%,)]*	173 (35.6%)
DMARD dose was lowered / stopped *[N (%)]*	
DMARD was thought to be the cause of the abnormality:	55 (11.3%)
DMARD dose was lowered to prevent (further) toxicity:	115 (23.7%)
Other actions *[N (%)]*	
Lifestyle advice:	16 (3.3%)
Additional diagnostics:	68 (14.0%)
Initiating/altering treatment:	27 (5.6%)
Informing/referral to other specialist/GP:	144 (29.6%)
Admission:	33 (6.8%)
Unknown:	25 (5.1%)

Table 2. Continued

Clinical outcome severity CTCAE, *missing: 25*	
Mild or moderate *[N (%)]*	360 (4.1%)
Severe or medically significant but not immediately life-threatening, or life threatening consequences *[N (%)]*	100 (20.5%)
Death related to adverse event *[N (%)]*	1 (0.2%), case vignette 4

* developed during a long-term patient-DMARD exposure period
** Due to signs/symptoms or recent history, interpretation by physician.
*** Physician did not take note of the laboratory abnormality anywhere in the electronic health record.

CTCAE-criteria:
Mild: Asymptomatic, clinical or diagnostic observation only, intervention not indicated
Moderate: Minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental Activities of Daily Living.
Severe or medically significant but not immediately life-threatening: Hospitalization or prolongation of hospitalization indicated; disabling, limiting self-care.

The main causes of new very abnormal laboratory results included renal and urinary disorders such as chronic kidney disease, acute kidney injury, and renal calculi. Other important contributors were medication-induced issues, intra/postoperative hemorrhage, and wound complications (*Table S8).*

Some aspects stood out in *Table 2* and may provoke inquiries; to clarify these aspects, several case vignettes are provided below in *Table 3*.

Table 3. Case vignettes

Vignette 1*: testing would have been performed on indication*

In a 53 year old patient using etanercept monotherapy (50mg every 2 weeks), lt-RLTM showed an ALT of 810 U/L (previous ALT result was 13U/L). In this case there was, however, an indication for laboratory testing as the patient had experienced intermittent pain in the right upper quadrant of the abdomen. An ultrasound was performed which showed gallstones and eventually the gallbladder was surgically removed.

Vignette 2: *a different cause than the DMARD was (eventually) deemed a more probable cause of the abnormal laboratory result.*

In a 49 year old patient using methotrexate 25mg weekly and leflunomide 10mg daily, lt-RLTM showed an ALT of 335 U/L (previous laboratory result was 17 U/L). The patient had experienced no symptoms indicating liver pathology warranting laboratory testing. After obtaining the lt-RLTM results the patient was sent for further analysis. Methotrexate and leflunomide as well as other comedication that could induce liver toxicity were discontinued. Further testing showed a hepatitis E infection, and both DMARDs were restarted after resolution.

Vignette 3 a and b: *no action was undertaken and/or abnormality was already known.*

a: In a 73 year old patient using hydroxychloroquine monotherapy (200mg daily), lt-RLTM showed an eGFR of 43ml/min/1.73m2 (previous eGFR result was 47ml/min/1.73m2). This patient was already under regular monitoring by an internal medicine specialist for chronic renal dysfunction (as well as hypertension and diabetes). Therefore, no action was undertaken upon retrieval of this lt-RLTM result.

b: In an 87 year old patient using a stable dose of methotrexate monotherapy (10mg weekly) for several years, lt-RLTM showed a declining platelet count 117 to 96.10^9/L. There was a slow and steady declining trend of not only the platelet count, but also in WBC count and Hb in the preceding two years before the very abnormal laboratory result. These findings were consistent with hypocellular bone marrow. Despite these findings, no proactive measures were taken, considering both the patient's age, absence of clinical sequelae, and the sustained remission of the RA under current therapy.

Vignette 4: *death associated with an adverse event (grade 5 clinical outcome severity)*

In a 69 year old patient using etanercept monotherapy (50mg every 10 days), lt-RLTM showed a decreased Hb level of 5.6mmol/L. This outcome was already known due to recent hospitalization for a blood transfusion. After that transfusion the patient developed phlebitis at the injection site for transfusion, leading to a disseminated septic arthritis of the wrist as the cause of death.

Discussion

Our results show that very abnormal lt-RLTM results are very rare and the ID of very abnormal results is comparable during use of DMARDs for which monitoring is and is not recommended. Additionally, in-depth review of these rare new very abnormal results revealed that they were often already known, occurred after dose increase, were not related to DMARD use or did not result in any action. Together, our findings strongly suggest that the current routine lt-RLTM practice leads to inappropriate testing. However, it is less straightforward what the optimal testing interval should be. A rational alternative strategy could be to monitor the first 6 months after start of a DMARD, and thereafter only 1/ in high risk patients (liver, renal, bone marrow diseases, earlier abnormalities), 2/ when increasing the dose, 3/ on indication based on signs/symptoms or recent events and 4/ at least every two years to pick up secular trends.

The present study is, to our knowledge, the first to provide a long-term and in-depth analysis of the value of current lt-RLTM strategies during regularly used DMARDs in RA. Furthermore, we are the first to manually review all new very abnormal laboratory results for indicators that provided insights into the (lack of) clinical value of testing. Alongside the methodology, this study's strengths lie in its use of recent data, large sample size and low risk of ascertainment bias due to nonrestrictive inclusion, and lt-RLTM data in all patients, regardless of received DMARD treatment.

Limitations of our study design include that we cannot conclude that the lack of serious adverse clinical outcomes is a sign of inappropriate testing, as monitoring was performed regularly and acted on if needed. This conclusion can only be drawn from a formal test treatment trial comparing routine monitoring to a less intensive monitoring schedule. However, for any test to result in meaningful changes in clinical outcomes, it is mandatory that new clinically relevant abnormal test results occur regularly, and with greater frequency when using drugs for which monitoring is recommended compared to those where monitoring is not recommended. Furthermore, these abnormal results should precede symptoms, be causally related to the drug use, and ultimately lead to changes in clinical decision-making. As these requirements seem virtually never met, we feel we can deduce that such a test treatment trial would yield negative results, as is generally the case in these types of studies (21,22). In addition, performing a formal non inferiority test treatment trial with these very low event rates and putatively very small non-inferiority margin would be a massive undertaking in terms of numbers of patients and years of follow-up making it not realistically feasible.

Another limitation is the large variety of DMARDs and treatment combinations, combined with the infrequent occurrence of very abnormal test results, as this poses challenges for conducting sub analyses based on each treatment or treatment combination. Also, the use of data up to 2020 resulted in the fact that janus kinase -inhibitors were underrepresented.

Generalizability of our results however seems robust in light of the typical patient demographics, disease- and treatment characteristics, although it might be limited to high income countries with a similar prevalence of comorbidities and polypharmacy as our Dutch population.

A specific limitation is the possible confounding by contraindication, the effect that people with risk for lab abnormalities would preferably be treated with DMARDS for which monitoring is not recommended due to their safety. However, the considerable overlap in patients between treatment groups guards against this.

Finally, our focus on long-term laboratory monitoring, and including only patients who used a DMARD > 6 months, has likely led to a 'healthy survivor effect'. Nevertheless, this effect does not bias our results, as our study only focusses on assessing the value of long-term monitoring, as it seems rational to routinely monitor toxicity in the first six months of drug use and we do not aim to generalize to patients who experience early treatment inefficacy or important adverse effects.

While our study did not exclude patients with other rheumatic conditions, it is possible that individuals with diagnoses such as systemic lupus erythematosus or other rheumatic diseases who also have received a diagnosis of RA may have been included in our study. This inclusion likely has had minimal impact on our results but may have led to additional abnormalities in laboratory test results. An interesting question would be whether the results of our present study can be extrapolated to other patient populations using the same drug classes (e.g. psoriatic arthritis, axial spondylarthritis, psoriasis, inflammatory bowel disease). In addition, future research should explore patients' and physicians' perspectives on this topic, including whether testing, even when yielding normal results, could enhance therapy adherence by reassuring patients of medication safety. Also, there is a need to identify potential legal concerns and practical barriers that may arise with the implementation of new practice guidelines.

Acknowledgements: We would like to acknowledge Michael Niemantsverdriet for his valuable research assistance.

Contributors: AdB, KB, NvH, RL, NdB, EU were involved in the conception and design of the study. EU, NdB, NvH, AdB, BvE, RP, TT were involved in identifying relevant data from the electronic health records. EU, NdB, NvH, AdB, BvE, RP contributed to the data analysis. Manual chart review was conducted by EU, who held consensus meetings with NvH and AdB. EU drafted the manuscript and all co-authors critically reviewed the manuscript and gave final approval for its submission.

Competing interests: None of the authors have any competing interests to disclose for the submitted work.

Ethics approval: This study has been judged to fall outside the scope of the Medical Research Involving Human Subjects Act (WMO). Therefore, Medical Ethical Review Board approval was not required (CMO: 2019-5443).

Data sharing statement: Additional unpublished data can be obtained from the corresponding author upon reasonable request.

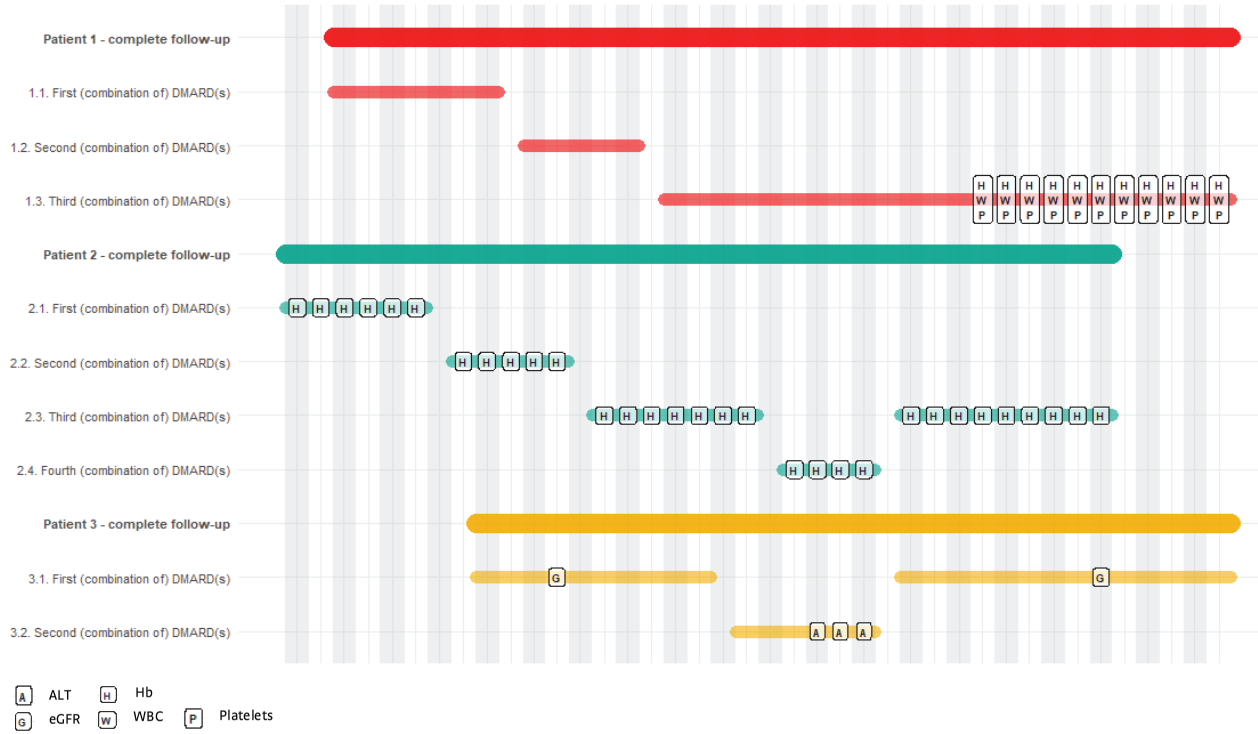

Figure S1. *Illustration of three example patients, several patient-DMARD exposure periods and abnormal laboratory results.*

Patient 1 developed a myelodysplastic syndrome during the third patient-DMARD exposure period, resulting in abnormal Hb, WBC- and platelet counts. Patient 2 had a chronic anemia with low Hb levels in all patient-DMARD exposure periods. Patient 3 showed a low eGFR in two patient-DMARD exposure periods and elevated liver enzymes in a third.

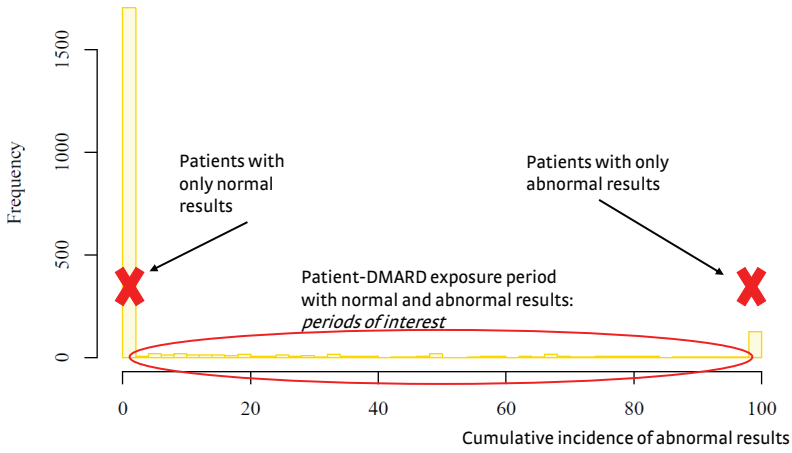

Figure S2. *Example* of plotted histogram for patient-DMARD exposure period consisting of only normal, only abnormal and variable results during lt-RLTM.*

*Example for eGFR during Methotrexate monotherapy

Table S1. Number of patients using a DMARD (combination), duration of DMARD use per patient and testing intervals per DMARD (combination)

DMARD(s)	Number of patients [N (%)]	Mean duration* of DMARD use per patient in years [mean (SD)]	Testing frequency [months (SD)]
Monitoring recommended			
Azathioprine	85 (1%)	2.6 (6)	4.3 (4)
Azathioprine & Adalimumab	41 (0.50%)	2.7 (3.3)	4 (2.2)
Azathioprine & Etanercept	47 (0.57%)	4.1 (5.5)	3.4 (1.4)
Leflunomide	300 (3.6%)	1.4 (2.9)	3.7 (2.5)
Leflunomide & Adalimumab	79 (0.95%)	0.9 (1.8)	3.3 (1.4)
Leflunomide & Etanercept	92 (1.1%)	1.8 (3.1)	3.8 (2)
Leflunomide & Hydroxychloroquine	56 (0.68%)	0.9 (1.7)	3 (1.5)
Leflunomide & Rituximab	63 (0.76%)	1.1 (2.1)	3.2 (1.6)
Methotrexate	2,259 (27.3%)	4.6 (6.8)	4.4 (3.3)
Methotrexate & Abatacept	42 (0.51%)	1.4 (3.5)	2.9 (1.8)
Methotrexate & Adalimumab	349 (4.2%)	1.6 (4)	3.6 (1.6)
Methotrexate & Etanercept	502 (6.1%)	2.9 (4.1)	3.7 (1.7)
Methotrexate & Hydroxychloroquine	1,217 (14.7%)	1.7 (3.2)	3.4 (1.7)
Methotrexate & Leflunomide	25 (0.30%)	0.9 (1.8)	3.6 (2)
Methotrexate & Rituximab	133 (1.6%)	1.6 (2.9)	3.4 (1.5)
Methotrexate & Sulfasalazine	120 (1.4%)	2.3 (4)	3.3 (1.9)
Methotrexate & Tocilizumab	48 (0.58%)	0.8 (2.2)	2.1 (1.4)
Rituximab	236 (2.8%)	3.1 (4.7)	4.1 (2.3)
Sulfasalazine	244 (2.9%)	3.3 (6.4)	4.5 (2.7)
Sulfasalazine & Hydroxychloroquine	54 (0.65%)	2 (4)	3.6 (1.3)
Tocilizumab	175 (2.1%)	1.7 (3.1)	2.8 (1.5)
Other combination	533 (6.4%)	2.3 (4.4)	4.6 (5.3)
Monitoring not recommended			
Abatacept	73 (0.88%)	1.1 (2)	2.6 (1.5)
Adalimumab	286 (3.5%)	2 (4.8)	4.3 (3.7)
Etanercept	463 (5.6%)	3.6 (5.2)	4.7 (3.3)
Golimumab	35 (0.42%)	2.1 (4)	3.5 (1.5)
Hydroxychloroquine	601 (7.3%)	2.9 (4.8)	5.4 (3.7)
Hydroxychloroquine & Adalimumab	28 (0.34%)	1.1 (2.1)	3.7 (1.1)
Hydroxychloroquine & Etanercept	50 (0.60%)	1.7 (2.7)	3.5 (1.5)
Infliximab	41 (0.50%)	3.1 (5.9)	3.7 (5.1)

Of note, if a patient-DMARD exposure period consisted of less than 25 patients, this subgroup was considered too small to be categorized as a separate group. Furthermore, combination therapy counts are not added in the monotherapy counts. Some patients may appear in both categories if they received both mono and combination therapy for the same medication.

*long-term medication use within cohort follow-up period.

Table S2. Number of abnormal lt-RLTM test results retrieved in patient-DMARD exposure periods during which monitoring was and was not recommended, split per DMARD

	Monitoring recommended				
	ALAT	eGFR	Hb	WBC	Platelets
Azathioprine	2 (0,3%)	74 (11%)	142 (21,1%)	22 (3,27%)	11 (1,63%)
Azathioprine & Adalimumab	0 (0%)	47 (10,68%)	56 (12,73%)	2 (0,45%)	0 (0%)
Azathioprine & Etanercept	0 (0%)	31 (6,14%)	110 (21,78%)	24 (4,75%)	25 (4,95%)
Leflunomide	1 (0,06%)	111 (6,52%)	327 (19,2%)	32 (1,88%)	62 (3,64%)
Leflunomide & Adalimumab	1 (0,32%)	26 (8,39%)	41 (13,23%)	3 (0,97%)	4 (1,29%)
Leflunomide & Etanercept	1 (0,16%)	33 (5,32%)	135 (21,77%)	8 (1,29%)	32 (5,16%)
Leflunomide & Hydroxychloroquine	0 (0%)	17 (7,56%)	38 (16,89%)	8 (3,56%)	2 (0,89%)
Leflunomide & Rituximab	2 (0,57%)	32 (9,12%)	69 (19,66%)	5 (1,42%)	0 (0%)
Methotrexate	123 (0,58%)	2601 (12,22%)	3026 (14,22%)	179 (0,84%)	418 (1,96%)
Methotrexate & Abatacept	0 (0%)	32 (9,22%)	40 (11,53%)	0 (0%)	0 (0%)
Methotrexate & Adalimumab	17 (0,64%)	291 (10,99%)	388 (14,65%)	9 (0,34%)	78 (2,94%)
Methotrexate & Etanercept	27 (0,57%)	268 (5,62%)	603 (12,64%)	52 (1,09%)	142 (2,98%)
Methotrexate & Hydroxychloroquine	49 (0,6%)	882 (10,81%)	1083 (13,27%)	82 (1%)	205 (2,51%)
Methotrexate & Leflunomide	2 (1,94%)	2 (1,94%)	23 (22,33%)	0 (0%)	3 (2,91%)
Methotrexate & Rituximab	3 (0,32%)	63 (6,77%)	204 (21,94%)	3 (0,32%)	2 (0,22%)
Methotrexate & Sulfasalazine	7 (0,64%)	101 (9,21%)	241 (21,97%)	4 (0,36%)	34 (3,1%)
Methotrexate & Tocilizumab	1 (0,26%)	18 (4,76%)	54 (14,29%)	9 (2,38%)	17 (4,5%)
Rituximab	21 (1,01%)	331 (15,89%)	317 (15,22%)	18 (0,86%)	67 (3,22%)
Sulfasalazine	2 (0,11%)	119 (6,27%)	339 (17,86%)	8 (0,42%)	31 (1,63%)
Sulfasalazine & Hydroxychloroquine	3 (0,85%)	26 (7,37%)	113 (32,01%)	9 (2,55%)	5 (1,42%)
Tocilizumab	6 (0,33%)	91 (4,94%)	230 (12,49%)	129 (7,01%)	169 (9,18%)
Other combination	27 (0,62%)	297 (6,87%)	715 (16,55%)	70 (1,62%)	76 (1,76%)

monitoring not recommended					
Abatacept	7 (1,17%)	50 (8,33%)	109 (18,17%)	2 (0,33%)	1 (0,17%)
Adalimumab	13 (0,67%)	201 (10,34%)	272 (14%)	18 (0,93%)	94 (4,84%)
Etanercept	36 (0,92%)	414 (10,55%)	561 (14,29%)	100 (2,55%)	110 (2,8%)
Golimumab	3 (1,7%)	4 (2,27%)	18 (10,23%)	1 (0,57%)	10 (5,68%)
Hydroxychloroquine	17 (0,51%)	508 (15,23%)	401 (12,02%)	43 (1,29%)	105 (3,15%)
Hydroxychloroquine & Adalimumab	0 (0%)	6 (4,76%)	14 (11,11%)	0 (0%)	0 (0%)
Hydroxychloroquine & Etanercept	0 (0%)	42 (16,41%)	29 (11,33%)	9 (3,52%)	10 (3,91%)
Infliximab	5 (0,87%)	33 (5,74%)	107 (18,61%)	5 (0,87%)	1 (0,17%)

Table S3. Number of very abnormal lt-RLTM test results retrieved in patient-DMARD exposure periods during which monitoring was and was not recommended, split per DMARD

	Monitoring recommended				
	ALAT	eGFR	Hb	WBC	Platelets
Azathioprine	0 (0%)	14 (2,08%)	5 (0,74%)	0 (0%)	0 (0%)
Azathioprine & Adalimumab	0 (0%)	6 (1,36%)	1 (0,23%)	0 (0%)	0 (0%)
Azathioprine & Etanercept	0 (0%)	10 (1,98%)	3 (0,59%)	0 (0%)	1 (0,2%)
Leflunomide	0 (0%)	20 (1,17%)	12 (0,7%)	0 (0%)	20 (1,17%)
Leflunomide & Adalimumab	0 (0%)	11 (3,55%)	6 (1,94%)	0 (0%)	0 (0%)
Leflunomide & Etanercept	0 (0%)	7 (1,13%)	2 (0,32%)	0 (0%)	1 (0,16%)
Leflunomide & Hydroxychloroquine	0 (0%)	1 (0,44%)	0 (0%)	0 (0%)	0 (0%)
Leflunomide & Rituximab	0 (0%)	9 (2,56%)	5 (1,42%)	0 (0%)	0 (0%)
Methotrexate	6 (0,03%)	469 (2,2%)	127 (0,6%)	2 (0,01%)	51 (0,24%)
Methotrexate & Abatacept	0 (0%)	6 (1,73%)	0 (0%)	0 (0%)	0 (0%)
Methotrexate & Adalimumab	1 (0,04%)	78 (2,94%)	11 (0,42%)	0 (0%)	10 (0,38%)
Methotrexate & Etanercept	0 (0%)	65 (1,36%)	27 (0,57%)	1 (0,02%)	27 (0,57%)
Methotrexate & Hydroxychloroquine	5 (0,06%)	161 (1,97%)	21 (0,26%)	1 (0,01%)	38 (0,47%)
Methotrexate & Leflunomide	1 (0,97%)	0 (0%)	0 (0%)	0 (0%)	3 (2,91%)
Methotrexate & Rituximab	0 (0%)	6 (0,65%)	2 (0,22%)	0 (0%)	0 (0%)
Methotrexate & Sulfasalazine	0 (0%)	9 (0,82%)	4 (0,36%)	0 (0%)	0 (0%)
Methotrexate & Tocilizumab	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Rituximab	0 (0%)	82 (3,94%)	19 (0,91%)	3 (0,14%)	15 (0,72%)
Sulfasalazine	0 (0%)	23 (1,21%)	12 (0,63%)	0 (0%)	3 (0,16%)
Sulfasalazine & Hydroxychloroquine	0 (0%)	3 (0,85%)	9 (2,55%)	0 (0%)	0 (0%)
Tocilizumab	1 (0,05%)	21 (1,14%)	22 (1,2%)	0 (0%)	53 (2,88%)
Other combination	0 (0%)	91 (2,11%)	40 (0,93%)	0 (0%)	2 (0,05%)

Monitoring not recommended					
Abatacept	0 (0%)	5 (0,83%)	6 (1%)	0 (0%)	0 (0%)
Adalimumab	2 (0,1%)	82 (4,22%)	28 (1,44%)	0 (0%)	63 (3,24%)
Etanercept	1 (0,03%)	143 (3,64%)	36 (0,92%)	2 (0,05%)	37 (0,94%)
Golimumab	0 (0%)	2 (1,14%)	0 (0%)	0 (0%)	2 (1,14%)
Hydroxychloroquine	3 (0,09%)	158 (4,74%)	20 (0,6%)	4 (0,12%)	14 (0,42%)
Hydroxychloroquine & Adalimumab	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Hydroxychloroquine & Etanercept	0 (0%)	12 (4,69%)	1 (0,39%)	0 (0%)	5 (1,95%)
Infliximab	0 (0%)	17 (2,96%)	3 (0,52%)	0 (0%)	0 (0%)

Comparative analysis

Table S4. Cumulative incidence with 95%-CI of patient-DMARD exposure periods varying from normal to abnormal during which monitoring was recommended

Medication	ALT	GFR	Hb	WBC count	Platelet count	Total patients
Azathioprine	2.5%	11.4%	19.2%	8.7%	3.7%	81
Azathioprine & Adalimumab	1.4%	10.1%	16.2%	2.9%	0%	69
Azathioprine & Etanercept	2.5%	9.0%	28.9%	12.7%	6.3%	79
Leflunomide	0.36%	7.6%	19.5%	3.6%	3.6%	278
Leflunomide & Adalimumab	0.69%	1.4%	9.2%	1.4%	0.69%	145
Leflunomide & Etanercept	1.8%	4.2%	15.3%	3.6%	5.4%	167
Leflunomide & Hydroxychloroquine	0%	4.8%	13.5%	4.8%	3.1%	127
Leflunomide & Rituximab	1.7%	6.1%	12.8%	3.4%	1.7%	116
Methotrexate	3.1%	15.2%	21.9%	3.3%	4.8%	2174
Methotrexate & Abatacept	1.5%	7.7%	20%	1.5%	0%	65
Methotrexate & Adalimumab	3.1%	10%	17.3%	1.5%	4.6%	481
Methotrexate & Etanercept	3.6%	9.1%	18.4%	5.4%	5.2%	642
Methotrexate & Hydroxychloroquine	2.9%	11.1%	16.6%	2.6%	3%	1572
Methotrexate & Leflunomide	3.4%	1.7%	8.5%	0%	0.8%	117
Methotrexate & Rituximab	1.5%	7.7%	25.5%	1.5%	1.5%	195
Methotrexate & Sulfasalazine	3.4%	9.1%	24%	2.8%	3.9%	178
Methotrexate & Tocilizumab	2.9%	11.6%	35.4%	16.2%	5.8%	69
Sulfasalazine	0.42%	9.7%	23.1%	2.9%	3.4%	238
Sulfasalazine & Hydroxychloroquine	1.2%	10.7%	23.8%	3.5%	2.4%	85
Combination other	1.8%	6.4%	22.5%	5.7%	3.9%	596

Table S5. Cumulative incidence with 95%-CI of variable patient-DMARD exposure periods varying from normal to abnormal during which monitoring was not recommended

Medication	ALT	GFR	Hb	WBC count	Platelet count	Total patients
Abatacept	2.8% (0.34-9.7)	10% (4.1-19.5)	22.4% (12.1-34.2)	0% (0-5.1)	0% (0-5)	72
Adalimumab	1.8% (0.59-4.2)	8.3% (5.4-12.2)	13.9% (10-18.6)	1.1% (0.22-3.1)	1.8% (0.59-4.1)	279
Etanercept	3.7% (2.1-5.9)	11.5% (8.6-14.8)	21.3% (17.5-25.4)	3.9% (2.3-6.2)	3.7% (2.1-5.9)	436
Golimumab	5.7% (0.7-19.2)	5.7% (0.7-19.2)	11.8% (3.3-27.4)	0% (0-10)	2.9% (0.07-14.9)	35
Hydroxychloroquine	0.56% (0.1-1.6)	12.3% (9.6-15.4)	15.2% (12.2-18.5)	3.2% (1.9-5)	4.3% (2.8-6.4)	534
Hydroxychloroquine & Adalimumab	0% (0-7.2)	2.1% (0.05-11.1)	10.2% (3.4-22.2)	0% (0-7.2)	0% (0-7.2)	49
Hydroxychloroquine & Etanercept	0% (0-4.3)	7.1% (2.7-14.9)	10.1% (4.5-19)	4.8% (1.3-11.9)	4.8% (1.3-11.7)	84
Infliximab	4.9% (0.6-16.5)	9.8% (2.7-23.1)	36.8% (21.8-54)	2.5% (0.06-13.2)	2.4% (0.06-12.9)	41
Rituximab	2.2% (0.71-5)	15.3% (10.9-20.6)	22.3% (17-28.3)	4.4% (2.1-7.9)	5.2% (2.7-8.9)	231
Tocilizumab	2.3% (0.63-5.8)	7.5% (4.1-12.5)	14.8% (9.8-21.1)	9.2% (5.4-4.6)	9.2% (5.3-14.5)	174

Table S6. Cumulative incidence with 95%-CI of variable patient-DMARD exposure periods varying to very abnormal during which monitoring was recommended

Medication	ALT	GFR	Hb	WBC count	Platelet count	Total patients
Azathioprine	0%	2.5%	3.8%	0%	0%	81
Azathioprine & Adalimumab	0%	5.8%	1.5%	0%	0%	69
Azathioprine & Etanercept	0%	1.3%	2.6%	0%	0%	79
Leflunomide	0%	2.2%	2.6%	0%	0.72%	278
Leflunomide & Adalimumab	0%	1.4%	0.71%	0%	0%	145
Leflunomide & Etanercept	0%	0%	1.2%	0%	1.2%	167
Leflunomide & Hydroxychloroquine	0%	0.79%	0%	0%	0%	127
Leflunomide & Rituximab	0%	3.5%	2.7%	0.86%	0%	116
Methotrexate	0.18%	4.3%	2.7%	0.09%	0.74%	2174
Methotrexate & Abatacept	0%	1.5%	1.7%	0%	0%	65
Methotrexate & Adalimumab	0.21%	3.3%	1.7%	0%	1%	481
Methotrexate & Etanercept	0%	2.5%	2.4%	0.16%	1.4%	642
Methotrexate & Hydroxychloroquine	0.25%	3.5%	1.1%	0.06%	1%	1572
Methotrexate & Leflunomide	0.85%	0%	0%	0%	0.85%	117
Methotrexate & Rituximab	0%	1.5%	1.1%	0%	0%	195
Methotrexate & Sulfasalazine	0%	2.8%	1.7%	0%	0%	178
Methotrexate & Tocilizumab	0%	0%	0%	0%	0%	69
Sulfasalazine	0%	1.7%	2.1%	0%	0.42%	238
Sulfasalazine & Hydroxychloroquine	0%	1.2%	5.9%	0%	0%	85
Combination, other	0.17%	2.5%	3.1%	0.17%	0.34%	596

Table S7. Cumulative incidence with 95%-CI of variable patient-DMARD exposure periods varying to very abnormal during which monitoring was not recommended

Medication	ALT	GFR	Hb	WBC count	Platelet count	Total patients
Abatacept	0%	1.4%	3%	0%	0%	72
Adalimumab	0.36%	3.3%	2.9%	0%	0.72%	279
Etanercept	0.23%	3%	3%	0.46%	1.4%	436
Golimumab	0%	2.9%	0%	0%	2.9%	35
Hydroxychloroquine	0.19%	4.5%	1.9%	0.56%	0.76%	534
Hydroxychloroquine & Adalimumab	2%	0%	0%	0%	0%	49
Hydroxychloroquine & Etanercept	0%	3.6%	0%	0%	1.2%	84
Infliximab	0%	9.8%	2.6%	0%	0%	41
Rituximab	0%	6.1%	4.1%	0.87%	0.43%	231
Tocilizumab	0%	2.3%	6.5%	0%	3.4%	174

Table S8. Number of cases and corresponding proportions of underlying disease pathology of the new very abnormal laboratory results (n=486)

Underlying pathology	*N (%)*
Renal and urinary disorders	129 (26.5%)
Injury, poisoning and procedural complications	89 (18.3%)
Infections and infestations	58 (11.9%)
Blood and lymphatic system disorders	50 (10.3%)
Neoplasms benign, malignant and unspecified	30 (6.2%)
Gastrointestinal disorders	23 (4.7%)
General disorders and administration site conditions	17 (3.5%)
Immune system disorders	16 (3.3%)
Cardiac disorders	11 (2.3%)
Metabolism and nutrition disorders	7 (1.4%,)
Reproductive system and breast disorders	5 (1%)
Respiratory, thoracic and mediastinal disorders	4 (0.82%)
Hepatobiliary disorders	4 (0.82%)
Congenital, familial and genetic disorders	2 (0.41%)
Unknown	41 (8.4%)

References

1. Ledingham J, Gullick N, Irving K, Gorodkin R, Aris M, Burke J, et al. BSR and BHPR guideline for the prescription and monitoring of non-biologic disease-modifying anti-rheumatic drugs. Rheumatology. 2017;56(6):865–8.
2. Holroyd CR, Seth R, Bukhari M, Malaviya A, Holmes C, Curtis E, et al. The British Society for Rheumatology biologic DMARD safety guidelines in inflammatory arthritis. Rheumatology. 2019;58(2):e3–42.
3. Rigby WFC, Lampl K, Low JM, Furst DE. Review of Routine Laboratory Monitoring for Patients with Rheumatoid Arthritis Receiving Biologic or Nonbiologic DMARDs. Int J Rheumatol. 2017;2017:1–15.
4. Tsakas JJ, Liew DFL, Adams CL, Hill CL, Proudman S, Whittle S, et al. Attitudes and practices in the laboratory monitoring of conventional synthetic disease modifying anti-rheumatic drugs by rheumatologists and rheumatology trainees. BMC Rheumatol. 2022;6(1):59.
5. Nakafero G, Grainge MJ, Card T, Mallen CD, Zhang W, Doherty M, et al. What is the incidence of methotrexate or leflunomide discontinuation related to cytopenia, liver enzyme elevation or kidney function decline? Rheumatology. 2021;60(12):5785–94.
6. Fraser SDS, Lin SX, Stammers M, Culliford D, Ibrahim K, Barrett R, et al. Persistently normal blood tests in patients taking methotrexate for RA or azathioprine for IBD: a retrospective cohort study. Br J Gen Pract. 2022;72(720):e528–37.
7. Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, et al. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Care Res Off J Am Coll Rheumatol. 2008;59(6):762–84.
8. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, et al. 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2012;64(5):625–39.
9. Singh JA, Saag KG, Bridges Jr SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol. 2016;68(1):1–26.
10. Singh JA. Treatment guidelines in rheumatoid arthritis. Rheum Dis Clin. 2022;48(3):679–89.
11. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020;79(6):685–99.
12. European Medicines Agency. The European regulatory system for medicines and the European Medicines Agency. 2014; Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Brochure/2014/08/WC500171674.pdf
13. Ranganathan P. Monitoring methotrexate toxicity: Are we being over-vigilant? Vol. 381, bmj. British Medical Journal Publishing Group; 2023.
14. Wood NAE, Farmer L, Croker R, Kyle S, Lewis T. Saving the planet with reduced routine DMARD blood monitoring frequency. bmj. 2023;382.
15. Shadeed A, Kattach L, Sam S, Flora K, Farah Z. Examining the safety of relaxed drug monitoring for methotrexate in response to the COVID-19 pandemic. Rheumatol Adv Pract. 2022;6(3):rkac100.
16. England N. Clinical guide for the management of Rheumatology patients during the coronavirus pandemic. Nhs. 2020;2:1–16.
17. UK Kidney Association. CKD stages [Internet]. [cited 2023 Jun 9]. Available from: https://ukkidney.org/health-professionals/information-resources/uk-eckd-guide/ckd-stages
18. Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. Lancet. 2013;381(9880):1845–54.
19. Sonsuz A, Bakkaloglu OK. Biomarkers in Liver Disease. Biomarkers in Medicine. 2022. 490–521 p.
20. U.S Department of Health and Human Services. Common Terminology Criteria for Adverse Events. Definitions. 2020;
21. Saquib N, Saquib J, Ioannidis JPA. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. Int J Epidemiol. 2015;44(1):264–77.
22. El Dib R, Tikkinen KAO, Akl EA, Gomaa HA, Mustafa RA, Agarwal A, et al. Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes. J Clin Epidemiol. 2017;84:61–9.

Chapter 7

General discussion

Main findings

Throughout this thesis, we have evaluated several diagnostic and prognostic tests used in RA care, including both laboratory tests and imaging modalities. We started with evaluating the value of routine radiographs of hands and feet at time of diagnosis (chapter 2). Thereafter, we investigated the diagnostic test accuracy of therapeutic drug monitoring in predicting response to a subsequent DMARD (chapter 3). Following that, we looked at the conceptual use of biomarkers in RA treatment in general and investigated the additional value of a hypothetical biomarker in predicting response to RA treatment (chapter 4). We then shifted focus to the benefits and risks of unexpected findings of FDG-PET/CT scans to measure arthritis in patients with RA (chapter 5). Lastly we explored toxicity monitoring and examined the value of long-term routine laboratory monitoring during long-term DMARD use (chapter 6).

Main finding 1: The prevalence of RA-associated erosions in patients with newly presenting arthritis suspected of RA was low, and rarely led to a change in diagnosis or prognosis, even in relevant subgroups (chapter 2).

Main finding 2: In contrast to earlier studies, our findings indicate that measuring adalimumab drug levels and anti-drug antibodies in RA patients who had previously failed adalimumab treatment does not predict their response to a subsequent (non) TNFi therapy (chapter 3).

Main finding 3: The use of a biomarker to predict response to a b/tsDMARD treatment in RA could theoretically add value to current treat-to-target clinical care. However, the gains in efficacy are modest, and cost gains are almost exclusively depending on a combination of successful early and proactive medication tapering that reduce high medication costs (chapter 4).

Main finding 4: Whole-body FDG-PET/CT scanning in RA patients for imaging of arthritis frequently yields incidental extra-articular findings that are false positive, but also misses cancers that do develop within 3 years after the scan (chapter 5).

Main finding 5: The likelihood of finding 'very abnormal' laboratory results with long-term routine laboratory toxicity monitoring was very low, and not different for DMARDs for which monitoring is recommended compared to those for which such a recommendation has not been issued. 'Very abnormal' laboratory results were almost all accompanied by indicator(s) suggesting inappropriate use of tests (chapter 6).

In summary, the diagnostic tests examined in this thesis have consistently demonstrated minimal additional clinical value within the specific contexts they were applied or intended for. Yet, the question arises: why are these tests still extensively used in clinical practice? And might this occur on a wider scale beyond the scope of this thesis?

Throughout this discussion, I will begin by exploring how we have arrived at a situation where many routine diagnostic tests are used despite offering minimal clinical benefit in their respective settings. Next, I will discuss why this widespread usage is problematic, and examine the barriers that hinder efforts to 'de-implement' these tests. I will then offer insights into how we can systematically evaluate the clinical value of existing tests within specific contexts, without the need for prospectively gathered data and trials, by presenting a structured guide supported by examples drawn research from my thesis.

Introducing new diagnostic tests

In the European Union, the Medical Devices Regulation (MDR) and In Vitro Diagnostic Devices Regulation (IVDR) prioritize ensuring the safety, performance, and quality of medical devices, including diagnostic tests (1–3). The IVDR outlines five distinct phases of diagnostic research that must be followed before a new diagnostic test can be implemented in clinical practice (Figure 1).

In phase 1 research will focus on evaluating validity of the test results: do test results in patients with the target disorder differ from those without?

Phase 2: If a test seems capable of distinguishing people with and without the target condition, its analytical performance is to be tested, defined by outcome measures such as sensitivity, specificity, accuracy, precision, repeatability and reproducibility.

Phase 3: If the test has passed phase 2 it will be further investigated within a clinically representative population (4–6). The study design used in phase three is a Diagnostic Test Accuracy (DTA) study where all patients who would be tested in routine care are enrolled and tested with the new index test. All patients should also undergo the reference standard test (either a gold standard or other tests that are already used in routine care). Assessment of the outcome should be done while blinded for both the results of the index test and the reference standard. Thereafter, the diagnostic accuracy of the index test can be determined. A DTA study could take form as a cohort study, a case-control study or a randomized design (7).

Phase 4: in this stage researchers will assess whether patient-related health outcomes (risk of disease, risk of death and/or quality of life (QoL) that follow after further testing and treatment that the test results induce will improve. The study design best suited to assess the diagnostic test in this fourth phase is a Test Treatment Trial (TTT) (8).

Phase 5: During or after phase 4 cost-effectiveness of the test will be evaluated by calculating incremental cost-effect ratios (ICER), costs per quality adjusted life-year (QALY) and Net Monetary Benefits (NMB). This could be done within TTT's or by means of modelling studies (6).

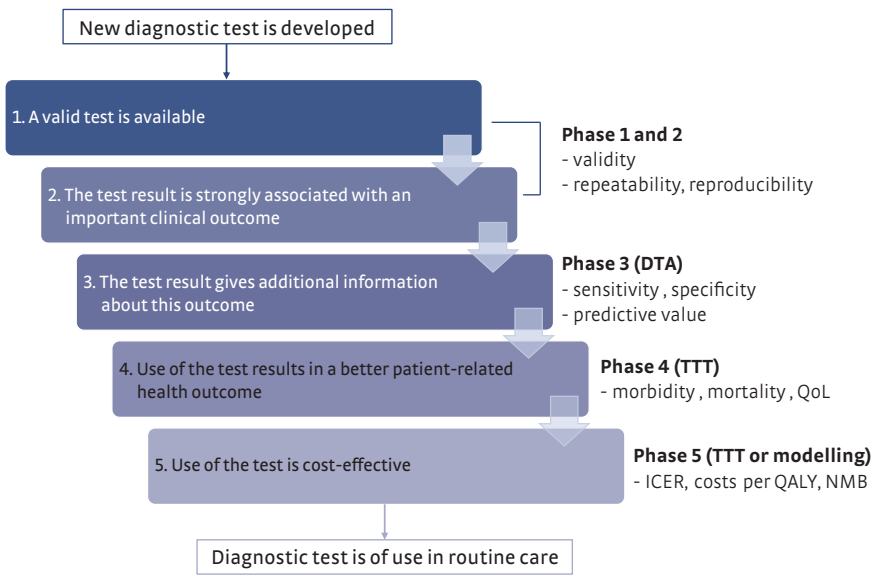

Figure 1. *summary of requirements for rational testing embedded in the phases of diagnostic research, along with corresponding outcomes and study designs for each phase.*

A difficulty seen in TTT's is the outcome measure: evaluation of patient-related outcomes differs per clinical path; when the test differentiates between a life-threatening and non-life threatening morbidity, patients will receive lifesaving treatments and patient-related benefit seems evident. However, when the target condition has a relatively benign clinical course, does not require immediate active treatment, or when the test is used for early detection of asymptomatic disease (screening), the patient-related benefits are less obvious (6). A common error in test treatment trials is relying solely on surrogate markers, without correlating them with

clinical symptoms. Such an approach simply assumes that improvements in surrogate markers unequivocally translate to better patient outcomes, which may not always be the case. Another disadvantage of TTT's is that they may only be conducted when there is certainty about the appropriateness of treatment. Furthermore, most trials focus on small group differences, making it challenging and costly to achieve sufficient statistical power to demonstrate significant differences between the two arms. Lastly, ethical issues may be raised when letting patients undergo 'unproven' diagnostic tests (9). Aforementioned problems are seen in outcomes of existing TTT's. To date, only about 1,000 of these trials have been conducted across all medical fields. Among them, 10% showed positive results and merely 2% showed positive results on reducing mortality when comparing the index test to the reference standard (10,11).

Ideally, a new diagnostic test would go through all five phases of research outlined by regulatory bodies like the EMA before being implemented, ensuring that the test demonstrates meaningful patient-related health outcomes. In practice, however, this comprehensive process is often not fully adhered to. Diagnostic tests are frequently implemented without completing all five phases, or studies may rely on surrogate outcomes rather than direct patient-related health benefits. This can occur due to precautionary reasons, such as the urgent need for diagnostic tools in certain medical situations, or because long-term trials with meaningful patient-related health outcomes are expensive and impractical.

Once a diagnostic test is integrated into routine care, it is often used for an extended period without sufficient ongoing evaluation of its diagnostic value. This lack of reflection can stem from practical challenges, including the high costs and logistical difficulties of continuous assessment, as well as regulatory gaps in enforcing these guidelines. Therefore, despite the existence of strict EMA guidelines, real-world implementation often deviates from this ideal, necessitating a closer examination of current practices and their implications for patient care.

The dynamic nature of test appropriateness

After implementation of a test, its appropriateness is not fixed, and therefore intermittent reflection of testing appropriateness is necessary. The dynamic nature of test appropriateness can be described by examples found within my thesis. These examples illustrate that, with improvements of care over time, less testing or monitoring is justifiable, which makes it worthwhile to intermittently evaluate the value of current testing strategies. For instance, at times tests are implemented to investigate abnormalities that were once prevalent. These abnormalities could then

become less common over time due to improved and pro-active treatment. This makes testing within the same population less relevant, despite the test itself not changing (radiographs of hands and feet, chapter 2). In other cases, as is seen with routine toxicity monitoring (chapter 5), a cautious approach is implemented as the long-term toxicity rates during use of a new treatment were unclear. If the long-term toxicity rates are never reviewed, (guidelines on) monitoring intervals will not be adjusted to what is strictly needed, and consequently a high rate of testing is sustained without any patient-related health gain. What adds to this argument is the fact that sometimes treatment doses can be reduced significantly over time without losing efficacy (12). Lower dosages of drugs will likely also lead to lower toxicity rates, decreasing the need for toxicity monitoring.

The evolution of clinical strategies may also redirect focus areas of research. A good example is the treat-to-target strategy now paramount in rheumatology. T2T implies pro-actively responding in terms of treatment kind and intensity in case of ineffective treatment. Such a clinical strategy with with proven effectiveness may significantly reduce the need for, and the theoretical impact of, therapy-guiding biomarkers in practice. Consequently, the allocation of time and resources for continuously and relentlessly exploring potential new therapy-guiding biomarkers (as discussed in chapter 4) may become less justified.

The aforementioned examples illustrate that the appropriate use of tests depends not only on the tests themselves but also on the context in which they are employed. This context is not static and is likely to change over time. It is also not uniform worldwide, as some healthcare systems exhibit less accessible healthcare, longer waiting times, increased travel distances, delays in initiating and/or modifying treatment plans, and fewer treatment options. Therefore, the added clinical value and appropriateness of a diagnostic test is influenced by the context in which it is used. On the other hand, excessive testing can also have significant effects on various aspects of our healthcare system.

Increasing pressures on healthcare systems

As individuals, we will all inevitably face health problems. When that time comes, we rely on the availability of timely, high-quality, local, and affordable care. However, the sustainability of our healthcare system is under increasing pressure due to developments such as aging, the emergence of ever more new healthcare technologies that should be accessible to all, and the rise in number of people with chronic illnesses due to technologies and changes in our lifestyles. The demand for healthcare is

outpacing its supply, leading to an increasing pressure on healthcare workers and jeopardizing quality of care (longer wait times and reduced accessibility to services). This trend is not sustainable; without intervention, projections indicate that one in three individuals will need to work in healthcare within forty years, compared to the current ratio of one in seven (13,14).

Overall, one contributing factor to the strain on healthcare systems is excessive testing, with routine tests serving as a major contributory element. They are often conducted without medical necessity as they are performed in the absence of symptoms and abnormalities are not expected. This practice carries the risk of overdiagnosis; an abnormality or condition is identified that would never have caused any symptoms or mortality risk. Consequently, treating such conditions can lead to unnecessary medical interventions. What begins as a simple routine test can trigger a series of additional tests and treatments. Thus, routine tests not only constitute a significant portion of diagnostic procedures but also significantly strain healthcare systems.

Ultimately, the most sustainable healthcare is reached by eliminating care that should never have been provided in the first place. The healthcare system should be stimulated to save scarce goods such as money and personnel by limiting its use to what is really needed. Only then will the workload remain manageable and healthcare accessible. For effective de-implementation of low value care, however, it is important that patients and healthcare providers recognize its importance and benefits of de-implementation of inappropriate testing.

What drives healthcare providers to conduct more tests than necessary?

Reasons for excessive use of tests can be categorized under three pillars (15):

- Environment/context: Guidelines and protocols, time constraints, physical vulnerabilities and language barriers, availability and ease of access to tests, contemporary medical practice, and new technology.
- Interpersonal: Pressure from patients and the doctor-patient relationship, as well as pressure from colleagues and medical culture.
- Intrapersonal: Fear of malpractice and litigation, knowledge and understanding, biases, and personal experiences.

These factors collectively contribute to the challenges physicians face. Specifically, growing time constraints and the increasing availability and ease of access to tests and new technologies make it increasingly difficult for physicians to avoid indulging in excessive testing.

Patients are able to extensively inform themselves about the conditions that might be responsible for their symptoms. However, this often unvalidated information usually does not provide the desired reassurance, but rather contributes to their concerns (16,17). A growing proportion of patients have strong opinions about which tests are necessary and which treatment would be most suitable (18). And while this could be a good thing, as we strive towards a patient-physician relationship in which shared decision making is performed, there is a second group of patients that visits healthcare providers not with the need for a diagnosis or treatment, but to rule out disease. A group of patients referred to as the 'worried-well' (19) has seen a significant increase in the desire for various diagnostic tests in recent years. However, this pursuit of 'diagnostic reassurance' often results in overdiagnosis, overtreatment, and increased healthcare consumption per patient, without providing additional reassurance (20). Physicians, driven by an inherent desire to alleviate patient fears and to avoid risking the doctor-patient relationship, may hesitate to refuse such requests. As Welch notes in his book, "physicians prefer to act rather than to wait," a tendency that reflects their inclination to order tests rather than delay (21). Additionally, communicating the absence of value in certain tests often requires more time and effort than simply ordering the tests themselves. Consequently, physicians find themselves drawn into fulfilling these demands, even when the benefit of the tests is questionable.

To assist physicians in avoiding inappropriately performing diagnostic tests, we need guidelines that are less defensive, clearly outlining recommended actions based on a priori probabilities (as detailed on previously described platforms). Phrases such as 'in case of doubt' should be avoided. The occasional possibility of missing a rare condition is a calculated risk that should not be personally attributed to physicians and we should not solely focusing on condemning the physician who did not immediately detect a tumour, but also attending to physicians who have ordered numerous unnecessary tests, along with their associated risks (22). This could be done within departments by reviewing which physician is ordering a high volume of tests and critically evaluate significant discrepancies in test ordering practices.

Gaining a deeper understanding of both patients' and physicians' perspectives on reducing the use of diagnostic tests is essential. What are the perceived advantages and disadvantages from their viewpoints? Why might they hesitate to embrace this

change, and what underlying fears drive their concerns? Addressing these worries and concerns is essential in achieving consensus, which is necessary for initiating change. Open discussions on this topic must be facilitated and prioritized to achieve equipoise effectively. Choosing Wisely campaigns play an active role in educating patients and physicians, stimulating conversations about the necessity of tests, treatments, and procedures (22). While this is a positive step, there is still a long way to go. Engaging healthcare providers and patients, however, is crucial. Without their active participation, systemic changes will not occur, even if studies indicate which tests could be de-implemented.

Navigating evaluation and de-implementation

Having established the widespread inappropriate use of tests and its significant impact on clinical practice, the next consideration is: where should we start evaluating tests currently in use? Assessing all existing tests is a significant task, requiring collaboration between several parties. To facilitate this collaboration and simplifying the process, quantifying the evaluation of test appropriateness is essential. Risk acceptability serves as an appropriate measure for this purpose.

Risk acceptability can be determined by offsetting the probability of the clinical outcome of interest occurring in the target population against the potential health gain from testing. It is crucial to base the probability of clinical outcomes on real data to avoid availability errors clouding judgment. Tests performed in populations where the target disease is rare and interventions have a low potential health impact should be prioritized for investigation. Among these, tests with the highest burden on patients, the healthcare system, and the environment should be given priority.

While expensive tests often attract attention, inexpensive tests that are frequently used can also accumulate significant costs and burdens over time. Importantly, the dynamics of risk acceptability differ between group-level and individual-level interactions. At the group level, it may be feasible to reduce costs by eliminating unnecessary routine lab monitoring from guidelines, potentially leading to cost savings without compromising health outcomes. However, managing the unnecessary ordering of high-impact tests, such as PET or MRI scans, presents a much more complex challenge. The unintended health consequences and costs associated with these tests underscore the need for careful consideration and targeted strategies to address unnecessary testing on an individual level.

To tackle excessive testing effectively, a structured de-implementation approach is essential. This approach requires the involvement of healthcare professionals, policy makers, healthcare administrators, and patients. Healthcare providers should be critical in ordering and interpreting tests, while policy makers and administrators are needed to develop and enforce new guidelines. Understanding patients' perspectives is also vital to address their expectations and concerns about diagnostic tests.

The de-implementation process begins with a comprehensive review of evidence. Data on (in)appropriateness should be gathered in a manner easily accessible to researchers, clinicians, and field experts involved in evidence review and clinical practice recommendations. An example of such accessibility is the appropriateness criteria platform provided by the American College of Radiology (23). This platform offers easy access, and both numeric and visual display of: 1. Investigated tests, 2. Study findings, and 3. Conclusion robustness (24). Efficient data gathering enables collaborative work, facilitates quick review for new clinical recommendations and aids to identify knowledge gaps. Data monitoring systems are necessary to track the impact on patient outcomes and healthcare costs, ensuring achievement of intended benefits. Additionally, continuous feedback mechanisms should be established to refine and adjust the de-implementation process as needed.

Integrating de-implementation strategies involves both local and broader efforts. Pilot projects within specific hospitals/institutions can assess the feasibility and impact of removing certain tests from routine use. National and regional policies should be developed to support the removal of low-value tests and encourage adherence to updated guidelines. On a global scale, sharing successful strategies and evidence can help other regions and countries implement similar practices.

A guide for assessing appropriateness of existing diagnostic tests

Once knowledge gaps are identified, existing diagnostic tests with uncertain clinical value can be evaluated. The following section will outline insights gained from my PhD research and provide guidance on how diagnostic tests that are integrated in current practice can be evaluated. Ideally, the evaluation of a tests value would employ a DTA or TTT design. However, as demonstrated by the studies in my thesis, similar outcomes can be achieved through less costly and more practical methods, by using existing data from routine clinical care. When working with existing data, we have noticed that it is important to consider several factors:

1. Use clearly defined research questions and operationalize these questions to capture the essence of what is tested. While this may seem obvious, it could pose challenges, particularly in the absence of prior research with similar methods or established frameworks. The following framework with requirements for rational testing and practical examples will may provide some guidance in this process.
2. Have a thorough understanding of the data and data registries. Be aware of systematic information gaps, such as missing data or the need to merge datasets. When working with big data, it's crucial to identify data pitfalls beforehand since they may not be apparent when dealing with large volumes of data.
3. Collaboration between data experts, clinicians and researchers is essential. These parties should be able to understand each other's needs to ensure successful data extraction and interpretation of results.

These factors underscore the necessity for collaboration between healthcare specialists experienced in clinical practice, and scientists that have an understanding and guard of the structure of large datasets. The former are able to identify pitfalls in current routine care, while the researchers oversee the process of collecting data, ensuring completeness and validity of datasets containing data from routine care.

Currently, there is no established framework for evaluating diagnostic tests that are integrated in clinical practice without making use of standard DTA/TTT designs. Therefore, I drafted a systematic guide for efficiently evaluating existing diagnostic tests by adapting the IVDR framework for the implementation of new diagnostic tests. Figure 2 provides an overview of this adapted framework.

A test should meet all requirements for rational testing to be suitable for the context in which it is applied. I will demonstrate the significance of each requirement through examples drawn from my PhD research that are listed below in textboxes. Before proceeding, I would like to clarify that our evaluation of tests solely focusses on routine testing. The mentioned tests may all hold value when used on indication based on signs or symptoms or in selected patient groups, but the recommendation for their routine implementation could be reconsidered.

Requirement 1: a valid test is available

In this scenario, it is assumed that a valid diagnostic test is both available and used in clinical care. Generally, this requirement is met for tests used in clinical practice. However, if for example a test is administered several times and yields different results each time without any change in the patient's condition, this suggests that the test may not be reliable. Similarly, if for example a biopsy is evaluated by different pathologists and produces varying diagnoses for the same sample, this further

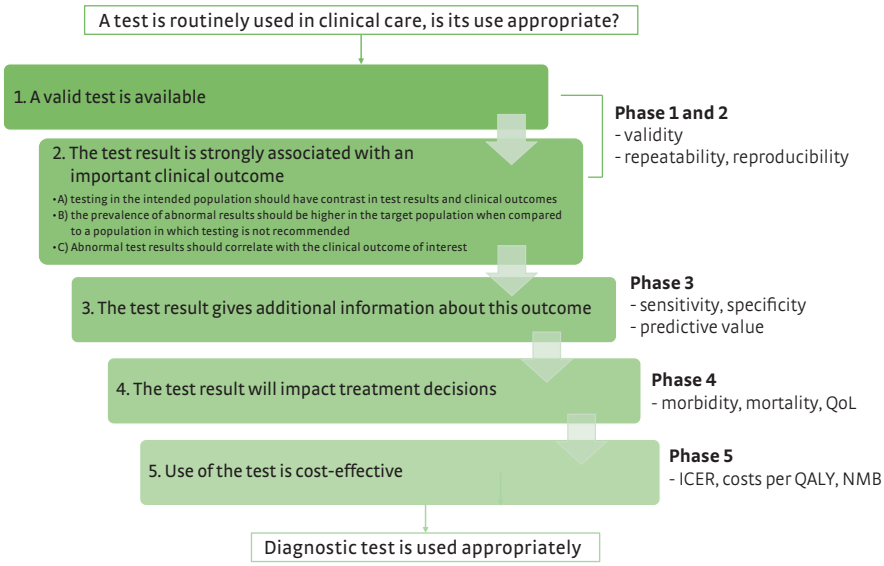

Figure 2. *framework for assessing implemented diagnostic tests for their appropriate use.*

suggests that the test may lack validity and may not be suitable for clinical decision-making. The diagnostic tests that were studied in my thesis had already shown to be valid, and were therefore not further tested on this requirement within my research.

Requirement 2: the test result is strongly associated with an important clinical outcome

A: testing in the intended population should have contrast in test results and clinical outcomes

If the target population is never or consistently affected by the clinical morbidity, and/or the test always or never yields abnormal results, it becomes impossible to establish a meaningful association between the test result and the clinical outcome. In such cases, the test loses its accuracy as a predictor of the clinical outcome.

The study in chapter 2 of my thesis shows that erosions are not prevalent in the target population, even within relevant subgroups in which we would suspect a higher risk for RA-associated erosions. Even if erosions were a perfect predictor for diagnosing RA or indicating a more severe disease course, routinely conducting radiographs of hands and feet at this stage of the disease is not beneficial if erosions are rarely or never detected on them.

B: the prevalence of abnormal results should be higher in the target population when compared to a population in which testing is not recommended

If the likelihood of the clinical outcome occurring in the target population is very low, resembling that of the non-affected general population, or a population in which testing is not recommended. In a population in which testing is not recommended, tests can also produce abnormal results. If the investigated test is not used as a screening tool in the general population, it's reasonable to avoid its routine use in a target population with a similar likelihood of obtaining abnormal test results.

In chapter 5 we found that very abnormal test results during routine laboratory monitoring during long-term DMARD use were rare. Additionally, very abnormal test results were present in similar rates in a control condition that was routinely tested even though this was not recommended.

C: abnormal test results should correlate with the clinical outcome of interest

Even when a test distinguishes between normal and abnormal results in the target population and the target population occasionally exhibits a clinical outcome, consistent correlation between test results and clinical outcomes is not guaranteed. Some abnormal test results could be distinguished as unintended findings indicating other health problems than the test intended to show. While it may be relevant to identify these abnormalities and consider routine testing or screening in specific patient groups where these abnormalities are commonly observed, maintaining routine testing in the whole target population is not justified.

In chapter 6 we found a large number of extra-articular abnormalities when scanning RA patients for arthritis. However, none of the suspected malignant lesions were confirmed or developed into a malignancy, and conversely, none of the patients who developed a subsequent malignancy had a positive scan.

Requirement 3: the test result gives additional information about the important clinical outcome

After confirming that the test results are associated with an important clinical outcome, the clinical route in which the test is used should be considered. Will patients that undergo this test receive additional testing before, at the same time or after the test that is under evaluation? Does it provide extra information compared to the other tests that are done? Or does it tell you exactly the same? If results of existing tests (e.g. physical examination or history taking) and results of the diagnostic test correlate perfectly in identifying the clinical outcome both now and in the future, using all the tests will not provide additional information. One of the tests in the clinical route can probably be eliminated as it does not add to the diagnostic- or treatment process.

In chapter 5, very abnormal laboratory outcomes during routine toxicity monitoring are often already known. For example, there was a previous hospitalization during which the same laboratory parameter was tested, or the patient is monitored by a general practitioner.

Requirement 4: the test results will impact treatment decisions

A useful test could lead to finding or excluding a clinical diagnosis, or aids in guiding treatment decisions (e.g. in selecting medications or indicating when treatment dosage should be altered). In case of monitoring it could also lead to other actions (i.e. referral, starting other medications) to make sure the patient will not endure toxic effects due to treatment.

In Chapter 2, the presence of RA-associated erosions on radiographs at initial presentation is aimed at providing diagnostic and prognostic certainty. However, diagnostic and prognostic criteria for RA typically reflect prolonged disease activity, with erosions developing in such cases. Therefore, detecting erosions on radiographs merely confirms the presence of longstanding disease activity, offering limited additional information. Moreover, treatment plans are unlikely to be altered based solely on radiographic erosions, as intensive treatment is already initiated based on other indicators of poor prognosis.

Requirement 5: use of the test is cost-effective

An ineffective test can never be cost-effective, but an effective test is still not necessarily cost effective. However, it's vital to recognize that cost-effectiveness isn't solely determined by test effectiveness; potential health gains and costs saved are crucial factors to consider.

In Chapter 4, we found that the cost-effectiveness of a therapy-guiding biomarker depends significantly on factors beyond its characteristics. Firstly, because current treat-to-target strategies for RA are already very effective, there is limited potential for additional health benefits from a new biomarker. Secondly, high costs of biologic agents used in RA treatment form a large part of overall expenses. A biomarker-guided approach can save costs by achieving low disease activity or remission faster, allowing for medication tapering. However, if the costs of biologic medications were to drop, for instance, due to the introduction of biosimilars, the cost-saving impact of the biomarker would also decrease.

Lastly, it is also important to consider the area where research towards future tests should focus. If new test were developed that is able to offer valid and timely results at a reasonable cost and with practical feasibility, would it significantly alter the diagnostic and/or therapeutic process? If not, should research continue to focus on this test?

The study in chapter 4 illustrates this problem as well: the use of a biomarker for prediction of response to a b- or tsDMARD treatment in RA can be of added value to RA care. However, gains in efficacy are modest due to current treat-to-target clinical care, even with a biomarker with optimal characteristics.

Overall, none of the diagnostic tests reviewed in my thesis met all the necessary requirements for rational testing, and the majority failed to fulfill several criteria. As meeting all the criteria is necessary for a test to be used appropriately (a chain of necessary causes), there is no need for a DTA study or a test-treatment trial in this scenario.

In conclusion, this thesis assessed the clinical utility of various diagnostic and prognostic tests in managing RA, finding that many routine tests, such as radiographs, therapeutic drug monitoring, biomarker testing, and long-term laboratory monitoring, offer minimal additional clinical value. Imaging techniques like FDG-PET/CT scans often yield incidental findings that are not clinically relevant. Despite these limitations, these tests are widely used, leading to overdiagnosis and increased strain on healthcare systems.

Excessive diagnostic testing is fuelled by the implementation of new tests without comprehensive test treatment trials and inadequate evaluation of their efficacy post-implementation. Additional factors include defensive guidelines, patient expectations, medical culture, and legal concerns.

Recommendations for future research:

1. Regularly re-evaluate the clinical utility and cost-effectiveness of diagnostic tests by means of a standardized framework (for which the last part of this discussion is a draft).
2. Establish a central database visualizing levels of evidence for tests, finding current knowledge gaps and allowing continuous updates on test evaluation.
3. Develop strategies for the de-implementation of low-value tests, supported by national and regional policies, and encourage adherence by drafting evidence-based guidelines.
4. Conduct pilot projects to assess the feasibility of reducing unnecessary tests in clinical practice and encourage the global adoption of successful strategies.
5. Investigate the attitudes of patients and healthcare providers towards diagnostic testing to identify barriers and facilitate change.

References

1. European Commission. Implementation Model for In-Vitro Diagnostic Medical Devices Regulation Step by Step Guide. Whitepaper. 2018;(Article 15):1–2.
2. US Food & Drug Administration. Overview of IVD Regulation. 2021.
3. Health and Youth Care Inspectorate. European regulations for medical devices and IVDs [Internet]. 2023 [cited 2024 May 20]. Available from: https://english.igj.nl/medical-technology/new-european-regulations-mdr-and-ivdr
4. Gluud C, Gluud LL. Evidence based diagnostics. Bmj. 2005;330(7493):724–6.
5. Sackett DL, Haynes RB. The architecture of diagnostic research. Bmj. 2002;324(7336):539–41.
6. Leeflang MMG, Allerberger F. How to: evaluate a diagnostic test. Clin Microbiol Infect. 2019;25(1):54–9.
7. Chassé M, Fergusson DA. Diagnostic accuracy studies. In: Seminars in nuclear medicine. Elsevier; 2019. p. 87-93.
8. Hot A, Bossuyt PM, Gerke O, Wahl S, Vach W, Zapf A. Randomized test-treatment studies with an outlook on adaptive designs. BMC Med Res Methodol. 2021;21(1):110.
9. Chan ECY. Promoting an ethical approach to unproven screening imaging tests. J Am Coll Radiol. 2005;2(4):311–20.
10. Siontis KC, Siontis GCM, Contopoulos-Ioannidis DG, Ioannidis JPA. Diagnostic tests often fail to lead to changes in patient outcomes. J Clin Epidemiol. 2014;67(6):612–21.
11. El Dib R, Tikkinen KAO, Akl EA, Gomaa HA, Mustafa RA, Agarwal A, et al. Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes. J Clin Epidemiol. 2017;84:61–9.
12. den Broeder AA, van Herwaarden N, van der Maas A, van den Hoogen FHJ, Bijlsma JW, van Vollenhoven RF, et al. Dose REduction strategy of subcutaneous TNF inhibitors in rheumatoid arthritis: design of a pragmatic randomised non inferiority trial, the DRESS study. BMC Musculoskelet Disord. 2013;14(1):1–9.
13. Wetenschappelijke Raad voor het Regeringsbeleid. Kiezen voor houdbare zorg. Mensen, middelen en maatschappelijk draagvlak. 2022.
14. Bestuur NVvR. Grenzen aan de zorg! 2021;
15. Lam JH, Pickles K, Stanaway FF, Bell KJL. Why clinicians overtest: development of a thematic framework. BMC Health Serv Res. 2020;20(1):1–11.
16. Kwakernaak J, Eekhof JAH, De Waal MWM, Barenbrug EAM, Chavannes NH. Patients' use of the internet to find reliable medical information about minor ailments: vignette-based experimental study. J Med Internet Res. 2019;21(11):e12278.
17. Bujnowska-Fedak MM, Węgierek P. The impact of online health information on patient health behaviours and making decisions concerning health. Int J Environ Res Public Health. 2020;17(3):880.
18. Winterberg DH, Krol L. Effectief geruststellen. Medisch Contact. 2005;
19. Garfield SR, Collen MF, Feldman R, Soghikian K, Richart RH, Duncan JH. Evaluation of an ambulatory medical-care delivery system. N Engl J Med. 1976;294(8):426–31.
20. Giard RW. Screening: careful considerations versus commercial medicine. Ned Tijdschr Geneeskd. 2003;147(39):1893–6.
21. Welch HG, Schwartz L, Woloshin S. Overdiagnosed: making people sick in the pursuit of health. beacon press; 2012.
22. AIBM foundation. Choosing Wisely®: A watershed moment in health care [Internet]. 2023 [cited 2024 May 20]. Available from: https://www.choosingwisely.org/#:~:text=The goal of the ABIM,and which ones are not
23. American College of Radiology. ACR AC Portal. 2024.
24. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Mak. 2007;27(5):696–713.

Dutch summary (Nederlandse samenvatting)

70% van alle medische beslissingen wordt beïnvloed door laboratoriumonderzoek, en 70% van je medische dossier bestaat uit laboratoriumgegevens. Sinds 1996 worden deze opvallende cijfers gebruikt om het belang van laboratoriumtests te benadrukken. Zonder te diep in te gaan op de wetenschappelijke onderbouwing van deze cijfers, geven deze uitspraken wel aan dat diagnostische tests een cruciale rol spelen in de moderne geneeskunde. Medische diagnostische en prognostische tests hebben een breed scala aan toepassingen en worden veelvuldig gebruikt. Voorbeelden hiervan zijn bloedtests, röntgenfoto's of PET-CT-scans, en weefselbiopsieën. Met de ontwikkeling van de moderne geneeskunde zijn talloze tests ontwikkeld en geïntegreerd in behandelrichtlijnen en protocollen voor patiëntenzorg.

Diagnostische tests hebben als voordeel dat ze een bijdrage leveren aan het aantonen of uitsluiten van een aandoening. Daarnaast kunnen tests helpen bij het stellen van een diagnose en inschatten van de prognose van een aandoening. Ook kunnen ze richtinggevend zijn bij het opstellen van behandelplannen. Het gebruik van veel tests brengt echter ook nadelen met zich mee. Het gebruik van veel tests brengt risico's zoals bloedingen of nierschade door contrast met zich mee, ook veroorzaakt het hoge kosten, druk op zorg en milieu en kan het leiden tot toevalsbevindingen. Ook kan er sprake zijn van onnauwkeurigheden in testuitslagen. Dit kan leiden tot vals-positieve of vals-negatieve resultaten, wat op zijn beurt kan resulteren in onnodige vervolg tests met nadelige gevolgen, zoals blootstelling aan straling van een scan en soms zelfs onnodige invasieve procedures zoals een biopt of operatie. Daarnaast brengt het ondergaan van tests een belasting met zich mee voor de patiënt. Patiënten moeten reizen, tijd vrijmaken en de (soms pijnlijke) tests ondergaan. Veelvuldig testen belast het medisch personeel, zet het de organisatie van de gezondheidszorg verder onder druk en heeft het een negatieve impact op het milieu. Hierom is het belangrijk om het gebruik van diagnostische en prognostische testen te beperken tot de tests die een duidelijke waarde toevoegen aan de patiëntenzorg.

In dit proefschrift werd de klinische toegevoegde waarde van verschillende diagnostische en prognostische tests geëvalueerd, met specifieke focus op tests die worden gebruikt in de routinematige zorg voor patiënten met reumatoïde artritis (RA). De belangrijkste uitkomsten zullen hieronder per hoofdstuk beschreven worden.

Hoofdstuk 2: Routinematige röntgenfoto's van handen en voeten

Voor de klinische diagnose van RA kan een reeks diagnostische methoden worden gebruikt. Allereerst arts-patiëntgesprek (anamnese) en lichamelijk onderzoek van belang om de duur en aard van symptomen en aanwezigheid van gewrichtsontste-

kingen (artritis) vast te stellen. Deze factoren zijn ook cruciaal in de 2010-criteria van het American College of Rheumatology (ACR)/European Alliance of Associations for Rheumatology (EULAR), samen met de aanwezigheid van verhoogde ontstekingswaarden in het bloed, en Reumatoïde Factor (RF) en anti-citrulline-antistoffen (anti-CCP), beide antistoffen passend bij RA. Deze criteria kunnen gebruikt worden ter ondersteuning van de diagnose RA. Het maken van röntgenfoto's van zowel handen als voeten wordt ook aanbevolen bij het onderzoek van patiënten met artritis die klinisch verdacht worden van RA. Deze aanbeveling werd aan de richtlijnen toegevoegd omdat het vinden van botafbraak door RA (erosies) kan helpen bij het stellen van de diagnose en bij het voorspellen of de ziekte ernstiger zal verlopen (prognose). Bij een ernstiger beloop zal er namelijk intensievere behandeling gestart moeten worden. De aanvullende diagnostische en prognostische waarde van specifiek de erosies die passen bij RA, onafhankelijk van de andere voorspellers die worden meegenomen in de criteria, is nog niet onderzocht. Resultaten van bestaande studies laten zien dat het routinematig maken van röntgenfoto's van handen en voeten misschien niet veel extra waarde biedt.

Om dit te onderzoeken hebben we een onderzoek uitgevoerd met data van patiënten die zich voor het eerst bij de reumatoloog meldden met artritis die verdacht was voor RA. Patiënten werden mochten meedoen aan het onderzoek als er een gewrichtsontsteking aanwezig was op het moment waarop de diagnose werd gesteld, de RF, anti-CCP en ontstekingswaarden in het bloed gemeten waren, RA werd genoteerd in het lijstje van mogelijke diagnosen (de differentiële diagnose), en röntgenfoto's van handen en voeten werden gemaakt. Uitkomsten van het onderzoek waren het aantal gevallen van een of meer erosies die passend waren bij RA, en hoe vaak de diagnostische of prognostische classificatie werd veranderd door de aanwezigheid van deze erosies. Patiënten bij wie de RF en anti-CCP niet verhoogd waren, patiënten zonder verhoogde ontstekingswaarden, en patiënten met een langere duur van symptomen werden geanalyseerd als subgroepen omdat er verwacht werd dat deze patiënten het meeste baat zouden hebben van het maken van de röntgenfoto's. Onze resultaten tonen aan dat de het aantal erosies op routinematige röntgenfoto's van handen en voeten bij patiënten die zich voor het eerst presenteren met artritis die verdacht was voor RA, zeer laag was. Het uitvoeren van routinematige röntgenfoto's leidt ook zelden tot een verandering in diagnose of prognose. Dit gold ook voor de relevante subgroepen. Hoewel het maken van röntgenfoto's bij patiënten met nieuw opkomende artritis, verdacht voor RA, waardevol kunnen zijn in specifieke gevallen en op basis van symptomen, zou de aanbeveling om deze röntgenfoto's routinematig uit te voeren heroverwogen moeten worden.

Hoofdstuk 3: Lab tests voor het voorspellen van de respons op een volgend reumamedicijn

Binnen de behandeling van RA wordt gebruik gemaakt van anti-reuma medicatie, de zogenoemde 'disease-modifying anti-rheumatic drugs' (DMARDs). Deze middelen remmen de gewrichtsontstekingen en voorkomen gewrichtsschade. Een van deze middelen is adalimumab, een tumornecrosefactor remmer (TNFi). Als adalimumab behandeling onvoldoende werkt, lijken zowel middelen van hetzelfde werkingsmechanisme (TNFi) niet-TNFi DMARDs op groepsniveau vergelijkbare opties als volgende stap in de behandeling. Resultaten uit eerdere onderzoeken suggereren echter dat aanwezigheid van lage waarden van adalimumab in het bloed en/of antilichamen tegen de adalimumab voorspelt dat een TNFi als volgend middel beter zal werken dan een middel met een ander werkingsmechanisme. Het idee is dat deze mensen eerder niet goed reageerden op het middel omdat ze niet voldoende werden blootgesteld aan de adalimumab.

Wij hebben onderzocht of deze theorie klopt. Dit deden we in RA-patiënten die minstens drie maanden adalimumab hadden gebruikt en dit middel zijn gestopt omdat het onvoldoende werkte. Ook moesten de patiënten hierna een ander reuma medicijn gestart zijn. Uit onze resultaten bleek dat zowel de bloedwaarden van adalimumab als de aanwezigheid van antilichamen geen voorspellende waarde had voor de kans van slagen van een volgend middel. Dit gold zowel voor de TNFi als een niet-TNFi. Dit werd gevonden voor zowel patiënten die meteen niet goed reageerden op de adalimumab als de groep patiënten bij wie het eerst een tijd goed ging, maar de adalimumab na een tijd toch niet voldoende werkte. Op basis van resultaten uit dit onderzoek is het testen van bloed waarden en antistoffen van adalimumab dus niet aan te raden in de dagelijkse praktijk.

Hoofdstuk 4: Aanvullende waarde en kosten effectiviteit van gebruik van een hypothetische biomarker bij het voorspellen van respons op een volgend reuma medicijn in patienten met RA

Er wordt veel onderzoek gedaan naar stofjes in het lichaam die een ziekte kunnen opsporen en vervolgen (biomarkers). De eerder genoemde RF, anti-CCP, ontstekingswaarden in het bloed, adalimumab bloedwaarden en -antistoffen zijn voorbeelden van biomarkers bij RA. RA wordt behandeld middels een zogenoemde 'treat-to-target' strategie waar een behandeldoel gesteld wordt en regelmatig wordt gecontroleerd of dit doel bereikt is. Zo niet, dan wordt de reuma medicatie aangepast totdat het doel bereikt word. Momenteel wordt deze strategie gestuurd door de ziekte activiteit: hoeveel gewrichtsontstekingen zijn er? Wat zijn pijnscores van patiënten? Hoe hoog zijn de ontstekingswaarden in het bloed? Een biomarker kan hier verder bij helpen doordat het de activiteit van de RA kan nauwkeuriger zou kunnen vaststellen en kan meten of de behandeling werkt.

Het is het belangrijk om te bepalen welke kenmerken (zoals kosten en nauwkeurigheid) een nieuwe biomarker zou moeten hebben om behandel uitkomsten te kunnen verbeteren zonder dat dit te veel kosten vergt. Door middel van een door de computer gesimuleerd model (Markov-model) werd een ziekte activiteit gestuurde treat-to-target strategie vergeleken met een strategie die gestuurd werd door een hypothetische biomarker. Hiervoor werd een zelf bedachte biomarker gebruikt waarvan de nauwkeurigheid van de test (testkarakteristieken) en kosten konden worden aangepast. Er werden vier scenario-analyses uitgevoerd, waarbij variaties in de kosten van de biomarker, de nauwkeurigheid van de biomarker, het aandeel patiënten dat de reuma medicatie kon afbouwen, en medicatiekosten werden meegenomen.

Resultaten toonden aan dat patiënten in de biomarker strategie met zeer optimistische aannames (met andere woorden: hij was heel nauwkeurig zonder veel te kosten) slechts 3 maanden extra zonder klachten of met weinig ziekte activiteit doorbrachten over een periode van 2 jaar. In de uitkomsten vanaf 2 jaar en later was er geen verschil tussen de groepen. Dit ging gepaard met een minimale kostenbesparing. De kostenbesparing werd vrijwel uitsluitend gedreven door het effect dat patiënten eerder lage tot geen ziekte activiteit meer hadden en hierom de dure reuma medicatie eerder kon afbouwen. Een biomarker voor de voorspelling van respons op DMARD-behandeling bij RA kan dus van toegevoegde waarde zijn voor de huidige treat-to-target zorg, maar de winst is bescheiden, en kostenbesparingen zijn zeer afhankelijk van vroegtijdig afbouwen en de hoge kosten van medicatie.

Hoofdstuk 5: FDG-PET/CT-scans

Het maken van PET scans, specifiek de 18FDG-PET gecombineerd met CT-scans (FDG-PET/CT) wordt niet aanbevolen bij RA, maar wordt soms wel door artsen gebruikt. Redenen voor het gebruik van FDG-PET/CT-scans zijn het aantonen van artritis of kunnen helpen bij behandelingsbeslissingen, omdat FDG-opname in aangetaste gewrichten ziekte activiteit kan weerspiegelen, en dit zou mogelijk beter voorspellend zijn dan klinische kenmerken van ziekteactiviteit (zoals gewrichtsontstekingen gevonden bij lichamelijk onderzoek of pijnklachten).

Een nadeel van FDG-PET/CT scans is dat ze onverwachte toevalsbevindingen geven (afwijkingen waar niet naar gezocht werd maar die wel gevonden zijn). Dit komt met name omdat het hele lichaam gescand moet worden, niet enkel de gewrichten. Ook is de FDG-PET/CT niet alleen gericht op ontstekingen, maar geeft cel activiteit aan, welke ook hoog is bij kanker. Veelgevonden toevalsbevindingen zijn dus verdacht voor kanker. De betrouwbaarheid van deze scans voor kanker in patiënten die geen klachten hebben die passend zijn bij kanker is nog niet goed onderzocht.

In de eerer uitgevoerde DRESS-studie werden FDG-PET/CT scans uitgevoerd om de mate van artritis bij patiënten met RA die met TNF-remmers werden behandeld te beoordelen. Patiënten werden in deze studie 3 jaar nauwgezet gemonitord door reumatologen. Dit bood een kans om de betrouwbaarheid van FDG-PET/CT-scans voor het diagnosticeren van kanker in een groep RA patiënten die geen kanker gerelateerde klachten hadden. Resultaten van dit onderzoek toonden dat gebruik van FDG-PET/CT-scans voor het beoordelen van artritis vaak toevalsbevindingen oplevert. De toevalsbevindingen waren vaak geen belangrijke bevindingen: de 9 mensen met op de scan een sterke verdenking op kanker bleken allemaal geen kanker te hebben, en alle 6 mensen die kanker ontwikkelden binnen 3 jaar na de scan hadden een negatieve FDG-PET/CT scan. We concluderen dat artsen en patiënten goed geïnformeerd moeten worden over de risico's van toevalsbevindingen bij het gebruik van FDG-PET/CT-scans voor wanneer ze worden ingezet voor het meten van artritis bij patiënten met RA.

Hoofdstuk 6: Langdurige routinematige toxiciteit monitoring bij gebruik van reumamedicatie

Behandeling met DMARDs kan gepaard gaan met bijwerkingen. Bij gebruik van deze middelen wordt geadviseerd om laboratoriumonderzoek te doen om een deel van deze bijwerkingen tijdig op te sporen. Bij langdurig gebruik (> 6 maanden) wordt geadviseerd iedere 3-6 maanden bloed onderzoek te verrichten. Dit wordt al jaren zo gedaan maar toch is de toegevoegde waarde van lange termijn routinematige laboratoriumonderzoek (lange-termijn monitoring) niet vastgesteld. Minder vaak bloedonderzoeken ondergaan zou de belasting voor patiënt en milieu kunnen verminderen en kosten kunnen besparen. Daarom hebben we de huidige lange-termijn monitoring strategieën beoordeeld door drie onderzoeksvragen te behandelen: Hoe vaak komen (zeer) afwijkende uitkomsten voor bij lange-termijn monitoring? Is er een verschil tussen patiënten die een DMARD gebruiken waarvoor lange-termijn monitoring wel of niet wordt aanbevolen? En wat zijn de kenmerken van zeer afwijkende laboratoriumtests?

Om deze vragen te beantwoorden hebben we een studie opgezet met gegevens van bijna 5.000 RA-patiënten in de Sint Maartenskliniek die samen meer dan 330.000 tests zijn ondergaan. Zeer abnormale resultaten werden nauwelijks gevonden, en de het voorkomen van zeer afwijkende tests vergelijkbaar tussen patiëntengroepen waarvoor monitoring wel of niet werd aanbevolen. De zeer abnormale resultaten traden meestal op na verhoging van de dosering van reuma medicatie, bij oudere patiënten, waren vaak al bekend of werden al vermoed, en werden voorafgegaan door langdurig al afwijkende bloed uitslagen. Ook werden ze meestal als niet gerelateerd aan DMARD-gebruik beschouwd of leidden niet tot verdere acties van de arts. Hierom hebben we geconcludeerd dat frequente bloed monitoring van reuma

medicatie na de eerste zes maanden van behandeling niet zinvol is. Wel zouden we monitoring adviseren na het ophogen van de dosis van de reuma medicatie, op basis van klachten passend bij bijwerkingen van het medicijn en in oudere patiënten.

List of publications

Publications

Ulijn, E., den Broeder, N., Wientjes, H. M., van Herwaarden, N., Meek, I., Tweehuysen, L., van der Maas, A., van den Bemt, B. J. F. & den Broeder, A. A. (2020). Therapeutic drug monitoring of adalimumab in RA: no predictive value of adalimumab serum levels and anti-adalimumab antibodies for prediction of response to the next bDMARD. *Annals of the Rheumatic Diseases*, *79*(7), 867-873.

Ulijn, E., den Broeder, A. A., Boers, N., Gotthardt, M., Bouman, C. A., Landewé, R., den Broeder, N. & van Herwaarden, N. (2022). Extra-articular findings with FDG-PET/CT in rheumatoid arthritis patients: more harm than benefit. *Rheumatology Advances in Practice*, *6*(1), rkac014.

Wientjes, M. H., **Ulijn, E.,** Kievit, W., Landewé, R. B., Meek, I., den Broeder, N., van Herwaarden, N., van den Bemt, B. J. F., Verhoef, L. M. & den Broeder, A. A. (2023). The added value of predictive biomarkers in treat-to-target strategies for rheumatoid arthritis patients: a conceptual modelling study. *Rheumatology*, *62*(8), 2700-2706.

Ulijn, E., den Broeder, N., Ten Cate, D., van Overdijk, K., Demirel, H., Landewé, R., van Herwaarden, N. & den Broeder, A. (2024). Limited Diagnostic and Prognostic Value of Routine Radiographs in Newly Presenting Arthritis Suspected of Rheumatoid Arthritis: A Retrospective Study. *Arthritis Care & Research*, *76*(4), 497-502.

Conference abstracts

Ulijn, E., den Broeder, A. A., Boers, N., Gotthardt, M., Bouman, C. A., Landewé, R., den Broeder, A. A. & van Herwaarden, N. Extra-articular findings with FDG-PET/CT in rheumatoid arthritis patients: more harm than benefit.
Poster presentation (in guided tour): EULAR congress 2022, Copenhagen

Ulijn, E., den Broeder, N., Ten Cate, D., van Overdijk, K., Demirel, H., Landewé, R., van Herwaarden, N. & den Broeder, A. Limited Diagnostic and Prognostic Value of Routine Radiographs in Newly Presenting Arthritis Suspected of Rheumatoid Arthritis: A Retrospective Study.
Oral presentation: EULAR congress 2023, Milan

Ulijn, E., den Broeder, N., Bevers, K., Pruijs, R., van Es, B., Tauber, T., Landewé, R., van Herwaarden, N., & den Broeder, A.
Poster presentation (by A. A. den Broeder): EULAR congress 2024, Vienna

PhD portfolio

Department: **RIHS**
PhD period: **01/10/2020 – 08/04/2025**
PhD Supervisor(s): **Dr A.A. den Broeder, Prof R.B.M. Landewé**
PhD Co-supervisor(s): **Dr N. van Herwaarden, Dr. N. den Broeder**

Training activities	Hours
Courses	
- Scientific writing in English (2021)	40.00
- Improving your statistical inferences (2021)	27.00
- Introduction to R (2023)	32.00
- Radboudumc - In the lead - Radboudumc introduction for PhD candidates (2023)	12.00
- Radboudumc - eBROK course (2023)	42.00
- Radboudumc - Scientific integrity (2023)	20.00
- Evalutation of medical tests (2023)	8.00
Conferences	
- EULAR copenhagen (2022)	32.00
- EULAR Milan (2023)	32.00
- NVR (2023)	10.00
- Oral presentation, MOSA conference, Maastricht	16.00
Other	
- Radboudumc - General Radboudumc introduction for research personnel (2023)	9.00
- Writing days, department of Rheumatology, Sint Maartenskliniek	16.00
- RIHS PhD retreat, 's Hertogenbosch	16.00
Total	**312.00**

Curriculum vitae

Evy Ulijn werd op 21 oktober 1996 geboren in Oss. Na een basisschool tijd in Curaçao en Singapore behaalde zij in 2015 haar (tweetalig) VWO diploma aan het Maaslandcollege te Oss. Datzelfde jaar startte ze met de bachelor opleiding Gezondheidswetenschappen aan de Universiteit Maastricht. Gedurende de bachelor opleiding koos Evy voor de richting Biologie en Gezondheid en heeft Evy een semester in het Zweedse Halmstad doorgebracht.

Na het behalen van haar bachelor diploma in 2018 is Evy in aanraking gekomen met onderzoek in de Sint Maartenskliniek, en is ze onder leiding van Alfons den Broeder en Nathan den Broeder gestart met haar eerste onderzoeksproject binnen de reumatologie, naast haar werkzaamheden als onderzoeksassistent. In 2019 is Evy gestart met de master Arts-Klinisch onderzoeker, eveneens aan de Universiteit Maastricht. Het doen van onderzoek beviel Evy dermate goed dat ze na het publiceren van haar eerste eigen artikel in samenspraak met haar uiteindelijke promotieteam besloten heeft om een promotietraject te starten.

Tijdens haar promotietraject werd Evy begeleid door Alfons den Broeder, Robert Landewé, Noortje van Herwaarden en Nathan den Broeder. De verschillende onderzoeksprojecten die het huidige gebruik van diagnostiek hebben geëvalueerd heeft Evy grotendeels naast de coschappen van haar master Arts-Klinisch Onderzoeker uitgevoerd. In 2023 studeerde Evy af als basisarts waarna ze een half jaar als artsonderzoeker in de Sint Maartenskliniek werkzaam is geweest.

Na het afronden van het grootste deel van haar promotietraject is Evy aan de slag gegaan als basisarts binnen de interne geneeskunde in het Jeroen Bosch Ziekenhuis. In 2025 start Evy met haar opleiding tot internist, aan de Radboud Universiteit.

Dankwoord

Dit proefschrift is tot stand gekomen met samenwerking en steun van velen. Daarom wil ik van deze gelegenheid gebruik maken om iedereen die gedurende dit promotietraject zijn of haar steentje heeft bijgedragen te bedanken.

Te beginnen bij mijn (co)promotoren Alfons den Broeder, Robert Landewé, Noortje van Herwaarden, Nathan den Broeder. Bedankt voor jullie (wetenschappelijke) adviezen, ondersteuning en steevast vertrouwen.
Alfons, vanaf het allereerste moment had jij in mij het volste vertrouwen. Ik kwam op een willekeurige dinsdagmiddag onaangekondigd je kantoor binnenlopen om te reageren op een al lang niet meer bestaande vacature. Dit hield je niet tegen om me een stuk of vijf voorstellen voor onderzoeksprojecten mee te geven en me er daags later te laten starten. Je creativiteit in het vinden van nieuwe en relevante onderzoeksvragen heeft me altijd bewonderd. Door je doortastendheid hebben we soms (na veelal onvermoeibaar vaak opnieuw beginnen) hele nieuwe onderzoeksmethoden bedacht om deze vragen te beantwoorden. Ik neem deze denkwijze mee in de rest van mijn carrière en nogmaals bedankt voor de steun en het vertrouwen wat ik van je heb mogen ontvangen.
Nathan, bedankt voor je fijne hulp en begeleiding al vanaf het begin van deze promotie (wat toen nog niet eens promotie was, en waar jij zelf ook pas net aan je promotie begonnen was). Ik heb veel moeten leren en ik ben na initieel intensieve begeleiding steeds zelfstandiger geworden. Toch had je er altijd oog voor of ik hulp nodig had of waar je me kon ondersteunen, wat zeker geen makkelijke taak is als iemand ver weg in Maastricht zit. Je begeleiding strekte verder dan alleen aan onderzoek gerelateerde taken, ook was je er voor vragen over laptops die weer eens niet werkten, het boeken van vliegtickets naar het congres in Kopenhagen, en hebben we na wat meer serieuze meetings vaak uitgebreid zitten keuvelen (iets wat ik goed kan). Hierdoor was het altijd een plezier om met je samen te werken!
Noortje, bedankt voor je precisie en doortastendheid, die onze onderzoeken naar een hoger niveau hebben getild. Je hebt me laten zien hoe je dit aanpakt, en ik heb daar veel van geleerd. Ook hielp je me met het plannen en stellen van prioriteiten wanneer ik het overzicht even kwijt was door de drukte. Ik heb veel bewondering voor jouw werklust en wil je bedanken voor je waardevolle begeleiding de afgelopen jaren.
Daarnaast wil ik vooral ook jullie flexibiliteit gedurende dit toch wel bijzondere promotietraject. Meermaals was ik bij Alfons en Noortje thuis welkom voor een promotieoverleg op enorm ongunstige tijdstippen (vrijdag avond na het eten), en kreeg ik er zelfs nog frietjes bij. Wellicht een voordeel van een promotieteam wat grotendeels familie van elkaar is. Zonder deze flexibiliteit vanuit jullie kant had dit promotietraject niet zo voorspoedig kunnen lopen.

Robert, ook jij was vanuit het verre zuiden (en soms noorden) van het land ondanks je eigen drukke schema altijd laagdrempelig beschikbaar voor overleg. Je hielp meermaals met een frisse blik en ook een kritische noot kon niet ontbreken. Dit heeft meermaals voor een mooie doorbraak in het onderzoeksproject gezorgd. Daarnaast bleef je, ook nadat ik de Sint Maartenskliniek verlaten heb, geïnteresseerd en faciliterend in de volgende stappen in mijn carrière. Dank voor je interesse en enthousiasme de afgelopen jaren.

Ook wil ik Arjen de Boer bedanken, voor zijn onuitputbare hulp, interesse en enthousiasme bij het aanleveren van relevante patiënt gerelateerde data. Vragen voor aanvullende data kwamen vanuit alle hoeken van de wereld en daardoor op uiteenlopende tijdstippen bij je binnen, desondanks werd ik altijd snel geholpen. Ik mocht je nooit noemen in de dankwoorden van mijn artikelen, maar bij deze doe ik het alsnog: bedankt voor de fijne samenwerking de afgelopen jaren. Ik heb hier enorm veel aan gehad.

Daarnaast wil ik mijn collega's van de Sint Maartenskliniek bedanken voor de leuke momenten de afgelopen jaren, ook al was het op de arts-onderzoekerskamer af en toe misschien zelfs iets té gezellig. Ik heb genoten van de pauzes met (haver)cappu's, een chaotisch en warm tripje naar de pannenkoekenboot ('a dream come true' voor sommigen van ons) en het samen sporten. Ook van jullie ondersteuning bij moeilijke momenten, het vieren van elkaars successen, en aan de leuke momenten als reisgenootjes naar cursussen en congressen heb ik enorm genoten. Bedankt voor jullie steun tijdens dit promotietraject, zowel op afstand als op het werk, en vooral voor de leuke tijd!

Verder wil ik de student-assistenten, en specifiek Stijn van der Plassche bedanken voor de hulp bij het (soms erg saaie) invoerwerk. Dank voor alle hulp bij het doornemen van talloze patiëntdossiers en voor de fijne samenwerking.

Ik wil Maroeska Roovers bedanken, mijn mentor tijdens dit promotietraject. Hoewel we elkaar slechts eens per jaar spraken, namen we altijd ruim de tijd, en voelde ik me vrij om alles te bespreken, van coschappen tot promotiezaken. Je begreep dat mijn traject anders verliep dan gebruikelijk en moedigde me aan niet op te geven als cursussen minder flexibel waren qua data en aanwezigheidseisen. Jouw steun en kritische blik hebben me door dit traject geholpen, en ik neem die waardevolle houding mee in mijn verdere carrière.

Marcel Flendrie, bedankt voor het begeleiden van mijn eerste stappen als basisarts op de polikliniek. In korte tijd heb je me waardevolle adviezen gegeven waar ik nog steeds vaak op terugval. Je zorgde voor een fijne, open sfeer waarin ik me als beginnend arts op mijn gemak voelde.

Leden van de manuscript commissie, ook jullie wil ik hartelijk bedanken voor de tijd en moeite die jullie hebben genomen voor het doornemen en beoordelen van mijn proefschrift.

Mike Kattenbelt, het is misschien niet gebruikelijk om een voormalig fysiotherapeut te bedanken in een proefschrift, maar zonder jouw aansporing was ik nooit naar de P&O van de Sint Maartenskliniek gestapt. Dit zetje, in een periode waarin ik niet wist wat ik met mezelf aan moest na mijn operatie, heeft geleid tot een bijzonder promotietraject en uiteindelijk tot dit proefschrift. Dank je wel!

Yiten, we kunnen soms urenlang bellen, en bij jou kan ik altijd mijn hart luchten. Ook ben je er altijd voor in om spontaan wat leuks te gaan doen. Dankjewel voor je steun en het luisterend oor dat je de afgelopen jaren voor me bent geweest.

Lude, we hebben elkaar door onze promotietrajecten heen weten te slepen, vaak met ontelbare memes als afleiding. Ook nu als AIOS kan ik bij jou terecht met mijn dagelijkse struggles en kunnen we daar altijd om blijven lachen – dankjewel!

Simon, als huisgenoten in Maastricht hebben we elkaar pas echt goed leren kennen. Ik wil je bedanken voor die leuke tijd en hoop dat we onze successen kunnen blijven delen, ook al wonen we nu op wat meer afstand van elkaar.

Victor, nu helemaal down under, bedankt voor de ontelbare zwemsessies, sportlessen, picknicks in het park en al die andere mooie momenten.

Anthony, Eric, Max, Ramon, Isa, Felicia, Nicole en nog vele andere lieve mensen die ik in Maastricht heb mogen leren kennen, ook jullie wil ik bedanken voor het luisterend oor wat jullie geboden hebben als ik het weer eens over een of ander onderzoeksproject had waarbij ik weer eens 'opnieuw moest beginnen'.

Fauve, Evy, Babs, Daphne en Max, samen hebben we de A-KO door gebikkeld. Dank voor het aanhoren van mijn promotieperikelen, maar ook voor het delen van de hoogte- en dieptepunten van de coschappen. We hebben samen mooie momenten gevierd – van in opleiding komen tot trouwen en kinderen krijgen – inmiddels ook met aanhang, wat het nóg gezelliger maakt.

Dan mijn paranimfen. Vera, bij jou thuis was ik altijd welkom om lekker onder je knots deken samen op de bank te hangen, te eten voor het zwemmen, en daarna soms te zwemmen, maar vaak ook gewoon door te kletsen op de badrand. We hebben lief en leed gedeeld tijdens het bestuursjaar, en ik heb genoten van de ontelbare keren samen sporten. Ik wil je heel erg bedanken voor alle steun de lol die we de afgelopen jaren hebben gehad, maar ook het luisterend oor wanneer ik dat nodig had. Ik had het voor geen goud willen missen!

Maike, we zijn ongeveer tegelijk met ons promotietraject begonnen, maar inmiddels ben jij al een doorgewinterde onderzoeker die ontzettend veel voor de onderzoeksgroep doet — ik denk niet dat de groep zomaar nog zonder je kan! Ondanks dat ik mijn promotie grotendeels op afstand heb uitgevoerd, is het ons toch gelukt om samen een artikel te schrijven, en ik vond het heel fijn om met jou samen te werken. En natuurlijk ook bedankt voor de gezellige tijd buiten het werk om, zoals op congres Kopenhagen en Milaan, de etentjes (onder andere op die legendarische pannenkoekenboot), en het fanatiek samen sporten. Bedankt voor al deze fijne momenten!

Opa, ik denk niet dat veel mensen kunnen zeggen dat hun opa nog promoveert. Na decennia ervaring als promovendus kon je me tijdens dit traject altijd ondersteunen met goede raad. En oma, bedankt voor de gezelligheid en steun aan de keukentafel, met koffie, koekjes en alles wat je voor me had. Jullie hebben me altijd gemotiveerd om door te gaan, zelfs wanneer het even moeilijk was. Dankjulliewel!

Lize, het is soms lastig bij te benen als je zus binnen één jaar zowel gaat promoveren als trouwen. Daarom wil ik je bedanken voor het geduldig aanhoren van mijn waarschijnlijk oersaaie verhalen over deze onderzoeken en de eindeloze gesprekken over de naderende bruiloft. Bedankt dat je er altijd voor me bent als ik wat ontspanning zoek, of we nu alpaca's knuffelen of slechte series kijken. Ook wil ik je bedanken voor je creativiteit bij het maken van de prachtige kaft van dit boekje en de posters voor congressen. Ik ben trots op je, en zodra je terug bent, dan proosten we samen op alles wat nog komen zal!

Mam, pap, jullie hebben beide een onderdeel gehad in mijn vroege opleiding als 'juf' en 'meester'. Naarmate mijn opleiding vordert, wordt het wel steeds lastiger om uit te leggen wat ik precies doe, waar het precies over gaat, en hoe lang het allemaal nog zal duren. Toch weet ik dat jullie trots zijn op alle stappen die ik zet, en weet ik dat jullie mij onvoorwaardelijk zullen blijven steunen. Bedankt voor alles wat jullie voor me hebben gedaan, voor jullie aandeel in dit proefschrift en in alle andere mooie dingen die ik tot nu toe heb mogen bereiken.

Alex, waar moet ik beginnen... Ik wil je bedanken voor de onvoorwaardelijke liefde en steun die je me de afgelopen 9,5 jaar hebt gegeven. Je hebt ellenlange verhalen aangehoord over de onderzoeken in dit proefschrift. Het feit dat je precies weet wat een DAS-score is en je 'adalimumab' zonder moeilijkheden kan uitspreken zegt denk ik al genoeg voor iemand die veel meer thuis is in onderzoek naar Malawiaanse vissen. Bedankt voor alle keren dat het eten klaarstond, de mooie vakanties om even bij te komen en ook alvast voor alle mooie dingen die nog gaan komen. Ik kan niet wachten om je het ja-woord te mogen geven en jou binnenkort te mogen bijstaan bij je eigen promotie. Ik hou van je en ben enorm trots op je!

Research data management

General information on data collection

For the research presented in this thesis, data was collected in the department of Rheumatology in the Sint Maartenskliniek. Data was extracted from electronic patient records with use of SAS or by manual data extraction and storage in electronic Case Report Forms (eCRF) in Castor EDC. Data was converged to STATA 13 or RStudio for further data analysis. Additionally, for the research presented in chapter 3, data was also collected in the department of Rheumatology in Radboud University Medical Center. Research Data Management was conducted according to the Findable, Accessible, Interoperable and Reusable (FAIR) principles. Within this section of my thesis I will provide a detailed description on how the FAIR principles were applied within the studies that were presented.

Ethics and privacy

The data and serum samples that were collected for studies presented in this thesis were obtained from human subjects. The studies presented in chapter 2, 3 and 6 were provided a waiver of ethical approval, numbers CMO 2020-6806, CMO 2019-5443, CMO 2022-15833. The study presented in chapter 4 makes use of data from two other studies that each received ethical approval (BIO-TOP: NL47946.091.14, RA inception cohort: CMO 2009/079).

The study presented in chapter 5 was a post-hoc analysis from data collected in the DRESS study. The DRESS study was performed at the Sint Maartenskliniek, and received ethical review board approval (number NL37704.091.11). Participants provided informed consent for collecting and processing the data by means of opt-in (chapter 5) or opt-out procedures (chapter 2, 3, 4). In the study presented in chapter 6, obtaining informed consent from all patients was deemed unfeasible by the ethical review board due to the large number of patients. All the studies involving human subjects were performed in accordance to the Declaration of Helsinki. Privacy of the participants in the studies was warranted by using encrypted and unique identification codes. The encryption keys were stored on a secure network drive separate from the study data and were only accessible to members of the study team.

FAIR principles

Included manuscripts were not published as open access. Both raw and processed data from all chapters are not archived in a Data Acquisition Collection or Research Documentation Collection in the Radboud Data Repository (RDR) as the data involved pseudonymised patient data of patients treated in a different hospital than the RadboudUMC and/or due to the fact that patients did not give permission to reuse or

share their data. The metadata for chapters 2, 3, 5 and 6 are published in Data Sharing Collections (DSC's) in the Radboud Data Repository. Contact details of the principal investigator and first author of the paper were added to the repositories. The datasets underlying these chapters are available for reuse for future research after a renewed permission by the participants is obtained. Chapter 4 is based on existing data, which was obtained from studies by W. Kievit et al and L. Tweehuysen et al. Data is available for reuse upon request from the original authors.

All the data that was obtained is stored on department servers and in Castor EDC that are only accessible by project members working at the Sint Maartenskliniek. Non-electronic data is stored in a filing cabinet with keyed lock to prevent unauthorized access to the documents, at the research department of the Sint Maartenskliniek.

Data was collected using structured electronical case report forms (Castor EDC) and stored in structured CSV file formats, resembling the METC filing format to ensure interoperability.

Data from the DRESS trial (chapter 6) will be stored for 25 years after study termination (2014). Data from the other studies will be saved for 15 years after study termination. Renewed permission of patients will be needed to re-use these data for future research.

Chapter	DAC	RDC	DSC	DSC licence
2	-	-	ru.rumc.xhandsfe_t0000098a_dsc_167	CC-BY-NC-ND 4.0
3	-	-	ru.rumc.tdmadali_t0000097a_dsc_044	CC-BY-NC-ND 4.0
5	-	-	ru.rumc.fdgpetct_t0000100a_dsc_311	CC-BY-NC-ND 4.0
6	-	-	ru.rumc.longterm_t0000099a_dsc_280	CC-BY-NC-ND 4.0

DAC = Data Acquisition Collection, RDC = Research Documentation Collection, DSC = Data Sharing Collection

Theses Sint Maartenskliniek

Ensink, C. (2025) Sensing the path to mobility - advancing gait rehabilitation with sensor technology.
Van der Togt, C. (2025) Further optimising treatment with biologicals and JAK-inhibitors in inflammatory arthritis: Focus on vaccine effectiveness and cost-effectiveness.
Van Rensch, P. (2025) Revision total knee arthroplasty. Optimizing the journey.
Zwijgers, E. (2025) Improving walking capacity after spinal cord injury.
Boekesteijn, R. (2024) Evaluating walking in lower-extremity osteoarthritis: Beyond the lab, towards the real world.
Den Broeder, N. (2024) More than tapering, less than full dose - Efficient use of biologics in the treatment of rheumatoid arthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Kuijpers, R. (2024) Adapt your step: Clinical assessment and training of walking adaptability in children with mild motor disorders. Radboud University Nijmegen, Nijmegen. The Netherlands.
Te Molder, M. (2024) The unhappy patient after TKA. A paradigm shift in assessing outcome. Radboud University Nijmegen, Nijmegen. The Netherlands.
Van Hal, T. (2024) A rheumatologist undercover: research of psoriatic arthritis at the dermatology clinic. Radboud University Nijmegen, Nijmegen. The Netherlands.
Veenstra, F. (2024). About gout. Studying potential targets for improvement of care. Radboud University Nijmegen, Nijmegen. The Netherlands.
De Jong, L.A.F. (2023). Effects of lower limb orthotic devices in people with neurological disorders. Radboud University Nijmegen, Nijmegen. The Netherlands.
Michielsens, C. (2023). Tapering strategies of biologics in inflammatory disorders. Radboud University Nijmegen, Nijmegen. The Netherlands.
Pouls, B. (2023). Supporting patients'medication management using eHealth. Test cases in rheumatology. Radboud University Nijmegen, Nijmegen. The Netherlands.
Stöcker, J. (2023). Accessible and effective non-pharmacological care for persons with systemic sclerosis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Huiskes, V. (2022). The synergistic role of patients and healthcare providers in reducing drug-related problems. Radboud University Nijmegen, Nijmegen. The Netherlands.
Marsman, D. (2022). Polymyalgia rheumatica. Clinical characteristics and new treatment opportunities. Radboud University Nijmegen, Nijmegen. The Netherlands.
Mulder, M. (2022). Going off-road. Exploring and mapping psoriatic arthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Alingh, J. (2021). Effect of robotic gait training on the post-stroke gait pattern. Evaluation of LOPES II. Radboud University Nijmegen, Nijmegen. The Netherlands.
Van Dijsseldonk, R. (2021). Step into the future: mobility after spinal cord injury. Radboud University Nijmegen, Nijmegen. The Netherlands.
Pelle, T. (2021). Beating osteoarthritis by e-self management in knee or hip osteoarthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Van Heuckelum, M (2020). Novel approaches to improve medication adherence in rheumatoid arthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Mathijssen, E. (2020). The voice of patients with rheumatoid arthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Bakker, S. (2019). Regional anesthesia and total knee arthroplasty. Anesthetic and pharmacological considerations. Radboud University Nijmegen, Nijmegen. The Netherlands.
Claassen, A. (2019). Strategies for patient education in rheumatic diseases. Radboud University Nijmegen, Nijmegen. The Netherlands.
Fenten, M. (2019). Optimizing locoregional anesthesia in fast track orthopaedic surgery. Radboud University Nijmegen, Nijmegen. The Netherlands.
Minten, M. (2019). On the role of inflammation and the value of low dose radiation therapy in osteoarthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.

Verhoef, L. (2019). Effective and efficient use of bDMARDs in rheumatoid arthritis. Radboud University Nijmegen, Nijmegen. The Netherlands.
Bekker, C. (2018). Sustainable use of medication. Medication waste and feasibility of redispensing, Utrecht University, Utrecht. The Netherlands.
Bikker, I.(2018). Organizing timely treatment in multi-disciplinary care. University of Twente, The Netherlands.
Bouman, C. (2018). Dose optimisation of biologic DMARDs in rheumatoid arthritis: long-term effects and possible predictors. Radboud University Nijmegen, The Netherlands.
Mahler, E. (2018). Contributors to the management of osteoarthritis. Utrecht University, The Netherlands.
Tweehuysen, L. (2018). Optimising biological treatment in inflammatory rheumatic diseases.
Predicting, tapering and transitioning. Radboud University Nijmegen, Nijmegen, The Netherlands.
Geerdink, Y. (2017). Getting a grip on hand use in unilateral cerebral palsy. Radboud University, Nijmegen, The Netherlands.
Remijn, L. (2017). Mastication in children with cerebral palsy. Radboud University, Nijmegen, The Netherlands.
Selten, E. (2017). Beliefs underlying treatment choices in osteoarthritis. Radboud University, Nijmegen, The Netherlands.
Van Hooff, M. (2017). Towards a paradigm shift in chronic low back pain? Identification of patient profiles to guide treatment. VU University Amsterdam, Amsterdam, The Netherlands.
Lesuis, N. (2016). Quality of care in rheumatology. Translating evidence into practice. Radboud University, Nijmegen, The Netherlands.
Luites, J. (2016). Innovations in femoral tunnel positioning for anatomical ACL reconstruction. Radboud University, Nijmegen, The Netherlands.
Pakvis, D. (2016). Survival, primary stability and bone remodeling assessment of cementless sockets. An appraisal of Wolff's law in the acetabulum. Radboud University, Nijmegen, The Netherlands.
Schoenmakers, K. (2016). Prolongation of regional anesthesia. Determinants of peripheral nerve block duration. Radboud University, Nijmegen, The Netherlands.
Altmann, V. (2015). Impact of trunk impairment on activity limitation with a focus on wheelchair rugby. Leuven University, Leuven, Belgium.
Bevers, K. (2015). Pathophysiologic and prognostic value of ultrasonography in knee osteoarthritis. Utrecht University, Utrecht, The Netherlands.
Cuperus, N. (2015). Strategies to improve non-pharmacological care in generalized osteoarthritis. Radboud University, Nijmegen, The Netherlands.
Kilkens, A. (2015). De ontwikkeling en evaluatie van het Communicatie Assessment & Interventie Systeem (CAIS) voor het aanleren van (proto-)imperatief gedrag aan kinderen met complexe ontwikkelingsproblemen. Radboud University, Nijmegen, The Netherlands.
Penning, L. (2015). The effectiveness of injections in cuffdisorders and improvement of diagnostics. Maastricht University, Maastricht, The Netherlands.
Stegeman, M. (2015). Fusion of the tarsal joints: outcome, diagnostics and management of patient expectations. Utrecht University, Utrecht, The Netherlands.
Van Herwaarden, N. (2015). Individualised biological treatment in rheumatoid arthritis. Utrecht University, Utrecht, The Netherlands.
Wiegant, K. (2015). Uitstel kunstknie door kniedistractie. Utrecht University, Utrecht, The Netherlands.
Willems, L. (2015). Non-pharmacological care for patients with systemic sclerosis. Radboud University, Nijmegen, The Netherlands.
Witteveen, A. (2015). The conservative treatment of ankle osteoarthritis. University of Amsterdam, Amsterdam, The Netherlands.
Zwikker, H. (2015). All about beliefs. Exploring and intervening on beliefs about medication to improve adherence in patients with rheumatoid arthritis. Radboud University, Nijmegen, The Netherlands.
Koenraadt, K. (2014). Shedding light on cortical control of movement. Radboud University, Nijmegen, The Netherlands.
Smink, A. (2014). Beating Osteoarthritis. Implementation of a stepped care strategy to manage hip or knee osteoarthritis in clinical practice. VU University Amsterdam, Amsterdam, The Netherlands.

Stolwijk, N. (2014). Feet 4 feet. Plantar pressure and kinematics of the healthy and painful foot. Radboud University, Nijmegen, The Netherlands.
Van Kessel, M. (2014). Nothing left? How to keep on the right track. Spatial and non-spatial attention processes in neglect after stroke. Radboud University, Nijmegen, The Netherlands.
Brinkman, M. (2013). Fixation stability and new surgical concepts of osteotomies around the knee. Utrecht University, Utrecht, The Netherlands.
Kwakkenbos, L. (2013). Psychological well-being in systemic sclerosis: Moving forward in assessment and treatment. Radboud University, Nijmegen, The Netherlands.
Severens, M. (2013). Towards clinical BCI applications: assistive technology and gait rehabilitation. Radboud University, Nijmegen, The Netherlands.
Stukstette, M. (2013). Understanding and treating hand osteoarthritis: a challenge. Utrecht University, Utrecht, The Netherlands.
Van der Maas, A. (2013). Dose reduction of TNF blockers in Rheumatoid Arthritis: clinical and pharmacological aspects. Radboud University, Nijmegen, The Netherlands.
Zedlitz, A. (2013). Brittle brain power. Post-stroke fatigue, explorations into assessment and treatment. Radboud University, Nijmegen, The Netherlands.
Beijer, L. (2012). E-learning based speech therapy (EST). Exploring the potentials of E-health for dysarthric speakers. Radboud University, Nijmegen, The Netherlands.
Hoogeboom, T. (2012). Tailoring conservative care in osteoarthritis. Maastricht University, Maastricht, The Netherlands.
Boelen, D. (2011). Order out of chaos? Assessment and treatment of executive disorders in brain-injured patients. Radboud University, Nijmegen, The Netherlands.
Heesterbeek, P. (2011). Mind the gaps! Clinical and technical aspects of PCL-retaining total knee replacement with the balanced gap technique. Radboud University, Nijmegen, The Netherlands.
Hegeman, J. (2011). Fall risk and medication. New methods for the assessment of risk factors in commonly used medicines. Radboud University, Nijmegen, The Netherlands.
Smulders, E. (2011). Falls in rheumatic diseases. Risk factors and preventive strategies in osteoporosis and rheumatoid arthritis. Radboud University, Nijmegen, The Netherlands.
Snijders, G. (2011). Improving conservative treatment of knee and hip osteoarthritis. Radboud University, Nijmegen, The Netherlands.
Vriezekolk, J. (2011). Targeting distress in rheumatic diseases. Utrecht University, Utrecht, The Netherlands.
Willems, P. (2011). Decision making in surgical treatment of chronic low back pain. The performance of prognostic tests to select patients for lumbar spinal fusion. Maastricht University, Maastricht, The Netherlands.
Aarts, P. (2010). Modified constraint-induced movement therapy for children with unilateral spastic cerebral palsy: the Pirate group intervention. Radboud University, Nijmegen, The Netherlands.
Groen, B. (2010). Martial arts techniques to reduce fall severity. Radboud University, Nijmegen, The Netherlands.
Van Koulil, S. (2010). Tailored cognitive behavioral therapy in fibromyalgia. Radboud University, Nijmegen, The Netherlands.
Van den Bemt, B. (2009). Optimizing pharmacotherapy in patients with rheumatoid arthritis: an individualized approach. Radboud University, Nijmegen, The Netherlands.
Van Nes, I. (2009). Balance recovery after supratentorial stroke. Influence of hemineglect and the effects of somatosensory stimulation. Radboud University, Nijmegen, The Netherlands.
Ruiter, M. (2008). Speaking in ellipses. The effect of a compensatory style of speech on functional communication in chronic agrammatism. Radboud University, Nijmegen, The Netherlands.
Baken, B. (2007). Reflexion on reflexes. Modulation during gait. Radboud University, Nijmegen, The Netherlands.
Gaasbeek, R. (2007). High tibial osteotomy. Treatment of varus osteoarthritis of the knee. Radboud University, Nijmegen, The Netherlands.
Koëter, S. (2007). Patellar instability. Diagnosis and treatment. Radboud University, Nijmegen, The Netherlands.

Langeloo, D. (2007). Monitoring the spinal cord during corrective spinal surgery: a clinical study of TES-MEP. Radboud University, Nijmegen, The Netherlands.
De Haart, M. (2005). Recovery of standing balance in patients with a supratentorial stroke. Radboud University, Nijmegen, The Netherlands.
Den Otter, R. (2005). The control of gait after stroke: an electromyographic approach to functional recovery. Groningen University, Groningen, The Netherlands.
Spruit, M. (2005). Surgical treatment of degenerative disc conditions of the lumbar spine. Biomechanical, clinical and radiological aspects. University Utrecht, Utrecht, The Netherlands.
Weerdesteyn, V. (2005). From the mechanisms of obstacle avoidance towards the prevention of falls. Radboud University, Nijmegen, The Netherlands.
Jongerius, P. (2004). Botulinum toxin type-A to treat drooling. A study in children with cerebral palsy. Radboud University, Nijmegen, The Netherlands.
Van de Crommert, H. (2004). Sensory control of gait and its relation to locomotion after a spinal cord injury. Radboud University, Nijmegen, The Netherlands.
Van der Linde, H. (2004). Prosthetic prescription in lower limb amputation. Development of a clinical guideline in the Netherlands. Groningen University, Groningen, The Netherlands.
Hendricks, H. (2003). Motor evoked potentials in predicting motor and functional outcome after stroke. University of Nijmegen, Nijmegen, The Netherlands.
Hosman, A. J. F. (2003). Idiopathic thoracic spinal deformities and compensatory mechanisms. University of Nijmegen, Nijmegen, The Netherlands.
Donker, S. (2002). Flexibility of human walking: a study on interlimb coordination. Groningen University, Groningen, The Netherlands.
Hochstenbach, J. (1999). The cognitive, emotional, and behavioural consequences of stroke. University of Nijmegen, The Netherlands.
De Kleuver, M. (1998). Triple osteotomy of the pelvis. An anatomical, biomechanical and clinical study. University of Nijmegen, Nijmegen, The Netherlands.
Van Balen, H. (1997). A disability-oriented approach to long-term sequelae following traumatic brain injury. Neuropsychological assessment for post-acute rehabilitation. University of Nijmegen, Nijmegen, The Netherlands.
Tromp, E. (1995). Neglect in action: a neuropsychological exploration of some behavioural aspects of neglect. University of Nijmegen, Nijmegen, The Netherlands.
Van Lankveld, W. (1993). Coping with chronic stressors of rheumatoid arthritis. University of Nijmegen, Nijmegen, The Netherlands.
Geurts, A. (1992). Central adaptation of postural organization to peripheral sensorimotor impairments. University of Nijmegen, Nijmegen, The Netherlands.
De Rooij, D. (1988). Clinical and serological studies in the connective tissue diseases. University of Nijmegen, Nijmegen, The Netherlands.