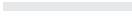


Exploring treat-to-target strategies and monitoring of disease activity with the HandScan in rheumatoid arthritis

Maxime Verhoeven



Exploring treat-to-target strategies and monitoring of disease activity with the HandScan in rheumatoid arthritis



Maxime Marie Antoinette Verhoeven

**Exploring treat-to-target strategies and monitoring of disease activity with the
HandScan in rheumatoid arthritis**

M.M.A. Verhoeven

Utrecht University, the Netherlands

ISBN/EAN: 978-94-6416-406-0

The research described in this thesis is co-funded by Roche B.V. Nederland, Demcon Hemics BV., and the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to simulate public-private partnerships.

Copyright © 2021 M.M.A. Verhoeven

All rights reserved. No part of this thesis may be reproduced, stored or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers.

Layout and design by Harma Makken, persoonlijkproefschrift.nl

Printing: Ridderprint | www.ridderprint.nl

Exploring treat-to-target strategies and monitoring of disease activity with the HandScan in rheumatoid arthritis

Het onderzoeken van ‘treat-to-target’ strategieën, en het monitoren van ziekteactiviteit met de HandScan in reumatoïde artritis
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus,
prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op donderdag 22 april 2021 des middags te 2.30 uur door

Maxime Marie Antoinette Verhoeven

geboren op 12 maart 1993
te Oldenzaal

Promotoren:

Prof. dr. F.P.J.G. Lafeber
Prof. dr. J.M. van Laar

Copromotoren:

Dr. P.M.J. Welsing
Dr. J. Tekstra

TABLE OF CONTENTS:

Chapter 1:	General introduction	9
------------	----------------------	---

PART I: Evaluation of (cost-)effectiveness of DMARD therapy approaches in treat-to-target strategies in (early) RA patients.

Chapter 2:	Effectiveness of remission-induction strategies for early rheumatoid arthritis: a systematic literature review.	19
Chapter 3:	Initiating tocilizumab, with or without methotrexate, compared with starting methotrexate with prednisone within step-up treatment strategies in early rheumatoid arthritis: an indirect comparison of effectiveness and safety of the U-Act-Early and CAMERA-II treat-to-target trials.	41
Chapter 4:	Validation of a prognostic multivariable prediction model for insufficient clinical response to methotrexate in early rheumatoid arthritis and its clinical application in Evidencio.	59
Chapter 5:	Is tocilizumab monotherapy as effective in preventing radiographic progression in rheumatoid arthritis as its combination with methotrexate?	79
Chapter 6:	Effectiveness and safety over 3 years after the 2-year U-Act-Early trial of the strategies initiating tocilizumab and/or methotrexate.	103
Chapter 7:	Effect on costs and quality-adjusted life-years of treat-to-target treatment strategies initiating methotrexate, or tocilizumab, or their combination in early rheumatoid arthritis.	129
Chapter 8:	Unravelling the cost of biological strategies in rheumatoid arthritis: a kaleidoscope of methodologies, interpretations and interests.	159

***PART II: Monitoring disease activity in treat-to-target strategies:
exploring usefulness of the HandScan.***

Chapter 9:	Utility of the HandScan in monitoring disease activity and prediction of clinical response in rheumatoid arthritis patients: an explorative study.	169
Chapter 10:	Development and validation of rheumatoid arthritis disease activity indices including HandScan (optical spectral transmission) scores.	185
Chapter 11:	Classification of RA into active or inactive disease, using a modified DAS with a HandScan score replacing joint counts.	203
Chapter 12:	A gender difference in HandScan scores in rheumatoid arthritis patients and controls?	215
Chapter 13:	Summary and general discussion	221
Addendum	Nederlandse samenvatting	234
	Dankwoord	242
	Curriculum vitae	245
	List of publications	246
	List of conference abstracts	248

Chapter 1:
General introduction

GENERAL INTRODUCTION

Rheumatoid arthritis (RA) is a relatively common systemic inflammatory disease, which is characterized by chronic inflammation of the synovial membrane.¹ This can result in joint damage, declined physical function and reduced quality of life.² The prevalence is about 1-2% in Dutch adults and it more often affects females.³ The disease has a high socio-economic impact as RA patients may be limited in their daily functioning due to disease-related stiffness, fatigue and/or pain, which may result in absence from (paid) work due to sick leave. In addition, presenteeism (i.e., being at work but working less efficiently) is observed more often in RA patients, compared to in healthy persons of the same age.⁴ In 2017, approximately 0.75% of all health care costs in the Netherlands (i.e., 654 million euro) was spent on healthcare for RA patients.⁵

To optimally treat RA patients, it is essential to reduce disease activity as soon after diagnosis as possible, preferably within the 'window of opportunity'.⁶ This refers to a time period starting at the onset of symptoms, and lasting presumably 3-6 months, in which initiation of treatment might result in improved disease related outcomes on the longer-term, due to prevention of irreversible joint damage, that may fuel disease activity.^{7,8} For instance, drug free remission is achieved in more patients starting treatment (very) early.⁷ In this, the optimal treatment should be based on shared decision making between patient and rheumatologist, taking factors like disease activity, progression of structural joint damage, safety aspects, comorbidities, personal circumstances, quality of life and cost into account.²

Medical treatment with disease modifying anti-rheumatic drugs (DMARDs) targets inflammation, and is associated with improved physical functioning and inhibition of progression of structural joint damage. Conventional synthetic (cs), biologic (b), and target synthetic (ts) DMARDs are used in RA patients.¹ Generally treatment with a single csDMARD, mostly methotrexate (MTX), is started directly after diagnosis according to international guidelines.^{2,9} Since it will take MTX therapy several weeks exert its full effect, it may be combined with short-term use of glucocorticoids (GCs) that are immediately effective, to 'bridge' the lag-time of csDMARD to become fully effective.² If the treatment target is not achieved within 3-6 months and poor prognostic factors (e.g., joint damage or high number of swollen joints) are present, the treatment strategy will have to be intensified.^{2,9} In that case, another csDMARD, a bDMARD (e.g., TNF-inhibitors) or a tsDMARD (e.g., small molecules such as JAK-inhibitors) can be added (i.e., more intensive treatment strategy). The bDMARDs are produced using live cell systems, and categorized based on their specific mechanism of action. bDMARDs and the recently introduced tsDMARDs are considered highly effective, but significantly more expensive compared to csDMARDs.

In the past two decades, several studies have shown that early and more intensive treatment strategies including (combinations of) cs- and bDMARDs, as well as treating to a predefined low level of disease activity, preferably remission, using protocolized step-up treatment (treat-to-target) and regularly monitoring disease activity (tight-control), significantly improve outcomes in RA patients.^{2,10}

Approximately 30-50% of early RA patients will need more intensive treatment, as they fail to respond sufficiently on MTX-monotherapy (with or without GC bridging).¹¹ Therefore, several studies tried to predict (in)sufficient response to MTX, however, resulting models are not implemented mainly due to their limited predictive ability. If MTX non-response or toxicity could be reliably predicted, then treatment with the bDMARD tocilizumab (TCZ, interleukin(IL)-6 receptor inhibitor) might be a suitable option for these non-responders since TCZ can be used with lower-dosed MTX, or even without MTX (i.e., TCZ-monotherapy). In contrast to other bDMARDs, like the most tumour necrosis factor inhibitors (TNFi) which are more effective in combination with MTX, no substantial loss in effectiveness is shown for TCZ-monotherapy compared to combination therapy with MTX.¹² Currently, TCZ, like other bDMARDs, is usually not prescribed as first-line treatment, mainly because of cost considerations and, to a lesser extent, safety concerns.²

In treat-to-target treatment strategies, disease activity is usually monitored by using the disease activity score assessing 28 joints (DAS28) as measuring instrument in the Netherlands.¹³ This is a composite measure, consisting of assessment of 28 joints for swelling and tenderness ('joint counts'; SJC28 and TJC28), an acute phase reactant (APR, e.g., erythrocyte sedimentation rate (ESR)) and patient global assessment (PGA) scored on a visual analogue scale (VAS, PGA-VAS).¹ The assessment of joint counts is rather time consuming, requires a trained healthcare professional and its inter-observer variety often is high.¹⁴ Regularly monitoring disease activity thus is both a key element, and a hurdle in the implementation of treat-to-target strategies in standard care in out-patient-clinics.^{15,16}

Other validated composite disease activity measures, which could also be used for monitoring disease activity, are the clinical disease activity index (CDAI) and the simplified disease activity index (SDAI). These indices have several similarities compared to DAS28, like an assessment of TJC28 and SJC28 and a PGA-VAS scored by the patient. In addition, an evaluator global assessment VAS scored by evaluator is included in CDAI and SDAI; an APR is included only in SDAI.¹⁷ An instrument to assess disease-modifying effects, specifically for longer-term disease duration, is radiography, to score radiographic progression of joint damage, typically on X-rays of joints of hands, wrists and feet.¹⁸

A relatively new medical device, using optical spectral transmission (OST) to measure inflammation quickly and objectively in hand and wrist joints, is the HandScan. The HandScan performs a measurement within 5 minutes, without taking time of a

healthcare professional. The OST-score (i.e., the score of the HandScan measurement) has been found to correlate with the grade of inflammation of hand joints as assessed by ultrasonography, coefficient $p=0.54$.¹⁹

Although the effectiveness of treatment for RA has been markedly improved in the last two decades, total costs of the disease have increased. A Scandinavian registry study has shown that total cost of RA between 1990 and 2010 increased with 32%.²⁰ The main cost driver of this increment is drug costs, which increased from 3% to 33% of total costs.²⁰ Given the availability and greater use of bDMARDs, which are considerably more expensive than csDMARDs, this is not surprising. On the other hand, due to better treatment, indirect costs (i.e., costs associated with productivity loss) decreased with 17%, from 75% to 58% of total costs. In addition, inpatient care decreased with 12%, to only 3% of total costs.²⁰ These results indicate that RA patients are more productive and less often hospitalized for their disease nowadays.²¹

Increase in medication costs underlines that benefits of modern treatment strategies, including drawbacks as expensive medication and regularly monitoring disease activity as well, should be weighed against the total costs of RA. Therefore, additional improvements in care for RA patients should contribute to control, not only of disease activity, but also of the cost effectiveness of treatment strategies. For example, tapering of medication when the treatment target is achieved, has been found to be a feasible option and has now been incorporated in European League Against Rheumatism (EULAR) recommendations for management of RA.^{22,23}

Improvements in care for RA patients might also be expected from exploring and evaluating effectiveness of treatment strategies using combinations of (relative cheap) csDMARDs compared to treatment strategies using the more costly bDMARDs, including protocolized tapering of DMARD treatment. If treatment response in individual patients would be predictable, further benefits might be expected from prediction of treatment response to enable a more stratified ‘personalized’ treatment strategy. If (combinations of) csDMARDs could be used instead of bDMARDs in specified subpopulations of RA without loss of effectivity, this would improve cost effectiveness of treatment, at least on the short term. Vice versa, the cost effectiveness of bDMARDs might improve when they are introduced early in the disease with a strict tapering strategy, and only patients really needing this treatment will remain on this treatment.

Besides facilitating every application of tight-control DMARD treatment strategies, (cost-)effectiveness of care for RA might be further improved if a novel disease measuring device, which is operated without the need for a healthcare professional, would result in less patient visits to the rheumatologist, only of patients with active disease, saving rheumatologists’ time.^{10,24}

Based on this background, the first part of this thesis is aimed to *evaluate the effectiveness and cost-effectiveness of DMARD therapy approaches in treat-to-target*

strategies in (early) RA patients, and the second part to explore whether monitoring of disease activity, a prerequisite for implementation of treat-to-target treatment strategy, can be performed using the HandScan without the need for a healthcare professional.

Outline

PART I: Evaluation of (cost-)effectiveness of DMARD therapy approaches in treat-to-target strategies in (early) RA patients.

1

In **chapter 2**, an overview of the literature is given of the effectiveness of intensive treatment strategies aiming at remission compared to that of standard of care of MTX therapy with or without GC bridging in early RA. **Chapter 3** reports the effectiveness and safety of initiating TCZ, with or without MTX, versus those of initiating MTX with a GC (long-term use of low-moderate dose, i.e., prednisone 10mg/day) within a treat-to-target strategy in early RA, analysing individual patient data of two clinical trials. Since it is known that some patients respond insufficiently to first-line treatment with MTX or are intolerant to MTX, and because specifically in this subgroup, a more intensive initial DMARD strategy could be of value, we performed an external validation of a prediction model for MTX non-response, described in **chapter 4**. Due to the in general more effective treatment nowadays, joint damage in RA is typically limited, making it more difficult to show differences on this outcome between therapies within a single trial. By using individual patient data of several randomized controlled trials, **chapter 5** reports the difference in effect of preventing radiographic progression of TCZ-monotherapy compared to TCZ+MTX, and identifies modifiers (i.e., baseline joint damage, disease activity, disease duration) of this effect.

The ultimate impact of treatment of a chronic disease like RA only becomes apparent over the longer-term. Therefore, the longer-term effectiveness and safety outcomes of patients initially treated with TCZ, MTX or their combination is reported in **chapter 6**, and the cost effectiveness over this period is reported in **chapter 7**. An editorial about different methodologies and interpretations of costing analyses of bDMARD strategies, in response to a paper on the budget impact of introducing an etanercept biosimilar, is provided in **chapter 8**.

PART II: Monitoring disease activity in treat-to-target strategies: exploring usefulness of the HandScan.

In **chapter 9**, the utility of the HandScan in monitoring disease activity and prediction of clinical response in RA patients treated within a treat-to-target strategy was assessed in an exploratory study. In extension of chapter 9, **chapter 10** describes the development

and (external) validation of a new disease activity index using the HandScan. The value of this new index, versus that of DAS28, to classify RA as active versus inactive, using the rheumatologist's clinical classification as reference, was investigated in **chapter 11**. The OST-score, obtained by the HandScan, might be influenced by gender as shown in this chapter. Therefore, in **chapter 12**, RA patients as well as controls are studied to explore whether OST-scores differ between males and females, and to determine which factors influence OST-scores, and which are accountable for the gender difference.

The results of all chapters are summarized and discussed in the final chapter, **chapter 13** and placed in a broader perspective, with suggestions for future research and care.

REFERENCES

1. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
2. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
3. Sloot, R. *et al.* *Reumatische aandoeningen in Nederland: Ervaringen en kengetallen*. (2016).
4. Allaure, S., Wolfe, F., Niu, J., Lavalley, M. & Michaud, K. Work disability and its economic effect on 55–64-year-old adults with rheumatoid arthritis. *Arthritis Care Res.* **53**, 603–608 (2005).
5. Rijksinstituut voor Volksgezondheid en Milieu. Reumatoïde artritis (RA) → Preventie & Zorg → Zorg. *Volksgezondheidenzorg.info* <https://www.volksgezondheidenzorg.info/onderwerp/reumatoïde-artritis-ra/preventie-zorg/zorg> (2020).
6. Raza, K. & Filer, A. The therapeutic window of opportunity in rheumatoid arthritis: does it ever close? *Ann. Rheum. Dis.* **74**, 793–794 (2015).
7. van Nies, J. A. B., Tsonaka, R., Fautrel, B. & van Mil-van der Helm, A. H. M. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann. Rheum. Dis.* **74**, 806–812 (2015).
8. Cush, J. J. Early rheumatoid arthritis - Is there a window of opportunity? *J. Rheumatol.* **30**, 1–7 (2007).
9. Singh, J. a *et al.* 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol.* **68**, 1–26 (2016).
10. Ramiro, S. *et al.* Is treat-to-target really working in rheumatoid arthritis? A longitudinal analysis of a cohort of patients treated in daily practice (RA BIODAM). *Ann. Rheum. Dis.* **79**, 453–459 (2020).
11. Teitsma, X. M. *et al.* Inadequate response to treat-to-target methotrexate therapy in patients with new-onset rheumatoid arthritis: development and validation of clinical predictors. *Ann. Rheum. Dis.* **77**, 1261–1267 (2018).
12. Emery, P., Sebba, A. & Huizinga, T. W. J. Biologic and oral disease-modifying antirheumatic drug monotherapy in rheumatoid arthritis. *Ann. Rheum. Dis.* **72**, 1897–1904 (2013).
13. Felson, D. T. *et al.* American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum.* **63**, 573–586 (2011).
14. Porter, D. *et al.* DAS28 and rheumatoid arthritis: the need for standardization. *Musculoskeletal Care* **9**, 222–227 (2011).

15. Jacobs, J. W. G., ten Cate, D. F. & van Laar, J. M. Monitoring of rheumatoid arthritis disease activity in individual patients: still a hurdle when implementing the treat-to-target principle in daily clinical practice. *Rheumatology* **54**, 959–961 (2015).
16. Pincus, T. & Segurado, O. G. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. *Ann. Rheum. Dis.* **65**, 820–822 (2006).
17. Smolen, J. S. & Aletaha, D. Scores for all seasons: SDAI and CDAI. *Clin. Exp. Rheumatol.* **32**, S75–S79 (2014).
18. van der Heijde, D. & Landewé, R. Should radiographic progression still be used as outcome in RA? *Clin. Immunol.* **186**, 79–81 (2018).
19. Besselink, N. J. *et al.* Optical spectral transmission to assess inflammation in hand and wrist joints of rheumatoid arthritis patients. *Rheumatology* **57**, 865–872 (2018).
20. Kalkan, A., Hallert, E., Bernfort, L., Husberg, M. & Carlsson, P. Costs of rheumatoid arthritis during the period 1990–2010: a register-based cost-of-illness study in Sweden. *Rheumatology* **53**, 153–160 (2014).
21. Hsieh, P. H. *et al.* Economic burden of rheumatoid arthritis: a systematic review of literature in biologic era. *Ann. Rheum. Dis.* **79**, S771–S777 (2020).
22. Burmester, G. R. & Pope, J. E. Novel treatment strategies in rheumatoid arthritis. *Lancet* **389**, 2338–2348 (2017).
23. Kerschbaumer, A. *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.* **79**, S744–S759 (2020).
24. Markusse, I. M. *et al.* Evaluating adherence to a treat-to-target protocol in recent-onset rheumatoid arthritis: reasons for compliance and hesitation. *Arthritis Care Res.* **68**, 446–453 (2016).

PART I

**Evaluation of (cost-)effectiveness of
DMARD therapy approaches in treat-to-
target strategies in (early) RA patients.**

Chapter 2:

**Effectiveness of remission-induction
strategies for early rheumatoid arthritis:
a systematic literature review.**

M.M.A. Verhoeven¹

P.M.J. Welsing¹

J.W.J. Bijlsma¹

J.M. van Laar¹

F.P.J.G. Lafeber¹

J. Tekstra¹

J.W.G. Jacobs¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands

Published in Curr Rheum Rep. 2019;21:24

ABSTRACT

Purpose of review

To review the effectiveness of remission-induction strategies compared to single csDMARD-initiating strategies according to current guidelines in early RA.

Recent findings

Twenty-nine studies, heterogeneous on e.g., specific treatment strategy and remission outcome used, were identified. Using DAS28-remission over 12 months, 13 (76%) of 17 remission-induction strategies showed significantly more patients achieving remission. Pooled relative 'risk' was 1.73 [95%CI 1.59-1.88] for (b)iological DMARD based remission-induction strategies, and 1.20 [95%CI 1.03-1.40] for combination csDMARD based remission-induction strategies compared to single csDMARD-initiating strategies. When additional glucocorticoid 'bridging therapy' was used in single csDMARD-initiating strategies, the higher proportion patients achieving remission in remission-induction strategies was no longer statistically significant (pooled RR 1.06 [95%CI 0.83-1.35]). For other remission outcomes, results were in line with above.

Summary

Remission-induction strategies are more effective in achieving remission compared to single csDMARD-initiating strategies, possibly more so in bDMARD based induction strategies. However, compared to single csDMARD-initiating strategies with glucocorticoids, induction strategies may not be more effective.

INTRODUCTION

In rheumatoid arthritis (RA), early initiation of disease modifying anti-rheumatic drug (DMARD) treatment, preferably within the ‘window of opportunity’, is thought to optimally prevent joint damage, improving long-term outcome and quality of life.^{1,2} Accordingly, current international guidelines advice to start treatment in early RA as soon as possible after diagnosis. Initial therapy is started with a conventional synthetic (cs)DMARD, most frequently methotrexate (MTX), in a ‘tight-controlled’ manner, aiming for low disease activity or, preferably, remission.^{1,3}

Initial MTX therapy is sometimes combined with short-term use of moderate-high dose glucocorticoids (GCs), which are then tapered as soon as possible: GC bridging therapy. The treatment strategy has to be intensified if the treatment target is not achieved within 6 months.^{1,3} This next step is often to add a biological (b) or targeted small molecule (ts) DMARD.^{4,5}

Previous research shows that approximately 30-50% of early RA patients need additional b/tsDMARD therapy.⁶

Patients who initiate a more intensive DMARD strategy as first-line treatment than that according to current guidelines as described above, have sometimes shown superior effectiveness outcomes, and achieve remission more often and earlier, sometimes also including sustained remission (SR) and even sustained drug free remission (sDFR), which may thus become achievable treatment targets.^{4,7}

Achieving remission earlier has been found to be related to improved long-term outcomes.⁷

Furthermore, SR and sDFR may become future treatment targets for early RA within the window of opportunity. This may lead to a paradigm shift towards the above described so called ‘remission-induction’ strategies.

For this reason it would be interesting to investigate the effectiveness of initiating in early RA more intensive treatment strategies, compared to single csDMARD-initiating strategies according to current guidelines; these more intensive strategies herein are designated remission-induction strategies.

The aim of the study is to provide a systematic summary of these remission-induction strategies and their effectiveness.

METHODS

Systematic literature search and study selection

A systematic review of the literature was performed according to current standards and reported according to the Preferred Items for Systematic Reviews and Meta-analyses (PRISMA) statement protocol.⁸ In October 2018 we performed a literature search in

Medline and Embase. The search combined terms relating to early RA, terms for cs-, b, and tsDMARD and remission, and publications limited to the last 5 year and English language. More details about the research question and search terms can be found in the Supplementary Data S1.

We defined more intensive, remission-induction strategies as initiating treatment with a bDMARD or a tsDMARD, both with or without a csDMARD, or initiating a csDMARD with moderately-high dosed GCs, with delayed tapering (not 'bridging therapy'), or starting ≥ 2 csDMARDs.

The single csDMARD-initiating strategy was defined as starting treatment with a single csDMARD, with or without GCs as bridging therapy, according to the current guidelines.

All titles and abstracts were screened by MMAV. If the reviewer was unsure about in-/excluding an abstract, it was discussed with one other co-author (PMJW) and one co-investigator (MdH) to reach consensus, and in case of remaining doubt based on title/abstract, the publication was included for full text evaluation. Full text screening was performed using the same strategy.

The following selection criteria were used; 1) human studies, 2) (very) early RA patients, 3) remission-induction strategy arm (according to definition of remission-induction strategy, see above), 4) single csDMARD-initiating strategy arm (according to definition of single csDMARD-initiating strategy, see above) and, 5) results presented regarding the comparison of a remission-induction strategy and a single csDMARD-initiating strategy on an outcome of remission.

Remission was defined as remission according to a validated disease activity index or the Boolean definition.¹

Randomized controlled trials (RCTs) as well as cohort studies with appropriate correction for multiple confounders were selected. Long-term extension studies of trials satisfying the above criteria were also selected to investigate long-term effects of remission-induction strategies on e.g., radiographic progression.

Data extraction and outcome assessment

The following data of studies was extracted: publication year, study design, patients' baseline characteristics (age, gender, rheumatoid factor (RF) status, health assessment questionnaire (HAQ), symptom duration, disease activity score assessing 28 joints (DAS28)), a description of the single csDMARD-initiating strategy and the remission-induction strategy, the number of patients per arm, a description of the remission outcome, the number of patients achieving remission per arm, a description of missing data and other remarks deemed necessary. In case of a study evaluating long-term outcomes of a remission-induction strategy, we extracted additional information (if available) for the follow-up duration, outcome for disease activity, medication use and radiographic progression.

A quality assessment of all selected publications was performed using "The Cochrane Collaboration's tool for assessing risk of bias".⁹ Information about random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome assessment and selective reporting was evaluated.

Statistics

Relative risks (RR) for achieving remission with 95% confidence intervals (CI) per study were calculated, separate for each remission outcome definition and graphically displayed in forest plots. When appropriate, results were pooled using a random-effect model according to the Mantel-Haenszel method. To explore the effects of specific remission-induction strategy used (e.g., use of a b/tsDMARD, the use of GC bridging therapy in the single csDMARD-initiating strategy) and the effect of symptom duration at start of therapy (within the window of opportunity, arbitrarily defined as symptom duration ≤ 3 months, versus outside the window of opportunity, arbitrarily defined as symptom duration >3 months),² group analyses were performed.

Outcomes of studies evaluating the longer term effectiveness of remission-induction strategies were only summarized descriptively.

All analyses were performed in Review manager version 5.3.¹⁰

RESULTS

After screening, 23 articles and 6 conference abstracts were included, involving 6319 patients treated according to a remission-induction strategy and 4647 according to a single csDMARD-initiating strategy (see flowcharts in Supplementary Figure S1). Four specific groups were defined based on characteristics of the drug regime and study duration, and comparisons made; 1) b/tsDMARD based remission-induction strategy versus single csDMARD-initiating strategy without GC bridging, 2) combination csDMARD based remission-induction strategy versus single csDMARD-initiating strategy without GC bridging, 3) remission-induction strategy (either combination csDMARD based strategy or bDMARD based strategy) versus single csDMARD-initiating strategy with GC bridging, and 4) studies evaluating long-term effects of remission-induction strategies (follow-up >4 years). An overview of patient and study characteristics of the included studies is shown in Table 1.

Table 1 Baseline patient and disease characteristic of included studies.

First author, publication year, reference	Design	Age in yrs. (SD)	Female (%)	RF+ score (SD)	HAQ score (SD)	Symptom duration (SD)	DAS28 (SD)	single csDMARD-initiating strategy	N in single csDMARD-induction strategy	Remission induction strategy	N in remission induction strategy	Time of assessment in yrs.	Treatment characteristics (both arms)
b/tsDMARD based remission-induction strategy versus single csDMARD-initiating strategy without GC bridging													
Atsumi, 2016 ¹	RCT	49 (11)	81	95	1.1 (0.7)	16 (11)	5.5 (1.2)	MTX+PBO	157	CZP+MTX	159	1	T2T
Bijlsma, 2016 ⁴	RCT	54 ⁽⁵⁾	67	72	1.2 (0.6)	4 ⁽⁵⁾	5.2 (1.1)	MTX+PBO	108	TCZ+MTX TCZ+PBO	106 103	0.5	T2T
Burmester, 2016 ¹²	RCT	50 (13)	78	89	1.6 (0.7)	26 (26)	6.7 (1.0)	MTX+PBO	289	TCZ+MTX (reduced dose) TCZ+PBO	291 290	1	T2T
Douglasos, 2014 ¹³	RCT	52 (14)	72	§	§	34 (22)	6.5 (1.0)	MTX	178	ETN+MTX	213	1	T2T
Emery, 2017 ¹⁴ Horslev- Petersen, 2014 ¹⁵	RCT	51 (14) 55 ⁽⁵⁾	77 66	97 72	1.6 (0.6) 1.1 ⁽⁵⁾	12 (17) 12 ⁽⁵⁾	6.7 (0.9) 5.6 [§]	MTX+PBO MTX+PBO	213 91	CZP+MTX ADA+MTX	655 89	1	T2T
Keystone, 2017 ¹⁶ Keystone, 2017 ^a ¹⁷	RCT	52 (14)	84	73	1.5 (0.7)	39 (44)	6.3 (0.9)	MTX+PBO	210	BARI + MTX BARI	215 159	1	T2T
Kirchgsner, 2018 ¹⁸	RCT	48 (12)	§	§	§	§	§	PBO	15	INF	15	1	no adjustments INF until 22w
Nam, 2014b ¹⁹ Smolen, 2015 ²⁰	RCT	48 (13) 49 (13)	76 78	55 97	1 (0.4) 1.7 (0.7)	28 ⁽⁵⁾ 26 (29)	4.1 (1.1) 6.3 (1.0)	MTX+PBO MTX+PBO	55 209	ETN+MTX ADA+MTX	55 210	1	T2T
Stamm, 2018 ²¹	RCT	53 (14)	73	35	0.9 (0.7)	10 (2)	4.9 (1.4)	MTX+PBO	36	INF+MTX	38	1	step up MTX
Takeuchi, 2014 ²²	RCT	54 (13)	81	84	1.2 (0.8)	16 (21)	6.6 (1.0)	MTX+PBO	163	ADA+MTX	171	0.5	T2T

Table 1 (continued)

First author, publication year, reference	Design	Age in yrs.	Female (%)	RF- score (SD)	HAQ score (SD)	Symptom duration (SD) in weeks (SD)	csDMARD-initiating strategy	single csDMARD-initiating strategy	N in single csDMARD-induction strategy	Remission-induction strategy	N in remission-induction strategy	Time of assessments in yrs.	Treatment characteristics (both arms)
Combination csDMARD based remission-induction strategy versus single csDMARD-initiating strategy without GC bridging													
Brunekreef, 2017 ²³	Cohort	59 (14)	62	65	§	§	§	MTX	297	MTX+HCQ+GCIV	156	1	T2T IM 80-120mg
Ma, 2014 ²⁴	RCT	54 (§)	68	87	1.6 (§)	§	5.8 (1.3)	MTX	87	CSA+MTX+GC	90	2	T2T bridg. 34W
Rannio, 2017 ²⁵	Cohort	57 (16)	67	71	0.9 (§)	24 (§)	4.2 (1.4)	MTX(+GC)	453	MTX+ SSZ+HCQ	158	1	T2T
Steunebrink, 2016 ²⁶ cohorts	2	59 (13)	62	54	1.1 (§)	§	4.7 (1.1)	MTX+PBO	128	MTX+HCQ	128	1	T2T
Remission-induction strategy (either combination csDMARD based strategy or bDMARD based strategy) versus single csDMARD-initiating strategy with GC bridging													
Akdemir, 2018 ²⁷	2 RCT	54 (14)	69	63	1.5 (0.7)	20 (§)	4.3 (0.8)**	MTX+GC	175	MTX+SSZ+GC	133	1	T2T bridg. 34W
DeJong, 2014 ²⁸	RCT	54 (14)	68	71	1 (0.7)	24 (13)	3.4 (1.0)**	MTX+GC	97	MTX+SSZ+HCQ+ GCIV	91	1	T2T IM 80-120mg or bridg.
Nam, 2014a ²⁹	RCT	53 (13)	69	55	1.4 (0.5)	5 (§)	3.8 (1.0)**	MTX+GCIV	57	MTX+INF	55	0.5	T2T IV 250mg
Stouten, 2017 ³⁰	RCT	§	§	§	§	§	§	MTX+GC	98	MTX+SSZ+GC MTX+LEF+GC	98 93	1	T2T bridg. 34W
Ter Wee, 2015 ³¹	RCT	52 (13)	69	59	1.4 (0.7)	24 (20)	5.4 (1.2)	MTX+GC	81	MTX+SSZ+GC	81	1	T2T bridg. 34W
Verschueren, 2017 ³²	Trial	52 (13)	71	58	1 (0.7)	3 (4)	4.7 (1.4)	MTX+GC MTX	141 172	MTX+SSZ+GC MTX+LEF+GC	98 90	1	T2T bridg. 34W

Table 1 (continued)

First author, publication year, reference	Design	Age in yrs. (SD)	Female (%)	RF+ (SD)	Symptom duration (SD) in weeks (SD)	DAS28 csDMARD- initiating strategy	N in single	Remission-induction strategy	N in remission-induction strategy	Time of assessment in yrs.	Characteristics (both arms)
Studies evaluating long-term effects of remission induction strategies (follow-up >4 years)											
Bergsma, 2017 ³³	RCT	55 (11)	55	61	1.3 (0.7) 21 [§]	4.2 (0.9) MTX+PBO	247	MTX+SSZ+GC	261	10	T2T bridg. 34W
Emery, 2016 ³⁴	RCT	§	§	§	§	§	MTX+PBO	160	GOL (reduced dose)+MTX	159	5 T2T
Keystone, 2014 ³⁵	RCT	§	§	§	1.3 (0.7) §	5.6 (1.7) MTX+PBO	164	GOL+MTX ADA+MTX	159	10	T2T
Konijn, 2017 ³⁶	RCT	57 (13)	67	§	§	§	MTX+GC	81	ADA+PBO MTX+SSZ+GC	196	10
Markusse, 2016 ³⁷	RCT	54 (14)	69	65	1.4 (0.7) 24 [§]	4.4 (0.9)*	MTX+GC	126	MTX+SSZ+GC INF+MTX	166	4 T2T bridg. 34W
Verhoeven, 2018 ³⁸	RCT	§	§	§	§	§	MTX+PBO	72	TCZ+MTX TCZ+PBO	81	4 T2T bridg. 28W

§= (spread of) variable not available; * = bDMARD based remission-induction strategy; ** = DAS44/assessing 44 joints); reduced dose= 0.5 of normal dose; RF= rheumatoid factor; HAQ= health assessment questionnaire; DAS28= disease activity score assessing 28 joints; RCT= randomized controlled trial; bridg= bridging therapy; im= intramuscular ;iv= intravenous; Ada= adalimumab; BarI= baricitinib; CSA= cyclosporine; CZP= certolizumab pegol; ETN= etanercept; GC= glucocorticoid; GOL= golimumab; HCQ= hydroxychloroquine; INF= infliximab; Lef= leflunomide; MTX= methotrexate; PBO= placebo; SSZ= sulfasalazine; TCZ= tocilizumab; T2T= treatment-to-target treatment strategy including step-up and step-down.

Several of the 29 studies used more than 1 remission definition; in all, 46 remission definitions were used, range 1-4 per study. Most studies used at least a definition of remission where remission had to be present ≥ 1 visit within 6 to 12 months follow-up and according to one of our remission outcome definitions, and we will describe the results based on these outcomes (Table 1). Seventeen studies defined remission as DAS28 <2.6, 12 studies used the Boolean remission definition, 7 studies used CDAI ≤ 2.8 and 10 studies used SDAI ≤ 3.3 , results are described separately below. Overall, for 32 of the 46 remission definitions (70%), a statistically significant effect in favor of remission-induction strategy was found.

DAS28-based remission

When DAS28 was used for remission definition, 13/17 (76%) studies showed a statistically significant effect in favor of the remission-induction strategy, Figure 1. The pooled RR of achieving remission for strategies using a bDMARD in the remission-induction strategy compared to the single csDMARD-initiating strategy without GC bridging was 1.73 [95%CI 1.59 - 1.88] versus 1.20 [95%CI 1.03 - 1.40] for studies which used a combination csDMARD based remission-induction strategy compared to the single csDMARD-initiating strategy without GC bridging. For studies using GC bridging in the single csDMARD-initiating strategy, no statistically significant additional effect for the remission-induction strategy was found (pooled RR 1.06 [95%CI 0.83 - 1.35]). One of them used a bDMARD in the remission-induction strategy arm.²⁹ One cohort study only provided an odds ratio (OR) for achieving remission in patients treated with a remission-induction strategy compared to a single csDMARD-initiating strategy, with or without additional GC use (without sufficient information to calculate an RR). Results were in favor of the remission-induction strategy (OR 1.82 [95%CI 1.01 - 3.29]).²⁵

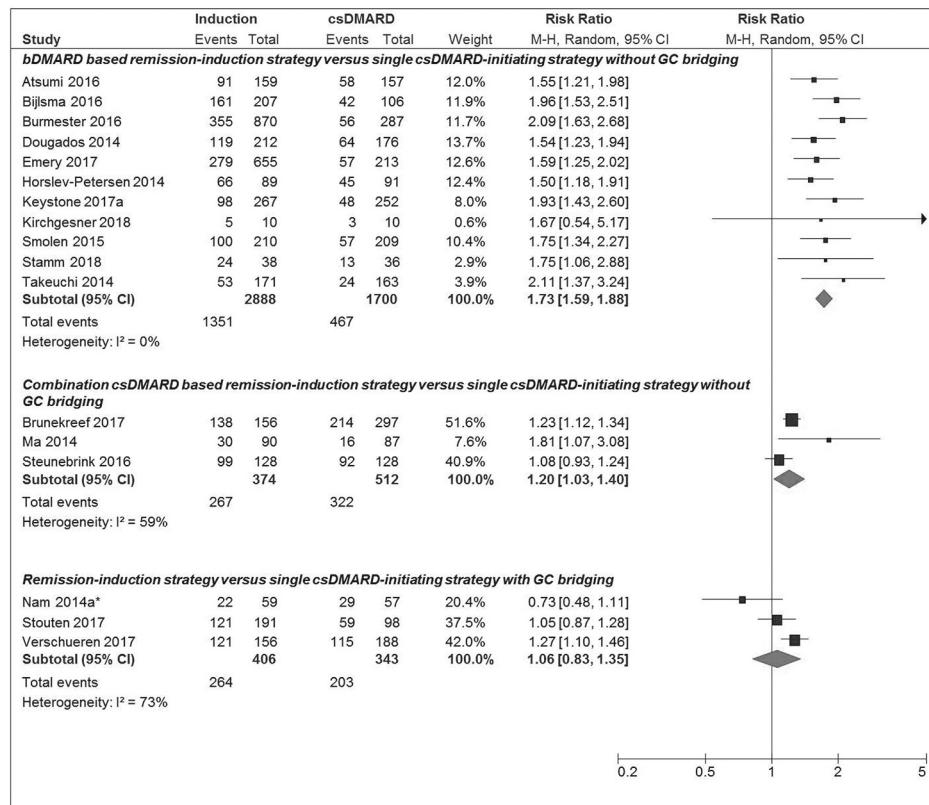


Figure 1 Forest plot of DAS28 remission outcome in individual studies comparing remission-induction strategies with single csDMARD-initiating strategies.

DAS28 remission= DAS28 <2.6; Induction= remission-induction strategy arm; csDMARD= single csDMARD-initiating strategy arm; M-H= Mantel-Haenszel; Random= random effect; CI= confidence interval; *= bDMARD in remission-induction strategy.

Boolean-based remission

For Boolean remission, 5/12 (42%) studies showed a statistically significant effect in favor of the remission-induction strategy. The pooled RR of achieving Boolean remission for the bDMARD based remission-induction strategy compared to the single csDMARD-initiating strategy without GC bridging was 1.75 [95%CI 1.40 - 2.20] versus 0.79 [95%CI 0.58 - 1.07] for the remission-induction strategy (1/5 bDMARD use in the remission-induction strategy)²⁹ compared to the single csDMARD-initiating strategy with GC bridging, Figure 2.

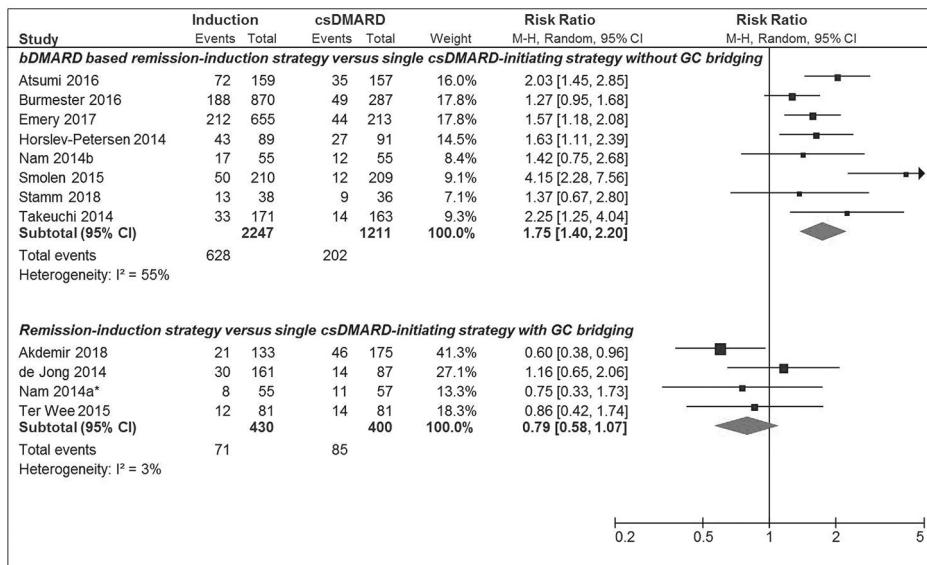


Figure 2 Forest plot of Boolean remission outcome in individual studies comparing remission-induction strategies with single csDMARD-initiating strategies.

Boolean remission= tender joint count ≤ 1 , swollen joint count ≤ 1 , CRP $\leq 1 \text{ mg/dL}$, patient global assessment ≤ 1 (on a 0-10 scale); Induction= remission-induction strategy arm; csDMARD= single csDMARD-initiating strategy arm; M-H= Mantel-Haenszel; Random= random effect; CI= confidence interval; *= bDMARD based remission-induction strategy.

CDAI-based remission

Only studies with b/tsDMARD use in the remission-induction strategy versus single csDMARD-initiating strategy without GC bridging were included in the analysis for CDAI remission. All studies (7/7, 100%) showed a statistical significant effect in favor of the remission-induction strategy arm. The pooled RR of achieving CDAI remission was 1.68 [95%CI 1.46 - 1.92], Figure 3.

SDAI-based remission

Nine studies with bDMARD use in the remission-induction strategy arm versus single csDMARD-initiating strategy without GC bridging, and one study using a bDMARD based remission-induction strategy versus a single csDMARD-initiating strategy with GC bridging were included in the analysis for SDAI remission.²⁹ A significant effect in favor of the remission-induction strategy was found in 7/10 (70%) studies, Figure 4. The pooled RR of achieving SDAI remission was 1.66 [95%CI 1.44 - 1.90] for bDMARD use in the remission-induction strategy arm versus the single csDMARD-initiating strategy without GC bridging arm. And for the single study where a remission-induction strategy was compared to a single csDMARD-initiating strategy with GC bridging, this was 1.10 [95%CI 0.60 - 2.05].

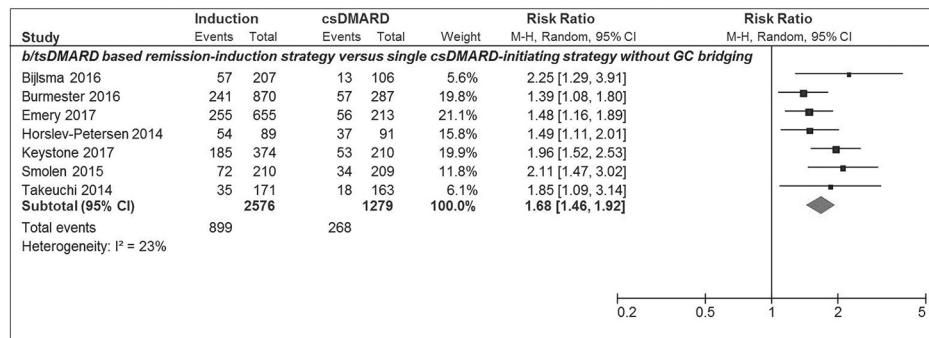


Figure 3 Forest plot of CDAI remission outcome in individual studies comparing remission-induction strategies with single csDMARD-initiating strategies.

CDAI remission= CDAI ≤ 2.8 ; Induction= remission-induction strategy arm; csDMARD= single csDMARD-initiating strategy arm; M-H= Mantel-Haenszel; Random= random effect; CI= confidence interval.

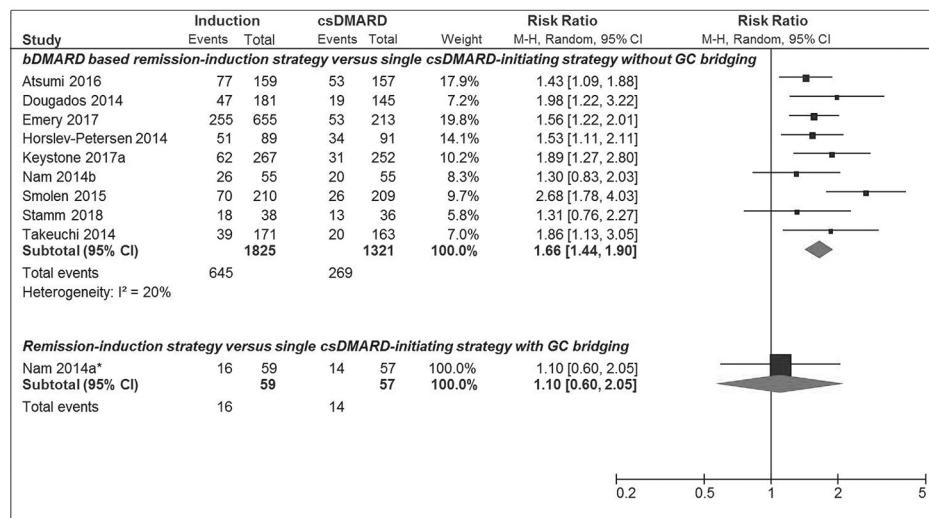


Figure 4 Forest plot of SDAI remission outcome in individual studies comparing remission-induction strategies with single csDMARD-initiating strategies.

SDAI remission= SDAI ≤ 3.3 ; Induction= remission-induction strategy arm; csDMARD= single csDMARD-initiating strategy arm; M-H= Mantel-Haenszel; Random= random effect; CI= confidence interval; *= bDMARD based remission-induction strategy.

Symptom duration

Regarding symptom duration, 6 studies started treatment “within the window of opportunity” (symptom duration ≤ 3 months). Another 9 studies started treatment “outside the window of opportunity” (symptom duration > 3 months; range 4 - 10 months). All studies reported the DAS28-based remission outcome, and 11/15 (73%) showed a statistical significant effect in favor of the remission-induction strategy. The

pooled RR of achieving remission for strategies within the window of opportunity was 1.43 [95%CI 1.15 - 1.77] versus 1.44 [95%CI 1.12 - 1.86] for studies outside the window of opportunity. Five studies used a single csDMARD-initiating strategy with GC bridging (i.e., 2 studies within and 3 studies outside), Supplementary Figure S2.

Longer term effectiveness of remission-induction strategies started in early RA

We found six studies evaluating the effect of a remission-induction strategy versus a single csDMARD-initiating strategy on the long term (4 to 10 years). In 4 studies, DAS remission was more often achieved in the initial remission-induction strategy compared to the single csDMARD-initiating strategy over time.^{33-35,37} In the remission-induction strategy arm, Boolean remission, as well as SDAI remission, was less often achieved in 1 of 2 studies with no difference in the other, compared to the single csDMARD-initiating strategy arm.^{34,36} No difference was found for CDAI remission, which was reported in only one study.³⁴ One study reported data about SR, which was achieved in almost all patients over time, without differences between the different strategy arms.³⁸ However, using (s)DFR as outcome, differences were shown in favor of the remission-induction strategies.^{37,38} No differences were found for radiographic progression over time between the different strategies.^{33-35,37} Details of these studies can be found in Table 1.

Risk of bias assessment

The risk of bias of the included studies was overall low. In general, 26/29 studies were RCTs, the remaining 3 were cohort studies. An overview of the risk of bias assessment is shown in Supplementary Table 1. The studies evaluating long-term effects of a remission-induction strategy compared to a single csDMARD-initiating strategy, after the initial RCT,³³⁻³⁸ treatment was according to the treating physician and standard care, without detailed information on the initial trial and attrition, prohibiting to fully assess all items of the risk of bias assessment. Further, moderate/high risk of bias was present in the seven studies evaluating short-term effects.^{23,25-27,30-32}

DISCUSSION

The current meta-analysis shows that a remission-induction strategy is more effective compared to a single csDMARD-initiating strategy, possibly specifically for a bDMARD based remission-induction strategy. However, this superior effect over single csDMARD-initiating strategy is limited and can no longer be detected statistically significantly when patients are treated initially also with GCs, short-term as 'bridging therapy'. Longer term follow-up studies showed conflicting results, but more a favorable outcome with regard to (s)DFR for the remission-induction strategy may be present.

No overall pooled effect estimate was given as studies were highly heterogeneous in study design regarding e.g., specific drug regimen and remission outcome used. We therefore defined groups of more homogeneous studies based on specific remission outcomes and characteristics of drug regimen. Results within these groups show that heterogeneity is typically low, and therefore we pooled the effect estimates. However, in some of these groups, heterogeneity was moderate, based on differences in study design, medical treatment, risk of bias and/or patient characteristics ($I^2 > 50\%$, see Figure 1 and 2).

One surprising finding was that the added value of a remission-induction strategy was found to be limited and non-statistically significant when compared to a single csDMARD-initiating strategy with GC bridging therapy. This may suggest that the current early start of therapy, including a treat-to-target approach with swift step-up treatment adjustments, achieves already very good results when the initial delay in treatment effectives covered by the bridging therapy.

Contrary to expectation, similar beneficial outcomes for patients treated within the window of opportunity were found when compared with those for patients treated outside the window of opportunity. However, only a limited number of studies reported data on symptom duration which is notoriously difficult to define, and our study was not specifically designed to test the window of opportunity hypothesis. Outside of our study, in some papers a difference in effectiveness of treatment has been shown in favor of patients treated within the window of opportunity.^{2,39}

In general, long-term effectiveness outcomes were not different between a remission-induction strategy and a single csDMARD-initiating strategy probably due to the widely applied the treat-to-target principle.¹

Results of our systematic literature review are in line with an earlier performed systematic literature review, which included only remission-induction strategies using a b/tsDMARD in the experimental arm.⁴⁰ We, uniquely include also combination csDMARD based remission-induction strategy arms, providing results applicable also for countries with limited availability of bDMARDs. Besides, we evaluated several established remission definitions according to validated disease activity indices and the Boolean definition.¹

No data on radiographic progression was reported, because of the limited study duration of most included studies; even over 2 years, radiographic progression is absent or modest at most in treat-to-target studies in early RA.^{41,42} Only some of the long-term extension studies reported on radiographic progression, but did not show any statistically significant differences.

The majority of all included studies, i.e., 20/29 (69%), were RCTs with no to moderate risk of bias. The longer term follow-up studies were follow ups of RCTs, in which the effectiveness was maintained, indicating the quality of keeping to the treat-to-target principle.

Conclusion

Remission-induction strategies initiated in early RA patients are more effective in achieving remission compared to single csDMARD-initiating strategies. However, their benefit compared to that of a single csDMARD-initiating therapy strategy with GC bridging seems to be limited.

REFERENCES

- Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76**, 960–977 (2017).
- Raza, K. & Filer, A. The therapeutic window of opportunity in rheumatoid arthritis: does it ever close? *Ann. Rheum. Dis.* **74**, 793–794 (2015).
- Singh, J. a *et al.* 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol.* **68**, 1–26 (2015).
- Bijlsma, J. W. J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
- Dougados, M. *et al.* Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann. Rheum. Dis.* **76**, 88–95 (2017).
- Teitsma, X. M. *et al.* Inadequate response to treat-to-target methotrexate therapy in patients with new-onset rheumatoid arthritis: development and validation of clinical predictors. *Ann. Rheum. Dis.* **77**, 1261–1267 (2018).
- Ajeganova, S. & Huizinga, T. Sustained remission in rheumatoid arthritis: latest evidence and clinical considerations. *Ther. Adv. Musculoskelet. Dis.* **9**, 249–262 (2017).
- Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* **339**, b2700 (2009).
- Higgins, J. P. T. *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* **343**, 1–9 (2011).
- Higgins, J. P. T. & Green, S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. (The Cochrane Collaboration, 2011, 2011).
- Atsumi, T. *et al.* The first double-blind, randomised, parallel-group certolizumab pegol study in methotrexate-naïve early rheumatoid arthritis patients with poor prognostic factors, C-OPERA, shows inhibition of radiographic progression. *Ann. Rheum. Dis.* **75**, 75–83 (2016).
- Burmester, G. R. *et al.* Tocilizumab in early progressive rheumatoid arthritis: FUNCTION, a randomised controlled trial. *Ann. Rheum. Dis.* **75**, 1081–1091 (2016).
- Dougados, M. R. *et al.* When to adjust therapy in patients with rheumatoid arthritis after initiation of etanercept plus methotrexate or methotrexate alone: findings from a randomized study (COMET). *J. Rheumatol.* **41**, 1922–1934 (2014).
- Emery, P. *et al.* Certolizumab pegol in combination with dose-optimised methotrexate in DMARD-naïve patients with early, active rheumatoid arthritis with poor prognostic factors: 1-year results from C-EARLY, a randomised, double-blind, placebo-controlled phase III study. *Ann. Rheum. Dis.* **76**, 96–104 (2017).

15. Horslev-Petersen, K. *et al.* Adalimumab added to a treat-to-target strategy with methotrexate and intra-articular triamcinolone in early rheumatoid arthritis increased remission rates, function and quality of life. The OPERA Study: an investigator-initiated, randomised, double-blind trial. *Ann. Rheum. Dis.* **73**, 654–661 (2014).
16. Keystone, E. C. *et al.* Time to achieve moderate/low disease activity and remission in RA patients on baricitinib compared to adalimumab, methotrexate, and placebo. *Arthritis Rheumatol.* **69**, (2017).
17. Keystone, E. C. *et al.* Achieving comprehensive disease control in patients with early and established rheumatoid arthritis treated with adalimumab plus methotrexate versus methotrexate alone. *RMD Open* **3**, (2017).
18. Kirchgesner, T. *et al.* Evaluation of MRI RAMRIS score and clinical response in patients with ACPA positive undifferentiated arthritistreated with infliximab versus placebo. *Ann. Rheum. Dis.* **77**, 1764 (2018).
19. Nam, J. L. *et al.* A randomised controlled trial of etanercept and methotrexate to induce remission in early inflammatory arthritis: the EMPIRE trial. *Ann. Rheum. Dis.* **73**, 1027–1036 (2014).
20. Smolen, J. S. *et al.* Attainment and characteristics of clinical remission according to the new ACR-EULAR criteria in abatacept-treated patients with early rheumatoid arthritis: new analyses from the abatacept study to gauge remission and joint damage progression in methotrexate (MTX)-naïve patients with early erosive rheumatoid arthritis (AGREE). *Arthritis Res. Ther.* **17**, 157 (2015).
21. Stamm, T. A. *et al.* Induction of sustained remission in early inflammatory arthritis with the combination of infliximab plus methotrexate: the DINORA trial. *Arthritis Res. Ther.* **20**, 174 (2018).
22. Takeuchi, T. *et al.* Adalimumab, a human anti-TNF monoclonal antibody, outcome study for the prevention of joint damage in Japanese patients with early rheumatoid arthritis: the HOPEFUL 1 study. *Ann. Rheum. Dis.* **73**, 536–543 (2014).
23. Brunekreef, T. & Bernelot Moens, H. Remission induction with methotrexate step-up therapy versus combination of hydroxychloroquine, methotrexate and triamcinolone: 3 year results. *Ann. Rheum. Dis.* **76**, 147–148 (2017).
24. Ma, M. H. Y., Scott, I. C., Dahanayake, C., Cope, A. P. & Scott, D. L. Clinical and serological predictors of remission in rheumatoid arthritis are dependent on treatment regimen. *J. Rheumatol.* **41**, 1298–1303 (2014).
25. Rannio, T. *et al.* Three out of four disease-modifying anti-rheumatic drug-naïve rheumatoid arthritis patients meet 28-joint Disease Activity Score remission at 12 months: results from the FIN-ERA cohort. *Scand. J. Rheumatol.* **46**, 425–431 (2017).
26. Steunebrink, L. M. M. *et al.* Initial combination therapy versus step-up therapy in treatment to the target of remission in daily clinical practice in early rheumatoid arthritis patients: results from the DREAM registry. *Arthritis Res. Ther.* **18**, 60 (2016).
27. Akdemir, G. *et al.* Comparison between low disease activity or das remission as treatment target in patients with early active rheumatoid arthritis. *RMD Open* **4**, (2018).
28. de Jong, P. H. *et al.* Randomised comparison of initial triple DMARD therapy with methotrexate monotherapy in combination with low-dose glucocorticoid bridging therapy: 1-year data of the tREACH trial. *Ann. Rheum. Dis.* **73**, 1331–1339 (2014).
29. Nam, J. L. *et al.* Remission induction comparing infliximab and high-dose intravenous steroid, followed by treat-to-target: a double-blind, randomised, controlled trial in new-onset, treatment-naïve, rheumatoid arthritis (the IDEA study). *Ann. Rheum. Dis.* **73**, 75–85 (2014).
30. Stouten, V. *et al.* Sustained effectiveness of methotrexate with step-down glucocorticoid remission induction (cobra slim) for early rheumatoid arthritis in a treat-to-target setting: 2-year results of the carera trial. *Ann. Rheum. Dis.* **76**, 147 (2017).

31. ter Wee, M. M. *et al.* Intensive combination treatment regimens, including prednisolone, are effective in treating patients with early rheumatoid arthritis regardless of additional etanercept: 1-year results of the COBRA-light open-label, randomised, non-inferiority trial. *Ann. Rheum. Dis.* **74**, 1233–1240 (2015).
32. Verschueren, P. *et al.* Effectiveness of methotrexate with step-down glucocorticoid remission induction (COBRA Slim) versus other intensive treatment strategies for early rheumatoid arthritis in a treat-to-target approach: 1-year results of CareRA, a randomised pragmatic open-label superiority trial. *Ann. Rheum. Dis.* **76**, 511–520 (2017).
33. Bergstra, A. S., Landewé, R. B., Huizinga, T. W. & Allaart, C. F. Rheumatoid arthritis patients with continued low disease activity have similar outcomes over 10 years, regardless of initial therapy. *Ann. Rheum. Dis.* **56**, 1721–1728 (2017).
34. Emery, P. *et al.* Efficacy and safety of subcutaneous golimumab in methotrexate-naïve patients with rheumatoid arthritis: five-year results of a randomized clinical trial. *Arthritis Care Res.* **68**, 744–752 (2016).
35. Keystone, E. C. *et al.* Longterm effect of delaying combination therapy with tumor necrosis factor inhibitor in patients with aggressive early rheumatoid arthritis: 10-year efficacy and safety of adalimumab from the randomized controlled PREMIER trial with open-label extension. *J. Rheumatol.* **41**, 5–14 (2014).
36. Konijn, N. P. C. *et al.* Similar efficacy and safety of initial COBRA-light and COBRA therapy in rheumatoid arthritis: 4-year results from the COBRA-light trial. *Rheumatol.* **56**, 1586–1596 (2017).
37. Markusse, I. M. *et al.* Long-term outcomes of patients with recent-onset rheumatoid arthritis after 10 years of tight controlled treatment: a randomized trial. *Ann. Intern. Med.* **164**, 523–531 (2016).
38. Verhoeven, M. *et al.* U-act-early trial 3 years follow-up. The longer effectiveness of treat-to-target strategies in early RA with tocilizumab, methotrexate, or their combination. *Ann. Rheum. Dis.* **77**, 560 (2018).
39. van Nies, J. A. B. *et al.* Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann. Rheum. Dis.* **74**, 806–812 (2015).
40. Singh, J. A. *et al.* Biologics or tofacitinib for people with rheumatoid arthritis naïve to methotrexate: a systematic review and network meta-analysis. *Cochrane Database Syst. Rev.* **2017**, CD012657 (2017).
41. Teitsma, X. M. *et al.* Radiographic joint damage in early rheumatoid arthritis patients: comparing tocilizumab- and methotrexate-based treat-to-target strategies. *Rheumatol.* **57**, 309–317 (2018).
42. Steunebrink, L. M. M. *et al.* Radiographic progression in early rheumatoid arthritis patients following initial combination versus step-up treat-to-target therapy in daily clinical practice: results from the DREAM registry. *BMC Rheumatol.* **2**, 1–10 (2018).

SUPPLEMENTARY FILE

Supplementary Data S1: Research question (PICO) and search terms

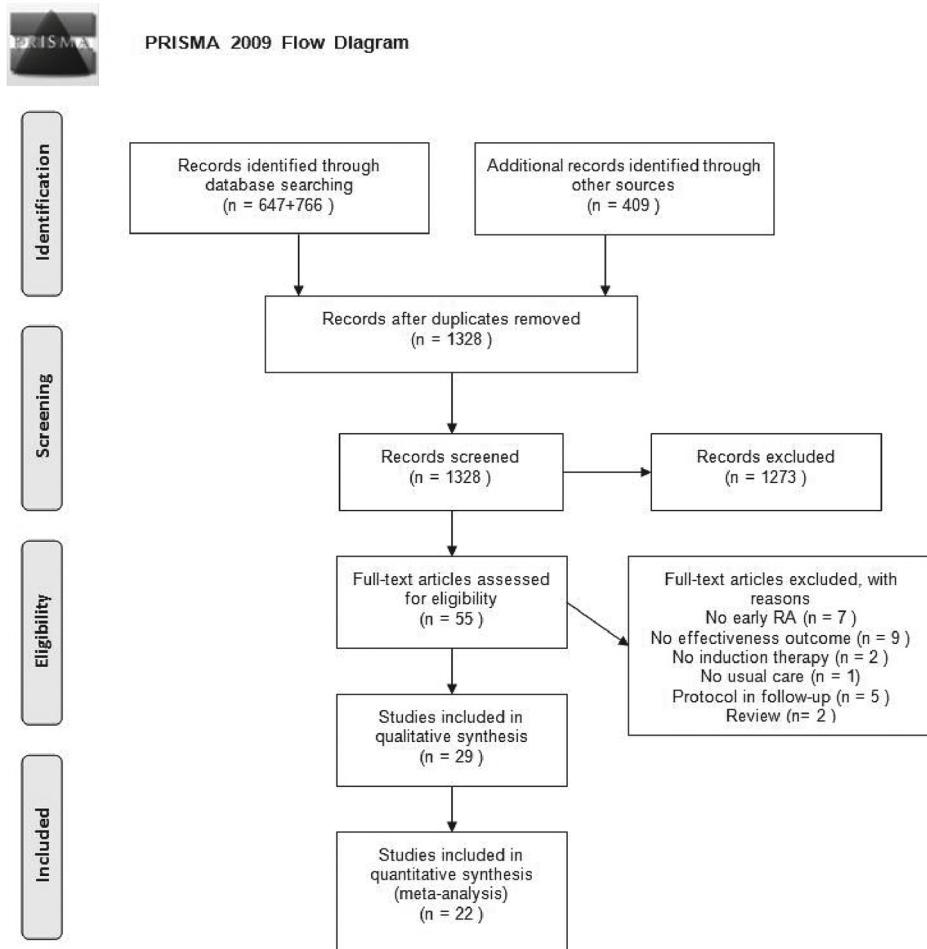
P= early RA patients

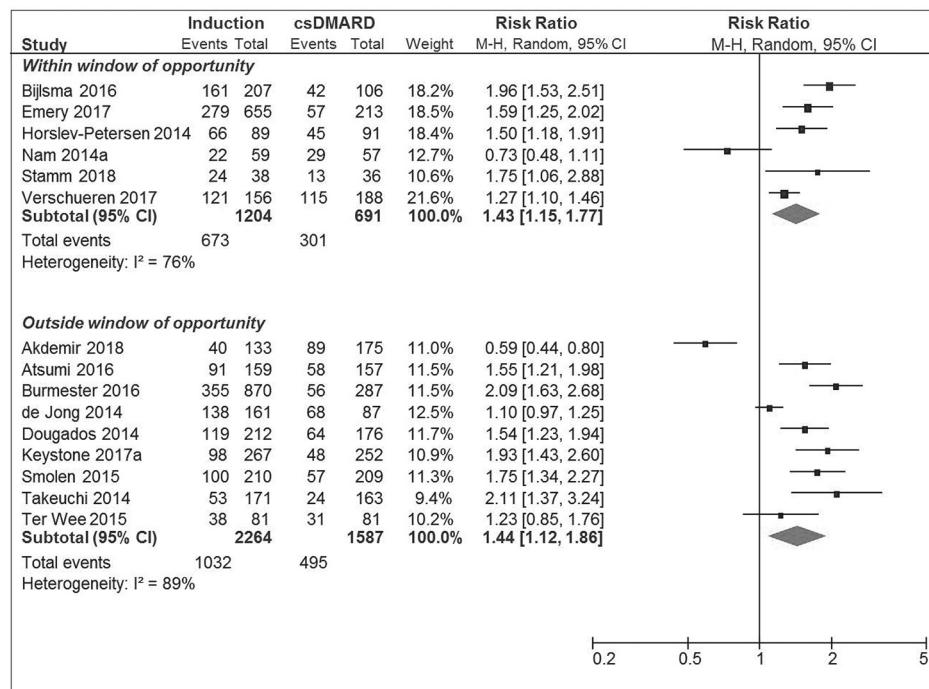
I= treatment with a bDMARD (with or without a csDMARD) or starting with a csDMARDs in addition to glucocorticoids in moderate-high dose or starting with a targeted small molecule, with or without a csDMARDs or combination csDMARDs therapy

C= control therapy (csDMARDs monotherapy in a step-up strategy, low-moderate GCs for short term)

O= remission(as defined by (validated disease activity indices or Boolean definition)

The following key search were used in title and/or abstract; 'early' OR 'dmard-naive' OR 'treatment-naive' OR 'dmard naive' OR 'new onset' OR 'new-onset' AND 'rheumatoid arthritis' OR 'ra' OR 'arthritis' OR 'arthritides' OR 'polyarthritis' OR 'polyarthritides' AND 'disease modifying' OR 'disease-modifying' OR 'antirheumatic' OR 'anti-rheumatic' OR 'anti rheumatic' OR 'biological' OR 'csdmards' OR 'csdmard' OR 'bdmards' OR 'bdmard' OR 'mtx' OR 'methotrexate' OR 'glucocorticoid' OR 'gc' OR 'prednisone' OR 'prednisolone' OR 'steroid' OR 'corticosteroid' OR 'adalimumab' OR 'certolizumab' OR 'etanercept' OR 'golimumab' OR 'infliximab' OR 'abatacept' OR 'rituximab' OR 'tocilizumab' OR 'sulfasalazine' OR 'Sulphasalazine' OR 'salazopyrin' OR 'hydroxychloroquine' OR 'leflunomide' OR 'tofacitinib' OR 'baricitinib' OR 'jak' OR 'janus-associated kinase' OR 'jak inhibitor' OR 'jak-inhibitor' OR 'plaquenil' AND 'remission' OR 'das28' OR 'cdai' OR 'sdai' OR 'disease activity' OR 'effectiveness' OR 'efficacy' OR 'response'.

**Supplementary Figure S1** Flowchart of included studies.



Supplementary Figure S2 Forest plot of DAS remission outcome in each individual study in which patients were treated within the window of opportunity (symptom duration ≤ 3 months) and treated outside the window of opportunity (symptom duration > 3 months).

DAS28 remission= DAS28 < 2.6 ; Induction= remission-induction strategy arm; csDMARD= single csDMARD-initiating strategy arm; M-H= Mantel-Haenszel; Random= random effect; CI= confidence interval; Nam 2014a= bDMARD in remission-induction strategy.

Supplementary Table 1 Quality assessment of individual studies.

First author, publication year	Sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome assessment	Selective reporting
Akdemir 2018	+	+	?	+	?	?
Atsumi 2016	+	+	+	+	+	+
Bergsma 2017	X	X	X	X	+	?
Bijlsma 2016	+	+	+	+	+	+
Brunekreef 2017	-	-	?	+	+	?
Burmester 2016	+	+	+	+	+	+
De Jong 2014	+	+	?	+	+	+
Dougados 2014	+	+	+	+	+	+
Emery 2016	X	X	X	X	+	+
Emery 2017	+	+	+	+	+	+
Horslev-Petersen 2014	+	+	+	+	+	+
Keystone 2014	X	X	X	X	+	+
Keystone 2017	+	+	+	+	+	?
Keystone 2017a	+	+	+	+	+	?
Kirchgsner 2018	+	+	+	+	?	?
Konijn 2017	X	X	X	X	+	+
Ma 2014	+	+	+	+	?	?
Markusse 2016	X	X	X	X	+	+
Nam 2014a	+	+	+	+	+	+
Nam 2014b	+	+	+	+	+	+
Rannio 2017	-	?	?	+	+	?
Smolen 2015	+	+	+	+	?	?
Stamm 2018	+	+	+	+	+	+
Steunebrink 2016	?	?	?	+	+	?
Stouten 2017	+	+	?	?	+	?
Takeuchi 2014	+	+	+	+	+	+
Ter Wee 2015	+	+	-	+	+	+
Verhoeven 2018	X	X	X	X	+	?
Verschueren 2017	+	+	-	?	+	+

X = not applicable; + = low risk; ? = moderate risk; - = high risk.

Chapter 3:

Initiating tocilizumab, with or without methotrexate, compared with starting methotrexate with prednisone within step-up treatment strategies in early rheumatoid arthritis: an indirect comparison of effectiveness and safety of the U-Act-Early and CAMERA-II treat-to-target trials.

M.M.A. Verhoeven¹

M.J.H. de Hair¹

J. Tekstra¹

J.W.J. Bijlsma¹

J.M. van Laar¹

A. Pethö-Schramm²

M.E.A. Borm³

E.J. ter Borg⁴

S.P. Linn-Rasker⁵

X.M. Teitsma¹

F.P.J.G. Lafeber¹

J.W.G. Jacobs¹

P.M.J. Welsing¹

¹*Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands;*

²*F. Hoffmann-La Roche, Basel, Switzerland;*

³*Roche Nederland BV, Woerden, the Netherlands;*

⁴*Department of Rheumatology, Sint Antonius Ziekenhuis, Nieuwegein, the Netherlands;*

⁵*Department of Rheumatology, Meander Medisch Centrum, Amersfoort, the Netherlands.*

ABSTRACT

Objective

Methotrexate (MTX), often combined with low-moderately dosed prednisone, is still the cornerstone of initial treatment for early RA. It is not known how this strategy compares with initial treatment with a biological. We therefore compared the effectiveness of tofacitinib (TCZ), or TCZ plus MTX (TCZ+MTX) with MTX plus 10mg prednisone (MTX+Pred), all initiated within a treat-to-target treatment strategy in early RA.

Methods

Using individual patient data of 2 trials, we indirectly compared tight-controlled treat-to-target strategies initiating TCZ (n=103), TCZ+MTX (n=106) or MTX+Pred (n=117), using initiation of MTX (n=227) as reference. Primary outcome was DAS28 over 24 months. To assess the influence of acute phase reactants (APRs), a disease activity composite outcome score without APR (i.e., m-CDAI) was analysed. Secondary outcomes were remission (several definitions), physical function and radiographic progression. Multi-level models were used to account for clustering within trials and patients over time, correcting for relevant confounders.

Results

DAS28 over 24 months was lower for TCZ+MTX than for MTX+Pred (mean difference: -0.62 [95%CI -1.14 to -0.10]). Remission was more often achieved in TCZ+MTX and in TCZ vs. MTX+Pred ($p=0.02/0.05$, respectively). Excluding APRs from the disease activity outcome score, TCZ-based strategies showed a slightly higher m-CDAI compared to MTX+Pred, but this was not statistically significant. Other outcomes were also not statistically significantly different between the strategies.

Conclusion

In early RA patients, although TCZ-based strategies resulted in better DAS28 and remission rates compared to MTX+Pred, at least part of these effects may be due to a specific effect of TCZ on APRs.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory joint disease often requiring lifelong treatment to reduce symptoms, improve physical function, and avoid structural joint damage. Implementation of treat-to-target treatment strategies has improved those outcomes. Guidelines recommend to start a conventional synthetic (cs) disease modifying anti-rheumatic drug (DMARD) as soon as the diagnosis is made and to swiftly adjust treatment when the treatment target is not achieved.¹ Methotrexate (MTX) is typically the initial treatment, often combined with glucocorticoids (GCs). MTX is well-tolerated, effective and inexpensive. GCs have been used for 70 years in RA and have proven to be effective and safe, if used in low-moderate dose.²⁻⁵ Combination treatment of csDMARDs with GCs improves disease control, physical function and radiographic progression.⁶⁻⁸ Similar beneficial effects have been shown for treatment with biological (b)DMARDs.¹

Earlier performed studies showed that first-line treatment initiating MTX was less effective compared to treatments initiating bDMARDs, importantly for example for achieving drug free remission (DFR).⁹ However, a bDMARD is typically not prescribed as first-line treatment, mainly because of cost considerations and to a lesser extent safety concerns. For this reason several studies attempted to predict bDMARD response on individual patient level, which proved not to be feasible so far.¹⁰ However, with tapering of originator bDMARDs and increasing number of less costly biosimilar approvals, these may in the future be more frequently considered as an initial treatment option in RA, as this may increasingly become more cost-effective. Therefore, it is important to better understand the relative effectiveness and safety of a first-line treatment strategy initiating a bDMARD compared to initiating a csDMARD with GCs. This is particularly important as the long-term impact of a potent treatment strategy might be more pronounced when started in early disease (i.e., within the window of opportunity).¹¹

Our department has previously performed the U-Act-Early and the second Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA-II) trials in early DMARD-naïve RA patients. In U-Act-Early the effectiveness and safety of treat-to-target treatment strategies initiating the bDMARD tocilizumab (TCZ) with MTX (TCZ+MTX) or with placebo (TCZ) were compared with those of a strategy initiating MTX with placebo (MTX).¹² Following the rationale that with longer-term use GCs act as a DMARD, in CAMERA-II effectiveness of a treat-to-target treatment strategy initiating MTX with 10mg prednisone (MTX+Pred) was compared with that of a strategy initiating MTX with placebo (MTX).⁶ The MTX strategies in both trials were similar. The U-Act-Early trial concluded that the TCZ strategy arms were more effective in achieving sustained remission (i.e., remission ≥ 24 weeks) and showed less erosive joint damage than the MTX strategy arm, with a similar safety profile.¹² The CAMERA-II trial concluded that

the MTX+Pred arm was more effective regarding achieving remission and inhibition of erosive joint damage than the MTX arm, with a numerically better safety profile.⁶

We hypothesised that TCZ-based treatment strategies in early RA patients would be more effective than a MTX+Pred-based treatment strategy, and would have similar safety. Therefore, we indirectly compared the effectiveness and safety of initial treatment strategies based on TCZ, with or without MTX, with those of an initial strategy based on MTX with 10mg prednisone, using individual patient data from the U-Act-Early and CAMERA-II trials.

METHODS

Patients

U-Act-Early and CAMERA-II were 2-year, double-blind, placebo-controlled randomized controlled trials in DMARD-naïve early RA patients. Patients who participated (1) fulfilled the 1987 American College of Rheumatology (ACR) criteria (CAMERA-II), or (for U-Act-Early) 2010 ACR/European League Against Rheumatism (EULAR) classification criteria, (2) had a disease duration <12 months and (3) had active disease (in U-Act-Early defined as disease activity score assessing 28 joints (DAS28) ≥ 2.6 ; not specified in CAMERA-II) at baseline.

In both trials, treatment followed a tight-control, treat-to-target strategy, aiming for (sustained) remission. In U-Act-Early remission was defined as DAS28 < 2.6 and ≤ 4 swollen joints on the 28 swollen joint count (SJC28). In CAMERA-II remission was defined as absence of swollen joints and at least 2 of the following: tender joint count (TJC) ≤ 3 , patient visual analogue score (VAS) of general health ≤ 20 mm, erythrocyte sedimentation rate (ESR) ≤ 20 mm/h.

In U-Act-Early, patients were randomized to a treatment strategy initiating oral MTX (starting dose 10mg/week), intravenous TCZ (8mg/kg every 4 weeks) or their combination (TCZ+MTX). In CAMERA-II, patients were randomized to a treatment strategy initiating oral MTX (starting dose 10mg/week) with or without 10mg prednisone (MTX+Pred) daily. In both trials, the MTX dose was increased in steps of 5mg, up to 30mg per week or up to the maximum tolerable dose until remission was achieved. In U-Act-Early when the treatment target was not achieved, placebo was replaced by the respective active treatment in the combination strategy arms initiating MTX or TCZ plus placebo, and in both trials, if the treatment target was still not achieved, a tumour necrosis factor inhibitor was started. More detailed information on the individual trials has been reported previously.^{6,12}

Assessments

For the current analyses, individual patient data from the U-Act-Early and CAMERA-II trials was used. DAS28 (using ESR) over 24 months was defined as the primary outcome. To account for the specific reduction of acute phase reactant (APR) by TCZ, a disease activity outcome not containing APR was used as secondary outcome. For this, we used a modification of the Clinical Disease Activity Index (m-CDAI)¹³ as one of the components of the CDAI (VAS physician) was not available in the CAMERA-II dataset. We used the same formula and rescaled the resulting score to the original range using the formula: $76 * (\text{TJC}28 + \text{SJC}28 + \text{VASpatient}) / 66$.

Other secondary effectiveness outcomes were (1) DAS28 remission ($\text{DAS}28 < 2.6$), (2) m-CDAI remission ($\text{m-CDAI} < 2.8$), (3) cumulative occurrence and persistence of remission as calculated with the continuity rewarded (ConRew) score,¹⁴ (4) physical function using health assessment questionnaire (HAQ) scores and (5) radiographic progression, all over 24 months. Radiographs of both trials were scored by the same professional reader, using the Sharp van der Heijde (SvdH) method. Safety outcomes were: occurrence of ≥ 1 event for infection and for elevated levels (above upper limit of normal) of alanine aminotransferase (ALT; ≥ 55 units per litre) and aspartate aminotransferase (AST; ≥ 40 units per litre),¹⁵ and drop out due to an adverse event (AE). These outcomes were selected because these were deemed clinically relevant and data was similarly collected in both trials.

Statistical analyses

Continuous variables were described using means with standard deviations (SD) or medians with interquartile ranges (IQR), as appropriate. Frequencies and proportions were calculated for categorical variables. Differences between the trial populations in baseline characteristics were evaluated and tested using t-test for continuous outcomes and chi-square test or Fisher exact test for categorical outcomes, or non-parametric alternatives where appropriate.

To account for missing data in relevant baseline data, multiple imputation was used. This was done separately for the dataset of each study (using the same approach and predictor variables), to allow for possible modification of the predictor effects in the imputation model by trial. We used predictive mean matching to impute data, except for smoking status and rheumatoid factor (RF) status for which logistic regression was used. For all imputations the following predictor variables were used: age, DAS28, HAQ, SvdH score and smoking status, all at baseline; gender, treatment strategy, RF-status, ConRew score, and occurrence of ≥ 1 event for elevated ALT, elevated AST, and infection, respectively, and drop out due to an AE. Multiple imputation was performed in R using the MICE package, yielding 40 imputed datasets.

To compare the DAS28 (as well as m-CDAI and HAQ) over time between the TCZ arms and the MTX+Pred arm, a linear mixed effects model was used. A random intercept at study level and at patient level was used to account for clustering of measurements within trials and within patients over time, respectively (for DAS28 as well as other time varying outcomes only). The MTX-strategy was similar in both trials and was used as reference arm in the analysis. The rationale of this approach is that within this reference arm, the effect of study (i.e., population) can be estimated assuming that differences in outcomes between patients from both trials can be regarded due to differences in trial population (as they are treated according to the same strategy). This is then used to correct for potential differences between studies and therefore, the three treatment strategy arms can be compared more validly. Treatment arm was categorised as TCZ, TCZ+MTX, MTX+Pred and MTX. Treatment arm and time were included in all analyses, and all models were corrected for DAS28 and HAQ at baseline, as well as for smoking status and RF-status and other variables that proved relevantly different between trials. We used a natural logarithm transformation for the m-CDAI in the analyses as this score was not normally distributed.¹⁶ A logistic mixed effect model was used for the binary outcomes (occurrence of remission and safety outcomes). For SvdH scores over time, a Poisson mixed effect model was used to handle the skewed distribution of these data. Results of the 40 imputed datasets were pooled using Rubin's Rule.¹⁷

A separate analysis over the short term only (i.e., first 3 months) was performed, for DAS28/m-CDAI based outcomes, to specifically compare the early response.

The influence of missing data imputation on the findings was investigated in a sensitivity analysis using the non-imputed data over 24 months.

Analyses were performed in SAS version 9.4 and R version 3.4.3.

RESULTS

In total 553 patients were included in both trials. During the trials, 146 patients dropped out (80 in U-Act-Early and 66 in CAMERA-II), mostly due to an adverse event. Overall dropout rates were not different between the trials, neither for safety nor for ineffectiveness reasons. Statistically significant differences in baseline characteristics between the trials were observed for alcohol use, RF-status, baseline DAS28 and its individual components, Table 1. An overview of baseline characteristics separately for each treatment arm is provided in Supplementary Table S1 and Supplementary Table S2 provides baseline characteristics separately for the MTX arms that were used as reference.

Table 1 Baseline characteristics.

	U-Act-Early n= 317	CAMERA-II n= 236	P-value
Female, n (%)	212 (67)	142 (60)	0.11
Age (yrs.), mean (SD)	53.4 (12.8)	53.7 (13.4)	0.79
BMI (mg/kg ²), mean (SD)	26.2 (4.3)	25.8 (4.0)	0.23
Alcohol use (≥ 1 unit/week), n (%)	192 (61)	84 (40)	<0.01
Current smoker, n (%)	90 (28)	69 (32)	0.34
Education level, n (%)			0.35
Low	136 (46)	77 (46)	
Moderate	112 (38)	53 (31)	
High	48 (16)	38 (23)	
Disease duration (days), median (IQR)	26 (16 - 43)	*	
RF-positive status, n (%)	231 (73)	118 (60)	<0.01
DAS28, mean (SD)	5.2 (1.1)	5.7 (1.3)	<0.01
CRP (mg/L), median (IQR)	9.0 (3.0 - 21.0)	17.0 (0.0 - 44.0)	0.06
ESR (mm per h), median (IQR)	25.0 (14.0 - 44.0)	31.0 (16.0 - 48.0)	0.02
SJC28, median (IQR)	6.0 (3.0 - 11.0)	10.0 (6.0 - 15.0)	<0.01
TJC28, median (IQR)	7.0 (4.0 - 11.0)	9.0 (5.5 - 16.0)	<0.01
HAQ score, median (IQR)	1.1 (0.6 - 1.5)	1.1 (0.6 - 1.5)	0.52
Any radiographic joint damage, n (%)	83 (26)	57 (24)	0.62

Outcomes are based on non-imputed data of the 2 trials. Reported baseline characteristics of CAMERA-II slightly differ from those of the original publication as we presented non-imputed data compared to imputed data in that publication.^{6,12}

* not collected, but at least <1 year according to the inclusion criteria.

SD= standard deviation; BMI= body mass index; RF= rheumatoid factor; DAS28= disease activity score assessing 28 joints [range 0-9.4; higher is more activity]; CRP= C-reactive protein in milligram per litre; IQR= interquartile range; ESR= erythrocyte sedimentation rate in millimetre per hour; SJC28= 28 swollen joint count; TJC28= 28 tender joint count; HAQ= health assessment questionnaire [range 0-3, 3= worst function]; any radiographic joint damage= total Sharp van der Heijde score ≥ 1 .

The average DAS28 over 24 months was statistically significantly lower in the TCZ+MTX arm compared to the MTX+Pred arm (-0.62 [95%CI -1.14 to -0.10]), Table 2 and Figure 1a. The analyses using the m-CDAI over 24 months as outcome showed no significant differences, but over the first 3 months significantly higher disease activity for TCZ arms compared to the MTX+Pred is found, Table 2 and Figure 1b.

DAS28 remission over 24 months was statistically significantly more often achieved in the TCZ arms compared to the MTX+Pred arm (TCZ+MTX: relative risk (RR) 1.11 [95%CI 1.02 to 1.22]; TCZ: RR 1.09 [95%CI 1.00 to 1.20]), as well over the first 3 months (TCZ+MTX: RR 1.12 [95%CI 1.02 to 1.22]; TCZ: RR 1.11 [95%CI 1.01 to 1.22]), Table 2 and Supplementary Figure S1a. M-CDAI remission was statistically significantly more often achieved in the MTX+Pred arm compared to the TCZ arms over the first 3 months (RR: TCZ+MTX vs. MTX+Pred: 0.90 [95%CI 0.81 to 0.99]; TCZ vs. MTX+Pred: 0.88 [95%CI 0.89

to 0.97]), but not over 24 months, Supplementary Figure S1b. No statistically significant differences were found between the TCZ arms and the MTX+Pred arm for any safety outcome, Table 2. Results for all primary and secondary outcomes are shown in Table 2, Figure 1 and, Supplementary Figure S1-S2.

Table 2 Differences in effectiveness and safety outcomes over 24 and 3 months between TCZ strategy arms and MTX+Pred strategy arm.

	TCZ+MTX vs. MTX+Pred	p-value	TCZ vs. MTX+Pred	p-value
Effectiveness				
DAS28, mean difference (95%CI)				
- Over 24 months	-0.62 (-1.14 to -0.10)	0.02	-0.44 (-1.00 to 0.12)	0.13
- Over first 3 months	-0.44 (-1.00 to 0.13)	0.14	-0.37 (-0.98 to 0.24)	0.24
m-CDAI, proportional mean difference (95% CI) ¹				
- Over 24 months	0.10 (-0.16 to 0.36)	0.47	0.24 (-0.04 to 0.52)	0.09
- Over first 3 months	0.25 (0.02 to 0.48)	0.04	0.35 (0.11 to 0.59)	0.01
DAS28 remission, RR (95%CI)				
- Over 24 months	1.11 (1.02 to 1.22)	0.02	1.09 (1.00 to 1.20)	0.05
- Over first 3 months	1.12 (1.02 to 1.22)	0.03	1.11 (1.01 to 1.22)	0.04
m-CDAI remission, RR (95%CI)				
- Over 24 months	0.96 (0.84 to 1.08)	0.52	0.90 (0.78 to 1.02)	0.12
- Over first 3 months	0.90 (0.81 to 0.99)	0.03	0.88 (0.89 to 0.97)	0.01
ConRew score, mean difference (95%CI)	3.92 (-1.12 to 8.97)	0.14	4.73 (-0.41 to 9.86)	0.08
HAQ score, mean difference (95%CI)	0.01 (-0.14 to 0.17)	0.86	0.09 (-0.07 to 0.25)	0.29
SvdH score, IRR (95%CI)	0.70 (0.32 to 1.57)	0.40	1.05 (0.48 to 2.29)	0.89
Safety				
Occurrence of ≥1x elevated ALT, RR (95%CI)	1.62 (0.66 to 3.97)	0.30	0.92 (0.33 to 2.53)	0.86
Occurrence of ≥1x elevated AST, RR (95%CI)	1.48 (0.35 to 6.17)	0.60	0.55 (0.11 to 2.86)	0.48
Occurrence of ≥1 infection, RR (95%CI)	0.99 (0.11 to 9.03)	0.97	1.05 (0.11 to 9.58)	0.96
Drop out due to AE, RR (95%CI)	1.18 (0.39 to 3.63)	0.77	1.30 (0.43 to 3.90)	0.65

Random effects for the variables in the model were evaluated and retained if the model fit was improved, as judged by a decrease in Bayesian information criterion or a change in one of the estimates of treatment effect of at least 10%. Random effects for the following variables were used in the models: DAS28, HAQ, RF-status, smoking status, all at baseline and time.

Negative values indicate lower values for continuous outcomes (i.e., DAS28, m-CDAI, HAQ) in TCZ strategy arms compared to the MTX+Pred strategy arm.

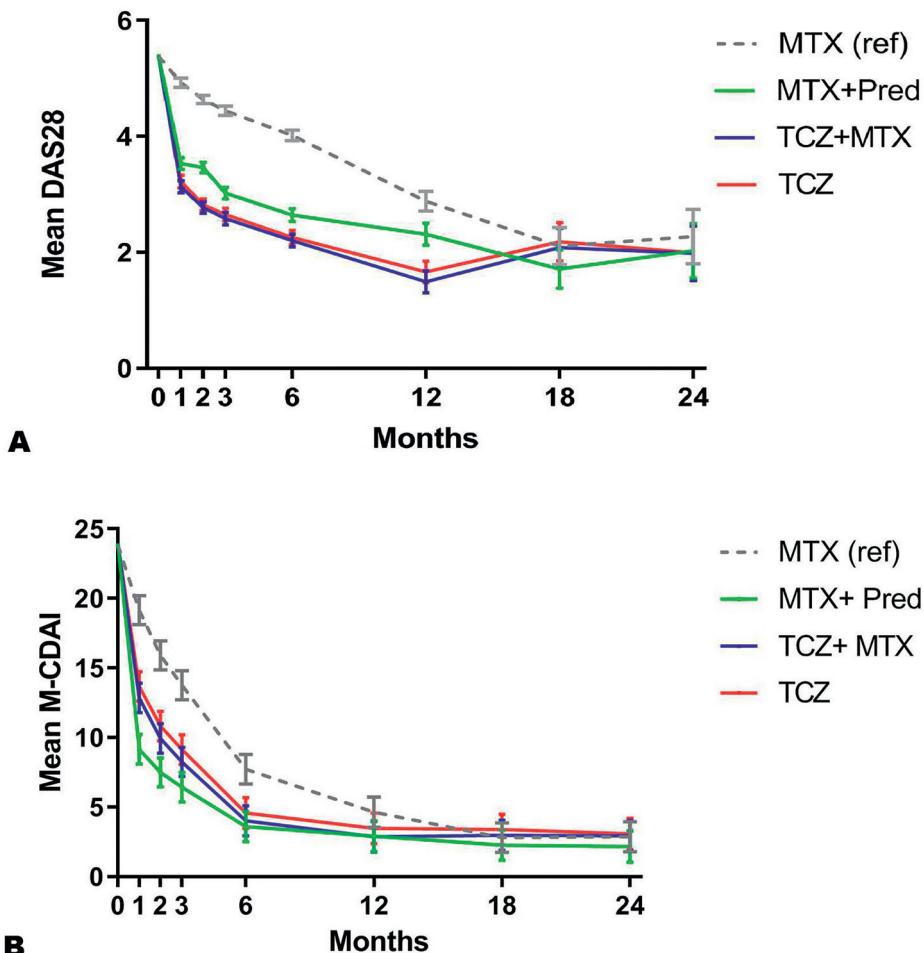
(I)RR >1 indicates higher occurrence of the outcome (i.e., DAS28 remission, m-CDAI remission, drop out due to AE) in TCZ strategy arms compared to the MTX+Pred strategy arm.

¹log transformed values. Multiplying the above regression coefficients by 100 leads to a difference on a modified percentage scale.¹⁶

Table 2 (continued)

Outcomes based on adjusted analyses using pooled outcomes of 40 imputed datasets.

TCZ+MTX= tocilizumab + methotrexate strategy arm; MTX+Pred= methotrexate + prednisone strategy arm, TCZ= tocilizumab + placebo strategy arm; DAS28= disease activity score assessing 28 joints; CI= confidence interval; RR= relative risk; m-CDAI= modified clinical disease activity index; ConRew= continuity rewarded based on DAS28 remission; HAQ= health assessment questionnaire; SvdH= Sharp van der Heijde; IRR= incidence rate ratio; ALT= alanine aminotransferase; AST= aspartate aminotransferase; AE= adverse event.

**Figure 1a** Mean (SE) DAS28 over time for each strategy arm as predicted by model.**Figure 1b** Mean (SE) modified CDAI in each strategy arm as predicted by model, based on back transformation to original scale.

DAS28= disease activity score assessing 28 joints; M-CDAI= modified CDAI; SE= standard error; MTX (ref)= methotrexate + placebo strategy used as reference arm in the analysis; MTX+Pred= methotrexate + prednisone strategy arm; TCZ+MTX= tocilizumab + methotrexate strategy arm; TCZ= tocilizumab + placebo strategy arm.

The sensitivity analyses using non-imputed data over 24 months were generally in line with the results based on imputed data and showed a statistically significantly lower DAS28 over 24 months in favour of the TCZ arms: TCZ+MTX vs. MTX+Pred: -0.51 [95%CI -0.84 to -0.19]; TCZ vs. MTX+Pred: -0.42 [95%CI -0.75 to -0.09]. No statistically significant differences were found for the other outcomes, see Supplementary Table S3.

DISCUSSION

Treatment strategies using TCZ, with or without MTX, resulted in better DAS28-based outcomes over 24 months compared to a strategy using MTX with 10mg prednisone, suggesting better disease control, although this could, at least partly, be driven by the specific APR-lowering effect of TCZ since APR (ESR) is a component of DAS28.¹⁸ Therefore, we analysed m-CDAI which does not include an APR. With this outcome, lower disease activity of TCZ-based strategies compared to MTX+Pred over 24 months could not be confirmed. Interestingly, over the first 3 months only, disease activity as measured with m-CDAI was statistically significantly lower in the MTX+Pred strategy. This raises the issue whether the specific APR-lowering effect of TCZ could clinically be relevant, e.g., with respect to cardiovascular diseases.^{19,20,21}

Despite the absolute differences found in DAS28-based outcomes, differences are likely not clinically relevant. TCZ-based treatment strategies as well as the combination of MTX+Pred in a treat-to-target treatment strategy result in a high proportion of patients being in remission (i.e., 80-60%) depending on remission definition in- or excluding APR. All point estimates were close to 1.00, indicating similar outcomes between treatment strategies. Besides, it is likely that the MTX+Pred treatment strategy is more cost-effective. Eventually, a new relevant treatment target which is becoming more feasible is DFR. The definition of DFR, however, still needs clarification as well as the optimal tapering strategy. In the U-Act-Early trial, DFR was defined as being free from DMARDs or placebo-DMARDs for at least 12 weeks while still in sustained remission (i.e., a DAS28 of less than 2.6 and a swollen joint count of four or fewer joints of the 28 joints assessed, during at least 24 weeks),¹² and a relevant proportion of patients achieved this. Unfortunately it was not possible to compare patients in DFR with the MTX+Pred arm due to different tapering protocols in the trials. For example, in U-Act-Early, MTX and TCZ were both tapered, whereas only MTX was tapered in CAMERA-II.^{6,12} It will be relevant to further investigate DFR as it is an important (future) outcome that may positively affect quality of life and work ability, even on longer term.²²

In the current study there was no clear difference regarding effectiveness and safety between the strategy starting TCZ(+MTX) or MTX+Pred, which might be different for established RA patients with a suboptimal response to MTX. Although in U-Act-Early and the present study,¹² effectiveness of TCZ was somewhat less than that of TCZ+MTX, for

patients with contraindications or intolerance for MTX, TCZ as monotherapy is effective. As currently TCZ (or another bDMARD) as initial therapy in early RA is typically not possible in clinical practice, the conclusion that MTX+Pred can be regarded a similarly effective and safe treatment strategy is reassuring.

A limitation of our study is the fact that this was not a head-to-head comparison and some relevant differences in baseline characteristics between trials were observed. However, all analyses were corrected for observed differences in baseline characteristics and for differences between trials by using the similar MTX arms as a reference strategy. Between the trials, there was a relevant difference in tapering protocols, with a more stringent approach to tapering in U-Act-Early. Furthermore, tapering in both trials was initiated only when being in remission over a certain time period, but the remission criteria were more strict in CAMERA-II; they were close to those of the ACR/EULAR Boolean remission criteria.²³ Tapering may have decreased disease control for the TCZ-based treatment strategies and hampered the adjustment based on reference arm. Less disease control during a later phase of the study for TCZ-based strategies can be seen in Figure 1a. Further, because VAS physician in CAMERA-II was missing, a modification of the CDAI was used which is not validated. The results based on this outcome should be interpreted with caution. Therefore, we thought it inappropriate to post-hoc switch to this measure as primary endpoint for the current study, and used the prespecified DAS28 which has been validated and is generally used in clinical trials, also in CAMERA-II and U-Act-Early. Apart from the difference of not containing an APR, m-CDAI weighs the other components differently from DAS28, which also might have influenced results.

Strengths of our study are our approach of using individual patient data, enabling calculation of uniform outcomes between the two trials, the fact that the trials were performed in a similar setting and included a similar treatment strategy arm, and the appropriate use of multiple imputation and mixed models. As such, handling the clustered structure of the data and correction for important confounding is a robust method of comparing the treatment strategies and optimally using available data.

Conclusion

In early RA, although tight-control strategies initiating TCZ resulted in slightly better mean DAS28 and remission rates, compared to a strategy initiating MTX with 10mg prednisone, at least part of these better effects may be due to the specific effect of TCZ on APRs, as apparent from m-CDAI results. MTX+Pred and TCZ (with or without MTX) appear to have similar clinical effects as initial treatment options in tight-control strategies in early RA.

REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76**, 960–977 (2017).
2. Strehl, C. *et al.* Defining conditions where long-term glucocorticoid treatment has an acceptably low level of harm to facilitate implementation of existing recommendations: viewpoints from an EULAR task force. *Ann. Rheum. Dis.* **75**, 952–957 (2016).
3. Hoes, J. N., Jacobs, J. W. G., Verstappen, S. M. M., Bijlsma, J. W. J. & van der Heijden, G. J. M. G. Adverse events of low- to medium-dose oral glucocorticoids in inflammatory diseases: a meta-analysis. *Ann. Rheum. Dis.* **68**, 1833–1838 (2009).
4. Caldwell, J. R. & Furst, D. E. The efficacy and safety of low-dose corticosteroids for rheumatoid arthritis. *Semin. Arthritis Rheum.* **21**, 1–11 (1991).
5. Buttigereit, F., Burmester, G. R., Straub, R. H., Seibel, M. J. & Zhou, H. Exogenous and endogenous glucocorticoids in rheumatic diseases. *Arthritis Rheum.* **63**, 1–9 (2011).
6. Bakker, M. F. *et al.* Low-dose prednisone inclusion in a methotrexate-based, tight control strategy for early rheumatoid arthritis: a randomized trial. *Ann Intern Med* **156**, 329–338 (2012).
7. Goekoop-Ruiterman, Y. P. M. *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.* **52**, 3381–3390 (2005).
8. Landewé, R. B. M. *et al.* COBRA combination therapy in patients with early rheumatoid arthritis: Long-term structural benefits of a brief intervention. *Arthritis Rheum.* **46**, 347–356 (2002).
9. Ajeganova, S. & Huizinga, T. Sustained remission in rheumatoid arthritis: latest evidence and clinical considerations. *Ther. Adv. Musculoskelet. Dis.* **9**, 249–262 (2017).
10. Lequerré, T., Rottenberg, P., Derambure, C., Cosette, P. & Vittecoq, O. Predictors of treatment response in rheumatoid arthritis. *Jt. Bone Spine* **86**, 151–158 (2019).
11. Raza, K. & Filer, A. The therapeutic window of opportunity in rheumatoid arthritis: does it ever close? *Ann. Rheum. Dis.* **74**, 793–794 (2015).
12. Bijlsma, J. W. J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
13. Smolen, J. S. & Aletaha, D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. *Arthritis Rheum.* **63**, 43–52 (2011).
14. Boers, M. *et al.* A new graph and scoring system simplified analysis of changing states: disease remissions in a rheumatoid arthritis clinical trial. *J. Clin. Epidemiol.* **63**, 633–637 (2010).
15. Genovese, M. C. *et al.* Transaminase levels and hepatic events during tocilizumab treatment pooled analysis of long-term clinical trial safety data in rheumatoid arthritis. *Arthritis Rheumatol.* **69**, 1751–1761 (2017).
16. Cole, T. J. & Altman, D. G. Statistics Notes: percentage differences, symmetry, and natural logarithms. *BMJ* **358**, 1–2 (2017).
17. Marshall, A., Altman, D. G., Holder, R. L. & Royston, P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med. Res. Methodol.* **9**, 57 (2009).
18. Schoels, M., Alasti, F., Smolen, J. S. & Aletaha, D. Evaluation of newly proposed remission cut-points for disease activity score in 28 joints (DAS28) in rheumatoid arthritis patients upon IL-6 pathway inhibition. *Arthritis Res. Ther.* **19**, 155 (2017).

19. Kay, J. *et al.* Clinical disease activity and acute phase reactant levels are discordant among patients with active rheumatoid arthritis: acute phase reactant levels contribute separately to predicting outcome at one year. *Arthritis Res. Ther.* **16**, R40 (2014).
20. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **365**, 2205–2219 (2011).
21. Nishimoto, N. & Mima, T. CHAPTER 10J - Tocilizumab. in *Rheumatoid Arthritis*. 367–371 (2009).
22. van Nies, J. A. B., Tsionaka, R., Fautrel, B. & van Mil-van der Helm, A. H. M. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann. Rheum. Dis.* **74**, 806–812 (2015).
23. Felson, D. T. *et al.* American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum.* **63**, 573–586 (2011).

SUPPLEMENTARY FILE

Supplementary Table S1 gives an overview of baseline characteristics for all strategy arms per trial.

Supplementary Table S1 Baseline characteristics for all treatment strategy arms per trial.

	U-Act-Early		CAMERA-II		
	TCZ+MTX n=106	TCZ n=103	MTX (ref) n= 108	MTX+Pred n= 117	MTX (ref) n= 119
Female, n (%)	65 (61)	78 (76)	69 (64)	70 (60)	72 (61)
Age (yrs.), mean (SD)	53.1 (11.8)	55.0 (12.9)	52.2 (13.7)	54.4 (13.8)	53.0 (13.0)
BMI (kg/m ²), mean (SD)	26.5 (4.8)	25.9 (3.9)	26.2 (4.3)	25.7 (4.1)	25.9 (4.0)
Alcohol use (\geq 1 unit/week), n (%)	64 (60)	63 (61)	65 (60)	43 (41)	41 (39)
Current smoker, n (%)	36 (34)	23 (22)	31 (29)	38 (32)	31 (28)
Education level, n (%)					
Low	45 (45)	30 (31)	48 (49)	35 (46)	41 (29)
Moderate	38 (38)	36 (37)	37 (37)	26 (33)	27 (47)
High	17 (17)	31 (32)	14 (14)	16 (21)	22 (24)
Disease duration (days), median (IQR)	24.5 (16.0 – 14.5)	25.5 (18.0 – 45.0)	27.0 (15.0 – 46.0)	*	*
RF-positive status, n (%)	75 (71)	69 (67)	87 (81)	51 (44)	67 (68)
DAS28, mean (SD)	5.1 (1.1)	5.3 (1.2)	5.1 (1.2)	5.8 (1.4)	5.6 (1.2)
CRP (mg/L), median (IQR)	8.0 (3.1 – 18.0)	9.0 (3.0 – 24.0)	9.6 (3.0 – 23.0)	17.0 (0.0 – 44.0)	17.0 (8.0 – 41.5)
ESR (mm per h), median (IQR)	23.5 (15.0 – 41.0)	27.0 (14.0 – 47.0)	25.0 (12.0 – 46.0)	31.0 (16.0 – 48.0)	30.0 (15.0 – 48.0)
SJC28, median (IQR)	6.5 (3.0 – 10.0)	7.0 (4.0 – 13.0)	6.0 (3.0 – 10.0)	11.0 (6.0 – 15.0)	9.0 (6.0 – 15.0)
TJC28, median (IQR)	6.0 (4.0 – 11.0)	7.0 (4.0 – 11.0)	7.0 (4.0 – 10.0)	10.0 (6.0 – 17.0)	9.0 (5.0 – 13.0)
HAQ score, median (IQR)	1.1 (0.6 – 1.6)	1.1 (0.9 – 1.6)	1.0 (0.6 – 1.4)	1.0 (0.4 – 1.5)	1.3 (0.8 – 1.6)
Any radiographic joint damage, n (%)	26 (25)	28 (27)	29 (27)	33 (28)	24 (20)

Outcomes are based on non-imputed data of the 2 trials. Reported baseline characteristics of CAMERA-II slightly differ from those of the original publication as we presented non-imputed data compared to imputed data in that publication.^{6,12}

* not collected in CAMERA-II, but at least <1 year according to the inclusion criteria.

TCZ+MTX= tocilizumab + methotrexate strategy arm; TCZ= tocilizumab + placebo strategy arm; MTX (ref)= methotrexate + placebo strategy reference arm; MTX+Pred= methotrexate + prednisone strategy arm; SD= standard deviation; BMI= body mass index; RF= rheumatoid factor; DAS28= disease activity score assessing 28 joints [range 0-9.4, higher is more activity]; CRP= C-reactive protein in milligram per litre; IQR= interquartile range; ESR= erythrocyte sedimentation rate in millimetre per hour; SJC28= 28 swollen joint count; TJC28= 28 tender joint count; HAQ= health assessment questionnaire [range 0-3, 3= worst function]; any radiographic joint damage= total Sharp van der Heijde score \geq 1.

In Supplementary Table S2 the baseline characteristics are shown, for each trial, of the MTX strategy arms, which were used as reference in the analyses. DAS28, SJC28, TJC28, RF-positivity and alcohol use in these strategy arms were statistically significantly different between the 2 trials. This is in line with the comparison of the trial population overall (as also expected due to randomization within trials).

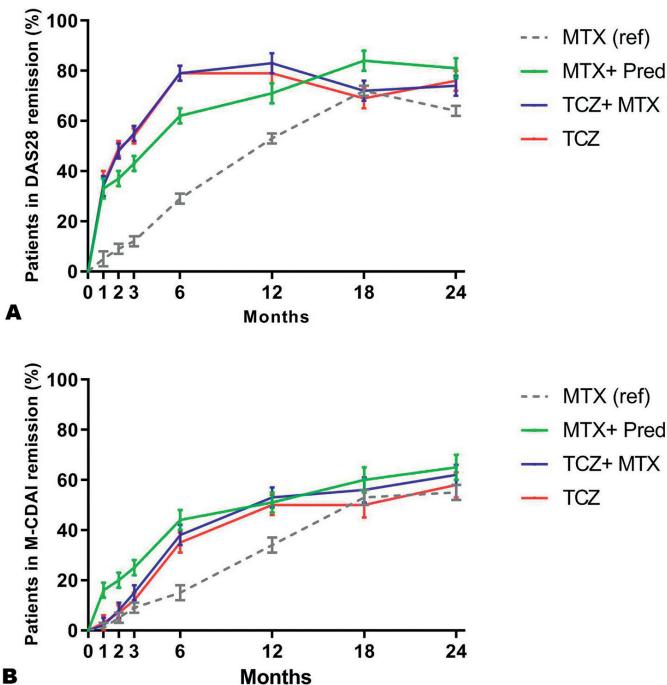
Supplementary Table S2 Baseline characteristics for the MTX strategy arms, used as reference in the analyses.

	U-Act-Early MTX (ref) n= 108	CAMERA-II MTX (ref) N= 119	P-value
Female, n (%)	69 (64)	72 (61)	0.68
Age (yrs.), mean (SD)	52.2 (13.7)	53.0 (13.0)	0.65
BMI (kg/m ²), mean (SD)	26.2 (4.3)	25.9 (4.0)	0.54
Alcohol use (≥ 1 unit/week), n (%)	65 (60)	41 (39)	<0.01
Current smoker, n (%)	31 (29)	31 (28)	1.00
Education level, n (%)			0.36
Low	48 (49)	41 (29)	
Moderate	37 (37)	27 (47)	
High	14 (14)	22 (24)	
Disease duration (days), median (IQR)	27.0 (15.0 – 46.0)	*	
RF-positive status, n (%)	87 (81)	67 (68)	0.04
DAS28, mean (SD)	5.1 (1.2)	5.6 (1.2)	<0.01
CRP (mg/L), median (IQR)	9.6 (3.0 – 23.0)	17.0 (8.0 – 41.5)	0.06
ESR (mm per h), median (IQR)	25 (12.0 – 46.0)	30.0 (15.0 – 48.0)	0.19
SJC28, median (IQR)	6.0 (3.0 – 10.0)	9.0 (6.0 – 15.0)	<0.01
TJC28, median (IQR)	7.0 (4.0 – 10.0)	9.0 (5.0 – 13.0)	0.02
HAQ score, median (IQR)	1.0 (0.6 – 1.4)	1.3 (0.8 – 1.6)	0.11
Any radiographic joint damage, n (%)	29 (27)	24 (20)	0.27

Outcomes are based on non-imputed data of the 2 trials. Reported baseline characteristics of CAMERA-II slightly differ from those of the original publication as we presented non-imputed data compared to imputed data in that publication.^{6,12}

* not collected in CAMERA-II, but at least <1 year according to the inclusion criteria.

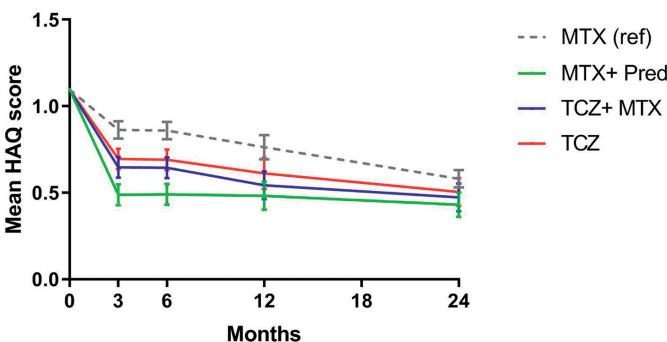
SD= standard deviation; BMI= body mass index, RF= rheumatoid factor; DAS28= disease activity score assessing 28 joints [range 0-9.4, higher is more activity]; CRP= C-reactive protein in milligram per litre; IQR= interquartile range; ESR= erythrocyte sedimentation rate in millimetre per hour; SJC28= 28 swollen joint count; TJC28= 28 tender joint count; HAQ= health assessment questionnaire [range 0-3, 3= worst function]; any radiographic joint damage= total Sharp van der Heijde score ≥ 1 .



Supplementary Figure S1a Percentage (SE) of patients achieving DAS28 remission for each strategy arm as predicted by model.

Supplementary Figure S1b Percentage (SE) of patients achieving modified CDAI remission in each strategy arm as predicted by model.

DAS28= disease activity score assessing 28 joints; SE= standard error; MTX (ref)= methotrexate + placebo strategy used as reference arm in the analysis; MTX+Pred= methotrexate + prednisone strategy arm; TCZ+MTX= tocilizumab + methotrexate strategy arm; TCZ= tocilizumab + placebo strategy arm, M-CDAI= modified CDAI.



Supplementary Figure S2 Mean (SE) HAQ scores over time for each strategy arm as predicted by model.

HAQ= health assessment questionnaire; SE= standard error; MTX (ref)= methotrexate + placebo strategy used as reference arm in the analysis; MTX+Pred= methotrexate + prednisone strategy arm; TCZ+MTX= tocilizumab + methotrexate strategy arm; TCZ= tocilizumab + placebo strategy arm.

In Supplementary Table S3 the results are shown of the sensitivity analyses using non-imputed data. The outcomes were generally in line with the outcomes based on imputed data. Statistical significant differences were shown for DAS28 over time, which were in favour of the TCZ strategy arms compared to the MTX+Pred strategy arm. Missing data was mostly found in the CAMERA-II trial (especially for the confounders RF status and smoking status), as this resulted in exclusion of patients with a more favourable prognosis in the CAMERA-II trial (as apparent for instance from the lower DAS28, SJC, TJC and HAQ in excluded patients 5.6 vs 5.7, 10 vs. 11, 10 vs. 11, and 1.06 vs. 1.11), more favourable outcomes were found in the TCZ treatment strategies (U-Act-Early) by using the non-imputed data. Therefore we consider the analysis imputing these baseline characteristics as the more valid results.

Supplementary Table S3 Difference in effectiveness and safety outcomes between strategy arms over 24 months based on non-imputed data.

	TCZ+MTX vs. MTX+Pred	TCZ vs. MTX+Pred
Effectiveness		
DAS28, mean difference (95%CI)	-0.51 (-0.84 to -0.19)	-0.42 (-0.75 to -0.09)
m-CDAI, proportional mean difference (95%CI) ¹	0.07 (-0.20 to 0.33)	0.22 (-0.05 to 0.49)
DAS28 remission, RR (95%CI)	1.10 (1.00 to 1.22)	1.09 (0.98 to 1.21)
m-CDAI remission, RR (95%CI)	0.97 (0.88 to 1.08)	0.92 (0.83 to 1.02)
ConRew, mean difference (95%CI)	3.00 (-2.68 to 8.65)	4.20 (-1.50 to 9.91)
HAQ score, mean difference (95%CI)	0.01 (-0.14 to 0.18)	0.08 (-0.08 to 0.24)
SvdH score, IRR (95%CI)	0.87 (0.36 to 2.11)	1.43 (0.60 to 3.44)
Safety		
Occurrence of ≥ 1 x elevated ALT, RR (95%CI)	1.89 (0.72 to 5.01)	1.09 (0.37 to 3.18)
Occurrence of ≥ 1 x elevated AST, RR (95%CI)	1.06 (0.25 to 4.60)	0.41 (0.08 to 2.13)
Occurrence of ≥ 1 infection, RR (95%CI)	2.62 (0.52 to 13.28)	2.61 (0.53 to 13.66)
Drop out due to AE, RR (95%CI)	1.22 (0.36 to 4.15)	1.34 (0.40 to 4.48)

Random effects for the variables in the model were evaluated and retained if the model fit was improved, as judged by a decrease in Bayesian information criterion or a change in one of the estimates of treatment effect of at least 10%. Random effects for the following variables were used in the models: DAS28, HAQ, RF-status, smoking status, all at baseline and time.

Negative values indicate lower values for continuous outcomes (i.e., DAS28, m-CDAI, HAQ) in TCZ-based strategy arms compared to the MTX+Pred arm.

(I)RR >1 indicate higher occurrence of the outcome (i.e., DAS28 remission, m-CDAI remission, drop out due to AE) in TCZ strategy arms compared to the MTX+Pred strategy arm.

¹log transformed values. Multiplying the above regression coefficients by 100 leads to a difference on a modified percentage scale.¹⁶

Outcomes based on adjusted analyses using non-imputed data.

TCZ+MTX= tocilizumab + methotrexate strategy arm; MTX+Pred= methotrexate + prednisone strategy arm; TCZ= tocilizumab + placebo strategy arm.

DAS28= disease activity score assessing 28 joints; CI= confidence interval; m-CDAI= modified clinical disease activity index; RR= relative risk; ConRew= continuity rewarded; HAQ= health assessment questionnaire; SvdH= Sharp van der Heijde; IRR= incidence rate ratio; ALT= alanine aminotransferase; AST= aspartate aminotransferase; AE= adverse event.

Chapter 4:

Validation of a prognostic multivariable prediction model for insufficient clinical response to methotrexate in early rheumatoid arthritis and its clinical application in Evidencio.

H.R. Gosselt^{1,2}

M.M.A. Verhoeven³

M.C.F.J. de Rotte⁴

S.M.F. Pluimj⁵

I.B. Muller¹

G. Jansen⁶

J. Tekstra³

M. Bulatović-Čalasan⁷

S.G. Heil²

F.P.J.G. Lafeber³

J.M.W. Hazes^{8*}

R. de Jonge^{1*}

*JMW and Rdj contributed equally to this manuscript as senior authors.

¹Department of Clinical Chemistry, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands;

²Department of Clinical Chemistry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands;

³Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands;

⁴Department of Clinical Chemistry, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, the Netherlands.

⁵Department of Paediatric Oncology, Prinses Máxima Centre for Paediatric Oncology, Utrecht, the Netherlands;

⁶Amsterdam Rheumatology and Immunology Center, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands;

⁷Department of Internal Medicine, University Medical Center Utrecht, Utrecht, the Netherlands;

⁸Department of Rheumatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.

ABSTRACT

Objective

Methotrexate (MTX) constitutes first-line therapy in rheumatoid arthritis (RA), yet approximately 30% of the patients do not benefit from MTX. Recently, we reported a prognostic multivariable prediction model for insufficient clinical response to MTX at 3 months of treatment in the ‘treatment in the Rotterdam Early Arthritis Cohort’ (tREACH), including baseline predictors: Disease Activity Score assessing 28 joints (DAS28), Health Assessment Questionnaire (HAQ), erythrocyte folate, single nucleotide polymorphisms (SNPs; ABCB1, ABCC3), smoking and BMI. The purpose of the current study was 1) to externally validate the model and 2) to enhance the model’s clinical applicability.

Methods

Erythrocyte folate and SNPs were assessed in 91 early disease modifying anti-rheumatic drug (DMARD)-naïve RA patients starting MTX in the external validation cohort (U-Act-Early). Insufficient response (DAS28>3.2) was determined after 3 months and non-response after 6 months of therapy. The previously developed prediction model was considered successfully validated in the U-Act-Early (validation cohort) if the area under the curve (AUC) of the receiver operating characteristic (ROC) was not significantly lower than in the tREACH (derivation cohort).

Results

The AUCs in U-Act-Early at 3 and 6 months were 0.75 (95%CI: 0.64-0.85) and 0.71 (95%CI: 0.60-0.82) respectively, similar to the tREACH. Baseline DAS28 >5.1 and HAQ >0.6 were the strongest predictors. The model was simplified by excluding the SNPs, while still classifying 73% correctly. Furthermore, interaction terms between BMI and HAQ and BMI and erythrocyte folate significantly improved the model increasing correct classification to 75%. Results were successfully implemented in Evidencio online platform assisting clinicians in shared decision-making to intensify treatment when appropriate.

Conclusion

We successfully externally validated our recently reported prediction model for MTX non-response and enhanced its clinical application thus enabling its evaluation in a clinical trial.

INTRODUCTION

Methotrexate (MTX) is the first-line therapy in rheumatoid arthritis (RA).¹ Although efficacious in a large proportion of patients, MTX is poorly effective in approximately 30% of early RA patients.^{2,3} Patients on MTX who do not show improvement at 3 months (insufficient responders) or do not reach the treatment target of low disease activity/remission at 6 months (non-responders) are switched to biologic disease modifying anti-rheumatic drug (bDMARD) therapies or novel targeted synthetic DMARD (tsDMARD) therapies, including e.g., Janus kinase (JAK) inhibitors,^{1,4,5} with or without concomitant MTX treatment.

To ensure that only patients unresponsive to MTX receive early (additional) treatment with b/tsDMARDs and those responsive to MTX are spared costly biologics or synthetic drugs, we and others have constructed models to predict MTX (non)-response.⁶⁻⁹ Our prognostic multivariable prediction model for the prediction of insufficient response, defined as: disease activity score assessing 28 joints (DAS28; DAS28>3.2) at 3 months of MTX therapy, was constructed in the treatment in the Rotterdam Early Arthritis Cohort (tREACH) and included clinical predictors (DAS28 and Health Assessment Questionnaire (HAQ)), life style predictors (smoking and BMI) and laboratory parameters involved in MTX metabolism (erythrocyte folate and single nucleotide polymorphisms: SNPs).⁹

This model classified 80% of patients correctly (area under the curve (AUC) of the receiver operating characteristic (ROC): 0.80 (95%CI: 0.73 – 0.86)) and was externally validated in the MTX-Rotterdam cohort showing a similar prognostic performance (AUC 0.80 (95%CI: 0.69 – 0.91)) even though BMI and smoking predictors were absent from this validation cohort.⁹

The aim of the current study was to validate the complete prediction model, including BMI and smoking status predictors, in an external early RA cohort (U-Act-Early) from a different geographic region and to enhance the model's applicability in clinical practice.¹⁰

METHODS

The methodology of this study followed transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines.¹¹

Patients

The external validation cohort consisted of 91 patients from the U-Act-Early cohort, a multicenter, double blind, placebo-controlled strategy trial, registered at ClinicalTrials.gov (number: NCT01034137).¹⁰ DMARD- and glucocorticoid (GC)-naïve early RA patients were eligible for inclusion once classified as RA patients according to the 1987 America

College of Rheumatology (ACR)¹² (n=7, 8%) or the 2010 ACR/EULAR classification criteria (n=84, 92%),¹³ and had a disease duration <12 months and active disease at baseline (DAS28≥2.6). Patients were randomly assigned to a treatment strategy with tocilizumab (TCZ) + placebo, MTX + placebo or their combination (TCZ+MTX) and treated to the target of sustained remission (i.e., a DAS28 <2.6 and swollen joint count of ≤4 joints of the 28 joints assessed, during ≥24 weeks). All 91 patients included were derived from the initial MTX + placebo strategy arm. The starting MTX dose was 10 mg/week orally and increased stepwise 5 mg every 4 weeks up to 30 mg/week until remission or the maximum tolerable dose. During the trial, GC use was not permitted. The tREACH (n=285) cohort was described earlier.⁹ Importantly, in the tREACH the optimal MTX dose of 25 mg/week was reached within three weeks (combined with other conventional synthetic (cs) DMARDs and/or GCs) and therapy was targeted to low disease activity (DAS28 ≤3.2) at 3 months. If this failed, step-up treatment with additional csDMARDs (sulfasalazine and/or hydroxychloroquine) or bDMARDs (i.e., TNF-alpha inhibitor) was initiated. In both cohorts, folic acid (10 mg/week) was prescribed during MTX treatment. U-Act-Early was approved by the medical ethics committee of the University Medical Center Utrecht (ML22497), and tREACH by the medical ethics committee of Erasmus Medical Center (MEC-2006-252). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written informed consent was obtained for all patients.

Outcome and clinical predictors

The primary outcome was insufficient MTX response after 3 months of treatment start, defined as DAS28>3.2, where DAS28 was based on the erythrocyte sedimentation rate (DAS28-ESR). All predictors were dichotomized prior to analyses. Cut-off values were defined previously in the tREACH cohort as: DAS28 >5.1, HAQ >0.6, erythrocyte folate <750 nmol/L, current smoking, BMI >25kg/m², ABCB1 rs1045642 (GG/GA vs AA) genotype, and ABCC3 rs4793665 (TC/CC vs TT) genotype.⁹ As erythrocyte-folate levels were slightly higher in U-Act-Early, new cut-off points were examined for erythrocyte-folate (deciles) and BMI (>30kg/m²) in U-Act-Early and tested for improvement of the model. The secondary outcome measure was non-response to MTX after 6 months of treatment, defined as DAS28-ESR>3.2.

Erythrocyte folate and genetic variants

In U-Act-Early erythrocyte folate and genetic variants included in the original prediction model⁹ were determined from EDTA whole blood samples stored at -80°C, as described elsewhere.^{14,15} DNA was obtained from whole blood using a MagNAPure Compact

(Roche Life Science, Almere, The Netherlands) and genotypes were determined for *ABCB1* rs1045642 and *ABCC3* rs4793665 using real-time PCR with Taqman, as described previously.¹⁵ Samples were tested for deviation from the Hardy-Weinberg equilibrium (HWE). SNPs were determined in the same lab and according to the same protocols as the tREACH study.⁹ Importantly, there were methodological differences in the measurement of folate between the tREACH and U-Act-Early cohorts. The Elecsys® Folate III assay (Ref 7027290190; Roche Diagnostics) has been re-standardized since 2017 in accordance with the WHO International Standard NIBSC Code 03/178. This resulted in 10% lower erythrocyte folate levels (U-Act-Early) compared to those quantified using the previous assay (tREACH). Furthermore, serum folate levels, required for folate correction in whole blood, were not available in U-Act-Early. As serum folate levels take up only a small part of the total folate concentrations, whole blood folate levels were corrected for the average serum folate concentration in the tREACH (25 nmol/L).

Statistical analysis

Clinical and laboratory parameters for the tREACH (derivation) and U-Act-Early (validation) cohorts were compared. The difference in DAS28 at 3 months (compared to baseline) was expressed as a mean with standard deviation (\pm SD) and assessed using a paired sample t-test. Differences between cohort variables were tested using an independent two-group t-test, if the assumptions of normal distribution (visual inspection) and equal variances (tested using the Levene's test) were met. If these assumptions were not met, non-parametric Mann Whitney U-test was performed. Differences in proportions were tested using a two-sample proportion test. Due to missing informed consents at the start of this validation study, eight subjects from the tREACH (derivation) cohort were excluded from analyses in this study. That is why the model as described previously was first re-analyzed on the tREACH data excluding these eight subjects, resulting in negligible differences in effect sizes compared to previous study.⁹ Next, to validate the prediction model in the external validation cohort (U-Act-Early), the predictors, DAS28 >5.1, HAQ >0.6, erythrocyte folate <750 nmol/L, current smoking, BMI >25kg/m², *ABCB1* rs1045642 genotype and *ABCC3* rs4793665 genotype were entered into a multivariable logistic regression and the probability for insufficient response was calculated using the pROC package in R according to the following formula:

$$P \text{ insufficient response } (\%) = \frac{e^{(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n)}}{1 + e^{(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n)}} \times 100,$$

where b0 represents the constant, b, b2 and bn represent the regression coefficients for each of the predictors x1, x2, xn. Subsequently, an ROC curve with AUC was constructed using the predicted probabilities and compared with results in the tREACH. The previously developed prediction model in tREACH (derivation) was considered successfully validated in the U-Act-Early (validation) if the AUC-ROC was not significantly

lower than in the tREACH. Goodness of fit between the predicted probabilities and observed values was tested using the Hosmer-Lemeshow test, where $p>0.05$ indicated that a model fit the data well. All analyses were performed in R studio (Version: 3.5.3, "2019-03-11"). P-values <0.05 were considered statistically significant. Subjects with missing data ($N=104$) were excluded from analyses (complete-case analysis).

To enhance the model's clinical applicability and thus facilitate its clinical implementation we applied the prediction model on the combined dataset of tREACH and U-Act-Early ($n=264$) using the step-up approach. Therefore, the model could be simplified using fewer predictors and possible two-way interactions could be examined in a combined cohort with more power. Statistically significant interactions ($p<0.05$) were added to the model. To simplify the model, we assessed individual contribution of variables to the predictive power of the model by sequential addition of predictors. Model fits were compared using the likelihood ratio test. Probability for insufficient response was calculated for each patient as well as the corresponding specificity, positive predictive value (PPV), sensitivity and negative predictive value (NPV) were determined, using the "pROC" package in R. The final prediction model was uploaded into the online platform 'Evidencio' providing a tool for clinicians to decide whether or not to start MTX combination therapy.

RESULTS

Cohort comparisons

In U-Act-Early (validation cohort), mean DAS28 decreased from 5.0 (± 1.1) to 3.6 (± 1.6) during the first 3 months ($p<0.001$). Mean DAS28 in tREACH (derivation cohort) was 5.0 (± 1.1), which decreased to a mean DAS28 of 3.1 (± 1.2 , $p<0.001$). Baseline DAS28 in U-Act-Early was similar to that of tREACH ($p=0.613$), Table 1. In U-Act-Early, 58 patients (64%) were categorized as insufficient MTX responders (i.e., $DAS28>3.2$ at 3 months) compared to 114 (43%) in tREACH ($p=0.006$). In U-Act-Early, 39 (44%) patients were classified as MTX non-responders (i.e., $DAS28>3.2$ at 6 months), which was not significantly different from 38% in tREACH, Table 1. Additionally, U-Act-Early consisted of significantly more rheumatoid factor positive patients (81%) compared to tREACH (65%) ($p=0.007$), whereas no significant differences were found for anti-citrullinated protein antibody (ACPA) positivity ($p=0.214$), Table 1. Despite the lower folate levels due to (international) re-standardization of the method, the erythrocyte-folate levels were still significantly higher in U-Act-Early compared to tREACH ($p=0.006$) and genotype GG/GA for *ABCB1* was significantly more frequent in tREACH ($p=0.016$), while genotypes for *ABCC3* were similar between cohorts, Table 1. Importantly, besides MTX, co-medication was prescribed in tREACH (derivation) but not in U-Act-Early (validation), Table 1.

Table 1 Descriptives of the derivation (tREACH) and external validation (U-Act-Early) cohorts.

	tREACH Derivation cohort N=277	U-Act-Early Validation cohort N=91	P-value
Clinical parameters			
DAS28>3.2 at 3 months	43%	64%	0.006**
DAS28>3.2 at 6 months	38%	44%	0.417
Gender, male	30%	36%	0.300
Age, mean ± SD	54 ± 14	53 ± 13	0.498
Baseline DAS28, mean ± SD	5.0 ± 1.1	5.0 ± 1.1	0.613
HAQ >0.6	76%	70%	0.330
Laboratory parameters			
Erythrocyte folate, median (IQR) [#]	862 (665 – 1163)	1020 (795 – 1221)	0.006**
ABCB1 GG/GA	73%	58%	0.016*
ABCC3 TC/CC	66%	67%	0.909
Rheumatoid factor positive	65%	81%	0.007**
ACPA positive	71%	79%	0.214
Life style parameters			
BMI, median (IQR)	25 (23 – 29)	25 (23 – 29)	0.950
Current smokers, n (%)	84 (33)	28 (31)	0.820
Co-medication			
Other DMARDs	56%	0%	<0.001***
Oral glucocorticoids	58%	0%	<0.001***
Parenteral glucocorticoids	28%	0%	<0.001***
Subcutaneous	0%	0%	1.000

Missing values tREACH: DAS28 at 3 months, n=13; DAS28 at 6 months, n=28; erythrocyte folate, n=78; rheumatoid factor, n=35; BMI, n=3; smoking status, n=21; HAQ, n=18; ABCB1, n=21; ABCC3, n=20. Missing values U-Act-Early: DAS28 at 6 months, n=2; rheumatoid factor, n=1; ACPA status, n=1.

Percentages shown are of valid data points. *P<0.05 was considered significant, **P<0.01, ***P<0.001.

[#]expressed in nmol/L

Validation of prediction model at 3 months

The model combining all predictors accomplished an AUC of 0.75 (95%CI: 0.64-0.85), Figure 1, in U-Act-Early (validation), which means that 75% of the patients could be classified correctly. The strongest predictor for insufficient response to MTX in U-Act-Early was baseline DAS28 >5.1 (p=0.008), Table 2. Odds ratios (ORs) for the predictors DAS28, HAQ, erythrocyte folate, BMI and smoking in U-Act-Early (validation) were in the same direction as in the tREACH (derivation), Table 2. The OR of the ABCC3 SNP was however in the opposite direction in U-Act-Early (OR=0.6, 95%CI: 0.23 – 1.79) compared to tREACH (OR=3.1, 95%CI: 1.39 – 6.94).

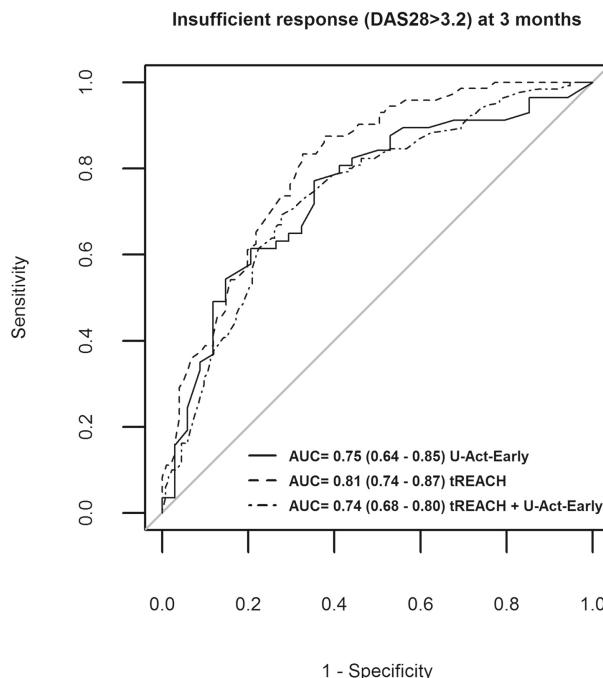


Figure 1 ROC curve for the prediction of insufficient response (DAS28>3.2) to MTX after 3 months of treatment.

Area under the curve (AUC) is reported as follows: AUC (95% confidence interval). Predictors were: baseline DAS28 >5.1, baseline HAQ >0.6, ABCB1 genotype, ABCC3 genotype, baseline erythrocyte folate, BMI >25kg/m² and current smoking.

Table 2 Validation of multivariable logistic regression models for insufficient response to MTX (DAS28>3.2) at 3 months of treatment in an external validation cohort (U-Act-Early).

	tREACH derivation cohort N= 173	U-Act-Early validation cohort N= 91
Predictors	OR (95%CI)	OR (95%CI)
DAS28 >5.1	3.7 (1.62 – 8.38)**	4.1 (1.44 – 11.82)**
HAQ >0.6	2.8 (1.15 – 7.00)*	2.1 (0.67 – 6.35)
ABCB1 GG/GA	2.4 (1.06 – 5.23)*	1.0 (0.37 – 2.75)
ABCC3 TC/CC	3.1 (1.39 – 6.94)**	0.6 (0.23 – 1.79)
Folate <750 nmol/L	2.1 (0.97 – 4.40)	3.4 (0.88 – 12.79)
Smoker	4.2 (1.91 – 9.42)**	1.3 (0.44 – 4.00)
BMI >25 kg/m ²	3.3 (1.52 – 7.21)**	1.6 (0.62 – 4.23)
AUC (95%CI)	0.81 (0.74 – 0.87)	0.75 (0.64 – 0.85)

Left column presents data from the derivation cohort (tREACH) and right of the external validation cohort (U-Act-Early).

Predictors that contributed significantly to the model were indicated with an asterisk, where *P<0.05 and **P<0.01.

OR= odds ratio; CI= confidence interval.

Similar results were found for the prediction of non-response at 6 months. An ROC curve was constructed with an AUC of 0.71 (95%CI: 0.60 – 0.82) in U-Act-Early (validation), which is similar to the predictive value of the tREACH model (derivation) at 6 months (AUC 0.75, 95%CI: 0.67 – 0.83), see Supplementary Figure S1.

Enhancement of model's clinical applicability

Next, the U-Act-Early and tREACH cohort were combined to increase power and enhance the model's clinical applicability. In this combined cohort the ORs for all predictors were greater than 1 and all predictors, except for the SNPs were significant, Supplementary Table S1. The combined model reached an AUC of 0.74 (95%CI: 0.68 – 0.80) at 3 months, Supplementary Table S1 and Figure 1.

Additionally, in this combined set, we investigated whether all predictors were required to reach 74% predictive power or whether the model could be further simplified. To do so, we analyzed changes in AUC upon sequential addition of predictors to the model. We started with the most readily available clinical predictors DAS28 >5.1 and HAQ >0.6, which generated an ROC with an AUC of 0.67 (95%CI: 0.61 – 0.74), Table 3.

Upon addition of smoking to the model, the AUC significantly increased ($p=0.01$) to 0.70 (95%CI: 0.64 – 0.76), followed by BMI, upon which the AUC further improved to 0.72 (95%CI: 0.66 – 0.78, $p=0.02$). Upon addition of erythrocyte folate to the model the AUC reached 0.73 (95%CI: 0.67 – 0.79, $p=0.02$). Addition of *ABCB1* and *ABCC3* genotypes did not significantly improve the model (AUC=0.74, 95%CI: 0.68 – 0.80, $p=0.12$), Table 3. Hence, the model could be simplified to a model where SNP genotypes were excluded resulting in a model with predictive power of 73%.

To fine-tune the model, all two-way interaction terms between predictors were tested. An interaction term between HAQ and BMI (OR= 3.68, 95%CI: 1.07 – 13.14) significantly contributed to the model. This means that a BMI >25kg/m² was associated with worse disease activity when HAQ values were >0.6. Furthermore, an interaction term between HAQ and erythrocyte folate (OR= 0.23, 95%CI 0.06 – 0.86) also significantly contributed to the model, indicating that low erythrocyte folate concentrations (<750 nmol/L) significantly predicted insufficient response when HAQ values were <0.6. Hence, interaction terms for HAQ and BMI and HAQ and erythrocyte folate were added to the model. Upon addition of these interaction terms to the model, the AUC of the final model, shown in Table 4, increased to 0.75 (95%CI: 0.69 – 0.81). As mentioned in the method section, we generated new cut-off values for erythrocyte folate and the BMI in the U-Act-Early cohort which, when included, did not result in higher AUCs.

Table 3 Logistic model building in combined datasets: U-Act-Early + tREACH.

Model	Predictors	Log likelihood	Chisquare	P	AUC (95%CI)
1	DAS28 + HAQ	-168.68			0.67 (0.61 – 0.74)
2	DAS28 + HAQ + smoking	-165.52	6.32	0.01*	0.70 (0.64 – 0.76)
3	DAS28 + HAQ + smoking + BMI	-162.98	5.08	0.02*	0.72 (0.66 – 0.78)
4	DAS28 + HAQ + smoking + BMI + erythrocyte folate	-160.43	5.11	0.02*	0.73 (0.67 – 0.79)
5	DAS28 + HAQ + smoking + BMI + erythrocyte folate + ABCC3	-158.57	3.71	0.05	0.74 (0.68 – 0.80)
6	DAS28 + HAQ + smoking + BMI + erythrocyte folate + ABCB1	-160.15	0.57	0.45	0.74 (0.68 – 0.80)
7	DAS28 + HAQ + smoking + BMI + erythrocyte folate + ABCC3 + ABCB1	-158.29	4.28	0.12	0.74 (0.68 – 0.80)

Each model was compared to the previous model.

Model 6 and 7 were compared to model 4.

*P-value<0.05 was considered significant. DAS28= DAS28>5.1; HAQ= HAQ >0.6; smoking= current smoking; BMI= BMI >25kg/m²; erythrocyte folate= erythrocyte folate <750 nmol/L; ABCC3= genotype TC or CC; ABCB1= genotype GG or GA.

Table 4 Final prediction model enhanced for clinical implementation.

	B	OR (95%CI)	p
Intercept	-1.67	0.19 (0.07 – 0.44)	<0.001***
Baseline DAS28 > 5.1	1.34	3.81 (2.12 – 6.99)	<0.001***
HAQ >0.6	0.44	1.56 (0.58 – 4.33)	0.383
BMI >25 kg/m ²	-0.34	0.71 (0.24 – 2.04)	0.528
Erythrocyte folate <750nmol/L	1.79	5.98 (2.00 – 19.09)	0.002**
Smoking (current smoker)	0.81	2.26 (1.25 – 4.16)	0.008**
HAQ >0.6 x BMI >25 kg/m ²	1.30	3.68 (1.07 – 13.14)	0.040*
HAQ >0.6 x Erythrocyte folate <750nmol/L	-1.46	0.23 (0.06 – 0.86)	0.031*
AUC		0.75 (95%CI: 0.69 – 0.81)	
Hosmer – Lemeshow test			0.634

The model was constructed in the combined dataset (tREACH + U-Act-Early, N=264).

The multiplication sign indicates that there is an interaction between two predictors.

*P-values < 0.05, **P-values < 0.01 and ***P-values<0.001.

AUC= area under the curve; β= beta coefficient of the final logistic regression model; OR (95%CI)= odds ratio with 95% confidence interval.

Model translation to the clinic

In order to apply the model in clinical practice, the prediction model was integrated into an online platform “Evidencio”.¹⁶ Using this tool, clinicians can easily enter DAS28, HAQ, erythrocyte folate, smoking and BMI for individual patients using sliding scales and buttons. The model then automatically calculates and presents a probability of

insufficient response to MTX for this specific patient. Specificity, positive predictive value (PPV), sensitivity and negative predictive values (NPV) were calculated for different cut-off values of these probabilities for insufficient response and are presented in Supplementary Data S1.

Our online model can be found in Evidencio using the following link: <https://www.evidencio.com/models/show/2191>. A patient-specific report presenting the results and model interpretation can be downloaded from Evidencio, Supplementary Data S1. An example is shown in Figure 2, where a patient with DAS28 score= 4.9, HAQ= 0.4, BMI= 22 kg/m², erythrocyte-folate= 720 nmol/L, and current smoking status= yes, has a probability of insufficient response of 71.7%, with corresponding specificity of 86% and PPV of 75%.

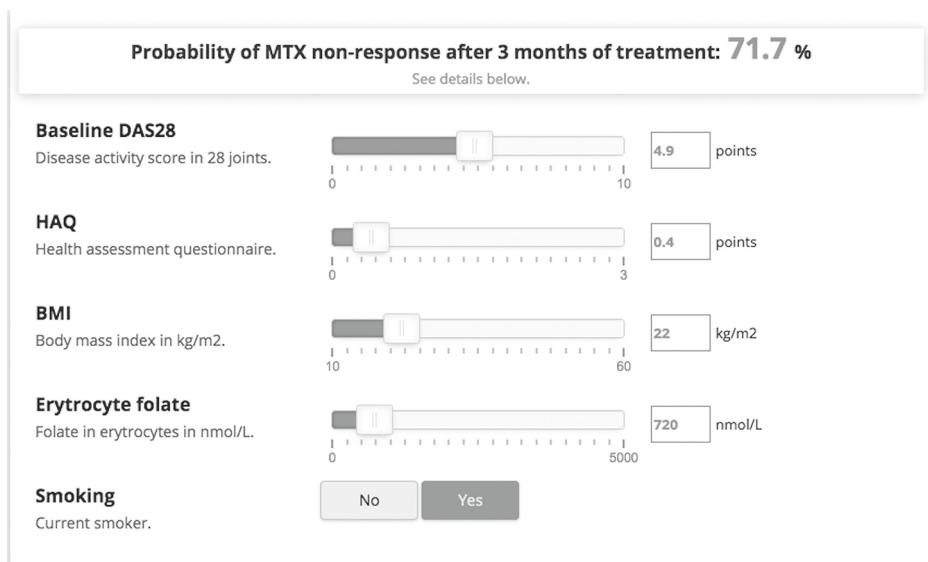


Figure 2 Example of online platform Evidencio for the implementation of the prediction model.

Values for each individual patient can be filled out using the buttons and slides.

Corresponding probability for insufficient response is automatically calculated using the prediction model.

DISCUSSION

We externally validated our previously developed prediction model for insufficient response to MTX therapy at 3 and 6 months after treatment initiation in early RA patients including all predictors DAS28 >5.1, HAQ >0.6, ABCB1 rs1045642 genotype, ABCC3 rs4793665 genotype, erythrocyte folate <750 nmol/L, current smoking and BMI >25 kg/m². To enhance clinical applicability and facilitate implementation, the

validated model was applied in the combined derivation and validation cohort. This model, without *ABCB1* rs1045642 genotype and *ABCC3* rs4793665 genotype, had an AUC of 0.75, meaning that it classified 75% of the insufficient responders correctly. Currently, according to the EULAR treatment guidelines MTX is the first-line therapy in RA. Treatment is only up-scaled after 3 to 6 months of insufficient response to MTX, despite evidence supporting a 'window of opportunity' for targeted treatment.^{1,17} This window of opportunity is a limited period between diagnosis and RA progression in which the disease could still be modified, radiographic damage/functional disability could be limited and progression could be slowed down upon early control of disease activity, for which sufficient treatment is required.¹⁷ Our prediction model could assist in identification of insufficient responders at diagnosis: for those with high probability of insufficient response to MTX, additional biologics or JAK-inhibitors could be prescribed, while for patients with low probabilities of insufficient response these expensive treatments could be spared. This distinction at diagnosis could save precious time for insufficient responders, allowing earlier control of disease activity resulting in better long-term outcomes.

We externally validated the model for the first time in its entirety as, besides the clinical and laboratory predictors, the lifestyle predictors (BMI and smoking) were also examined in the U-Act-Early cohort (as opposed to the initial validation in the MTX-Rotterdam cohort which lacked the life-style predictors).⁹ The strongest predictor was high disease activity at baseline (DAS28 >5.1) confirming previous findings.^{7,18} Due to differences in treatment intensities (i.e., MTX dose and co-medication) between the derivation and the validation cohort, we investigated whether the model was applicable at 6 months despite step-up treatments after the 3-month mark. Indeed 71% of the non-responders to MTX were classified correctly, which was similar to the 75% in the tREACH derivation cohort at 6 months. In the combined cohort, all predictors except for *ABCB1* and *ABCC3* genotypes significantly contributed to the predictive power of the model. Addition of *ABCB1* and *ABCC3* genotypes to the model showed only minimal improvement, resulting in an absolute change in AUC of 0.01, which was not statistically significant. A meta-analysis on the relationship between *ABCB1* genotype and response to MTX in 2014 RA patients showed an association between this genotype and response to MTX, yet our patient group was too small to validate this result.⁹ Another recent GWAS study did not show a relationship between *ABCB1* or *ABCC3* and treatment response.¹⁹ Since the differences in predictive power were minor and the effect of *ABCC3* genotype pointed in opposite directions in the two cohorts, possibly indicating a spurious finding, we excluded both genotypes from the model.

In agreement with our study, increased BMI (obese >30 kg/m²) was previously found to be associated with insufficient response to MTX in RA patients.^{20,21} It is postulated that the effect of BMI on non-response to MTX could be due to the release

of proinflammatory adipokines (e.g., leptin, interleukin-6 and or TNF-alpha) from the adipose tissue.²⁰ Also the predictive power of smoking was in accordance with previous studies.^{6,7,20,22}

So far, several prediction models for MTX non-response have been proposed which resulted in AUCs ranging from 0.65 – 0.85.^{6-8,22,23} Different outcome measures at different time points and combination therapies complicate comparison between models and their validation. However, the best-performing models all included clinical parameters and laboratory parameters, which is in line with our findings.^{6,8} We also showed that clinical predictors (DAS28 >5.1 and HAQ >0.6) alone classified fewer insufficient responders correctly (67%) compared to the model combining clinical, life style (BMI and smoking) and laboratory predictors (erythrocyte folate), which classified 73% of patients correctly. Most clinical predictors and life style predictors are easy to assess. Erythrocyte folate may not be available in every laboratory, however the assay is relatively easy to assess.²⁴

Strengths of this study are that both derivation and validation studies were prospectively designed and that patients in the external validation cohort were included from different districts in the Netherlands. Limitations are that the size of the external validation cohort was limited, however the number of cases in both the internal (tREACH) and external (U-Act-Early) cohort were similar. In addition, the model was validated in an MTX monotherapy group, while it was designed in a combination (GC and csDMARD) therapy group. Commonly, however MTX is co-prescribed with a short course of GCs (prednisone) as MTX's optimal effect ensues after 8-12 weeks.¹ Despite differences in co-medication between the cohorts, the prediction model had similar predictive value and OR for predictors were in the same direction in both cohorts, indicating that co-medication did not affect the prediction of response to MTX. Another limitation is that the smoking status was assessed using questionnaires; possibly biasing the results as smoking behavior could be underreported or underestimated. In future studies, cotinine, the degradation product of nicotine, could be quantified as an objective measure for smoking status which can easily be determined in serum.²⁵

Furthermore, we showed that the online platform Evidencio provides an easy tool for implementation of the prediction model in clinical practice. Evidencio is freely available so that the data can be uploaded to automatically validate the model in specific cohorts. In addition, using the Evidencio platform clinicians can directly use the model in their practice. When a new patient is diagnosed with RA, patient's information on DAS28, HAQ, erythrocyte-folate, BMI and smoking status can be provided to Evidencio. Subsequently, a probability of insufficient response to MTX with corresponding specificity, positive predictive value (PPV), sensitivity and negative predictive value (NPV) are provided by the tool and may help clinicians and patients in shared-decision making on step-up treatment with bDMARDs or tsDMARDs. The choice of a cut-off depends on

the clinical goal. Taking into consideration the “window of opportunity”¹⁷ for optimal treatment, we consider it crucial to adequately treat insufficient MTX responders with additional bDMARDs/tsDMARDs. Therefore, our goal for this prediction model was to identify as many insufficient responders as possible, while at the same time attempting to restrict the use of bDMARDs/tsDMARDs to those patients who really need them, hence to avoid misclassification of sufficient responders. Considering this, a cut-off probability of 70% (of insufficient response) could be chosen. At this cut-off, 75% of patients classified as insufficient responder match actual insufficient responders (PPV) and could be treated with additional bDMARDs/tsDMARDs. Additionally, at this cut-off 86% of all sufficient responders would be correctly classified as such (specificity) and could be spared additional treatment.

The importance of erythrocyte-folate for the predictive power of the model implies that this model is specific in predicting insufficient response to MTX, as MTX is structurally similar to folate. Hence, low erythrocyte folate levels are possibly a surrogate measure for poor MTX absorption, transportation and MTX accumulation in the cell, as described previously.¹⁴ However, it is possible that a certain proportion of insufficient responders to MTX are difficult-to-treat RA patients who are poorly responsive to various b/ts DMARDs.^{26,27} So far, we cannot identify difficult-to-treat RA patients in advance. Furthermore, as recently argued, treatment strategies could be more important than specific drugs, implying that these patients could still benefit from quicker and more aggressive treatment to reach a certain treatment target when earlier identified as insufficient responders.^{28,29}

Conclusion

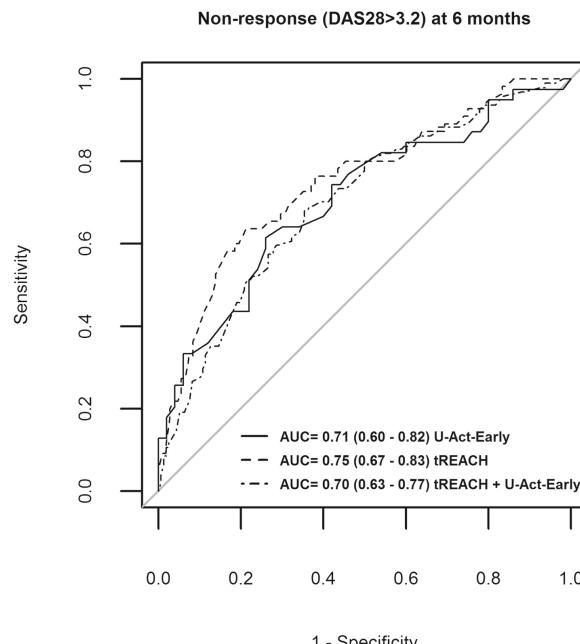
We successfully externally validated our previously published prognostic prediction model of insufficient response to MTX, which correctly classified 75% of insufficient responders at 3 months and 71% of non-responders at 6 months of treatment. The model can be used in clinical practice to identify insufficient responders to MTX with the goal of treating them with an additional biologic or JAK-inhibitor as early as possible to reduce disease activity and limit joint damage. Application of the tool by means of a clinical trial is warranted.

REFERENCES

- Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
- Brown, P. M., Pratt, A. G. & Isaacs, J. D. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. *Nat. Rev. Rheumatol.* **12**, 731–742 (2016).
- Bakker, M. F. *et al.* Low-dose prednisone inclusion in a methotrexate-based, tight control strategy for early rheumatoid arthritis: a randomized trial. *Ann. Intern. Med.* **156**, 329–339 (2012).
- Taylor, P. C. Clinical efficacy of launched JAK inhibitors in rheumatoid arthritis. *Rheumatol.* **58**, i17–i26 (2019).
- Burmester, G. R. & Pope, J. E. Novel treatment strategies in rheumatoid arthritis. *Lancet* **389**, 2338–2348 (2017).
- Wessels, J. A. M. *et al.* A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. *Arthritis Rheum.* **56**, 1765–1775 (2007).
- Teitsma, X. M. *et al.* Inadequate response to treat-to-target methotrexate therapy in patients with new-onset rheumatoid arthritis: development and validation of clinical predictors. *Ann. Rheum. Dis.* **77**, 1261–1267 (2018).
- Sergeant, J. C. *et al.* Prediction of primary non-response to methotrexate therapy using demographic, clinical and psychosocial variables: results from the UK rheumatoid arthritis medication study (RAMS). *Arthritis Res. Ther.* **20**, 147 (2018).
- de Rotte, M. C. *et al.* Development and validation of a prognostic multivariable model to predict insufficient clinical response to methotrexate in rheumatoid arthritis. *PLoS One* **13**, e0208534 (2018).
- Bijlsma, J.W.J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
- Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med.* **13**, g7594 (2015).
- Arnett, F. C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
- Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).
- de Rotte, M. C. *et al.* Association of low baseline levels of erythrocyte folate with treatment nonresponse at three months in rheumatoid arthritis patients receiving methotrexate. *Arthritis Rheum.* **65**, 2803–2813 (2013).
- de Rotte, M. C. *et al.* ABCB1 and ABCC3 gene polymorphisms are associated with first-year response to methotrexate in juvenile idiopathic arthritis. *J. Rheumatol.* **39**, 2032–2040 (2012).
- van Steenbeek, C. D. *et al.* Facilitating validation of prediction models: a comparison of manual and semi-automated validation using registry-based data of breast cancer patients in the Netherlands. *BMC Med. Res. Methodol.* **19**, 117 (2019).
- Burgers, L. E., Raza, K. & van der Helm-van Mil, A. H. Window of opportunity in rheumatoid arthritis—definitions and supporting evidence: from old to new perspectives. *RMD Open* **5**, e000870 (2019).
- Smolen, J. S. *et al.* Predictors of disease activity and structural progression after treatment with adalimumab plus methotrexate or continued methotrexate monotherapy in patients with early rheumatoid arthritis and suboptimal response to methotrexate. *Ann. Rheum. Dis.* **77**, 1566–1572 (2018).

19. Taylor, J. C. *et al.* Genome-wide association study of response to methotrexate in early rheumatoid arthritis patients. *Pharmacogenomics J.* **18**, 528–538 (2018).
20. Levitsky, A. *et al.* Obesity is a strong predictor of worse clinical outcomes and treatment responses in early rheumatoid arthritis: results from the SWEFOT trial. *RMD Open* **3**, e000458 (2017).
21. Jawaheer, D. *et al.* Gender, body mass index and rheumatoid arthritis disease activity: results from the QUEST-RA study. *Clin. Exp. Rheumatol.* **28**, 454–461 (2010).
22. Ektimmerman, F. *et al.* Validation of a clinical pharmacogenetic model to predict methotrexate nonresponse in rheumatoid arthritis patients. *Pharmacogenomics* **20**, 85–93 (2019).
23. López-Rodríguez, R. *et al.* Evaluation of a clinical pharmacogenetics model to predict methotrexate response in patients with rheumatoid arthritis. *Pharmacogenomics J.* **18**, 539–545 (2018).
24. Pfeiffer, C. M., Zhang, M. & Jabbar, S. Framework for laboratory harmonization of folate measurements in low- and middle-income countries and regions. *Ann NY Acad Sci.* **1414**, 96–108 (2018).
25. Vartiainen, E., Seppälä, T., Lillsunde, P. & Puska, P. Validation of self reported smoking by serum cotinine measurement in a community-based study. *J. Epidemiol. Community Health* **56**, 167–170 (2002).
26. Roodenrijns, N. M. T. *et al.* Characteristics of difficult-to-treat rheumatoid arthritis: results of an international survey. *Ann. Rheum. Dis.* **77**, 1705–1709 (2018).
27. de Hair, M. J. H., Jacobs, J. W. G., Schoneveld, J. L. M. & van Laar, J. M. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. *Rheumatology* **57**, 1135–1144 (2018).
28. Chatzidionysiou, K. & Sfikakis, P. P. Low rates of remission with methotrexate monotherapy in rheumatoid arthritis: Review of randomised controlled trials could point towards a paradigm shift. *RMD Open* **5**, e000993 (2019).
29. Drosos, A. A., Pelechas, E. & Voulgari, P. V. Treatment strategies are more important than drugs in the management of rheumatoid arthritis. *Clin. Rheumatol.* **39**, 1363–1368 (2020).

SUPPLEMENTARY FILE



Supplementary Figure S1 ROC curve for the prediction of non-response (DAS28>3.2) to MTX after 6 months of treatment.

Area under the curve (AUC) is reported as follows: AUC (95% confidence interval). Predictors were: baseline DAS28 >5.1, baseline HAQ >0.6, ABCB1 genotype, ABCC3 genotype, baseline erythrocyte folate, BMI >25kg/m² and current smoking.

Supplementary Table S1 Effect sizes of multivariable logistic regression models for insufficient response to MTX (DAS28>3.2) at 3 months of treatment in the combined data set.

U-Act-Early + tREACH Combination cohorts N=264	
Predictors	OR (95%CI)
DAS28 >5.1	3.6 (1.98 – 6.47)**
HAQ >0.6	2.1 (1.08 – 3.90)*
ABCB1 GG/GA	1.2 (0.71 – 2.17)
ABCC3 TC/CC	1.8 (0.99 – 3.11)
Folate <750 nmol/L	1.9 (1.06 – 3.50)*
Smoker	2.2 (1.23 – 4.00)**
BMI >25 kg/m ²	1.9 (1.11 – 3.31)*
AUC (95%CI)	0.74 (0.68 – 0.80)

Model shown is from the combined data set (U-Act-Early + tREACH). OR= odds ratio, CI= confidence interval. Predictors that contributed significantly to the model were indicated with an asterisk, where *P<0.05 and **P<0.01.

Supplementary Data S1: Evidencio Report “Prediction of 3 months MTX non-response (DAS28>3.2) in early rheumatoid arthritis



Model-id: 2191 | Version: 1.28 | Revision date: 2020-06-09

PREDICTION OF 3 MONTHS MTX NON-RESPONSE (DAS28>3.2) IN EARLY RHEUMATOID ARTHRITIS

Validated prediction model to identify DMARD-naïve rheumatoid arthritis patients with high risk of insufficient response to MTX.

Research authors: *Gosselet HR, Verhoeven MMA, de Rotte MCFJ, Pluijm SMF, Muller IB, Jansen G, Tekstra J, Bulatović-Čalasan M, Heil SG, Lafeber FPJG, Hazes JMW, and de Jonge R.*

RESULT

2020-06-30 11:22

Probability of MTX non-response after 3 months of treatment: 71.7 %

Based on the following parameters:

Baseline DAS28	4.9 points
HAQ	0.4 points
BMI	22 kg/m ²
Erythrocyte folate	720 nmol/L
Smoking	Yes

Disclaimer: Calculations alone should never dictate patient care, and are no substitute for professional judgement.

OUTCOME STRATIFICATION

Result interval 60 to 80

Based on a probability cut-off of 70% risk of insufficient response to methotrexate, calculated sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV) were:

Sensitivity: 43% Specificity: 86% PPV: 75% NPV: 61%

Result interval 70 to 90

Based on a probability cut-off of 80% risk of insufficient response to methotrexate, calculated sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV) were:

Sensitivity: 8% Specificity: 96% PPV: 67% NPV: 52%

CONDITIONAL INFORMATION

An interaction term was activated and added to the model:

Based on the provided input, a two-way interaction term between HAQ and erythrocyte folate (OR = 0.23, 95% CI 0.06 – 0.86) was automatically added to the model. This interaction term was found to significantly contribute to the model, meaning that low erythrocyte folate concentrations (<750 nmol/L) significantly predicted insufficient response when HAQ values were <0.6.

RESULT INTERPRETATION

How this model should be used:

This prediction model could assist in identification of insufficient responders at diagnosis. For patients with high probability of insufficient response to MTX, additional biologics or JAK-inhibitors could be prescribed. For those with low probabilities of insufficient response, these expensive treatments could be spared. This distinction at diagnosis could save precious time for insufficient responders, allowing earlier control of disease activity resulting in better long-term outcomes.

Model performance:

Discriminative power of the model was assessed through evaluating the area under the receiver operating characteristic curve (AUC). The AUC of the model was 0.75 (95% CI: 0.69 – 0.81), indicating that the model correctly classified patients in 75% of the cases.

Goodness-of-fit between the predicted probabilities and observed values was tested using the Hosmer-Lemeshow test. The associated P-value was 0.634, indicating good model fit.

Decisions on appropriate risk cut-offs:

Taking into consideration the “window of opportunity” for optimal treatment we consider it crucial to adequately treat insufficient MTX responders with additional bDMARDs/tsDMARDs. Therefore, our goal for this prediction model was to identify as many insufficient responders as possible, while at the same time attempting to restrict the use of bDMARDs/tsDMARDs to those patients who really need them, hence to avoid misclassification of sufficient responders. Considering this, a cut-off probability of 70% (of insufficient response) could be chosen.

At this cut-off, 75% of patients classified as insufficient responder match actual insufficient responders (PPV) and could be treated with additional bDMARDs/tsDMARDs. Additionally, at this cut-off 86% of all sufficient responders would be correctly classified as such (specificity) and could be spared additional treatment.

Cut-off values for probabilities of insufficient response (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
10	1	0	49	NA
20	96	15	62	80
30	88	48	62	81
40	87	51	63	80
50	73	71	71	73
60	58	79	73	66
70	43	86	75	61
80	8	96	67	52
90	1	100	1	51

Probabilities of insufficient response were calculated according to the prediction model. Model performance measures were calculated for different cut-off values (column 1). PPV = positive predictive value, NPV = negative predictive value.

Figure 1. Model performance given different risk cut-offs for insufficient response (Gosset et al, 2020)

Calculations alone should never dictate patient care, and are no substitute for professional judgement. See our full disclaimer at: <https://www.evidencio.com/disclaimer>.

Chapter 5:

Is tocilizumab monotherapy as effective in preventing radiographic progression in rheumatoid arthritis as its combination with methotrexate?

An analysis using individual patient data from multiple clinical trials.

M.M.A. Verhoeven¹

J. Tekstra¹

J.W.G. Jacobs¹

J.W.J. Bijlsma¹

J.M. van Laar¹

A. Pethö-Schramm²

M.E.A. Borm³

F.P.J.G. Lafeber¹

P.M.J. Welsing¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht,
Utrecht University, the Netherlands;

²F. Hoffmann-La Roche, Basel, Switzerland;

³Roche Nederland BV, Woerden, the Netherlands.

Epub ahead of print in Arthritis Care Res. 30 November 2020

ABSTRACT

Objective

To compare the effect of preventing radiographic progression (in its 3 components) of tocilizumab (TCZ) monotherapy with those of TCZ combined with methotrexate (TCZ+MTX), and to evaluate possible effect modifiers.

Methods

Randomized trials comparing TCZ-monotherapy with TCZ+MTX combination therapy regarding radiographic progression were analyzed on individual patient (n=820) data level for early as well as established RA patients using mixed-effects models. Outcomes were: not having radiographic progression after 2 years (i.e., preventing radiographic progression), respectively in total Sharp van der Heijde (SvdH) score, in erosion score and in joint space narrowing (JSN) score. Effect modification by baseline joint damage, disease duration and DAS28 was studied.

Results

Overall, TCZ+MTX was more effective in preventing radiographic progression regarding total SvdH scores compared to TCZ-monotherapy. However, in early RA patients with more joint damage (RR 1.02 vs. 0.91 for the less damage group), or a lower DAS28 (RR 1.04 vs. 0.92) at baseline, this advantage disappeared. In established RA, the advantage of TCZ+MTX over TCZ in preventing radiographic progression disappeared with a longer disease duration at baseline (RR 1.04 vs. 0.83). Results for erosion scores as outcome were in line, but were less clear for JSN.

Conclusion

Combination therapy with TCZ+MTX is more effective in preventing radiographic progression compared to TCZ-monotherapy, but the effectiveness of TCZ-monotherapy may approximate the effectiveness of TCZ+MTX in early RA patients with more joint damage and/or a lower DAS28 at baseline, and in established RA patients with longer disease duration.

INTRODUCTION

Joint damage is a negative outcome of rheumatoid arthritis (RA), which results in declined physical function and quality of life.¹ However, preventing radiographic progression of joint damage has improved over the last decades by early intensive treat-to-target treatment strategies with disease modifying anti-rheumatic drugs (DMARDs), including biological (b)DMARDs.¹ Consequently, detection of differences in preventing radiographic progression between effective treatment strategies is challenging, especially early in the disease.²

Nevertheless, joint damage remains an important outcome, since it objectively reflects irreversible damage,³ and the 'disease-modifying' effect of treatment strategies.⁴

Methotrexate (MTX) is the most frequently used first-line DMARD in RA, but patients may need to switch to, or to combine MTX with a (b)DMARD because of adverse events and/or insufficient response to MTX. It is a clinical fact that a relevant subgroup of patients doesn't adhere to MTX treatment due to side effects, aversion or inadequate efficacy.⁴ Tocilizumab (TCZ) may then be a suitable option since it can be used with lower doses of MTX, or even without MTX (i.e., as TCZ-monotherapy), while still being effective regarding disease activity as well as preventing radiographic progression.⁵

However, it is not clear whether the combination of TCZ+MTX has a better effect on preventing radiographic progression than TCZ-monotherapy, although this would be relevant knowledge for clinical decision making.

The aim of our study, using individual patient data (IPD), was to determine the effect of preventing radiographic progression of TCZ-monotherapy compared to TCZ+MTX combination therapy, on different components of radiographic progression, and to identify possible effect modifiers.

We hypothesized that, in general, a more intensive strategy (i.e., TCZ+MTX) would increase the effectiveness of preventing radiographic progression of treatment. However, this may vary among subgroups regarding disease phase and severity.

METHODS

Randomized controlled trials (RCTs) in RA with at least a treatment arm with TCZ administered intravenously as monotherapy as well as a treatment arm with TCZ in combination with MTX, and assessing radiographs of hands and feet, at baseline and after two years were identified, and IPD was obtained (see Supplementary Data S1). RCTs published until January 31, 2020 were selected. The Medical Research Involving Human Subjects Act (WMO) was not applicable to this study as it concerned re-analysis of existing data.

In total, four RCTs with TCZ intravenously at least in one treatment arm as monotherapy and in one treatment arm combined with MTX (i.e., TCZ+MTX) were identified; namely, ACT-RAY (n=553),⁶ FUNCTION (n=1164),⁷ SURPRISE (n=105),⁸ and U-Act-Early (n=317).⁹ The sharp van der Heijde (SvdH) score was used to measure radiographic progression in all trials. One RCT used a tight-control treat-to-target approach, indicating that treatment could continuously (4-weekly interval) be intensified when the treatment target was not achieved.⁹ Two RCTs were performed in early RA,^{7,9} and two tapered TCZ in case of remission.^{8,9} In 2 RCTs, a stable dose of oral glucocorticoid (GC) use (≤ 10 mg/day prednisone or equivalent) was permitted next to study treatment.^{6,7,10} Of 3 RCTs (n=2034), IPD could be obtained.^{6,7,9} All 3 RCTs are registered at ClinicalTrials.gov. (ACT-RAY: NCT00810199, FUNCTION: NCT01007435, and U-Act-Early: NCT01034137), and all patients gave written informed consent. In short, inclusion criteria were RA according to the classification criteria, age above 18, and a moderate-to-active disease. Early RA patients were DMARD-naïve; established RA patients had responded insufficiently to MTX. Radiographs were assessed by a single reader in FUNCTION and U-Act-Early, and by two independent readers in ACT-RAY; of the latter study, the average score was used in our analysis. For detailed information see Supplementary Data S2. The percentage of missing data on radiographic outcome in the individual trials is reported in their respective publications and was on average 29%. Patients with missing radiographic data were not different from patients with radiographic data present regarding joint damage, disease duration and disease activity, all at baseline, and nor were different between treatments arms within trials, hence no imputation of missing data was performed.

Statistical analysis

The primary endpoint was defined as: not having radiographic progression (versus having any radiographic progression, i.e., a score >0) after two years, termed as “prevention of radiographic progression”; primary analyses were performed on total scores, and secondary on erosions and joint space narrowing (JSN) scores. The different TCZ-regimens that were compared consisted of TCZ8mg/kg q4w i.v. + MTX median 15mg/weekly (TCZ+MTX) and TCZ8mg/kg q4w i.v. (TCZ).

Due to heterogeneity of patient populations and study results, we decided to analyze data of trials in early, MTX-naïve RA patients and trials in established RA patients separately. All analyses were according to the intention-to-treat principle and adjusted for gender and age. Treatment effect modification by baseline joint damage, disease duration and DAS28 was explored using these factors as covariates as well as in interaction terms with treatment in the models.

Logistic mixed effect models with random intercept and random effect of treatment, both at study level, were used to analyze the data. If analysis results indicated that

baseline joint damage, disease duration or DAS28 were possible treatment effect modifiers (i.e., $p \leq 0.20$ for interaction term), stratified analyses were performed for these factors to better interpret the interaction using log-binomial regression (to obtain relative chances) and forest plots. Stratification was based on median scores of effect modifiers. Patients with a score at or below this cut-off were classified in the low-level subgroup. Patients with scores above this cut-off were classified in the high-level subgroup. Relative chances (relative risk; RR) of preventing radiographic progression were calculated, and graphically illustrated per (sub)group. Furthermore, the absolute difference in the risk of preventing radiographic progression was calculated per subgroup. This was calculated using the RR and the rate of no-progression in the TCZ+MTX group as reference. By multiplying these, the rate of no-progression in the TCZ group is calculated. The difference between these rates of no-progression is the absolute risk difference.¹

All analyses were performed with SAS v9.4. All tests were two-sided and $p \leq 0.05$ was considered statistically significant.

RESULTS

Table 1 shows characteristics of all patients included in the analyses ($n=1506$). In total 1089 patients were classified as having early RA and 417 patients as having established RA. Baseline DAS28 was generally high (reflecting active disease), although slightly lower in U-Act-Early, most likely due to the inclusion criteria being less strict. The median change over two years in total SvdH score, as well as scores for erosions and JSN, were 0 in all trials and treatment arms, Table 1. The maximum change in total SvdH score was 33 in U-Act-Early, 31 in FUNCTION, and 23 in ACT-RAY, respectively (data not shown).

¹ An example, the relative chance of preventing radiographic progression of 0.91 for TCZ vs. TCZ+MTX can be translated into an absolute risk difference using the average percentage of patients treated with TCZ+MTX who have no-progression in this group (reference rate: 91%). Using the RR and this reference rate, the percentage of patients treated with TCZ with no-progression would be $0.91 * 91\% = 83\%$, and the absolute risk difference ($91\% - 83\% = 8\%$).

Table 1 Patients' characteristics per individual randomized controlled trial.

Trial	Early	Established	
	U-Act-Early ⁹	FUNCTION ⁷	ACT-RAY ⁶
	n= 232	n= 857	n= 417
Female, n (%)	161 (69)	677 (79)	345 (83)
Age (years), mean (SD)	53.9 (12.4)	50.0 (13.0)	52.8 (12.1)
BMI (kg/m ²), mean (SD)	26.0 (4.4)	27.5 (6.2)	26.2 (5.1)
Duration RA (years), median (IQR)	0.1 (0.0-0.1)	0.6 (0.1-1.1)	5.5 (2.3-11.4)
Baseline DAS28, mean (SD)	5.1 (1.1)	6.7 (1.1)	6.3 (1.0)
RF positive, n (%)	166 (72)	776 (91)	-
Baseline total SvdH score, median (IQR)	0 (0-1)	1.5 (0.5-5.5)	29 (17.5-50.5)
Baseline Erosion SvdH score, median (IQR)	0 (0-0)	1 (0-3.5)	15.5 (10-24)
Baseline JSON SvdH score, median (IQR)	0 (0-0)	0 (0-1.5)	13.5 (7-27)
Δ total SvdH score, median (IQR)	0 (0-1)	0 (0-0)	0 (0-0.5)
Δ Erosion SvdH score, median (IQR)	0 (0-1)	0 (0-0)	0 (0-0)
Δ JSON SvdH score, median (IQR)	0 (0-0)	0 (0-0)	0 (0-0.5)
Initial treatment, n (%)			
- MTX monotherapy	74 (32)	211 (25)	-
- TCZ+MTX 8mg/kg	78 (34)	220 (25)	215 (52)
- TCZ+MTX 4mg/kg	-	211 (25)	-
- TCZ-monotherapy 8mg/kg	80 (34)	215 (25)	202 (48)
GC-users, n (%)	0 (0)	429 (37)	212 (51)
Baseline total SvdH score of 0, n (%)			
- MTX monotherapy	53 (72)	55 (26)	-
- TCZ+MTX 8mg/kg	58 (74)	55 (25)	0 (0)
- TCZ+MTX 4mg/kg	-	44 (21)	-
- TCZ-monotherapy 8mg/kg	57 (71)	54 (25)	0 (0)
Δ total SvdH score per initial treatment, median (IQR)			
- MTX monotherapy	0 (0-0)	0 (0-0)	-
- TCZ+MTX 8mg/kg	0 (0-0)	0 (0-0)	0 (0-0.5)
- TCZ+MTX 4mg/kg	-	0 (0-0.5)	-
- TCZ-monotherapy 8mg/kg	0 (0-1.5)	0 (0-0)	0 (0-0)
Δ total SvdH score of 0, n (%)			
- MTX monotherapy	39 (53)	135 (64)	-
- TCZ+MTX 8mg/kg	61 (78)	185 (84)	129 (60)
- TCZ+MTX 4mg/kg	-	152 (72)	-
- TCZ-monotherapy 8mg/kg	55 (68)	172 (80)	111 (55)

n= number of patients; SD= standard deviation; BMI= body mass index; RA= rheumatoid arthritis; DAS28= disease activity score assessing 28 joints [range 0-9.4, higher is more active disease]; RF= rheumatoid factor; SvdH= Sharp van der Heijde [higher means more radiographic joint damage present]; JSON= joint space narrowing; GC= glucocorticoids; Δ= change from baseline to 2 years; MTX= methotrexate; TCZ= tocilizumab.

Early RA

Overall, in early RA patients, TCZ-monotherapy resulted in less prevention of radiographic progression (total SvdH scores) than TCZ+MTX combination therapy (RR 0.96, 95%CI 0.90 – 1.03). However, these effects were modified by baseline joint damage (TCZ vs. TCZ+MTX [$p<0.01$]), and DAS28 (TCZ vs. TCZ+MTX [$p=0.04$]), see Supplementary Table S1. In subgroups, the advantage of TCZ+MTX over TCZ on total SvdH scores disappeared with a high-level of baseline joint damage (RR for preventing progression of TCZ versus TCZ+MTX 1.02, 95%CI 0.87 – 1.18, versus RR 0.91, 95%CI 0.81 – 1.02 in the subgroup with a low-level of baseline damage) or with a low-level of DAS28 at baseline (RR 1.04, 95%CI 0.93 – 1.17 versus RR 0.92, 95%CI 0.83 – 1.03 in the subgroup with a high-level of baseline DAS28), see Figure 1 and Supplementary Table S2. Translating these results to absolute differences in the chance of preventing radiographic progression in the subgroups (see for calculation details footnote 1), resulted in an absolute risk difference between TCZ versus TCZ+MTX of 8% in patients with a low-level of baseline joint damage, versus only 1% in the subgroup with a high-level of baseline joint damage. In the subgroup with a low-level of disease activity at baseline, this was 3% versus 7% in the subgroup with a high-level of disease activity at baseline, see supplementary Table S3.

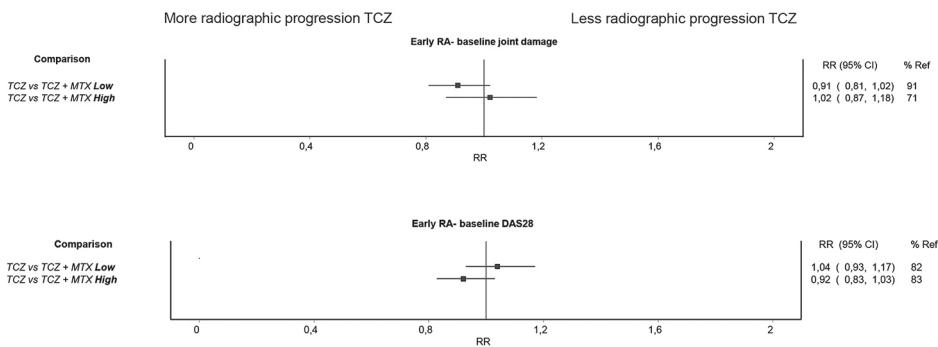


Figure 1 Relative chance of preventing radiographic progression in early RA.*

Number of patients in low-level baseline joint damage subgroup; 160 in TCZ vs. 166 in TCZ+MTX, number of patients in high-level baseline joint damage subgroup; 129 in TCZ vs. 132 in TCZ+MTX.

Number of patients in low-level baseline DAS28 subgroup; 149 in TCZ vs. 145 in TCZ+MTX, number of patients in high-level baseline DAS28 subgroup; 146 in TCZ vs. 153 in TCZ+MTX.

*Relative chances (95%CI) are based on stratified analyses, controlling for age, gender and DAS28 at baseline. A RR above 1 is associated with less radiographic progression for TCZ.

Low/high levels of baseline joint damage (SvdH score $\leq 1/ > 1$) or disease activity (DAS28 $\leq 6.37/ > 6.37$) were based on their respective median values in the data.

RA= rheumatoid arthritis; SvdH= Sharp van der Heijde; TCZ= tocilizumab; MTX= methotrexate; RR= relative chance; % Ref= proportion of patients with no progression (i.e., based on raw data) in the reference group, i.e., TCZ+ MTX group.

Outcomes for erosions were partly in line with those of total SvdH scores, with a RR of 1.03 (95%CI 0.91 – 1.17) for the subgroup with a high-level of baseline joint damage versus a RR of 1.02 (95%CI 0.93 – 1.10) for the subgroup with a low-level of baseline DAS28, see Supplementary Figure S1. This indicates that the advantage of TCZ+MTX over TCZ also seemed to disappear in these subgroups. In the subgroup with a low-level of joint damage at baseline, the absolute risk difference was 5%, versus only 2% in the subgroup with a high-level of baseline joint damage, see supplementary Table S3.

Regarding JSN, outcomes were less clear regarding the overall advantage of TCZ+MTX, however effect modification was observed for baseline joint damage ($p=0.20$), see Supplementary Table S1. In the subgroup with a low-level of joint damage at baseline the absolute risk difference was 3%, versus 8% in the subgroup with a high-level of baseline joint damage, see supplementary Table S3.

Established RA

Overall, in established RA patients, TCZ-monotherapy resulted in less prevention of radiographic progression (total SvdH scores) than TCZ+MTX combination therapy (RR 0.96, 95%CI 0.87 – 1.07). However, these effects were modified by baseline joint damage (TCZ vs. TCZ+MTX [$p=0.08$]), and disease duration (TCZ vs. TCZ+MTX [$p=0.04$]), see Supplementary Table S4. In subgroups, the advantage of TCZ+MTX over TCZ on total SvdH scores disappeared with a high-level of baseline disease duration (RR for preventing progression of TCZ versus TCZ+MTX 1.04, 95%CI 0.84 – 1.30 versus RR 0.83, 95%CI 0.64 – 1.06 in the low-level baseline disease duration subgroup), see Figure 2 and Supplementary Table S5. Translated to absolute differences in the chance of preventing radiographic progression between treatment regimens, the absolute risk difference was 10% in the subgroup with a low-level of baseline disease duration, versus only 2% in the subgroup with a high-level of baseline disease duration, see supplementary Table S6.

Outcomes for erosions were in line with those of total SvdH scores, with a RR of 1.04 (95%CI 0.87 – 1.25) for the high-level baseline disease duration subgroup, see Supplementary Figure S2. In the subgroup with a low-level of baseline disease duration the absolute risk difference was 10%, versus only 3% in the subgroup with a high-level of baseline disease duration, see supplementary Table S6.

For JSN as outcome, results were less clear regarding the overall advantage of TCZ+MTX, however effect modification was observed ($p=0.12$), see Supplementary Table S4. In the subgroup with a low-level of baseline disease duration the absolute risk difference was 7%, versus 0% in the subgroup with a high-level of baseline disease duration, see supplementary Table S6.

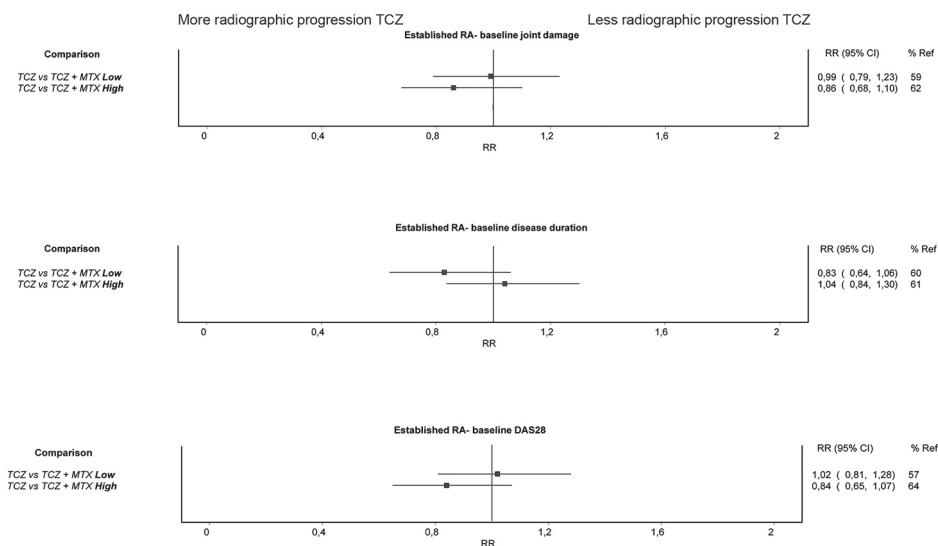


Figure 2 Relative chance of preventing radiographic progression in established RA.* Number of patients in low-level baseline joint damage subgroup; 96 in TCZ vs. 110 in TCZ+MTX, number of patients in high-level baseline joint damage subgroup; 105 in TCZ vs. 106 in TCZ+MTX. Number of patients in low-level baseline disease duration subgroup; 103 in TCZ vs. 105 in TCZ+MTX, number of patients in high-level baseline disease duration subgroup; 99 in TCZ vs. 110 in TCZ+MTX. Number of patients in low-level baseline DAS28 subgroup; 103 in TCZ vs. 104 in TCZ+MTX, number of patients in high-level baseline DAS28 subgroup; 99 in TCZ vs. 111 in TCZ+MTX.
*Relative chances (95%CI) are based on stratified analyses, controlling for age, gender and DAS28 at baseline. A RR above 1 is associated with less radiographic progression for TCZ.
Low/high levels of baseline joint damage (SvdH score ≤28.5/>28.5), disease duration (≤5.46 years/>5.46 years) or disease activity (DAS28 ≤6.37/>6.37) were based on their respective median values in the data.
RA= rheumatoid arthritis; SvdH= Sharp van der Heijde; TCZ= tocilizumab; MTX= methotrexate; RR= relative chance; % Ref= proportion of patients with no progression (i.e., based on raw data) in the reference group, i.e., TCZ+MTX group.

DISCUSSION

In general, TCZ-monotherapy was found to have less effect on preventing radiographic progression than TCZ+MTX combination therapy, however, this effect was found to vary between patients depending on joint damage, disease duration and disease activity, all at baseline.

When analyzing these modifying factors within the subgroups, we found the effectiveness of TCZ-monotherapy approximates that of TCZ+MTX for early RA patients with lower DAS28 (i.e., low-level subgroup) or (more) joint damage (i.e., high-level subgroup) at baseline. The ‘window of opportunity’ hypothesis¹¹ implies that RA is more susceptible for treatment during the first 6 months after disease onset. When symptoms are mild and slowly progressing (low disease activity), a considerable (unnoticed) delay

in diagnosis might result, and damage might already have occurred. These patients might have passed this 'window of opportunity', and thus MTX treatment might have less (additional) effect.¹² Actually, indeed in IPD of this study, early RA patients with lower baseline DAS28 more often had joint damage at baseline (Table S7), suggesting a late diagnosis of RA because of the milder symptoms.

For established RA patients, effectiveness of TCZ-monotherapy approximates that of TCZ+MTX for the subgroup with longer disease duration at baseline (i.e., high-level subgroup). This might indicate that, with a longer disease duration more often MTX is (no longer) effective, when a csDMARD has shown to be insufficiently effective, consequently the effectiveness of the combination treatment predominantly relies on the added bDMARD, here TCZ.¹³ Baseline GC use may affect radiographic progression, however, this was not different between subgroups for disease duration (55% in high-level group vs. 58% in low-level), as well as for the subgroups, indicating that baseline GC use probably has not biased the difference we found between the (sub)groups in our study.

Overall results were in line for total SvdH and erosions scores, but less clear for JSON scores. This may be due to the fact that JSON is a slow process, occurring far beyond the 2 year follow-up, and can additionally be influenced by genetic and mechanical factors (e.g., osteoarthritis),¹⁴ whereas erosion formation is mainly inflammatory driven and more specific for RA.

The current study naturally has limitations. First, radiographs were assessed by different readers in the original trials, and radiographs were not re-assessed specifically for the current analysis; a substantial portion of radiographic data was missing, although the influence on our subgroup analysis was probably limited, as patients with and without information on radiographic progression were not different regarding baseline joint damage, disease duration and disease activity (as well as regarding age, gender and rheumatoid factor positivity). Second, although we used IPD of multiple studies, the total sample size is still relatively low for detecting effect modification. For the SURPRISE trial, unfortunately data sharing was not possible due to legal considerations.⁸ However, despite the relatively low sample size, effect modifiers were detected. Even when we tested both interactions in the model (i.e., baseline SvdH*treatment & baseline DAS28*treatment) in early RA, both predictors were still modifiers. Outcomes of the analyses, based on 4 strata (Supplementary Table S7), were in line with reported outcomes. Third, radiographic progression is nowadays less, due to the availability of better treatment, resulting in absolute chances of preventing radiographic progression varying between 12% and 0%, indicating that differences in preventing radiographic progression may not be clinically relevant in all subgroups. However, using combined data may thus be a suitable means to identify relevant treatment effect estimates in subgroups. We also considered to apply the minimal clinically important difference as

cut-off for radiographic progression,¹⁵ however, as only a few patients met this criterion, meaningful analyses were not possible.

Despite these limitations, the current study used IPD of multiple RCTs, containing information of over 1500 RA patients, which provided an exclusive opportunity to explore radiographic progression in patients treated with TCZ with or without MTX in more detail.

Conclusion

For the majority of patients, TCZ combination therapy with MTX is more effective in preventing radiographic progression compared to TCZ-monotherapy. However, in early RA patients with more joint damage and/or lower DAS28 at baseline and in established RA patients with longer disease duration, the effectiveness of TCZ-monotherapy might approximate that of TCZ+MTX. In these specific subgroups, TCZ+MTX combination therapy might have no additional advantage regarding prevention of radiographic progression.

REFERENCES

1. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
2. van der Heijde, D. & Landewé, R. Should radiographic progression still be used as outcome in RA? *Clin. Immunol.* **186**, 79–81 (2018).
3. Carpenter, L. *et al.* Have radiographic progression rates in early rheumatoid arthritis changed? A systematic review and meta-analysis of long-term cohorts. *Rheumatol.* **55**, 1053–1065 (2016).
4. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
5. Combe, B., Lula, S., Boone, C. & Durez, P. Effects of biologic disease-modifying anti-rheumatic drugs on the radiographic progression of rheumatoid arthritis: a systematic literature review. *Clin. Exp. Rheumatol.* **36**, 658–667 (2018).
6. Huizinga, T. W. J. *et al.* Clinical and radiographic outcomes at 2 years and the effect of tocilizumab discontinuation following sustained remission in the second and third year of the ACT-RAY study. *Ann. Rheum. Dis.* **74**, 35–43 (2015).
7. Burmester, G. R. *et al.* Tocilizumab combination therapy or monotherapy or methotrexate monotherapy in methotrexate-naïve patients with early rheumatoid arthritis: 2-year clinical and radiographic results from the randomised, placebo-controlled FUNCTION trial. *Ann. Rheum. Dis.* **76**, 1279–1284 (2017).
8. Kaneko, Y. *et al.* Tocilizumab discontinuation after attaining remission in patients with rheumatoid arthritis who were treated with tocilizumab alone or in combination with methotrexate: results from a prospective randomised controlled study (the second year of the SURPRISE study). *Ann. Rheum. Dis.* **77**, 1268–1275 (2018).
9. Bijlsma, J. W. J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).

10. Kremer, J. M. *et al.* Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate responses to methotrexate at one year - the lithe study. *Arthritis Rheum.* **63**, 609–621 (2011).
11. van Nies, J. A. B., Tsonaka, R., Fautrel, B. & van Mil-van der Helm, A. H. M. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann. Rheum. Dis.* **74**, 806–812 (2015).
12. Stack, R. J. *et al.* Delays in help seeking at the onset of the symptoms of rheumatoid arthritis: a systematic synthesis of qualitative literature. *Ann. Rheum. Dis.* **71**, 493–497 (2012).
13. Buckley, F., Finckh, A., Huizinga, T. W. J., Dejonckheere, F. & Jansen, J. P. Comparative efficacy of novel DMARDs as monotherapy and in combination with methotrexate in rheumatoid arthritis patients with inadequate response to conventional DMARDs: a network meta-analysis. *J. Manag. care Spec. Pharm.* **21**, 409–423 (2015).
14. van der Heijde, D. Erosions versus joint space narrowing in rheumatoid arthritis: what do we know? *Ann. Rheum. Dis.* **70**, 116–118 (2011).
15. Bruynesteyn, K. *et al.* Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence. *J. Rheumatol.* **29**, 2306–2312 (2002).

SUPPLEMENTARY FILE

Supplementary Data S1: Search strategy

The intent of the current study will be to enhance knowledge of joint-sparing effects in RA patients treated with different TCZ-regimens.

PICO question:

P= RA patients

I= treatment with tocilizumab monotherapy (intravenously, without methotrexate)

C= treatment with tocilizumab combination therapy (intravenously, with methotrexate)

O= radiographic progression

Selection criteria:

1. Randomized controlled trial
2. Patients were randomly allocated to treatment arm
3. Tocilizumab treatment in at least one arm with monotherapy as well as one arm with combination therapy with MTX (TCZ+MTX)
4. Radiographs performed at baseline and after 2 years
5. RA patients, according to the 2010 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) criteria or 1987 revised ACR criteria
6. Active disease at baseline
7. Aged above 18 years

Supplementary Data S2: Included studies

ACTRAY:

Design:

International, multicenter, 2 arm randomized (TCZ+MTX vs. TCZ+Placebo) double blind placebo controlled parallel group study of 2 year duration with possible treatment adjustments at weeks 24, 36, 52, 64, 76 , 88 and 100.

Inclusion:

Men and women, ≥18 years of age, body weight ≤150kg, with moderate to severe RA, experiencing active disease, and had responded inadequately to MTX as defined by DAS28 >4.4 at baseline. At screening the DAS28 was equal or greater than 4.0.

Randomization:

TCZ+MTX: enrolled = 279, dosed = 277

TCZ+placebo: enrolled = 277, dosed = 276

FUNCTION:

Design:

Four-arm, double blinded, double-dummy randomized, parallel-group, pivotal PhaseIII, multi-center trial.

Inclusion:

Men and women, ≥ 18 years, with moderate-to-severe active early RA (duration ≤ 2 years at baseline) who had not previously been treated with MTX or biologic agent.

Randomization:

TCZ+MTX 8mg/kg: enrolled = 291

TCZ+placebo 8mg/kg: enrolled = 292

TCZ+MTX 4mg/kg: enrolled = 292

placebo+MTX: enrolled = 289

U-Act-Early:

Design:

A 2-year multicenter, three-parallel-arm, randomized, double-blind, double-dummy, treat-to-target strategy study of patients with newly diagnose RA.

Inclusion:

Men and women, ≥ 18 years, with early RA (duration ≤ 1 years at baseline) patients with an active disease (DAS28 > 2.6).

Randomization:

TCZ+MTX 8mg/kg: enrolled = 106

TCZ+placebo 8mg/kg: enrolled = 103

placebo+MTX: enrolled = 108

Supplementary Table S1 Comparative effects of different TCZ-regimens on preventing radiographic progression in early RA, results of mixed model analysis.

	total SvdH	Erosions	JSN
TCZ vs. TCZ+MTX*	Relative Risk (95%CI)	Relative Risk (95%CI)	Relative Risk (95%CI)
TCZ vs. TCZ+MTX*			
Treatment (TCZ)	5.77 (3.48 to 8.05)	8.61 (5.89 to 11.34)	1.80 (0.00 to 4.74)
Baseline DAS28	1.07 (0.81 to 1.34)	1.00 (0.72 to 1.29)	1.22 (0.85 to 1.59)
Disease duration (weeks)	1.02 (0.97 to 1.06)	1.03 (0.98 to 1.08)	1.00 (0.95 to 1.06)
Baseline SvdH	0.96 (0.93 to 0.99)	0.97 (0.94 to 1.00)	0.95 (0.92 to 0.99)
Gender (female)	1.31 (0.85 to 1.78)	1.72 (1.22 to 2.22)	1.30 (0.67 to 1.94)
Age	0.98 (0.96 to 1.00)	0.99 (0.97 to 1.01)	0.96 (0.93 to 0.98)
Baseline SvdH * treatment	1.03 (1.00 to 1.07)	1.02 (0.99 to 1.05)	1.02 (0.99 to 1.06)
Baseline DAS28 * treatment	0.69 (0.33 to 1.05)	0.66 (0.26 to 1.07)	0.78 (0.31 to 1.25)

Number of patients is 295 in TCZ and 298 in TCZ+MTX.

TCZ= tocilizumab; MTX= methotrexate; DAS28= disease activity score assessing 28 joints; SvdH= Sharp van der Heijde; JSN= joint space narrowing.

Outcome variable is not having radiographic progression over 2 years (yes/no), intercept for models (regression coefficient) total SvdH 1.76 (95%CI -0.28 to 3.80); for erosions 1.35 (95%CI -0.88 to 3.58); for JSN 3.65 (95%CI -3.50 to 10.80).

P-value for interaction term Baseline SvdH * treatment based on total SvdH score: TCZ vs. TCZ+MTX [p=0.06].

P-value for interaction term baseline DAS28 * treatment based on total SvdH score: TCZ vs. TCZ+MTX [p=0.04].

Supplementary Table S2 Comparative treatment effects of different TCZ-regimens on preventing radiographic progression in early RA, stratified analyses.

	total SvdH	Erosions	JSN	total SvdH	Erosions	JSN	Relative Risk (95%CI)							
							TCZ vs. TCZ+MTX ¹	TCZ vs. TCZ+MTX ¹						
Effect in patients with low baseline damage[#]														
Effect in patients with high baseline damage^{\$}														
Treatment (TCZ)	0.91 (0.81 to 1.02)	0.95 (0.86 to 1.05)	0.97 (0.87 to 1.07)	1.02 (0.87 to 1.18)	1.03 (0.91 to 1.17)	0.96 (0.87 to 1.06)								
Effect in patients with low baseline disease activity[#]														
Treatment (TCZ)	1.04 (0.93 to 1.17)	1.02 (0.93 to 1.11)	0.98 (0.89 to 1.08)	0.92 (0.83 to 1.03)	0.97 (0.88 to 1.07)	0.96 (0.88 to 1.05)								

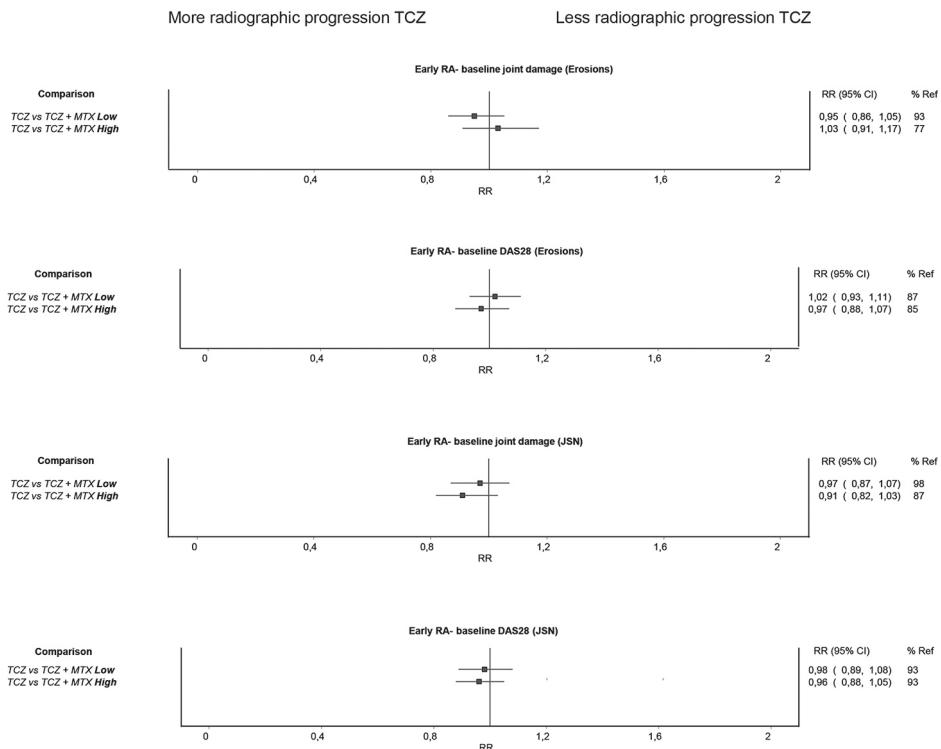
Number of patients in low level baseline joint damage subgroup; 160 in TCZ vs. 166 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 129 in TCZ vs. 132 in TCZ+MTX.

Number of patients in low level baseline DAS28 subgroup; 149 in TCZ vs. 145 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 146 in TCZ vs. 153 in TCZ+MTX.

TCZ= tocilizumab, MTX= methotrexate, DAS28=disease activity score assessing 28 joints, SvdH= Sharp van der Heijde, JSN= joint space narrowing.

^{\$}Subgroup baseline SvdH score/DAS28 > median,[#] Subgroup baseline SvdH score/DAS28 ≤ median

¹= cut-off SvdH score of 1, baseline DAS28 of 6.3



Supplementary Figure S1 Relative chance of preventing radiographic progression in early RA.*

Number of patients in low level baseline joint damage subgroup; 160 in TCZ vs. 166 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 129 in TCZ vs. 132 in TCZ+MTX.

Number of patients in low level baseline DAS28 subgroup; 149 in TCZ vs. 145 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 146 in TCZ vs. 153 in TCZ+MTX.

*Relative chances (95%CI) are based on stratified analyses, controlling for age, gender and DAS28 at baseline. A RR above 1 is associated with less radiographic progression for TCZ.

Low/high levels of baseline joint damage (SvdH score $\leq 1/ > 1$) or disease activity (DAS28 $\leq 6.37/ > 6.37$) were based on their respective median values in the data.

RA= rheumatoid arthritis; SvdH= Sharp van der Heijde; TCZ= tocilizumab; MTX= methotrexate; RR= relative chance; % Ref= proportion of patients with no progression (i.e., based on raw data) in the reference group, i.e., TCZ+ MTX group.

Supplementary Table S3 average absolute chance (proportion) of early RA patients with no radiographic progression at baseline per treatment arm, and absolute risk difference between treatment arms.

Modifier	Group Comparison	Relative chance	Absolute risk in TCZ+MTX arm (%)	Absolute risk in TCZ-monotherapy arm (%)	Absolute risk difference (%)
Total SvdH scores					
Baseline joint damage	Low	tcz vs tcz+mtx	0.91	91	83
	High	tcz vs tcz+mtx	1.02	71	72
Baseline DAS28	Low	tcz vs tcz+mtx	1.04	82	85
	High	tcz vs tcz+mtx	0.92	83	76
Erosion scores					
Baseline joint damage	Low	tcz vs tcz+mtx	0.95	93	88
	High	tcz vs tcz+mtx	1.03	77	79
Baseline DAS28	Low	tcz vs tcz+mtx	1.02	87	89
	High	tcz vs tcz+mtx	0.97	85	82
Joint space narrowing scores					
Baseline joint damage	Low	tcz vs tcz+mtx	0.97	98	95
	High	tcz vs tcz+mtx	0.91	87	79
Baseline DAS28	Low	tcz vs tcz+mtx	0.98	93	91
	High	tcz vs tcz+mtx	0.96	93	89

Number of patients in low level baseline joint damage subgroup; 160 in TCZ vs. 166 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 129 in TCZ vs. 132 in TCZ+MTX.

Number of patients in low level baseline DAS28 subgroup; 149 in TCZ vs. 145 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 146 in TCZ vs. 153 in TCZ+MTX.

Low= low level subgroup; High= high level subgroup; DAS28= disease activity score assessing 28 joints; TCZ= tocilizumab, MTX= methotrexate.

Interpretation: The relative chance of not having radiographic progression of 0.91 for TCZ vs. TCZ+MTX for patients with low baseline joint damage (TCZ vs TCZ+MTX Low) can be translated to an absolute risk difference using the average percentage of patients treated with TCZ+MTX who have no-progression in this group (reference group) of 91%. Using the RR and reference rate of no-progression would be $0.91 \times 91\% = 83\%$ for patients treated with TCZ, and the absolute risk difference thus 8%. For patients with high baseline damage, the reference rate of no-progression is 71%, rate of no-progression with TCZ of $1.02 \times 71\% = 72\%$, and an absolute difference of 1%.

Supplementary Table S4 Comparative effects of different TCZ-regimens on preventing radiographic progression in established RA, results of mixed model analysis.

	total SvdH Relative Risk (95%CI)	Erosions Relative Risk (95%CI)	JSN Relative Risk (95%CI)
TCZ vs. TCZ+MTX*			
Treatment (TCZ)	0.85 (0.44 to 1.66)	1.02 (0.50 to 2.07)	0.69 (0.35 to 1.37)
Baseline DAS28	1.00 (0.81 to 1.24)	1.05 (0.83 to 1.31)	0.94 (0.75 to 1.18)
Disease duration (years)	0.99 (0.94 to 1.03)	1.02 (0.97 to 1.08)	1.02 (0.98 to 1.07)
Baseline SvdH	1.01 (0.99 to 1.02)	1.00 (0.98 to 1.01)	0.99 (0.98 to 1.00)
Gender (female)	1.94 (1.14 to 3.29)	2.20 (1.29 to 3.75)	1.00 (0.57 to 1.73)
Age	1.00 (0.98 to 1.01)	0.99 (0.98 to 1.01)	1.00 (0.98 to 1.02)
Baseline SvdH * treatment	0.99 (0.97 to 1.00)	0.99 (0.98 to 1.01)	0.99 (0.98 to 1.01)
Disease duration (years) * treatment	1.07 (1.00 to 1.14)	1.02 (0.95 to 1.09)	1.06 (0.99 to 1.14)

Number of patients is 202 in TCZ and 215 in TCZ+MTX.

TCZ= tocilizumab; MTX= methotrexate; DAS28= disease activity score assessing 28 joints; SvdH= Sharp van der Heijde; JSN= joint space narrowing

Outcome variable is not having radiographic progression over 2 years (yes/no), intercept for models (regression coefficient) for total SvdH -0.70 (95%CI -2.26 to 1.26); for erosions -0.54 (95%CI -2.60 to 1.52); for JSN 1.43 (95%CI -1.26 to 4.12).

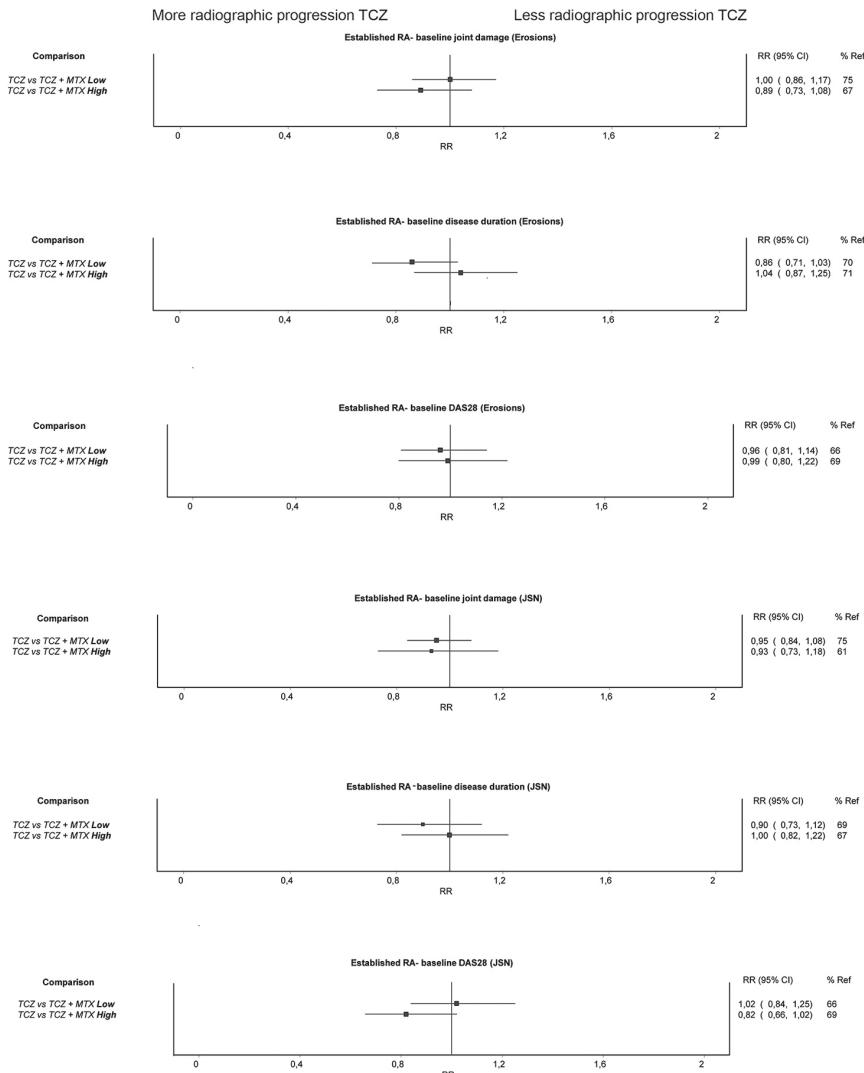
P-value for interaction term Baseline SvdH * treatment based on total SvdH score: TCZ vs. TCZ+MTX [p=0.08].

P-value for interaction term Disease duration * treatment based on total SvdH score: TCZ vs. TCZ+MTX [p=0.04].

Supplementary Table S5 Comparative treatment effects of different TCZ-regimens on preventing radiographic progression in established RA, stratified analyses.

TCZ vs. TCZ+MTX¹	total SvdH	Erosions	JSN	total SvdH	Erosions	JSN
	Relative Risk (95%CI)					
Effect in patients with low baseline damage[#]						
Treatment (TCZ)	0.99 (0.79 to 1.23)	1.00 (0.86 to 1.17)	0.95 (0.84 to 1.08)	0.86 (0.68 to 1.10)	0.89 (0.73 to 1.08)	0.93 (0.73 to 1.18)
Effect in patients with shorter disease duration[#]						
Treatment (TCZ)	0.83 (0.64 to 1.06)	0.86 (0.71 to 1.03)	0.90 (0.73 to 1.12)	1.04 (0.84 to 1.30)	1.04 (0.87 to 1.25)	1.00 (0.82 to 1.22)
Effect in patients with low baseline disease activity[#]						
Treatment (TCZ)	1.02 (0.81 to 1.28)	0.96 (0.81 to 1.14)	1.02 (0.84 to 1.25)	0.84 (0.65 to 1.07)	0.99 (0.80 to 1.22)	0.82 (0.66 to 1.02)
Number of patients in low level baseline joint damage subgroup; 96 in TCZ vs. 110 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 105 in TCZ vs. 106 in TCZ+MTX.						
Number of patients in low level baseline disease duration subgroup; 103 in TCZ vs. 105 in TCZ+MTX, number of patients in high level baseline disease duration subgroup; 99 in TCZ vs. 110 in TCZ+MTX.						
Number of patients in low level baseline DAS28 subgroup; 103 in TCZ vs. 104 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 99 in TCZ vs. 111 in TCZ+MTX.						
TCZ= tocilizumab; MTX= methotrexate; DAS28= disease activity score assessing 28 joints; SvdH= Sharp van der Heijde; JSN= joint space narrowing.						
¹ Subgroup baseline SvdH score/disease duration in years/DAS28 > median; [#] Subgroup baseline SvdH score/ disease duration in years/DAS28 ≤ median						

TCZ vs. TCZ+MTX; prevention of radiographic progression in RA



Supplementary Figure S2 Relative chance of preventing radiographic progression established RA.*

Number of patients in low level baseline joint damage subgroup; 96 in TCZ vs. 110 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 105 in TCZ vs. 106 in TCZ+MTX.

Number of patients in low level baseline disease duration subgroup; 103 in TCZ vs. 105 in TCZ+MTX, number of patients in high level baseline disease duration subgroup; 99 in TCZ vs. 110 in TCZ+MTX.

Number of patients in low level baseline DAS28 subgroup; 103 in TCZ vs. 104 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 99 in TCZ vs. 111 in TCZ+MTX.

*Relative chances (95%CI) are based on stratified analyses, controlling for age, gender and DAS28 at baseline. A RR above 1 is associated with less radiographic progression for TCZ.

Low/high levels of baseline joint damage (SvdH score $\leq 28.5 / > 28.5$), disease duration (≤ 5.46 years / > 5.46 years) or disease activity (DAS28 $\leq 6.37 / > 6.37$) were based on their respective median values in the data.

RA= rheumatoid arthritis; SvdH= Sharp van der Heijde; TCZ= tocilizumab; MTX= methotrexate; RR= relative chance; % Ref= proportion of patients with no progression (i.e., based on raw data) in the reference group, i.e., TCZ+ MTX group.

Supplementary Table S6 average absolute chance (proportion) of established RA patients with no radiographic progression at baseline per treatment arm, and absolute risk difference between treatment arms.

Modifier	Group	Comparison	Relative chance	Absolute risk in TCZ+MTX	Absolute risk in TCZ- arm (%)	Absolute monotherapy risk arm (%)	Absolute risk difference (%)
Total SvdH scores							
Baseline joint damage	Low	tcz vs tcz+mtx	0.99	59	58	58	1
	High	tcz vs tcz+mtx	0.86	62	53	53	9
Baseline disease duration	Low	tcz vs tcz+mtx	0.83	60	50	50	10
	High	tcz vs tcz+mtx	1.04	61	63	63	2
Baseline DAS28	Low	tcz vs tcz+mtx	1.02	57	58	58	1
	High	tcz vs tcz+mtx	0.84	64	54	54	10
Erosion scores							
Baseline joint damage	Low	tcz vs tcz+mtx	1	75	75	75	0
	High	tcz vs tcz+mtx	0.89	67	60	60	7
Baseline disease duration	Low	tcz vs tcz+mtx	0.86	70	60	60	10
	High	tcz vs tcz+mtx	1.04	71	74	74	3
Baseline DAS28	Low	tcz vs tcz+mtx	0.96	66	63	63	3
	High	tcz vs tcz+mtx	0.99	69	68	68	1
Joint space narrowing scores							
Baseline joint damage	Low	tcz vs tcz+mtx	0.95	75	71	71	4
	High	tcz vs tcz+mtx	0.93	61	57	57	4
Baseline disease duration	Low	tcz vs tcz+mtx	0.9	69	62	62	7
	High	tcz vs tcz+mtx	1	67	67	67	0
Baseline DAS28	Low	tcz vs tcz+mtx	1.02	66	67	67	1
	High	tcz vs tcz+mtx	0.82	69	57	57	12

Number of patients in low level baseline joint damage subgroup; 96 in TCZ vs. 110 in TCZ+MTX, number of patients in high level baseline joint damage subgroup; 105 in TCZ vs. 106 in TCZ+MTX.

Number of patients in low level baseline disease duration subgroup; 103 in TCZ vs. 105 in TCZ+MTX, number of patients in high level baseline disease duration subgroup; 99 in TCZ vs. 110 in TCZ+MTX.

Number of patients in low level baseline DAS28 subgroup; 103 in TCZ vs. 104 in TCZ+MTX, number of patients in high level baseline DAS28 subgroup; 99 in TCZ vs. 111 in TCZ+MTX.

Low= low level subgroup; High= high level subgroup; DAS28= disease activity score assessing 28 joints; TCZ= tocilizumab; MTX= methotrexate.

Interpretation: The relative chance of not having radiographic progression of 0.99 for TCZ vs. TCZ+MTX for patients with low baseline joint damage (TCZ vs TCZ+MTX Low) can be translated to an absolute risk difference using the average percentage of patients treated with TCZ+MTX who have no-progression in this group (reference group) of 59%. Using the RR and reference rate of no-progression would be $0.99 \times 59\% = 58\%$ for patients treated with TCZ, and the absolute risk difference thus 1%. For patients with high baseline damage, the reference rate of no-progression is 62%, rate of no-progression with TCZ of $0.86 \times 62\% = 53\%$, and an absolute difference of 9%.

Supplementary Table S7 Comparative treatment effects of different TCZ-regimens on preventing radiographic progression in early RA, stratified analyses.

Strata		TCZ, n	TCZ+MTX, n	RR (95%CI)*
1) high baseline joint damage AND high baseline DAS28	73	83	0.97 (0.79-1.18)	
2) high baseline joint damage AND low baseline DAS28	56	49	1.04 (0.87-1.26)	
3) low baseline joint damage AND high baseline DAS28	76	62	0.88 (0.70-1.11)	
4) low baseline joint damage AND low baseline DAS28	90	104	0.97 (0.85-1.11)	

*Relative chances (95%CI) are based on stratified analyses, controlling for age, gender and DAS28 at baseline. A RR above 1 is associated with less radiographic progression for TCZ.

Low/high levels of baseline joint damage (SvdH score ≤1/>1) and disease activity (DAS28 ≤6.37/>6.37) were based on their respective median values in the data.

TCZ= tocilizumab; MTX= methotrexate; DAS28= disease activity score assessing 28 joints; SvdH= Sharp van der Heijde; RR= relative 'chance'.

Chapter 6:

Effectiveness and safety over 3 years after the 2-year U-Act-Early trial of the strategies initiating tocilizumab and/or methotrexate.

M.M.A. Verhoeven¹

J. Tekstra¹

P.M.J. Welings¹

A. Pethö-Schramm²

M.E.A. Borm³

G.A.W. Bruyn⁴

R. Bos⁵

E.N. Griep⁶

R. Klaasen⁷

J.M. van Laar¹

F.P.J.G. Lafeber¹

J.W.J. Bijlsma¹

M.J.H. de Hair¹

J.W.G. Jacobs¹

¹*Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht,
Utrecht University, the Netherlands;*

²*F. Hoffmann-La Roche, Basel, Switzerland;*

³*Roche Nederland BV, Woerden, the Netherlands;*

⁴*Department of Rheumatology, MC Group, Lelystad, the Netherlands;*

⁵*Department of Rheumatology, Medical Center Leeuwarden, Leeuwarden, the Netherlands;*

⁶*Department of Rheumatology, Antonius Hospital, Sneek, the Netherlands;*

⁷*Department of Rheumatology, Meander Medical Center, Amersfoort, the Netherlands.*

ABSTRACT

Objective

U-Act-Early was a 2-year, randomized placebo controlled, double-blind trial, in which DMARD-naïve early RA patients were treated to the target of sustained remission (SR). The 2 strategies initiating tocilizumab (TCZ), with and without methotrexate (MTX), were more effective than the strategy initiating MTX. The aim of the current study was to determine longer-term effectiveness in daily clinical practice.

Methods

At the end of U-Act-Early, patients were included in a 3-year post-trial follow-up (PTFU), in which treatment was according to standard care and data was collected every 3 months during the first year and every 6 months thereafter. Primary end point was DAS28 over time. Mixed effects models were used to compare effectiveness between initial strategy groups, correcting for relevant confounders. Between the groups as randomized, proportions of patients were tested for DMARD use, SR and radiographic progression of joint damage.

Results

Of patients starting U-Act-Early, 226/317 (71%) participated in the PTFU. Over the total 5 years, mean DAS28 was similar between groups ($p>0.20$). During U-Act-Early, bDMARD use decreased in both TCZ initiation groups and increased in the MTX initiation group, but during follow-up this trend did not continue. SR was achieved at least once in 99% of patients. Of the 226 patients, only 30% had any radiographic progression over 5 years, without significant differences between the groups.

Conclusion

Although on short-term the strategies initiating TCZ yielded the most clinical benefit, on longer-term differences in important clinical outcomes between the strategies disappeared, probably due to continuation of the treat-to-target principle.

INTRODUCTION

The current guidelines for the treatment of rheumatoid arthritis (RA) recommend treating to target, aiming for remission, to reduce the risk of disability and improve long-term outcomes. With this aim, generally, a conventional synthetic disease modifying anti-rheumatic drug (csDMARD) is started immediately after diagnosis. If the treatment target is not achieved within 3-6 months, another csDMARD or a biological (b)DMARD is added.¹

Several placebo-controlled randomized controlled trials (RCTs) in early RA patients compared the effects of initial treatment strategies using a bDMARD to those of initial treatment strategies with only (a) csDMARD(s), mostly methotrexate (MTX).²⁻⁸ In the short term (<1 year), disease control and outcomes, including progression of joint damage and physical function, were better in strategies initiating a bDMARD with or without a csDMARD. The first RCT of initiation of the bDMARD tocilizumab (TCZ) in DMARD-naïve early RA patients showed that more patients receiving TCZ(+/- MTX) achieved remission at 24 and 52 weeks compared with patients receiving MTX.⁶ However, this was not a treat-to-target design like the U-Act-Early trial. In this 2-year trial involving early DMARD- and glucocorticoid (GC)-naïve RA patients with the treatment target of sustained remission (SR), efficacy and safety of step-up treatment strategies initiating treatment with TCZ, MTX or their combination were compared. When the treatment target was achieved, medication was tapered and eventually stopped, if patients remained in remission. Outcomes of U-Act-Early were in line with those of other RCTs showing improved effectiveness for strategies initiating/adding a bDMARD from the start of therapy, compared with strategy groups not including a bDMARD from the start.⁹ For example, SR was achieved earlier in the TCZ strategy groups in the U-Act-Early trial.¹⁰

However, starting TCZ immediately upon diagnosis in early RA patients does not accord with the current disease management recommendations. To justify initiation of (expensive) bDMARDs directly after diagnosis, not only clear advantages in the first months, but also a longer-term effectiveness would be warranted. Therefore, patients in the U-Act-Early trial were followed for another 3 years, to determine longer-term effectiveness. We hypothesized that during this follow-up period, the disease activity score assessing 28 joints (DAS28, based on erythrocyte sedimentation rate (ESR)) would remain similar (continuing the treat-to-target approach) between the initial strategy groups, and that radiographic progression would be less in patients who had initiated TCZ (+/- MTX) compared with those who had initiated MTX. Moreover, based on the finding that during the 2-year trial TCZ use steadily increased in the MTX initiation strategy group and decreased in the 2 TCZ initiation strategy groups, we hypothesized

that during follow-up, bDMARD use would further increase in the MTX initiation strategy group and further decrease in the TCZ initiation strategy groups.

METHODS

This was an observational open label multicentre 3-year post-trial follow-up (PTFU) of the U-Act-Early trial. Post-trial treatment was left to the discretion of the treating rheumatologist and patients were followed during routine clinical practice, which had a focus on treat-to-target.

Patients who participated in the U-Act-Early trial were eligible for participation and were asked to participate at the end of U-Act-Early between Q2 2012 and Q3 2014. The only exclusion criteria were being lost-to-follow up, unwillingness to give informed consent and having had too many serious protocol violations (e.g., >4 times non-adherence to protocol over a period of 1 year). Nineteen of the initial 21 hospitals in the Netherlands participating in the U-Act-Early trial also participated in the PTFU. Detailed information on the U-Act-Early trial has been reported.¹⁰

Data collection

After the initial monthly follow-up during the U-Act-Early trial, in the PTFU period data was collected every 3 months in the first year and every 6 months thereafter up to 3 years. At every visit, all components of DAS28, physical function with the Dutch Health Assessment Questionnaire (HAQ) and information on use of non-steroidal anti-inflammatory drugs (NSAIDs), GCs, csDMARDs and bDMARDs were assessed, next to the occurrence of adverse events (AE) and, serious AE (SAE). Remission was defined as DAS28 <2.6 AND ≤4 swollen joints of 28 assessed joints, and SR was defined as being in remission for ≥24 weeks. Sustained drug free remission (sDFR) was defined as having tapered and stopped all DMARDs AND being in remission for ≥3 months. Radiographs from baseline U-Act-Early, end of U-Act-Early (2 years) and last available time point (5 years or end of follow-up) were scored in chronological order by an experienced professional reader, according to the Sharp van der Heijde (SvdH) method. To make optimal use of available x-rays, if the x-ray was not taken at the 5 year time point, but ≥3 year, the 5 year progression was estimated by extrapolation using the following formula; change SvdH score between baseline and last available x-ray/(date last available x-ray - date baseline x-ray/365)*5. The institutional review boards of the participating centers confirmed that the Medical Research Involving Human Subjects Act (WMO) was not applicable to this study, and all patients gave written informed consent.

Outcomes

The primary end point was DAS28 over 5 years. The secondary end point was medication use (NSAID use, GCs, csDMARDs and bDMARDs) during the 3-year PTFU. Other secondary end points were number of patients achieving SR and sDFR, cumulative duration of SR and sDFR, change in SvdH score, HAQ scores, and the number of patients with occurrence of any AE, or SAE, all variables evaluated over 5 years.

Statistics

Data of all patients enrolled in this PTFU were used for analyses. Continuous variables were described using means with standard deviations (SD) or medians with interquartile ranges (IQR), where appropriate. Frequencies and proportions were calculated for categorical variables. Differences between the initial strategy groups in baseline characteristics were evaluated and tested using one-way ANOVA for continuous outcomes and chi-square test or Fisher's exact test for categorical outcomes. Mixed model analyses were used to assess differences over time between the initial treatment strategy groups in continuous outcomes, with a random intercept and fixed effects for treatment arm, visit-week, the interaction visit-week*treatment arm, correcting for DAS28 at baseline and center (i.e., both the stratification factors used for randomization). Differences between the initial treatment strategy groups in proportions of patients using medication were tested with a Cochran-Mantel-Haenszel (CMH) test. Differences between the initial treatment strategy groups in proportions of patients achieving SR and sDFR and cumulative duration of all SR and sDFR periods in individual patients were tested with CMH test and linear regression analysis, respectively, correcting for baseline DAS28 (for CMH; DAS28 <5.1 or ≥5.1, as this cut-off was used for the randomization) and center. Differences between the initial treatment strategy groups in median change in radiographic scores were tested with the van Elteren test, correcting for baseline DAS28 and center. The statistical analyses were performed in SAS version 9.4 and R version 3.4.3. All tests were two-sided; a p-value ≤0.05 was considered statistically significant.

Sensitivity analysis

To evaluate the generalizability of our findings, a sensitivity analysis was performed for the primary end point, including also patients (n= 91) who had not participated in the PTFU, only in the 2-year U-Act-Early trial. Missing data for all patients in DAS28 and swollen joint count (SJC)28 over time were imputed using a joint multivariate model of 2-level data. The package "jomo" in R was used for multiple imputation; this package also handles categorical data and uses cluster-specific covariance matrices.¹¹ The variables treatment arm, visit-week, duration of participation, reason for drop out, and DAS28, HAQ, age, disease duration, gender and rheumatoid factor (RF) status (positive/negative) at baseline were used in the imputation model. Patient was used

as cluster variable, to account for repeated measurements over time within patients. The number of imputed datasets was based on the percentages of missing values over time. By using Rubin's rule, analysis results based on the imputed datasets were pooled to provide an overall result.¹² A second sensitivity analysis was performed to exclude the influence of acute phase reactants (APRs) on the outcome, as it is known that TCZ reduces APRs specifically.¹³ However, validated disease activity indices without APR include the visual analogue scale (VAS) physician, which was not assessed during the PTFU. As an alternative, the unweighted components of the DAS28, except for ESR, were analysed separately using linear mixed effects models to study the differences in outcomes between the initial strategy groups over time using observed data. Square root transformation was applied to skewed SJC28 and tender joint count (TJC)28 data.

RESULTS

Of the 317 patients who had started in U-Act-Early, 226 patients were included in the PTFU, of whom 85% completed the study, Figure 1. During this period one patient died due to a brain stem infarction (TCZ strategy group) and one patient due to squamous cell carcinoma of the cervix (MTX strategy group); these complications were deemed not to be related to the treatment. The patient disposition during the U-Act-Early trial is shown in Supplementary Figure S1.

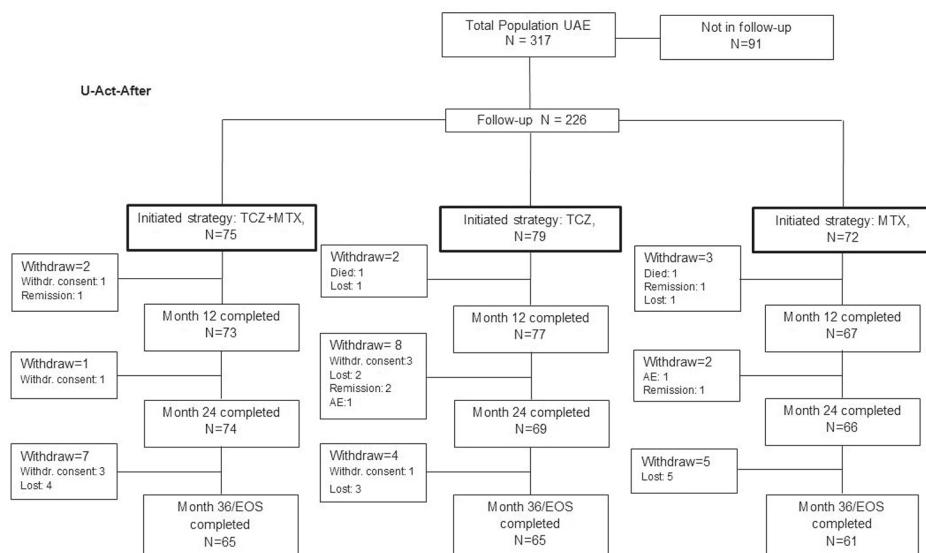


Figure 1 Patient disposition during the 3-year post-trial follow-up.

UAE= U-Act-Early; TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; Withdr. consent= withdrawal of consent; AE= adverse event; Lost= lost to follow-up; EOS= end of study; follow-up= post-trial follow-up.

Baseline characteristics of the patients included in the PTFU were not statistically significantly different between the initial strategy groups at start of the U-Act-Early trial. Table 1. Additional information about patients (not) included in the PTFU is presented in Supplementary Table S1.

Table 1 Baseline characteristics at start of U-Act-Early of patients included in the post-trial follow-up study (n=226).*

	TCZ+ MTX (n=75)	TCZ (n=79)	MTX (n=72)	P-value
Female, n (%)	47 (63)	61 (77)	48 (67)	0.13
RF +, n (%)	53 (71)	55 (71)	58 (82)	0.21
Anti CCP +, n (%)	51 (68)	54 (69)	56 (79)	0.28
RF- and/or anti-CCP+, n (%)	55 (73)	62 (79)	64 (89)	0.06
Caucasian, n (%)	71 (95)	77 (97)	71 (99)	0.47
Smoking status, n (%)				0.81
- Never smoked	29 (39)	34 (43)	28 (39)	
- Quit smoking	21 (28)	24 (30)	25 (35)	
- Current smoking	25 (33)	21 (27)	19 (26)	
Age (years), mean (SD)	53.8 (11.2)	55.5 (11.6)	53.7 (12.9)	0.63
Symptom duration (days), median (IQR)	27 (18 - 43)	25 (19 - 43)	28 (16 - 46)	0.96
DAS28, mean (SD)	5.1 (1.1)	5.3 (1.1)	5.0 (1.2)	0.33
HAQ, median (IQR)	1.3 (0.6 - 1.6)	1.3 (0.9 - 1.8)	1.0 (0.5 - 1.4)	0.17
SvdH score, median (IQR)	0 (0 - 0)	0 (0 - 2)	0 (0 - 1)	0.24

*according to their initial treatment strategy as randomized.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; RF= rheumatoid factor; += positive; anti-CCP= anti-cyclic citrullinated peptide antibodies; SD= standard deviation; IQR= interquartile range; DAS28= disease activity score assessing 28 joints; HAQ= health assessment questionnaire score; SvdH= Sharp van der Heijde score.

The following results are based on data of the 226 patients who were included in the PTFU.

DAS28

In line with the results of the U-Act-Early trial, the mean DAS28 over time was statistically significantly lower for the TCZ initiation strategy groups compared with the MTX initiation strategy group at year 1 and 2 of U-Act-Early. There were no statistically significant differences in average DAS28 over time between the initial treatment strategy groups during the 3-year PTFU period, neither over the total duration of 5 years, see Table 2 and Figure 2.

Table 2 Effectiveness outcomes over 5 years, results based on data of patients included in the post-trial follow-up, n=226.**Mean difference in DAS28 between initial treatment strategy groups over 5 years and at the end of every year***

Period	Strategy comparison	Mean difference	95%CI of mean difference
Over 5 years	TCZ+MTX vs. MTX	- 0.11	- 0.32 to 0.10
	TCZ vs. MTX	- 0.12	- 0.32 to 0.09
	TCZ+MTX vs. TCZ	0.00	- 0.20 to 0.21
1	TCZ+MTX vs. MTX	- 0.75	- 0.89 to - 0.61
	TCZ vs. MTX	- 0.65	- 0.79 to - 0.51
	TCZ+MTX vs. TCZ	- 0.10	- 0.23 to 0.04
2	TCZ+MTX vs. MTX	- 0.63	- 0.77 to - 0.49
	TCZ vs. MTX	- 0.55	- 0.69 to - 0.40
	TCZ+MTX vs. TCZ	- 0.08	- 0.23 to 0.06
3	TCZ+MTX vs. MTX	0.07	- 0.19 to 0.33
	TCZ vs. MTX	0.09	- 0.17 to 0.34
	TCZ+MTX vs. TCZ	- 0.02	- 0.27 to 0.24
4	TCZ+MTX vs. MTX	0.09	- 0.17 to 0.35
	TCZ vs. MTX	0.10	- 0.16 to 0.36
	TCZ+MTX vs. TCZ	- 0.01	- 0.27 to 0.24
5	TCZ+MTX vs. MTX	0.11	- 0.16 to 0.37
	TCZ vs. MTX	0.12	- 0.14 to 0.38
	TCZ+MTX vs. TCZ	- 0.01	- 0.27 to 0.25

Sustained (drug free) remission over 5 years

	TCZ+ MTX (n=75)	TCZ (n=79)	MTX (n=72)	P-value
Number of patients (%) achieving SR at least once	75 (100)	77 (98)	72 (100)	0.15 ¹
Cumulative duration of SR in weeks, median (IQR)	216 (152-251)	190 (135-240)	172 (129-202)	<0.01 ²
Number of patients (%) achieving sDFR at least once	26 (35)	19 (26)	14 (19)	0.10 ¹
Cumulative duration of sDFR in weeks, median (IQR)	119 (76-157)	107 (53-157)	83 (37-146)	0.27 ²

Median (IQR) [min to max] change in SvdH scores over 5 years

	TCZ+MTX (n=75)	TCZ (n=79)	MTX (n=72)	Strategy comparisons; p-value³
Total SvdH score	0 (0-1) [-2 to 15]	0 (0-1) [-2 to 30]	0 (0-2) [0 to 35]	TCZ+MTX vs. MTX; 0.41 TCZ vs. MTX; 0.05 TCZ+MTX vs. TCZ; 0.67
Erosion score	0 (0-0) [0 to 9]	0 (0-1) [-2 to 17]	0 (0-1) [0 to 35]	TCZ+MTX vs. MTX; 0.62 TCZ vs. MTX; 0.80 TCZ+MTX vs. TCZ; 0.62
JSD score	0 (0-0) [-2 to 8]	0 (0-0) [-1 to 13]	0 (0-1) [0 to 6]	TCZ+MTX vs. MTX; 0.03 TCZ vs. MTX; 0.11 TCZ+MTX vs. TCZ; 0.31

Table 2 (continued)**Mean difference in HAQ scores between initial treatment strategy groups over 5 years and at the end of every year[#]**

Period	Strategy comparisons	Mean difference	95%CI of mean difference
Over 5 years	TCZ+MTX vs. MTX	- 0.05	- 0.22 to 0.11
	TCZ vs. MTX	0.03	- 0.14 to 0.20
	TCZ+MTX vs. TCZ	- 0.08	- 0.25 to 0.08
1	TCZ+MTX vs. MTX	- 0.10	- 0.26 to 0.05
	TCZ vs. MTX	- 0.04	- 0.20 to 0.11
	TCZ+MTX vs. TCZ	- 0.06	- 0.21 to 0.09
2	TCZ+MTX vs. MTX	- 0.09	- 0.25 to 0.06
	TCZ vs. MTX	- 0.03	- 0.19 to 0.12
	TCZ+MTX vs. TCZ	- 0.06	- 0.21 to 0.09
3	TCZ+MTX vs. MTX	- 0.05	- 0.23 to 0.12
	TCZ vs. MTX	0.02	- 0.15 to 0.20
	TCZ+MTX vs. TCZ	- 0.08	- 0.25 to 0.09
4	TCZ+MTX vs. MTX	- 0.05	- 0.23 to 0.12
	TCZ vs. MTX	0.02	- 0.15 to 0.20
	TCZ+MTX vs. TCZ	- 0.08	- 0.25 to 0.09
5	TCZ+MTX vs. MTX	- 0.05	- 0.23 to 0.12
	TCZ vs. MTX	0.03	- 0.15 to 0.20
	TCZ+MTX vs. TCZ	- 0.08	- 0.25 to 0.09

The treatment strategy groups are according to the initial randomization at start of U-Act-Early. All analyses were corrected for baseline DAS28 category (DAS28 <5.1 or ≥5.1) and centre. Outcomes are based on mixed model analyses with random intercept for repeated measurements, and fixed effects for treatment arm, visit-week, interaction visit-week*treatment arm.

¹= Cochran-Mantel-Haenszel test, ²= linear regression, ³= van Elteren test.

#= no statistically significant differences between the groups.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; 95%CI= 95 percent confidence interval; SR= sustained remission; IQR= interquartile range; sDfR= sustained drug free remission; min= minimum; max= maximum; SvdH= Sharp van der Heijde score; JSN= joint space narrowing.

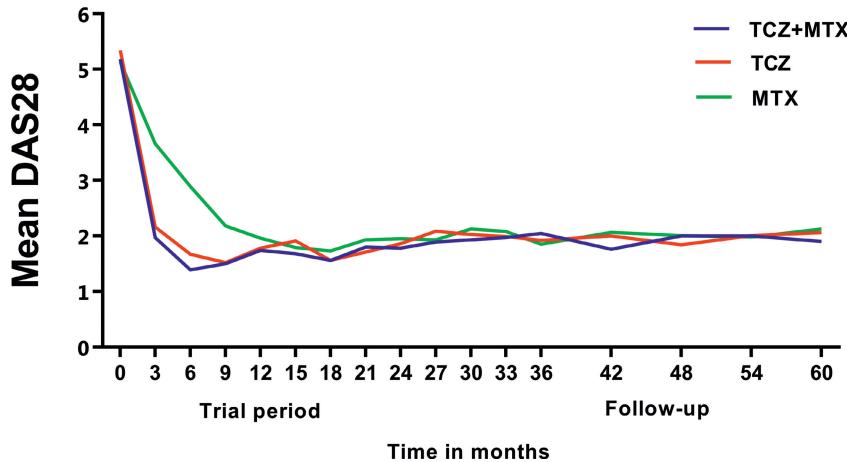


Figure 2 DAS28 over 5 years as predicted by model.

UAE= U-Act-Early; TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; follow-up= post-trial follow-up.

Medication use during the 3-year post-trial follow-up

Any medication for RA was used by 75 patients (100%) in the TCZ+MTX initiation strategy group, 77 (96%) in the TCZ initiation strategy group and 71 (99%) in the MTX initiation strategy group, respectively ($p=0.39$). NSAIDs were used by 68 (91%) patients in the TCZ+MTX initiation strategy group, 72 (91%) in the TCZ initiation strategy group and 65 (90%) in the MTX initiation strategy group, respectively ($p=0.98$). The number of patients using GCs was numerically higher (although not statistically significantly) in the MTX initiation strategy group ($n=47$, 67%) compared with the TCZ+MTX ($n=41$, 55%) and TCZ ($n=39$, 49%) initiation strategy groups ($p=0.09$). csDMARDs were used by more patients in the MTX initiation strategy group ($n=65$, 90%) compared with the TCZ+MTX ($n=56$, 75%) and TCZ ($n=58$, 73%) initiation strategy groups ($p=0.02$). In contrast, the number of patients using bDMARDs was numerically higher in the TCZ+MTX ($n=40$, 53%) and TCZ ($n=36$, 46%) initiation strategy groups than in the MTX initiation strategy group ($n=26$, 36%), although not statistically significantly ($p=0.11$). Any other bDMARD than TCZ was used by 8% of all patients (8% TCZ+MTX, 8% TCZ and 10% MTX, respectively). Additional information about medication use is provided in Supplementary Table S4 and Figures S2-6.

Sustained (drug free) remission

Over 5 years, SR was achieved at least once by 224 of the 226 patients (99%) and the proportions of patients achieving SR were not statistically significantly different ($p=0.15$; Table 2) between the initial strategy groups. The cumulative duration of SR was

statistically significantly longer in the TCZ initiation strategy groups compared with the MTX initiation strategy group ($p<0.01$; Table 2). However, including only the PTFU period, no differences were found between the strategy groups (median [IQR] duration 121 [74-154], 116 [53-145] and 109 [65-140] weeks in TCZ+MTX, TCZ and MTX, respectively ($p=0.62$)). Fifty-nine of the 226 patients (26%) achieved sDFR at least once during the 5 years, without statistically significant differences between the initial strategy groups, nor for the cumulative duration of sDFR, Table 2.

Radiographic progression

Of the 226 patients, only 30% had any radiographic progression over 5 years, without significant differences between the groups ($p=0.09$), Figure 3. The median changes in both total SvdH score and erosion score over 5 years were 0 in all strategy groups, Table 2. The median change in JSON score was statistically significantly lower for the TCZ+MTX initiation strategy group compared with the MTX initiation strategy group (IQR: TCZ+MTX 0-0, MTX 0-1; $p=0.03$).

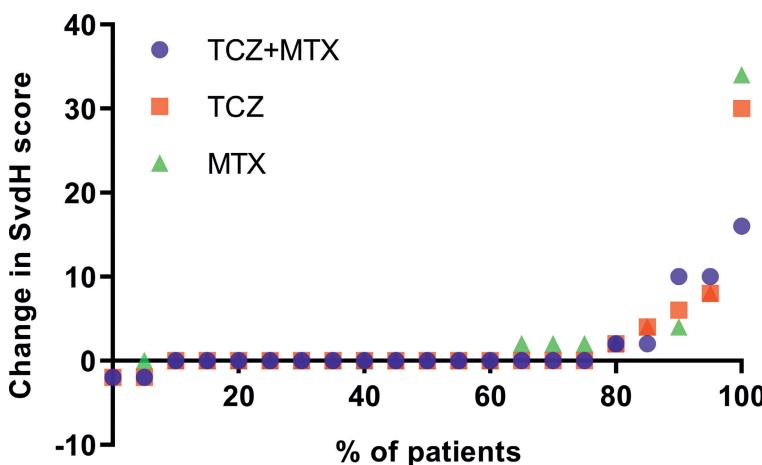


Figure 3 Cumulative probability plot for absolute change in total Sharp van der Heijde score over 5 years.

Change in SvdH score is based on radiographs from baseline U-Act-Early and last available time point (5 years or end of follow-up if earlier than 5 years). Median duration of radiographs from start of U-Act-Early was 5 years in all initial treatment strategies.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; SvdH= Sharp van der Heijde score.

Physical function

No statistically significant differences in HAQ scores over 5 years were observed between the strategy groups, Table 2.

Subgroup analyses for (non-) severe RA are shown in Supplementary files (Table S2 and S3). In general, results are in line with the overall results, probably except for total SvdH score.

Safety

No statistically significant differences over 5 years were found in safety outcomes between strategy groups as initially randomized, Table 3. Of the patients in the TCZ+MTX initiation strategy group, 21 (28%) experienced at least one SAE compared with 23 (29%) in the TCZ initiation strategy group and 15 (21%) in the MTX initiation strategy group, $p=0.47$. Infections, benign/malignant neoplasms and cardiac disorders were the most common SAEs and occurred at least once 15, 8 and 8 times, respectively. Occurrence of at least one serious infection was numerically more frequent in the TCZ initiation strategy groups compared with the MTX initiation strategy group (7 in TCZ+MTX, 5 in TCZ vs. 1 in MTX), $p=0.11$.

Table 3 Safety outcomes over 5 years, results based on data of patients included in the post-trial follow-up, n=226.

	TCZ+MTX (n=75)	TCZ (n=79)	MTX (n=72)	P-value ¹
≥1 AE, n (%)	75 (100)	78 (99)	72 (100)	1.00
AE rate per 100 patient-years	336	340.96	382	
Treatment was given for AE, %	73	70	72	0.13
≥1 SAE, n (%)	21 (28)	23 (29)	15 (21)	0.47
SAE rate per 100 patient-years	7.1	10.5	6.5	
≥1 serious infection, n (%)	7 (9)	5 (6)	1 (1)	0.11

¹= extended Fisher exact test

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; AE= adverse event; SAE= serious adverse event.

Sensitivity analysis

For the primary end point, DAS28 over 5 years, the results of the sensitivity analysis using data of all patients who were included in U-Act-Early (n=317), after imputation of missing PTFU data, were in line with the findings of our main analysis with DAS28 as end point, see Supplementary Table S5.

The sensitivity analysis, analysing the individual unweighted components of the DAS28 separately (ESR excluded), to eliminate the influence of TCZ on APR in the primary outcome yielded statistically significant differences between the treatment strategies for U-Act-Early trial period, but not for the PTFU, see Supplementary Table S3.

DISCUSSION

This is the first long-term PTFU evaluating the effectiveness and safety of tight-controlled treat-to-target treatment strategies using a bDMARD from start of therapy in early DMARD- and GC-naïve RA patients up to 5 years. The high effectiveness during the trial period of 2 years was maintained in a real-world setting without differences between the treatment strategy groups over the 3-year PTFU period. This suggests that an early start of any tight-controlled treat-to-target treatment strategy is also effective on longer-term. The cumulative duration of SR over 5 years was significantly longer for the TCZ initiation strategy groups compared with the MTX initiation strategy group. For the TCZ+MTX initiation strategy group, the median SR duration was about 4 years (216 weeks), which may have an important positive impact on daily activities and quality of life. However, this effect may be biased as patients in the MTX initiation strategy group had a considerable delay in achieving remission compared with the TCZ initiation strategy groups. Including only the PTFU period, no differences for SR were found between the strategy groups, but the duration of SR was numerically longer in the TCZ initiation strategy groups. For future research, a cost effectiveness analysis is warranted to determine whether early initiation of TCZ is cost-effective over a period up to 5 years, compared with MTX initiation.

During the trial, medication was tapered if the patients remained in SR. In the open label PTFU, the numbers of patients using a bDMARD remained numerically higher in the TCZ initiation strategy groups compared with the MTX initiation strategy group, and the number of patients using csDMARDs remained higher in the MTX initiation strategy group. This outcome and the long duration of SR as mentioned above suggest that bDMARD tapering is not yet widely applied in a real-world setting. The most frequently used bDMARD was TCZ in all strategy groups. Physicians were not completely familiar with the effects of TCZ tapering, probably therefore TCZ was tapered cautiously in the real-world setting during the follow-up phase.

Our study also provides evidence for the safety over 5 years of the treatment strategies initiating TCZ, MTX or their combination in a step-up treat-to-target treatment strategy compared according to their randomized groups. The effectiveness and safety results are in line with those of other PTFU studies, with TCZ use, performed in established RA patients.^{14–16}

The limitations of this study include the non-protocolized follow-up, during the PTFU. However, this non-protocolized follow-up gives a realistic view of treatments in a daily practice setting. Not all patients who participated in the U-Act-Early trial could be included in the PTFU, but the sensitivity analysis suggests that results based on the patients with observed data is generalizable to the original U-Act-Early population. Another limitation is the number of missing radiographs at the end of the follow-up

period up to 5 years. Although during the follow-up radiographs available at a previous visit could be used for our analyses, still 17% of radiographs in the TCZ+MTX group, 33% in the TCZ group and 21% in the MTX group were missing. Finally, quality of life was not assessed, as well as data about comorbidity.

Conclusion

The high effectiveness of the treat-to-target treatment strategies during the U-Act-Early trial period of 2 years was maintained during the 3-year PTFU in a real-world setting. Although during the 2-year trial period TCZ-based initiation strategies yielded the most clinical benefit, in the following 3 years differences in important clinical outcomes between initial strategies disappeared, probably due to continuation of the treat-to-target principle, frequently leading to a convergence of treatment regimens. bDMARD use remained higher in the strategy groups initiating TCZ early, possibly due to the lack of a protocolized tapering strategy after the U-Act-Early trial. All treatment strategies proved to be safe.

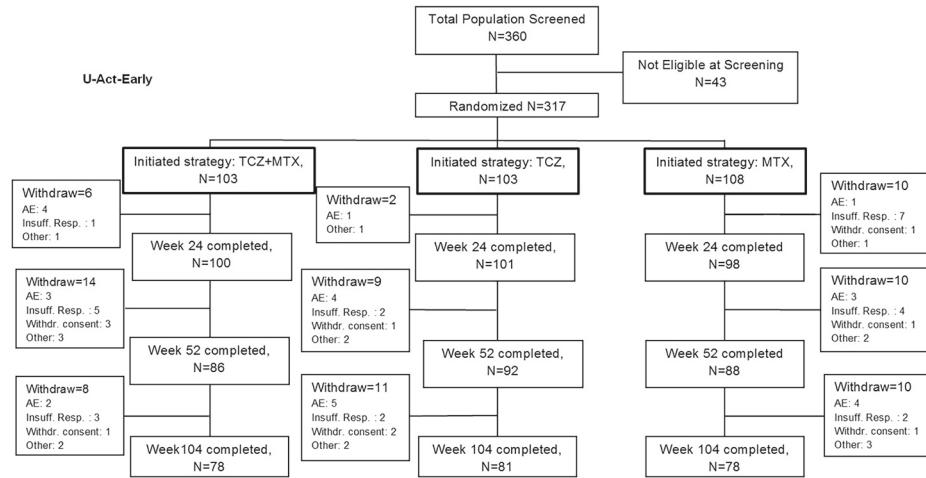
REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76**, 960–977 (2017).
2. St.Clair, E. W. *et al.* Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum.* **50**, 3432–3443 (2004).
3. Quinn, M. A. *et al.* Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal. *Arthritis Rheum.* **52**, 27–35 (2005).
4. Tak, P. P. *et al.* Sustained inhibition of progressive joint damage with rituximab plus methotrexate in early active rheumatoid arthritis: 2-year results from the randomised controlled trial IMAGE. *Ann. Rheum. Dis.* **71**, 351–357 (2012).
5. Detert, J. *et al.* Induction therapy with adalimumab plus methotrexate for 24 weeks followed by methotrexate monotherapy up to week 48 versus methotrexate therapy alone for DMARD-naïve patients with early rheumatoid arthritis: HIT HARD, an investigator-initiated study. *Ann. Rheum. Dis.* **72**, 844–850 (2013).
6. Burmester, G. R. *et al.* Tocilizumab in early progressive rheumatoid arthritis: FUNCTION, a randomised controlled trial. *Ann. Rheum. Dis.* **75**, 1081–1091 (2016).
7. Goekoop-Ruiterman, Y. P. M. *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.* **52**, 3381–3390 (2005).
8. Takeuchi, T. *et al.* Adalimumab, a human anti-TNF monoclonal antibody, outcome study for the prevention of joint damage in Japanese patients with early rheumatoid arthritis: the HOPEFUL 1 study. *Ann. Rheum. Dis.* **73**, 536–543 (2014).
9. Singh, J. A. *et al.* Biologics or tofacitinib for people with rheumatoid arthritis naïve to methotrexate: a systematic review and network meta-analysis. *Cochrane Database Syst. Rev.* **2017**, CD012657 (2017).

10. Bijlsma, J. W. J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
11. Quartagno, A. M., Carpenter, J. & Quartagno, M. M. *Package ‘jomo’*. (2018).
12. Marshall, A., Altman, D. G., Holder, R. L. & Royston, P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med. Res. Methodol.* **9**, 57 (2009).
13. Smolen, J. S. & Aletaha, D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. *Arthritis Rheum.* **63**, 43–52 (2011).
14. Nishimoto, N. *et al.* Long-term safety and efficacy of tocilizumab, an anti-IL-6 receptor monoclonal antibody, in monotherapy, in patients with rheumatoid arthritis (the STREAM study): evidence of safety and efficacy in a 5-year extension study. *Ann. Rheum. Dis.* **68**, 1580–1584 (2009).
15. Kremer, J. M. *et al.* Clinical efficacy and safety maintained up to 5 years in patients with rheumatoid arthritis treated with tocilizumab in a randomised trial. *Clin. Exp. Rheumatol.* **34**, 625–633 (2016).
16. Jones, G. *et al.* Five-year efficacy and safety of tocilizumab monotherapy in patients with rheumatoid arthritis who were methotrexate- and biologic-naïve or free of methotrexate for 6 months: the AMBITION study. *J. Rheumatol.* **44**, 142–146 (2016).

SUPPLEMENTARY FILE

Supplementary Figure S1 shows the patients disposition during the U-Act-Early trial. Eighty (80) patients dropped out due to various reasons (e.g., AE, insufficient response, withdrawal of consent or other).



Supplementary Figure S1 Flowchart, patient disposition during U-Act-Early trial.
TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; Withdr. consent= withdraw of consent; AE= adverse event; Insuff. Resp.= insufficient response.

Supplementary Table S1 provides information about patients who were included and those not included in the post-trial follow-up. No significant differences were shown at start of U-Act-Early between those 2 groups. However, a worse effectiveness and safety profile and shorter follow-up time (drop-out because of ineffectiveness or SAE) were present in patients not included in the post-trial follow-up compared with patients included in the post-trial follow-up at last visit of the U-Act-Early trial, Supplementary Table S1.

Supplementary Table S1 Characteristics of patients included and not included in the post-trial follow-up, at start of U-Act-Early and at last visit of U-Act-Early.

	At start of U-Act-Early			At last visit of U-Act-Early		
	Included in follow-up n=226	Not included in follow-up n=91	P-value	Included in follow-up n=226	Not included in follow-up n=91	P-value
Female, n (%)	156 (69)	56 (62)	0.20	DAS28, mean (SD)	1.94 (0.99)	<0.01
RF+, n (%)	166 (74)	63 (69)	0.38	Remission, n (%)	180 (81)	<0.01
Anti CCP+, n (%)	161 (74)	62 (68)	0.51	HAQ, mean (SD)	0.59 (0.56)	0.01
RF- and/or anti-CCP+, n (%)	181 (80)	68 (75)	0.26	SvdH Score, median (IQR)	0 (0 - 3)	0 (0 - 5)
Caucasian, n (%)	219 (97)	85 (93)	0.02	Safety, n (%)	224 (99)	0.27
Age (years), mean (SD)	54.3 (11.9)	51.1 (14.8)	0.09	- AE	224 (99)	<0.01
				- SAE	27 (12)	
Smoking status, n (%)		0.96		Follow-up time in days, median (IQR)	104 (103 - 106)	<0.01
- Never smoked	91 (40)	38 (42)			55 (32 - 92)	
- Quit smoking	70 (31)	28 (31)				
- Current smoking	65 (29)	25 (27)				
Symptom duration (days), median (IQR)	26 (18 - 43)	27 (16 - 46)	0.83	Follow-up time in days, median (IQR)	104 (103 - 106)	<0.01
DAS28, mean (SD)	5.2 (1.1)	5.4 (1.1)	0.25		(32 - 92)	
HAQ, median (IQR)	1.1 (0.6 - 1.6)	1.1 (0.8 - 1.5)	0.98	Follow-up time in days, median (IQR)	104 (103 - 106)	<0.01
SvdH score, median (IQR)	0 (0 - 1)	0 (0 - 1)	0.25		(32 - 92)	

RF= rheumatoid factor; + = positive; anti-CCP= anti-cyclic citrullinated peptide antibodies; SD= standard deviation; IQR= interquartile range; DAS28= disease activity score assessing 28 joints; HAQ= health assessment questionnaire; SvdH= Sharp van der Heijde score; AE= adverse event; SAE= serious adverse event.

To analyse if effects of therapy were different for severe and non-severe RA, we first defined subgroups as severe (baseline DAS28>5.1) vs. non-severe (DAS28≤5.1); this classification is based on 1) literature showing that baseline DAS28 is an important prognostic factor for RA severity and 2) the fact that randomisation in U-Act-Early was performed by high (DAS28>5.1) or low disease activity (DAS28≤5.1) enabling a valid subgroup analysis. This resulted in 2 subgroups of n=112 severe and n=114 non-severe patients, allowing analyses, although with limited power. Supplementary Table S2 and S3 provide outcomes of Table 2 in the manuscript, stratified by severe and non-severe RA. Thereafter, we tested for significance the interaction between treatment (TCZ+MTX, TCZ, MTX) and RA severity. For DAS28 and HAQ over 5 years, we used the same model as in the primary analysis. For total SvdH score, we used a zero inflated Poisson regression with total SvdH score as outcome, and treatment, RA severity, and the interaction term treatment*RA severity as covariates, controlling for baseline SvdH score. For DAS28 and HAQ, we found no statistically significant interaction between treatment and RA severity, but for total SvdH score the interaction was statistically significant ($p<0.001$). The relative rates for radiographic progression were in favour of the TCZ strategy groups in the severe subgroup (TCZ+MTX vs. MTX relative rate (RR) (95% confidence interval) 0.32 (0.19 to 0.53); TCZ vs. MTX 0.38 (0.26 to 0.55)), whereas no significant differences were shown in the non-severe subgroup (TCZ+MTX vs. MTX 1.07 (0.78 to 1.48); TCZ vs. MTX 1.24 (0.87 to 1.78)). This indicates that the joint protective effect of TCZ regimens, in contrast to that of MTX, is slightly stronger in severe RA than in non-severe RA.

Supplementary Table S2 Effectiveness outcomes over 5 years, results based on data of patients included in the post-trial follow-up severe RA subgroup, n=112.

Mean difference in DAS28 between initial treatment strategy groups over 5 years*

Period	Strategy comparison	Mean difference	95%CI of mean difference
Over 5 years	TCZ+MTX vs. MTX	- 0.05	- 0.35 to 0.25
	TCZ vs. MTX	- 0.19	- 0.48 to 0.11
	TCZ+MTX vs. TCZ	0.14	- 0.15 to 0.43

Sustained (drug free) remission over 5 years

	TCZ+ MTX (n=38)	TCZ (n=40)	MTX (n=34)	P-value
Number of patients (%) achieving SR at least once	38 (100)	40 (100)	34 (100)	1.00 ¹
Cumulative duration of SR in weeks, median (IQR)	186 (118-243)	184 (126-242)	144 (93-202)	<0.03 ²
Number of patients (%) achieving sDFR at least once	12 (32)	10 (25)	6 (18)	0.40 ¹
Cumulative duration of sDFR in weeks, median (IQR)	95 (71-154)	139 (53-159)	60 (30-123)	0.26 ²

Supplementary Table S2 (continued)

Median (IQR) [min to max] change in SvdH scores over 5 years				
	TCZ+ MTX (n=38)	TCZ (n=40)	MTX (n=34)	Strategy comparisons; p-value^a
Total SvdH score	0 (0-0) [0 to 5]	0 (0-2) [0 to 30]	0 (0-4) [0 to 35]	TCZ+MTX vs. MTX; 0.16 TCZ vs. MTX; 0.31 TCZ+MTX vs. TCZ; 0.11
Erosion score	0 (0-0) [0 to 3]	0 (0-1) [0 to 17]	0 (0-1) [0 to 35]	TCZ+MTX vs. MTX; 0.19 TCZ vs. MTX; 0.79 TCZ+MTX vs. TCZ; 0.06
JSN score	0 (0-0) [0 to 5]	0 (0-1) [0 to 13]	0 (0-2) [0 to 6]	TCZ+MTX vs. MTX; 0.12 TCZ vs. MTX; 0.35 TCZ+MTX vs. TCZ; 0.15

Mean difference in HAQ scores between initial treatment strategy groups over 5 years[#]			
Period	Strategy comparisons	Mean difference	95%CI of mean difference
Over 5 years	TCZ+MTX vs. MTX	- 0.01	- 0.25 to 0.24
	TCZ vs. MTX	- 0.06	- 0.30 to 0.19
	TCZ+MTX vs. TCZ	0.05	- 0.19 to 0.29

The treatment strategy groups are according to the initial randomization at start of U-Act-Early. All analyses were corrected for centre. Outcomes are based on mixed model analyses with random intercept for repeated measurements, and fixed effects for treatment arm, visit-week, interaction visit-week*treatment arm.

¹= Cochran-Mantel-Haenszel test, ²= linear regression, ³= van Elteren test, [#]= no statistically significant differences between the groups, Severe RA= baseline DAS28>5.1.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; 95%CI= 95 percent confidence interval; SR= sustained remission; IQR= interquartile range; sDFR= sustained drug free remission; min= minimum; max= maximum; SvdH= Sharp van der Heijde score; JSN= joint space narrowing.

Supplementary Table S3 Effectiveness outcomes over 5 years, results based on data of patients included in the post-trial follow-up non-severe RA subgroup, n=114.

Mean difference in DAS28 between initial treatment strategy groups over 5 years[#]				
Period	Strategy comparison	Mean difference	95%CI of mean difference	
Over 5 years	TCZ+MTX vs. MTX	- 0.15	- 0.42 to 0.21	
	TCZ vs. MTX	- 0.08	- 0.35 to 0.19	
	TCZ+MTX vs. TCZ	-0.07	- 0.42 to 0.13	

Sustained (drug free) remission over 5 years				
	TCZ+ MTX (n=37)	TCZ (n=39)	MTX (n=38)	P-value
Number of patients (%) achieving SR at least once	37 (100)	37 (95)	38 (100)	0.14 ¹
Cumulative duration of SR in weeks, median (IQR)	224 (163-254)	203 (141-232)	182 (142-205)	<0.01 ²

Supplementary Table S3 (continued)

Number of patients (%) achieving sDFR at least once	14 (38)	9 (23)	8 (21)	0.20 ¹
Cumulative duration of sDFR in weeks, median (IQR)	136 (101-163)	102 (33-132)	85 (62-152)	0.24 ²
Median (IQR) [min to max] change in SvdH scores over 5 years				
	TCZ+ MTX (n=37)	TCZ (n=39)	MTX (n=38)	Strategy comparisons; p-value ³
Total SvdH score	0 (0-2) [-2 to 15]	0 (0-1) [-2 to 9]	0 (0-2) [0 to 10]	TCZ+MTX vs. MTX; 0.97 TCZ vs. MTX; 0.08 TCZ+MTX vs. TCZ; 0.31
Erosion score	0 (0-2) [0 to 9]	0 (0-0) [-2 to 6]	0 (0-1) [0 to 5]	TCZ+MTX vs. MTX; 0.52 TCZ vs. MTX; 0.12 TCZ+MTX vs. TCZ; 0.23
JSN score	0 (0-0) [-2 to 8]	0 (0-0) [-1 to 9]	0 (0-1) [0 to 6]	TCZ+MTX vs. MTX; 0.14 TCZ vs. MTX; 0.19 TCZ+MTX vs. TCZ; 0.97
Mean difference in HAQ scores between initial treatment strategy groups over 5 years*				
Period	Strategy comparisons	Mean difference	95%CI of mean difference	
Over 5 years	TCZ+MTX vs. MTX	- 0.15	- 0.36 to 0.54	
	TCZ vs. MTX	0.04	- 0.16 to 0.25	
	TCZ+MTX vs. TCZ	- 0.19	- 0.40 to 0.01	

The treatment strategy groups are according to the initial randomization at start of U-Act-Early. All analyses were corrected for centre. Outcomes are based on mixed model analyses with random intercept for repeated measurements, and fixed effects for treatment arm, visit-week, interaction visit-week*treatment arm.

¹= Cochran-Mantel-Haenszel test, ²= linear regression, ³= van Elteren test, *= no statistically significant differences between the groups, non-severe RA= baseline DAS28≤5.1.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; 95%CI= 95 percent confidence interval; SR= sustained remission; IQR= interquartile range; sDFR= sustained drug free remission; min= minimum; max= maximum; SvdH= Sharp van der Heijde score; JSN= joint space narrowing.

Supplementary Table S4 shows the number of patients using any medication for RA during the post-trial follow-up. The use of GCs and csDMARDs were most frequently used in the MTX initiation strategy group, where the bDMARDs were most frequently used in the TCZ initiation strategy groups.

According to the protocol, GC was not permitted during U-Act-Early. During the post-trial follow-up, the median dosage of GC, in those patients that used oral GC, was somewhat higher in the TCZ+MTX initiation strategy group. However, no relevant difference between the groups could be observed, and proportions of patients using GCs were higher in the MTX initiation strategy group, see Supplementary Figure S2. GC injections were given in 17% of patients of the TCZ+MTX and TCZ groups, and in 12% of patients of the MTX group.

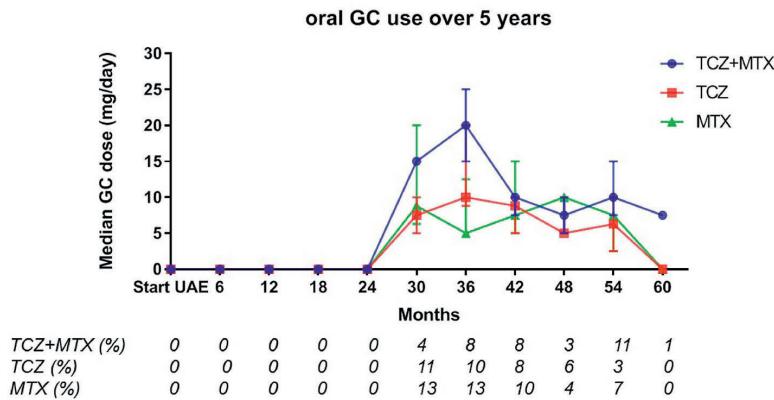
The mean dosage of MTX was higher in the MTX initiation strategy group during the trial period compared with the TCZ initiation strategy groups. During the post-trial follow-up, the mean dosage of MTX increased in the TCZ initiation strategy groups, but it remained higher in the MTX initiation strategy group, Supplementary Figure S3. Regarding to TCZ use, the mean dosage TCZ is more or less equal in the initial treatment strategy groups after 1 and 2 years. However, during the post-trial follow-up, TCZ use increased in the TCZ initiation strategy groups and decreased in the MTX initiation strategy group, Supplementary Figure S4.

During U-Act-Early, starting after 6 months, TCZ use decreased in both TCZ strategy groups and increased in the MTX strategy group, but it remained higher in both TCZ strategy groups up till 24 months. During the post-trial follow-up, TCZ as well as total bDMARD use remained numerically higher in both TCZ strategy groups. Any other bDMARD than TCZ was used in 8% of all patients (8% TCZ+MTX, 8% TCZ and 10% MTX, respectively), Supplementary Figure S5. A reduced dose of bDMARD was mostly used in the TCZ strategy groups, where a full dose of bDMARD was mostly used in the MTX initiation strategy group. However, the total dose of bDMARD use was still higher in the TCZ initiation strategy groups, Supplementary Figure S6.

Supplementary Table S4 Number of patients using RA medication during 3-year post-trial follow-up, (n=226).

	TCZ+MTX strategy group, n=75	TCZ strategy group, n=79	MTX strategy group, n=72	P-value
Any medication for RA, n (%)	75 (100)	77 (96)	71 (99)	0.39
NSAID, n (%)	68 (91)	72 (91)	65 (90)	0.98
GCs, n (%)	41 (55)	39 (49)	48 (67)	0.09
csDMARDs, n (%)	56 (75)	58 (73)	65 (90)	0.02
- MTX, n (%)	38 (51)	48 (61)	51 (71)	
bDMARDs, n (%)	40 (53)	36 (46)	26 (36)	0.11
- TCZ, n (%)	38 (51)	30 (38)	21 (29)	

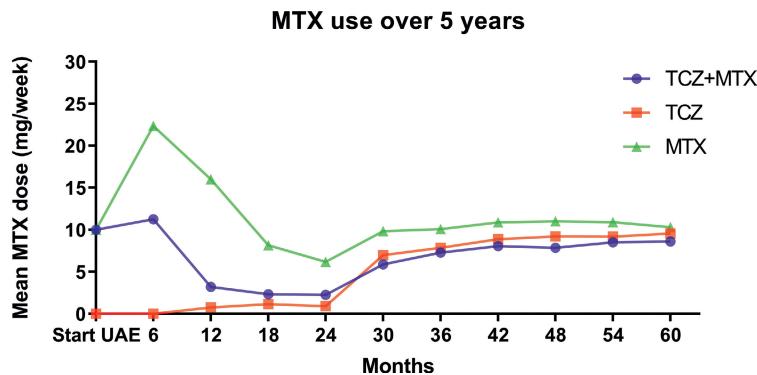
TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; RA= rheumatoid arthritis; NSAID= non-steroidal anti-inflammatory drugs; GCs= glucocorticoids; csDMARDs= conventional synthetic disease modifying anti-rheumatic drugs; MTX= methotrexate; bDMARDs= biological disease modifying anti-rheumatic drugs; TCZ= tocilizumab.



Supplementary Figure S2 The median oral GC dosages in each strategy over time, (n=226).

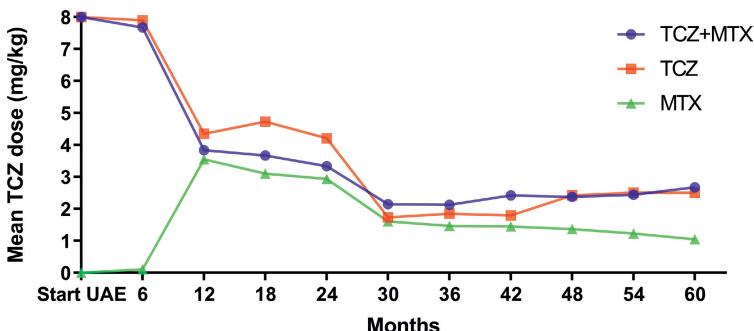
TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; GC= glucocorticoid, dose in prednisone-equivalents.

Proportions of patients using GC (oral) at certain time point. Median GC dose in GC users with lower and upper quartile.



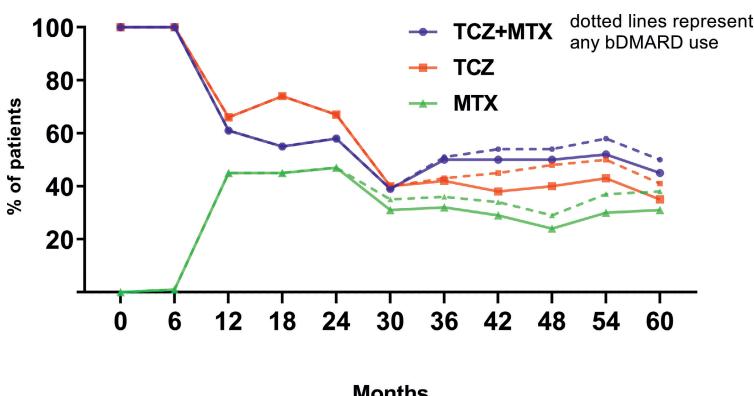
Supplementary Figure S3 The course of MTX dosages in each strategy over time, (n=226).

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group.



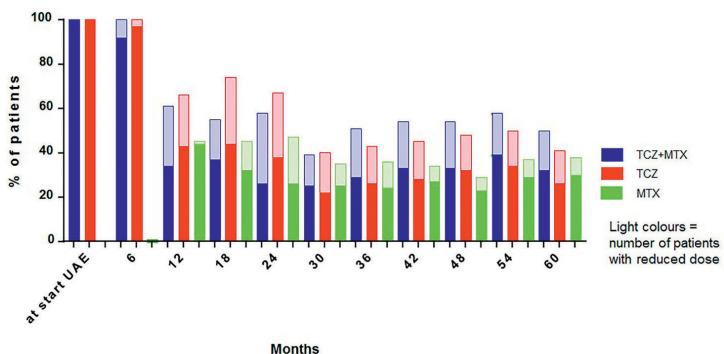
Supplementary Figure S4 The course of TCZ dosages in each strategy over time, (n=226).

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group.



Supplementary Figure S5 Any bDMARD use over 5 years, (n=226).

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group.



Supplementary Figure S6 bDMARD use: Full or Reduced dose, (n=226).

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group.

Supplementary Table S5 shows the outcome of the sensitivity analysis using all patients who were included in U-Act-Early (n=317, based on imputed data). The TCZ strategy groups showed an improved mean difference in DAS28 compared with the MTX strategy group during the first year (TCZ+MTX vs MTX; -0.87 [95%CI-1.15;-0.59], TCZ vs. MTX; -0.85 [95%CI-1.13;-0.57]) and during the second year (TCZ+MTX vs MTX; -0.53 [95%CI-0.91;-0.14], TCZ vs. MTX; -0.55 [95%CI-0.91;-0.20]). No statistical significant differences were shown during the post-trial follow-up period between the initial strategy groups.

The second sensitivity analysis shows outcomes of the components of the DAS28 using data of all patients included in the follow-up study (n=226). An improvement in VAS was found between TCZ+MTX vs. MTX (mean difference -0.47 [95%CI -0.90;-0.04]) during the first year. The TCZ strategy groups showed an improved mean difference in square root transformed values for SJ28 during the first year (TCZ+MTX vs MTX; -0.31 [95%CI-0.43;-0.18], TCZ vs. MTX; -0.16 [95%CI-0.29;-0.03]) and during the second year (TCZ+MTX vs MTX; -0.26 [95%CI -0.37;-0.14], TCZ vs. MTX; -0.13 [95%CI -0.25;-0.01]). Besides, an improved mean difference in square root transformed values for TJC28 were shown for the TCZ strategy groups during the first year (TCZ+MTX vs MTX; -0.24 [95%CI -0.51;-0.08], TCZ vs. MTX; -0.17 [95%CI -0.33;-0.01]) and during the second year (TCZ+MTX vs MTX; -0.21 [95%CI -0.37;-0.05]).

Supplementary Table S5 Sensitivity analyses based on imputed data (n=317) for DAS28 over time and based on observed data (n=226) for components of DAS28.

Period	Strategy comparisons	Mean differences	95%CI of mean differences	P-value
Change in DAS28 between initial treatment strategy groups over 5 years and at the end of every year (n=317)				
1	TCZ+MTX vs. MTX	- 0.87	- 1.15 to - 0.59	<0.01
	TCZ vs. MTX	- 0.85	- 1.13 to - 0.57	<0.01
2	TCZ+MTX vs. MTX	- 0.53	- 0.91 to - 0.14	0.01
	TCZ vs. MTX	- 0.55	- 0.91 to - 0.20	<0.01
3	TCZ+MTX vs. MTX	- 0.37	- 0.83 to 0.08	0.12
	TCZ vs. MTX	- 0.41	- 0.83 to 0.00	0.06
4	TCZ+MTX vs. MTX	- 0.32	- 0.80 to 0.16	0.20
	TCZ vs. MTX	- 0.37	- 0.80 to 0.06	0.11
5	TCZ+MTX vs. MTX	- 0.37	- 0.77 to 0.24	0.31
	TCZ vs. MTX	- 0.41	- 0.78 to 0.13	0.17
Change in VAS between initial treatment strategy groups at the end of every year (n=226)				
1	TCZ+MTX vs. MTX	-0.47	-0.90 to -0.04	0.03
	TCZ vs. MTX	-0.21	-0.63 to 0.22	0.34
	TCZ+MTX vs. TCZ	-0.26	-0.68 to 0.16	0.22
2	TCZ+MTX vs. MTX	-0.38	-0.80 to 0.04	0.08
	TCZ vs. MTX	-0.14	-0.55 to 0.28	0.28
	TCZ+MTX vs. TCZ	-0.24	-0.65 to 0.17	0.26
3	TCZ+MTX vs. MTX	0.16	-0.34 to 0.65	0.55
	TCZ vs. MTX	0.27	-0.22 to 0.76	0.28
	TCZ+MTX vs. TCZ	-0.11	-0.60 to 0.37	0.65

Supplementary Table S5 (continued)

Period	Strategy comparisons	Mean differences	95%CI of mean differences	P-value
4	TCZ+MTX vs. MTX	0.17	-0.33 to 0.67	0.50
	TCZ vs. MTX	0.28	-0.22 to 0.78	0.27
	TCZ+MTX vs. TCZ	-0.11	-0.60 to 0.38	0.66
5	TCZ+MTX vs. MTX	0.19	-0.32 to 0.69	0.47
	TCZ vs. MTX	0.29	-0.21 to 0.80	0.25
	TCZ+MTX vs. TCZ	-0.11	-0.60 to 0.39	0.68
Change in $\sqrt{SJC28}$ between initial treatment strategy groups at the end of every year (n=226)				
1	TCZ+MTX vs. MTX	-0.31	-0.43 to -0.18	<0.01
	TCZ vs. MTX	-0.16	-0.29 to -0.03	0.01
	TCZ+MTX vs. TCZ	-0.15	-0.27 to -0.02	0.02
2	TCZ+MTX vs. MTX	-0.26	-0.37 to -0.14	<0.01
	TCZ vs. MTX	-0.13	-0.25 to -0.01	0.03
	TCZ+MTX vs. TCZ	-0.12	-0.12 to 0.14	0.04
3	TCZ+MTX vs. MTX	0.05	-0.08 to 0.18	0.46
	TCZ vs. MTX	0.04	-0.09 to 0.17	0.58
	TCZ+MTX vs. TCZ	0.01	-0.12 to 0.14	0.85
4	TCZ+MTX vs. MTX	0.06	-0.08 to 0.19	0.40
	TCZ vs. MTX	0.04	-0.09 to 0.17	0.54
	TCZ+MTX vs. TCZ	0.02	-0.11 to 0.15	0.81
5	TCZ+MTX vs. MTX	0.06	-0.07 to 0.20	0.35
	TCZ vs. MTX	0.05	-0.09 to 0.18	0.51
	TCZ+MTX vs. TCZ	0.02	-0.11 to 0.15	0.77
Change in $\sqrt{TJC28}$ between initial treatment strategy groups at the end of every year (n=226)				
1	TCZ+MTX vs. MTX	-0.24	-0.51 to -0.08	<0.01
	TCZ vs. MTX	-0.17	-0.33 to -0.01	0.04
	TCZ+MTX vs. TCZ	-0.07	-0.24 to 0.09	0.37
2	TCZ+MTX vs. MTX	-0.21	-0.37 to -0.05	<0.01
	TCZ vs. MTX	-0.15	-0.31 to 0.01	0.06
	TCZ+MTX vs. TCZ	-0.06	-0.21 to 0.10	0.46
3	TCZ+MTX vs. MTX	-0.01	-0.20 to 0.18	0.93
	TCZ vs. MTX	-0.04	-0.23 to 0.15	0.70
	TCZ+MTX vs. TCZ	0.03	-0.16 to 0.22	0.76
4	TCZ+MTX vs. MTX	-0.00	-0.20 to 0.19	0.98
	TCZ vs. MTX	-0.03	-0.23 to 0.16	0.72
	TCZ+MTX vs. TCZ	0.03	-0.16 to 0.22	0.74
5	TCZ+MTX vs. MTX	0.00	-0.19 to 0.20	0.98
	TCZ vs. MTX	-0.03	-0.22 to 0.16	0.75
	TCZ+MTX vs. TCZ	0.03	-0.16 to 0.22	0.72

The treatment strategy groups are according to the initial randomisation at start of U-Act-Early. All analyses were corrected for baseline DAS28 category (DAS28 <5.1 or ≥5.1) and centre. Outcomes are based on mixed model analyses with random intercept for repeated measurements, and fixed effects for treatment arm, visit-week, interaction visit-week*treatment arm.

TCZ+MTX= tocilizumab + methotrexate initiation strategy group; TCZ= tocilizumab + placebo-methotrexate initiation strategy group; MTX= methotrexate + placebo-tocilizumab initiation strategy group; 95%CI= 95 percent confidence interval; DAS28= disease activity score assessing 28 joints; VAS= visual analogue scale; SJC= swollen joint count; TJC= tender joint count.

Chapter 7:

Effect on costs and quality-adjusted life-years of treat-to-target treatment strategies initiating methotrexate, or tocilizumab, or their combination in early rheumatoid arthritis.

5 year economic evaluation.

M.M.A. Verhoeven¹

J. Tekstra¹

J.M. van Laar¹

A. Pethö-Schramm²

M.E.A. Borm³

J.W.J. Bijlsma¹

J.W.G. Jacobs¹

F.P.J.G. Lafeber¹

P.M.J. Welsing¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands;

²F. Hoffmann-La Roche, Basel, Switzerland;

³Roche Nederland BV, Woerden, the Netherlands.

Epub ahead of print in The Journal of Rheumatology August 2020

ABSTRACT

Objective

To evaluate the cost effectiveness over 5 years of treat-to-target treatment strategies in early DMARD-naïve RA initiating tocilizumab (TCZ) +/- methotrexate (MTX) versus initiating MTX.

Methods

Data on resource use were collected with questionnaires at baseline, 3, 6, 12, 24 months and yearly thereafter, and were converted to costs using Dutch reference prices. QALYs were calculated using EQ5D-5L, with utility based on Dutch tariff or estimated by HAQ. To account for missing cost- and QALY data and for sample uncertainty, first bootstraps (10,000 samples) were obtained. Secondly, single imputation using chained equations nested within these bootstrap samples was performed. An economic evaluation was performed for TCZ+MTX and TCZ, compared to MTX as initial treatment in a treat-to-target strategy from a healthcare and a societal perspective over 5 years. Several sensitivity analyses were performed.

Results

Mean differences in QALYs were small and not significant (TCZ+MTX vs. MTX: 0.06 (95%CI -0.02 to 0.13); TCZ vs. MTX: -0.03 (95%CI -0.05 to 0.11)). Limited savings in indirect non-healthcare costs and productivity loss costs (for TCZ only) were observed but these did not compensate for the higher medication costs. Sensitivity analyses did not materially change these findings, although lower-priced TCZ, or reserving TCZ as initial therapy for prognostically unfavorable RA patients improved cost effectiveness considerably, but did not individually lead to a strategy being cost-effective.

Conclusion

Based on our analyses, early initiation of TCZ(+MTX), is not cost-effective compared to MTX initiation in a step-up treat-to-target treatment strategy over 5 years in early RA patients.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease of mainly the joints, in which lifelong treatment is required to reduce symptoms, improve physical function, and avoid structural joint damage. Over the last decades, early treatment with disease modifying anti-rheumatic drugs (DMARDs) according to the treat-to-target principle together with the use of newer, more effective, biological (b)DMARDs improved these outcomes.¹

Due to their high costs compared to conventional synthetic (cs)DMARDs, bDMARDs are reserved as second-line treatment after a first DMARD (mostly methotrexate (MTX)) has shown to be insufficiently effective. Typically, bDMARDs are added to the existing csDMARD therapy. Some studies have evaluated TNF-inhibitors (TNFi) as initial bDMARD therapy in RA, and have shown that initial treatment with this bDMARD, though effective, is currently not cost-effective over 1-2 years.²⁻⁴ Simulation model based analyses over a longer term (i.e., 5-year to life-time horizon) also conclude that this is not a cost-effective strategy.⁵⁻⁷

Tocilizumab (TCZ) is a non-TNFi bDMARD, targeting the IL-6 receptor. A recent trial and its post-trial follow-up (PTFU) investigated the effect of TCZ as first-line treatment, in early RA patients. U-Act-Early was a 2-year multicentre, double-blind, randomized, placebo-controlled trial in early (DMARD-naïve) RA patients treated to the target of remission (i.e., DAS28<2.6 with ≤4 swollen joints). Patients were assigned to step-up treatment strategies starting with TCZ, MTX or their combination (TCZ+MTX).⁸ If the treatment target was not achieved, MTX (in TCZ-monotherapy group) or TCZ (in MTX group) was added. When patients achieved and remained in remission for ≥24 weeks, medication was tapered and finally stopped.⁸ Patients were followed for 3 years after the trial (i.e., PTFU), during which treatment was at the discretion of the treating rheumatologist.⁹

Results of U-Act-Early and its PTFU showed a high effectiveness of all treatment strategies. Almost all patients achieved (sustained) remission over follow-up, which was achieved earlier in the TCZ strategy groups. Due to active tapering during the trial, we observed that bDMARD use decreased in TCZ-based strategies and increased in the MTX-based strategy, and TCZ use was almost similar in both arms after two years. High effectiveness and higher sustained (drug free) remission rates compared to previous cost effectiveness bDMARD-studies were achieved in U-Act-Early, as well as in its PTFU, possibly due to the patient population with very early RA, treated within the 'window of opportunity'. Furthermore, TCZ-monotherapy was equally effective as TCZ+MTX as initial strategy,⁹ in contrast to most TNFi. Since some patients do not tolerate MTX, TCZ-monotherapy may have a positive influence on quality of life and productivity. Therefore, the question could be raised whether using TCZ (+/- MTX) as initial therapy in a strict treat-to-target strategy with active tapering in case of sustained remission

might be cost-effective in very early RA patients compared to MTX as initial strategy, taking into account costs related to medical consumption, other direct healthcare costs, indirect non-healthcare costs, as well as costs related to productivity loss.

The current study aimed to compare these costs and quality-adjusted life years (QALYs), as a generic measure of effectiveness, of initial treatment with TCZ+MTX, TCZ and MTX in a treat-to-target treatment strategy over the 5-year observation period.

METHODS

Detailed methods of U-Act-Early and PTFU are reported elsewhere.^{8,9} In the current study we evaluated the cost effectiveness of TCZ(+MTX) compared to MTX. For this purpose we used individual patient data of U-Act-Early and its PTFU, hence we performed a trial-based cost-utility analysis (i.e., using empirical study data). Patients visited the out-patient-clinic every month during the 2-year trial period, every 3 months during the first year of the PTFU and every 6 months thereafter. The medical ethics research committee of the University Medical Center Utrecht approved the trial, ID number: ML28388. For the PTFU, the institutional review boards of the participating centers confirmed that the Medical Research Involving Human Subjects Act (WMO) was not applicable. In both studies, all patients gave written informed consent.

QALYs

Effectiveness was expressed as QALYs, measured at baseline, 3, 6, 12 and 24 months using the EuroQol 5-Dimensions 5-Levels (EQ-5D5L) with results expressed as a utility score. Utility is a score ranging from 0 (death) to 1 (full health), and was based on the Dutch tariff.¹⁰ For the post-trial period, during which EQ-5D5L was not measured, and in case of missing EQ-5D5L data, utility was estimated from HAQ and age.¹¹ See Supplementary Data S1 for details.

Medication costs

Medication use was based on recorded dose and duration of (non-placebo) MTX and TCZ, and recorded use (yes/no) for NSAIDs. If patients dropped out during U-Act-Early (about 25%, n=80), we assumed that TCZ was discontinued, treatment with MTX (15mg per week) was continued and a TNFi (i.e., adalimumab 40mg subcutaneously biweekly) was added until the end of the trial, as advised by the U-Act-Early study protocol. If the reason for drop-out was infection or malignancy (about 34% of dropouts, n=27), we assumed that treatment was continued with MTX (15mg per week), and that, for safety reasons, another csDMARD (e.g., sulfasalazine 1000mg twice a day) was added, instead of a TNFi, in line with the U-Act-Early protocol and clinical guidelines during the inclusion period (i.e., 2010-2012). For patients completing U-Act-Early, but not included

in PTFU, medication use (per category) during the first 3 months after U-Act-Early ('first 3 months of PTFU') were assumed equal to the use in the last 3 months of U-Act-Early and further set as 'missing' and imputed for the analysis (see below).

To calculate medication costs, medication use was multiplied by national prices, separately for bDMARD, csDMARD, and NSAIDs, see Supplementary Table 1.¹² If the administration route of the medication was intravenous, we calculated additional costs for day care. For NSAIDs, we assumed the most frequently used NSAID in the trial (i.e., naproxen) in a dose of 0.75 gram per day.

Other direct healthcare costs and indirect non-healthcare costs

Other direct healthcare related resource use (e.g., physician visit, hospital admissions, non RA medication) and indirect non-healthcare related resource use (e.g., travel cost, purchase of a stair lift, etc.) typically over the last 3 months were obtained via healthcare resource use questionnaires, as used in previous studies.^{3,4,10,13}

Productivity loss costs

Data about productivity loss was obtained with the work participation questionnaire as filled out at baseline, 3, 6, 12 and 24 months during the trial, and yearly during PTFU. Costs related to productivity loss were calculated as incidental work loss hours caused by sickness as well as structural reduction in working hours, including discontinuation of paid work. The productivity loss costs related to structural reduction in work were calculated compared to the productivity hours per week at baseline. Working hours lost were valued by the average wage of Dutch citizens by gender to calculate the total productivity loss costs, following the human capital approach.¹⁰ Productivity loss costs were also calculated, applying the friction cost approach using only structural productivity loss hours during the first 15 weeks (i.e., friction period 2017).¹⁰ If the patient reached the retirement age of 65 years, we assumed that patients retired and productivity loss costs no longer occurred.

Costs were calculated using 2017 reference prices in euro (€) and in line with the Dutch costing manual see Supplementary Table 1.¹⁰

Discounting

Costs were discounted using a discount rate of 4% per year, and QALYs using a discount rate of 1.5% per year, according to Dutch guidelines for economic evaluations in healthcare.¹⁰

Missing data and data imputation

To obtain yearly costs and QALYs, linear interpolation was used over scheduled visits within a year. To account for remaining missing cost (12%/22% during trial/PTFU) and QALY

(13%/20%) values in a year, as well as (population) uncertainty in outcomes, we used the following approach, which has been suggested to be optimal in this situation.¹⁴ As first step, 10,000 bootstrap samples (with replacement) were taken per treatment strategy arm. In the second step, single imputation of the missing yearly QALY and cost estimates was performed per bootstrap sample (see Supplementary Data S2 for details).

Analysis

Baseline (inclusion in U-Act-Early) characteristics of the strategy groups were described for the 'intention-to-treat' (ITT) population and the population as included in the PTFU. Mean values (2.5 and 97.5 percentiles) of costs per category and QALYs were calculated for the treatment strategy groups as well as for the differences of TCZ+MTX and TCZ, both compared to MTX over a 2-year (trial period) and 5-year (trial + PTFU) time horizon.

Incremental cost effectiveness ratios (ICER: calculated as difference in costs between two strategies divided by difference in effectiveness, expressed as € per QALY gained) were estimated. A healthcare perspective (i.e., including only healthcare cost) as well as societal perspectives (i.e., including non-healthcare cost like costs related to productivity) were used for this estimation. The primary analysis uses the ITT population of U-Act-Early (n=317) and the follow-up period of 5 years. Results were illustrated in cost effectiveness planes. The distribution of simulations over the quadrants in these planes was also calculated and illustrates if TCZ(+MTX) is cost-effective compared to MTX (e.g., south east quadrant indicates lower costs with higher effectiveness (TCZ(+MTX) is dominant compared to MTX)). Cost effectiveness acceptability curves (CEACs) were constructed to illustrate the probability of TCZ(+MTX) being cost-effective compared to MTX at different willingness to pay (WTP; price society is willing to pay for a gain of 1 QALY) thresholds.

Sensitivity and scenario analyses

To evaluate the impact of missing data, a complete case analysis was performed (i.e., using data from PTFU population) as sensitivity analyses. To investigate the sensitivity of the results to treatment effect, we assumed QALYs were 0.05 higher and lower, respectively, in the TCZ strategies. We also performed a sensitivity analysis using discount rates of 4% per year for both costs and QALYs, as is generally in line with international guidelines.

Moreover the following scenario analyses were performed. As TCZ has recently become available as a subcutaneous injection, making an admission day no longer necessary, we performed a scenario analysis assuming that TCZ is given subcutaneously instead of intravenously, resulting in a reduction of 23% of the (total) price (Supplementary Table 1). Furthermore, a scenario analysis was performed in which, in

addition to subcutaneous administration, reductions in drug price of 10%, respectively 30%, were assumed. Finally, we performed an analysis in which only patients with a baseline DAS28>5.1 (i.e., more severe disease; a stratification factor for randomization in U-Act-Early) were included.⁸ All sensitivity/scenario analyses are based on discounted costs and QALYs (unless otherwise specified, with 4% for costs and 1.5% for QALYs, respectively) and over the 5-year time horizon.

RESULTS

Overall no relevant differences between the treatment groups (as initially randomized) were found at baseline, Table 1. Ninety-one (91, 29%) patients did not participate in the PTFU. Only for mean baseline working hours per week, differences were found between patients included in the PTFU and patients not included in the PTFU (22.3 vs. 26.7 (p=0.03)), Supplementary Table 2. Approximately 80% of all patients was employed at baseline, and worked on average 24 hours weekly.

Table 1 baseline characteristics of patients included in the U-Act-Early trial.

	TCZ+MTX (n=106)	TCZ (n=103)	MTX (n=108)
Employed, n (%)	85 (80)	79 (77)	87 (81)
Working hours per week, mean (SD)	24.0 (15.3)	22.1 (15.5)	24.6 (15.8)
Female, n (%)	65 (61)	78 (76)	69 (64)
RF +, n (%)	75 (71)	68 (66)	86 (80)
Anti CCP +, n (%)	72 (68)	67 (65)	84 (78)
RF- and/or anti-CCP+, n (%)	79 (75)	77 (75)	93 (86)
Age (years), median (IQR)	53.0 (46.0-60.0)	55.0 (47.0-63.0)	53.0 (44.5-62.0)
Symptom duration (days), median (IQR)	24.5 (16.0-41.5)	25.5 (18.0-45.0)	27.0 (15.0-46.0)
DAS28, mean (SD)	5.2 (1.1)	5.3 (1.1)	5.1 (1.2)
HAQ, mean (SD)	1.1 (0.67)	1.3 (0.66)	1.1 (0.59)
included in post-trial follow-up, n (%)	75 (71)	79 (77)	72 (67)

TCZ+MTX= initial tocilizumab + methotrexate strategy group; TCZ= initial tocilizumab + placebo-methotrexate strategy group; MTX= initial methotrexate + placebo-tocilizumab strategy group; RF= rheumatoid factor; += positive; anti-CCP= anti-cyclic citrullinated peptide antibodies; SD= standard deviation; IQR= interquartile range; DAS28= disease activity score assessing 28 joints [range 0-9.4, higher is more activity]; HAQ= health assessment questionnaire score [range 0-3, 3= worst function].

Figure 1 and Supplementary Table 3 shows the cost per category in each year, and over 2 and 5 years. Total medication costs were, as expected, higher in the TCZ strategy groups compared to the MTX strategy group over 2 and 5 years (TCZ+MTX and TCZ vs. MTX; 27,900 and 29,100 vs. 13,200; 41,200 and 43,700 vs. 26,600, respectively). After 2 years, medication costs decreased in all groups, and only in the last year (year 5) mean medication costs (i.e., total and TCZ costs) were less for the TCZ strategy groups.

Regarding productivity loss costs using the human capital approach, the highest costs over 5 years were observed in the TCZ+MTX strategy group: 51,700, vs. 39,900 (TCZ) and 46,500 (MTX), respectively. Using the friction cost approach for productivity loss costs differences between TCZ and MTX-based strategy groups were small. Direct healthcare costs were highest in the TCZ+MTX group and indirect non-healthcare costs were lowest in this group over 5 years. A similar number of QALYs was attained in all treatment groups over 5 years. Supplementary Table 4 gives an overview of undiscounted cost and QALYs.

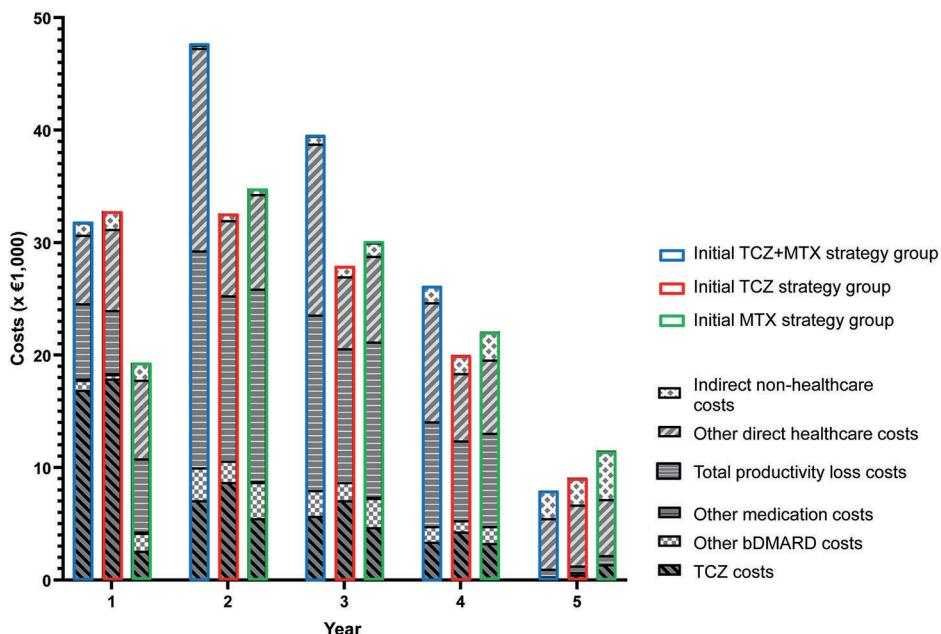


Figure 1 Cost (€ x 1,000) over time per initial treatment strategy in means.

Outcomes based on single imputation nested in 10,000 bootstraps; costs expressed in euros. TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; other medication costs= all RA medication costs, including NSAID and csDMARDs, excluding bDMARDs; productivity loss costs= costs related to work loss or being less productive; direct healthcare costs= all costs related to healthcare, also for other diseases (RA medication costs excluded); indirect non healthcare costs= costs related to patient and family (e.g., travel cost; buying stair lift, etc.).

Mean differences between the treatment strategies with 95% percentile ranges per cost category and QALYs are shown in Table 2.

For TCZ+MTX, over 5 years, direct healthcare costs and medication costs as well as total cost are significantly (i.e., the 95% percentile range of the difference does not include 0) higher compared to those of MTX and indirect non-healthcare costs significantly lower. Productivity loss costs and QALYs are numerically higher without

being significant. Regarding the 2 to 5-year evaluation period the difference in total cost increases and the QALY difference remains the same, decreasing the cost effectiveness of TCZ+MTX compared to MTX.

For TCZ, direct healthcare costs (RA medication costs excluded), and productivity loss costs are lower compared to MTX (not significant), but total cost are significantly higher. Over a 5-year evaluation period the difference in total cost remains constant and significantly higher, except if productivity loss costs are calculated using the human capital approach, which results in a smaller not significant increase in costs compared to those of MTX. Differences in QALYs are not significant, but the small numerical advantage over the 2-year period is lost over the 5-year period, leading to a small disadvantage in QALYs for TCZ vs. MTX.

TCZ+MTX has an ICER between 526,674 and 594,021 per QALY gained compared to MTX (depending on perspective taken and method of calculating productivity loss costs) and is inferior in 23% of all bootstrap samples over 5 years. TCZ has a negative ICER (between -426,967 and -149,241 depending on the perspective taken and method of calculating productivity loss costs) as on average QALYs were lost over 5 years, and TCZ is inferior in between 47% and 65% of all bootstrap samples over 5 years. Taken the societal perspective, using human capital approach, TCZ dominated (i.e., more QALYs and less costs) in 8% of all bootstrap samples, and in 18% TCZ was found less expensive, but also less effective, Figure 2. Outcomes based on undiscounted values are shown in Supplementary Table 5.

The probability of TCZ(+MTX) being a cost-effective intervention over 5 years, using different WTP thresholds for a QALY, was in general low, but somewhat higher for TCZ compared to TCZ+MTX, see Figure 3 and Supplementary Figure 1-2.

Table 2 Mean (2.5-97.5 percentile) difference in costs (€, rounded to the nearest hundreds) and QALYs, and ICER with distribution in cost effectiveness plane (%).

	TCZ+MTX vs. MTX	TCZ vs. MTX
Over 2 years		
Medication costs	14,800 (12,000 to 17,600)	15,900 (13,100 to 18,600)
Direct healthcare costs	8,700 (4,100 to 13,400)	-1,400 (-4,700 to 1,800)
Indirect non healthcare costs	-600 (-1,200 to 0)	200 (-700 to 1,100)
Productivity loss costs [#]	2,500 (-4,200 to 9,300)	-3,200 (-9,900 to 3,400)
Productivity loss costs ^{\$}	400 (-1,100 to 1,900)	-300 (-1,900 to 1,200)
Total cost and effect		
Healthcare perspective	23,500 (18,000 to 29,000)	14,500 (10,000 to 18,900)
Societal perspective [#]	25,300 (17,200 to 33,300)	11,400 (3,600 to 19,300)
Societal perspective ^{\$}	23,200 (17,800 to 28,600)	14,400 (9,600 to 19,000)
QALYs	0.06 (-0.02 to 0.13)	0.03 (-0.05 to 0.11)

Table 2 (continued)

	TCZ+MTX vs. MTX	TCZ vs. MTX
ICER ^{HC}	425,088	539,531
% (SE , SW, <u>NW</u> , NE)	(0,0,7,93)	(0,0,26,74)
ICER [#]	458,954	425,144
% (SE , SW, <u>NW</u> , NE)	(0,0,7,93)	(0,0,26,74)
ICER ^{\$}	420,829	533,521
% (SE , SW, <u>NW</u> , NE)	(0,0,7,93)	(0,0,26,74)
Over 5 years		
Medication costs	14,600 (9,300 to 19,900)	17,200 (11,700 to 22,500)
Direct healthcare costs	19,900 (9,700 to 30,300)	-2,700 (-9,700 to 3,800)
Indirect non healthcare costs	-4,100 (-5,400 to -2,800)	-2,800 (-4,300 to -1,400)
Productivity loss costs [#]	5,200 (-8,400 to 18,700)	-6,600 (-19,900 to 6,800)
Productivity loss costs ^{\$}	1,500 (-1,900 to 4,200)	0 (-3,300 to 2,900)
Total cost and effect		
Healthcare perspective	34,500 (22,200 to 46,600)	14,500 (5,100 to 23,500)
Societal perspective [#]	35,600 (18,700 to 52,200)	5,000 (-10,500 to 20,800)
Societal perspective ^{\$}	31,600 (19,600 to 43,600)	11,500 (1,700 to 20,900)
QALYs	0.06 (-0.10 to 0.22)	-0.03 (-0.20 to 0.1)
ICER ^{HC}	575,982	-426,967
% (SE , SW, <u>NW</u> , NE)	(0,0,23,77)	(0,0,65,35)
ICER [#]	594,021	-149,241
% (SE , SW, <u>NW</u> , NE)	(0,0,23,77)	(8,18,47,27)
ICER ^{\$}	526,674	-337,609
% (SE , SW, <u>NW</u> , NE)	(0,0,23,77)	(1,0,65,34)

Outcomes based on single imputation nested in 10,000 bootstraps, costs expressed in euro's.

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; Medication costs= all RA medication costs; Direct healthcare costs= all costs related to healthcare, also for other diseases (RA medication costs excluded); Indirect non healthcare costs= costs related to patient and family (e.g., travel cost; buying stair lift; etc.); Productivity loss costs= costs related to work loss or being less productive; [#]= using human capital approach; ^{\$}= using friction cost approach, only counting costs for a period of absence up to 85 days; Healthcare perspective= direct healthcare cost + medication cost; Societal perspective= direct healthcare costs + indirect not healthcare costs + productivity loss costs + medication costs; QALY= quality-adjusted life years; ICER= incremental cost effectiveness ratio (using societal perspective according to human capital approach); HC= healthcare perspective; **SE**= south east: gain in QALY, less expensive (i.e., TCZ dominant); SW= south west: loss in QALY, less expensive; NW= north west: loss in QALY, more expensive (i.e., TCZ inferior); NE= north east: gain in QALY, more expensive.

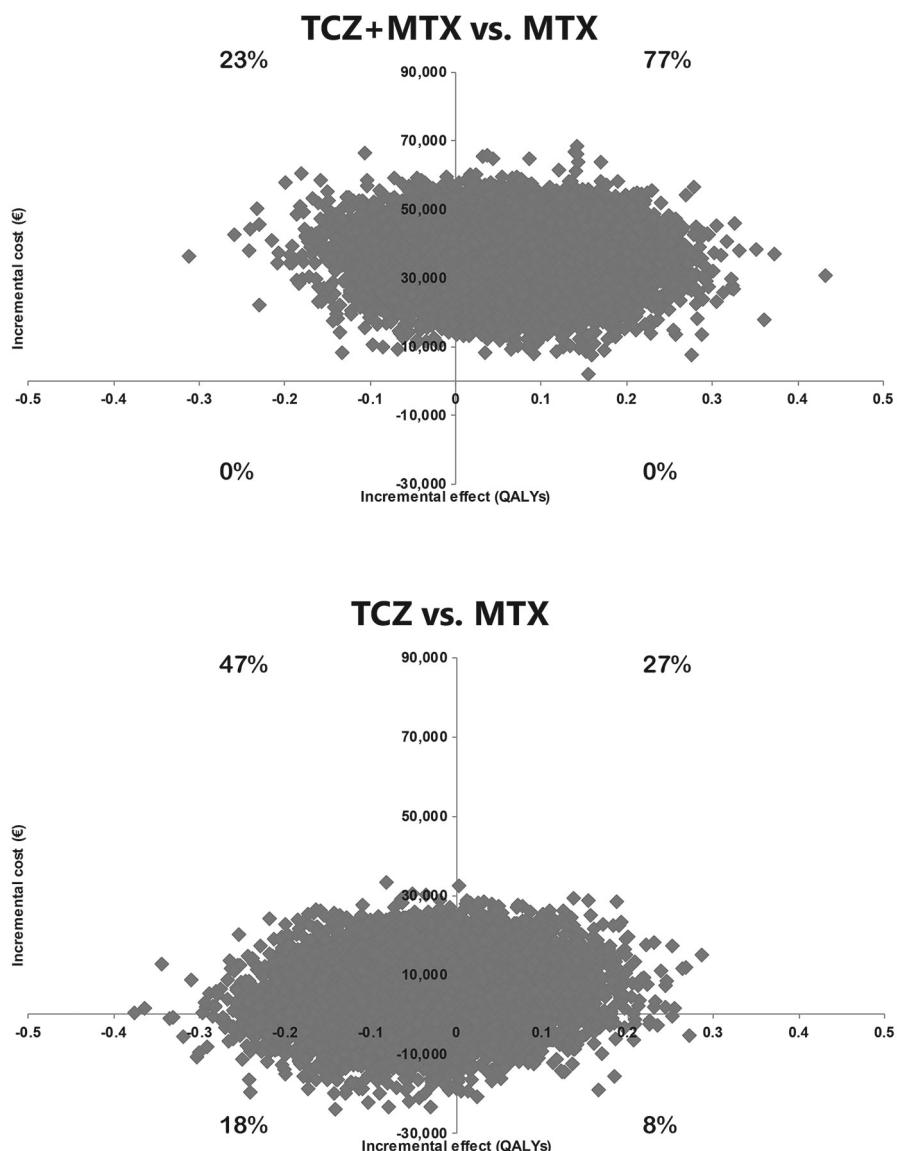


Figure 2 Cost effectiveness planes for the TCZ-based initiation treatment strategy groups versus the MTX initiation strategy group over 5 years, using the societal perspective (human capital approach).

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group.

Example: TCZ dominated (i.e., more QALYs and less costs) in 8% of all bootstrap samples (8% 'chance' that the intervention is cost-effective) compared to MTX.

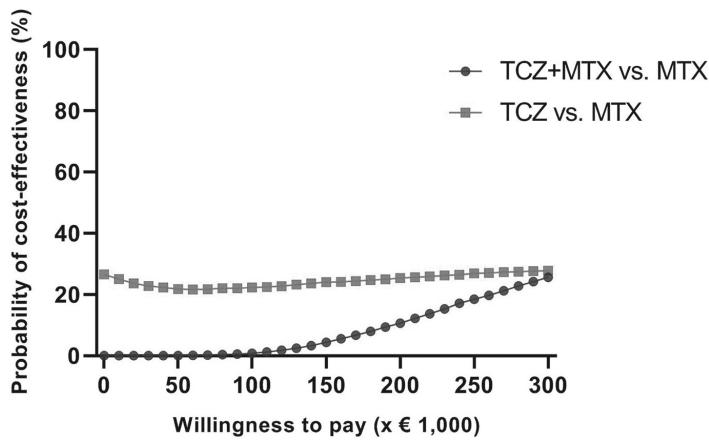


Figure 3 Cost effectiveness acceptability curve for the TCZ-based initiation treatment strategy groups versus the MTX initiation strategy group over 5 years, using the societal perspective (human capital approach).

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group.

Example: Using a willingness to pay of €50,000, the probability that TCZ+MTX or TCZ is more cost-effective compared to MTX is 0 or 20%, respectively.

Sensitivity and scenario analyses

Results of the sensitivity/scenario analyses are shown in Table 3. Overall results confirmed that the probability that TCZ strategies are cost-effective is low. The exception may be if a cost reduction of 30% for subcutaneously administrated TCZ is assumed. In this case the cost savings are considerable and may compensate for the limited QALY loss for TCZ compared to those of MTX (taking a societal perspective and using a human capital approach). TCZ is found to be dominant in 25% of the bootstrap samples, and in 49% less expensive, but also less effective in this case. Assuming a QALY gain of 0.05, the average QALY difference between TCZ and MTX become positive, and TCZ is dominated in 15% of all bootstrap samples, and in 11% TCZ was found less expensive, but also less effective. In the subgroup of patients with a DAS28>5.1 at start of treatment, TCZ is found to be dominant in 21% of the bootstrap samples, and in 24% less expensive, but still somewhat less effective (mean difference in QALY -0.01).

Outcomes based on undiscounted values are shown in Supplementary Table 6. Supplementary Figure 3 and 4 illustrate the CEACs at different willingness to pay, for our base case and all sensitivity analyses.

Table 3 Mean (2.5-97.5 percentile) difference in costs (€, rounded to the nearest hundreds) and QALYs, and ICER with distribution in cost effectiveness plane (%).

Analysis	Over 5 years	TCZ+MTX vs. MTX	TCZ vs. MTX
base case	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	35,600 (18,700 to 52,200) 0.06 (-0.10 to 0.22) 594,021 (0,0,23,77)	5,000 (-10,500 to 20,800) -0.03 (-0.20 to 0.13) -149,241 (8,18,47,27)
Sensitivity analyses			
complete case	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	30,400 (9,100 to 51,200) 0.01 (-0.14 to 0.16) 3,586,015 (0,0,46,54)	10,100 (-9,900 to 29,700) -0.16 (-0.32 to 0.00) -64,263 (0,16,81,3)
QALY +0.05 intervention	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	35,600 (18,700 to 52,200) 0.12 (-0.04 to 0.28) 296,907 (0,0,7,93)	5,000 (-10,500 to 20,800) 0.02 (-0.14 to 0.19) 194,208 (15,11,27,47)
QALY -0.05 intervention	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	35,600 (18,700 to 52,200) 0.00 (-0.16 to 0.16) -849,242,215 (0,0,50,50)	5,000 (-10,500 to 20,800) -0.09 (-0.25 to 0.07) -53,908 (3,24,63,10)
Discounting 4% for costs and QALYs	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	35,600 (18,700 to 52,200) 0.06 (-0.09 to 0.22) 595,474 (0,0,22,78)	5,000 (-10,500 to 20,800) -0.03 (-0.18 to 0.13) -186,561 (9,18,45,28)
Scenario analyses			
TCZ-SC	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	31,700 (14,900 to 48,000) 0.06 (-0.10 to 0.22) 528,655 (0,0,23,77)	-200 (-15,400 to 15,100) -0.03 (-0.20 to 0.13) 4,539 (16,35,30,19)
TCZ-SC -10%	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	30,500 (13,700 to 46,800) 0.06 (-0.10 to 0.22) 508,807 (0,0,23,77)	-1,700 (-16,700 to 13,400) -0.03 (-0.20 to 0.13) 51,233 (19,40,26,15)
TCZ-SC -30%	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	28,100 (11,400 to 44,500) 0.06 (-0.10 to 0.22) 468,847 (0,0,23,77)	-4,900 (-2,000 to 9,900) -0.03 (-0.20 to 0.13) 145,244 (25,49,16,10)
subgroup DAS28>5.1	Societal perspective [#] QALYs ICER [#] % (SE , SW, <u>NW</u> , NE)	34,400 (11,100 to 57,400) 0.07 (-0.17 to 0.30) 509,695 (0,0,29,71)	1,400 (-21,100 to 23,600) -0.01 (-0.25 to 0.24) -257,000 (21,24,28,27)

Outcomes based on single imputation nested in 10,000 bootstraps, costs expressed in euro's.

Table 3 (continued)

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; Societal perspective= direct healthcare costs + indirect non healthcare costs + productivity loss costs + medication costs; #= using human capital approach; QALY= quality-adjusted life years; ICER= incremental cost effectiveness ratio (using societal perspective according to human capital approach); **SE**= south east: gain in QALY, less expensive (i.e., TCZ dominant); SW= south west: loss in QALY, less expensive; NW= north west: loss in QALY, more expensive (i.e., TCZ inferior); NE= north east: gain in QALY, more expensive; SC= subcutaneous.

DISCUSSION

We hypothesized that initiating a TCZ-based strategy using a strict treat-to-target approach and including a clear tapering strategy when in sustained remission might become cost-effective. Our results do not confirm this hypothesis: we found that treat-to-target treatment strategies initiating TCZ as first treatment after diagnosis of RA are not cost-effective compared to a treatment strategy initiating MTX. Our results are in line with those of previous research in which the cost effectiveness of early initiation of bDMARDs compared to MTX initiation was analyzed in early (DMARD-naïve) RA patients.⁷

The observed decrease in medication use during U-Act-Early, however, did not decrease further in the PTFU.⁹ If a tapering protocol had been included in the PTFU, this might have led to a further decrease in TCZ use and might have increased cost effectiveness of TCZ strategies compared to MTX. Interestingly, our study showed that medication costs did decrease over time but remained higher for TCZ strategies over time, the only exception was the final (5 year) observation where medication cost became lower in both TCZ strategy groups compared to MTX. As specifically in year 5 total medication costs, productivity loss costs and direct healthcare costs are lower in TCZ-based treatment strategies, we hypothesize that this may be due to the fact that these patients need less intensive treatment due to their initial intensive treatment strategy, during the window of opportunity. However, to definitely establish this an evaluation over a longer time horizon would be needed. Furthermore, the differences between TCZ(+MTX) and MTX regarding productivity loss costs were lower than expected. Interestingly, productivity loss costs were saved only in the TCZ group, not in the TCZ+MTX group, compared to MTX. This may indicate that omitting MTX may be an advantage regarding productivity effects of treatment, as MTX is associated with (mild) adverse events.¹⁵ In line with this, also limited saving in overall healthcare costs and indirect non-healthcare costs were made in the TCZ group. We did not include presentism in these costs (i.e., costs of being less productive during working hours) which may have led to a higher impact of productivity on cost savings.

However, sensitivity/scenario analyses show that effect would need to be substantial to materially influence the results. Specifically, sensitivity analyses showed that only

with a combination of a substantial TCZ price reduction, and treatment of only the subgroup of early RA patients with a DAS28>5.1 at start of treatment, TCZ-based strategies may become cost-effective. Results of these analyses were based on the human capital approach, as it is internationally the most widely used method, and accounts most extensively for productivity effects in our sensitivity/scenario analyses, where the aim was to investigate if TCZ-based strategies could be cost-effective in some of the sensitivity/scenario analyses. No sensitivity/scenario analyses were performed for the type or costs of follow-up treatment in patients who dropped out because the number of patients dropping out as well as the duration of this follow-up treatment were similar between treatment strategy arms, indicating that these analyses would not have resulted in differences in costs regarding the comparison of the treatment strategies. Overall no differences in QALYs were shown in our study, which is in line with the majority of earlier performed studies as shown in a systematic review.⁷ The, unexpected, numerically lower number of QALYs obtained with TCZ versus MTX over 5 years, may be due to the fact that TCZ was not self-administered. The (additional) hospital visits for TCZ infusions may have been a burden for patients, resulting in QALY losses. We would expect a gain in QALY for early RA patients if treated with TCZ subcutaneously, compared to intravenously.¹⁶

The current study has several limitations. Firstly, there was considerable missing data, which we handled by resampling using a large number of bootstraps (i.e., 10,000) and imputation. This two-step method was previously described as an optimal way to handle uncertainty due to both sampling and missing of data.¹⁴ Secondly, our study had a reasonable short 5-year time horizon, whereas a life-long time horizon would be ideal to establish the final effect of short term intervention in early RA, given that it is a chronic disease. Although extension of the time horizon to life-time would have been possible with a health economic modeling approach (e.g., discrete event simulation, Markov modeling) using our patient data as part of the input, this was beyond the main aim of this paper, but would be interesting for future research. Thirdly, it was not possible to calculate the exact NSAID costs. We assume that our choice for naproxen, as one of the three most used NSAIDs in the Netherlands, with limited price differences, results in a plausible estimate of NSAID costs for this population in the Netherlands. Furthermore, the combination of NSAID and other csDMARD costs (i.e., other medication costs) are only maximally 1% of total medication costs. Therefore, it is unlikely that a more precise estimate of NSAID costs would influence our results.

Strengths of the current study are the availability of 'real-life' data on quality of life and resource use, preventing the need for models to estimate outcomes. Besides, the 317 patients included in the current analysis were treated according to the treat-to-target approach, which enhance the generalizability for RA patients in the Netherlands and other European countries.

Conclusion

Based on our analyses, early initiation of TCZ, with or without MTX, is not cost-effective compared to MTX initiation in a step-up treat-to-target treatment strategies over 2 or 5 years in early RA patients. Based on sensitivity analyses, subcutaneous administration of lower-priced TCZ, may be cost-effective compared to MTX in the subgroup of patients with high disease activity at start of treatment.

REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76**, 960–977 (2017).
2. Eriksson, J. K. *et al.* Cost-effectiveness of infliximab versus conventional combination treatment in methotrexate-refractory early rheumatoid arthritis: 2-year results of the register-enriched randomised controlled SWEFOT trial. *Ann. Rheum. Dis.* **74**, 1094–1101 (2015).
3. van der Hout, W. B. *et al.* Cost-utility analysis of treatment strategies in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum.* **61**, 291–299 (2009).
4. de Jong, P. H. P. *et al.* Best cost-effectiveness and worker productivity with initial triple DMARD therapy compared with methotrexate monotherapy in early rheumatoid arthritis: cost-utility analysis of the TREACH trial. *Rheumatology* **55**, 2138–2147 (2016).
5. Schipper, L. G. *et al.* Original article treatment strategies aiming at remission in early rheumatoid arthritis patients: starting with methotrexate monotherapy is cost-effective. *Rheumatology* **50**, 1320–1330 (2011).
6. Spalding, J. R. & Hay, J. Cost effectiveness of tumour necrosis factor- α Inhibitors as first-line agents in rheumatoid arthritis. *Pharmacoeconomics* **24**, 1221–1232 (2006).
7. Joensuu, J. T., Huoponen, S., Aaltonen, K. J. & Konttinen, Y. T. The cost-effectiveness of biologics for the treatment of rheumatoid arthritis: a systematic review. *PLoS Med.* e0119683 (2015)
8. Bijlsma, J. W. J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
9. Verhoeven, M. M. A. *et al.* Effectiveness and safety over 3 years after the 2-year U-Act-Early trial of the strategies initiating tocilizumab and / or methotrexate. *Rheumatology*, **59**, 2325–2333 (2019).
10. Institute for Medical Technology Assessment. *Kostenhandleiding*. (2016).
11. Pennington, B. & Davis, S. Mapping from the health assessment questionnaire to the EQ-5D : the impact of different algorithms on cost-effectiveness results. *Value Heal.* **17**, 762–771 (2014).
12. Zorginstituut Nederland. Medicijnkosten.nl.
13. Nair, S. C. *et al.* Economic evaluation of a tight-control treatment strategy using an imaging device (HandScan) for monitoring joint inflammation in early rheumatoid arthritis. *Clin. Exp. Rheumatol.* **33**, 831–836 (2015).
14. Brand, J., van Buuren, S., le Cessie, S. & van den Hout, W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. *Stat. Methods Med. Res.* **38**, 210–220 (2019).
15. Salliot, C. & van der Heijde, D. Long-term safety of methotrexate monotherapy in patients with rheumatoid arthritis: a systematic literature research. *Ann. Rheum. Dis.* **68**, 1100–1104 (2009).
16. Burmester, G. R. *et al.* Efficacy and safety of subcutaneous tocilizumab versus intravenous tocilizumab in combination with traditional DMARDs in patients with RA at week 97 (SUMMACTA). *Ann. Rheum. Dis.* **75**, 68–74 (2016).

SUPPLEMENTARY FILE

Supplementary Data S1

QALYs

Effectiveness was expressed as QALYs, measured at baseline, 3, 6, 12 and 24 months using the EuroQol 5-Dimensions 5-Levels (EQ-5D5L) with results expressed as a utility score. Utility is a score ranging from 0 (death) to 1 (full health), and was based on the Dutch tariff.¹⁰ For the post-trial period, during which EQ-5D5L was not measured, and in case of missing EQ-5D5L data, utility was estimated from HAQ and age, using the following formula; $EQ-5D5L = 0.82 - 0.179*HAQ - 0.019*HAQ^2 + 0.002*age$. This formula is based on a prediction model reported previously,¹ which was re-estimated in our own dataset using linear regression, to obtain optimal EQ-5D5L estimations for our study. QALYs were calculated as the area under the utility curve using EQ-5D5L measurements (year 1 and 2) or the estimated EQ-5D5L (year 3, 4 and 5) with linear interpolation.

¹= Pennington B, Davis S. Mapping from the Health Assessment Questionnaire to the EQ-5D : The Impact of Different Algorithms on Cost-Effectiveness Results. Value Health. 2014;17:762-71.

Supplementary Data S2

Missing data and data imputation

Costs (per category, over the last 3 months) were collected at baseline, 6, 12, 24, and 60 months, and in addition utility (and medication costs) at 3, 18, 36, 48 months. To obtain yearly costs and QALYs, linear interpolation was used over scheduled 3 monthly visits within a year. For the costs estimates over the 3rd and 4th year, interpolation between month 24 and 60 (and 36 and 48 for medication costs and utility) was used. Remaining missing information for costs and QALYs per year was considerable. During the trial phase, 12% (n=38) and 13% (n=41) of patients had missing information in at least one year for costs and QALYs, respectively, and during the PTFU period, on average 22% (n=50) and 20% (n=45) of patients/yearly estimates, respectively. As these missing might not be completely 'missing at random', we imputed these values using multivariable imputation from chained equations. To account for missing cost and QALY values in a year, as well as (population) uncertainty in outcomes, we used the following approach, which has been suggested to be optimal in this situation.¹ As first step, 10,000 bootstrap samples (with replacement) were taken per treatment strategy arm. In the second step, single imputation of the missing yearly QALY and cost estimates was performed per bootstrap sample. The variables DAS28, HAQ, utility, age, gender, rheumatoid factor status (RF), anti-cyclic citrullinated peptide status (anti-ccp), seropositivity for RF and/or anti-ccp, all at baseline, DAS28 and HAQ at year 1 and at year 2, and the estimates of QALYs and costs (per category) over all years (except for the specific variable being imputed) were used as predictors. This resulted in

10,000 bootstrap samples per treatment arm with complete data reflecting both sample uncertainty as well as the uncertainty due to data imputation.

¹= Brand J, van Buuren S, le Cessie S, van den Hout W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Stat Methods Med Res. 2019;38:210–20.

Supplementary Table 1 reference prices in €.

Productivity costs		
Average gross wage per hour		Amount
- male		38.78
- female		32.33
Medication costs		
Variable	Dose	Costs
Methotrexate	15 tablets of 2.5mg each	2.49
Tocilizumab ¹	1.5 vial of 20mg/ml	614.15
Tocilizumab	1 syringe of 0.9ml injection fluid	297.50
Hydroxychloroquine	50 tablets of 200mg each	6.32
Adalimumab	1 syringe of 0.4 ml injection fluid	570.92
Etanercept	1 syringe of 1ml injection fluid	229.59
Leflunomide	30 tablets of 10mg each	33.97
Sulfasalazine	60 tablets of 500mg each	3.69
Rituximab ²	15 vials of 10mg/ml	3790.45
Glucocorticoids (prednisolone)	60 tablets of 2.5 mg each	16.33
Glucocorticoids ³ (methylprednisolone)	7.5 vials of 40mg/ml	16.84
NSAIDs	2 tablets of 220mg each	0.30
Day admission	price 2009 (267€) * 1.10 = price 2014 price 2014 * 1.023 = price 2017	300.46
Day admission ≤1 hour	price 2009 (82€) * 1.10 = price 2014 price 2014 * 1.023 = price 2017	92.27
co-medication rituximab	- 1000 mg paracetamol (orally) - 2 mg tavegil (iv) - 100 mg methylprednisolone (Solu-Medrol) iv.	4.44
Direct healthcare costs		
Variable	Frequency	Costs
General day admission	per day	476
X-ray, echo	per time	83.50
MRI-scan	per time	222
CT-scan	per time	142.5
general surgery	per time	1300
Hospital day, academic center	per day	643
Hospital day, general hospital	per day	443

Supplementary Table 1 (continued)

Hospital day, rehabilitation center	per day	460
Hospital day, nursing home	per day	168
Hospital day, intensive care	per day	1600.5
Specialist	per visit	97.25
Specialist in training	per visit	121.5
Psychotherapist	per visit	94
General practitioner	per visit	66
Social worker	per visit	65
Paramedical care	per visit	35
Indirect non healthcare costs		
Variable	Frequency	costs
Care at home	per hour	50
Home help	per hour	23

¹Additinoal costs for day admission, ² Additional costs for day admission and co-medication, ³if intra articular + €9.92 and if intra muscular + €2.75

Supplementary Table 2 Baseline characteristics of U-Act-Early patients (not) included in the post-trial follow-up.

	TCZ+MTX (n=75)	TCZ (n=79)	MTX (n=72)	P-value	Included in follow- up n=226	Not included in follow- up n=91	P-value
Employed, n (%)	60 (80)	60 (76)	56 (78)	0.83	176 (78)	75 (82)	0.37
Working hours per week, mean (SD)	23.3 (15.2)	21.1 (16.0)	22.6 (16.8)	0.49	22.3 (16.0)	26.7 (13.9)	0.03
Female, n (%)	47 (63)	61 (77)	48 (67)	0.13	156 (69)	56 (62)	0.20
RF +, n (%)	53 (71)	55 (71)	58 (82)	0.21	166 (74)	63 (69)	0.38
Anti CCP +, n (%)	51 (68)	54 (69)	56 (79)	0.28	161 (74)	62 (68)	0.51
RF- and/or anti-CCP+, n (%)	55 (73)	62 (79)	64 (89)	0.06	181 (80)	68 (75)	0.26
Age (years), mean (SD)	53.8 (11.2)	55.5 (11.6)	53.7 (12.9)	0.63	54.3 (11.9)	51.1 (14.8)	0.09
Symptom duration (days), median (IQR)	27 (18-43)	25 (19-43)	28 (16-46)	0.96	26 (18-43)	27 (16-46)	0.83
DAS28, mean (SD)	5.1 (1.1)	5.3 (1.1)	5.0 (1.2)	0.33	5.2 (1.1)	5.4 (1.1)	0.25
HAQ, median (IQR)	1.3 (0.6-1.6)	1.3 (0.9-1.8)	1.0 (0.5-1.4)	0.17	1.1 (0.6-1.6)	1.1 (0.8-1.5)	0.98

TCZ+MTX= initial tocilizumab + methotrexate strategy group; TCZ= initial tocilizumab + placebo-methotrexate strategy group; MTX= initial methotrexate + placebo-tocilizumab strategy group; RF= rheumatoid factor, + = positive; anti-CCP= anti-cyclic citrullinated peptide antibodies; SD= standard deviation; IQR= interquartile range; DAS28= disease activity score assessing 28 joints; HAQ= health assessment questionnaire score.

Supplementary Table 3 Cost (€, rounded to the nearest hundreds) and QALYs over time per initial treatment strategy in means (2.5-97.5 percentiles), using discounted values.

	TCZ+MTX (n=106)	TCZ (n=103)	MTX (n=108)
Total medication costs			
- year 1	17,900 (17,000 to 18,800)	18,400 (17,500 to 19,200)	4,400 (3,500 to 5,300)
- year 2	10,000 (8,800 to 11,300)	10,700 (9,500 to 12,000)	8,800 (7,400 to 10,100)
- year 3	8,100 (7,100 to 9,100)	8,700 (7,700 to 9,700)	7,300 (6,200 to 8,300)
- year 4	4,800 (4,200 to 5,400)	5,300 (4,700 to 5,900)	4,800 (4,000 to 5,600)
- year 5	300 (200 to 500)	700 (300 to 1,100)	1,400 (800 to 2,100)
- over 2 years	27,900 (26,100 to 29,800)	29,100 (27,200 to 30,900)	13,200 (11,100 to 15,200)
- over 5 years	41,200 (37,800 to 44,600)	43,700 (40,400 to 47,200)	26,600 (22,500 to 30,700)
TCZ costs			
- year 1	16,900 (15,800 to 18,000)	17,900 (17,000 to 18,800)	2,600 (1,900 to 3,400)
- year 2	7,100 (5,700 to 8,500)	8,700 (7,300 to 10,100)	5,500 (4,200 to 6,900)
- year 3	5,700 (4,600 to 6,900)	7,100 (6,000 to 8,200)	4,700 (3,600 to 5,800)
- year 4	3,400 (2,700 to 4,000)	4,300 (3,600 to 5,000)	3,300 (2,500 to 4,000)
- year 5	100 (0 to 300)	500 (200 to 900)	1,400 (800 to 2,000)
- over 2 years	24,000 (21,700 to 26,300)	26,600 (24,500 to 28,800)	8,100 (6,200 to 10,100)
- over 5 years	33,300 (29,300 to 37,200)	38,500 (34,600 to 42,500)	17,400 (13,500 to 21,500)
Other bDMARD costs			
- year 1	900 (500 to 1,500)	400 (100 to 800)	1,600 (900 to 2,400)
- year 2	2,900 (1,800 to 3,900)	1,900 (1,100 to 2,900)	3,200 (2,200 to 4,300)
- year 3	2,300 (1,500 to 3,200)	1,600 (900 to 2,300)	2,600 (1,700 to 3,500)
- year 4	1,400 (900 to 1,900)	1,000 (600 to 1,400)	1,500 (1,000 to 2,000)
- year 5	200 (0 to 300)	200 (0 to 400)	0 (0 to 0)
- over 2 years	3,800 (2,400 to 5,300)	2,300 (1,300 to 3,500)	4,800 (3,200 to 6,700)
- over 5 years	7,700 (4,900 to 10,600)	5,100 (2,900 to 7,400)	8,900 (6,000 to 12,100)
Other medication costs			
- year 1	100 (100 to 100)	100 (100 to 100)	100 (100 to 200)
- year 2	0 (0 to 0)	0 (0 to 0)	100 (0 to 100)
- year 3	0 (0 to 0)	0 (0 to 0)	100 (0 to 100)
- year 4	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)
- year 5	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)
- over 2 years	100 (100 to 200)	100 (100 to 100)	200 (200 to 200)
- over 5 years	200 (200 to 200)	200 (100 to 200)	300 (300 to 300)
Total productivity loss costs*			
- year 1	6,700 (5,500 to 8,000)	5,600 (4,300 to 7,000)	6,500 (5,200 to 7,800)
- year 2	19,300 (15,700 to 23,100)	14,700 (11,300 to 18,300)	17,100 (13,600 to 20,700)
- year 3	15,600 (12,700 to 18,600)	11,900 (9,100 to 14,800)	13,800 (11,000 to 16,800)
- year 4	9,300 (7,500 to 11,100)	7,100 (5,500 to 8,900)	8,300 (6,600 to 10,100)
- year 5	700 (300 to 1,100)	600 (100 to 1,200)	800 (400 to 1,300)
- over 2 years	26,100 (21,300 to 31,000)	20,300 (15,700 to 25,100)	23,600 (18,900 to 28,400)
- over 5 years	51,700 (42,100 to 61,600)	39,900 (30,800 to 49,600)	46,500 (37,100 to 56,200)

Supplementary Table 3 (continued)

	TCZ+MTX (n=106)	TCZ (n=103)	MTX (n=108)
Total productivity loss costs^s			
- year 1	2,500 (2,100 to 3,000)	2,300 (1,800 to 2,800)	2,500 (2,100 to 3,000)
- year 2	3,500 (2,800 to 4,300)	3,100 (2,400 to 3,800)	3,200 (2,500 to 3,900)
- year 3	2,900 (2,400 to 3,500)	2,500 (2,000 to 3,100)	2,600 (2,100 to 3,200)
- year 4	1,900 (1,500 to 2,300)	1,700 (1,300 to 2,100)	1,700 (1,300 to 2,000)
- year 5	500 (200 to 900)	500 (100 to 1,100)	400 (200 to 600)
- over 2 years	6,100 (5,000 to 7,200)	5,300 (4,300 to 6,500)	5,700 (4,600 to 6,800)
- over 5 years	11,400 (9,300 to 13,700)	10,100 (8,000 to 12,400)	10,300 (8,300 to 12,500)
Direct healthcare costs			
- year 1	6,100 (5,500 to 6,700)	7,200 (6,200 to 8,500)	7,000 (6,200 to 7,900)
- year 2	18,000 (14,200 to 21,900)	6,700 (5,500 to 8,200)	8,400 (6,100 to 10,900)
- year 3	15,200 (12,200 to 18,300)	6,400 (5,400 to 7,500)	7,600 (5,800 to 9,600)
- year 4	10,600 (8,900 to 12,400)	6,000 (5,300 to 6,700)	6,500 (5,400 to 7,600)
- year 5	4,500 (3,900 to 5,000)	5,400 (4,900 to 5,900)	5,000 (4,600 to 5,400)
- over 2 years	24,100 (20,300 to 28,000)	14,000 (12,300 to 15,900)	15,400 (13,000 to 18,200)
- over 5 years	54,400 (45,800 to 63,200)	31,800 (28,400 to 35,500)	34,500 (29,200 to 40,400)
Indirect non-healthcare costs			
- year 1	1,100 (800 to 1,400)	1,600 (1,000 to 2,300)	1,500 (1,100 to 2,000)
- year 2	300 (200 to 500)	600 (500 to 800)	500 (400 to 700)
- year 3	700 (600 to 800)	900 (800 to 1,100)	1,200 (1,000 to 1,300)
- year 4	1,400 (1,300 to 1,500)	1,600 (1,400 to 1,700)	2,500 (2,200 to 2,800)
- year 5	2,400 (2,200 to 2,500)	2,400 (2,100 to 2,600)	4,300 (3,700 to 4,800)
- over 2 years	1,400 (1,100 to 1,800)	2,200 (1,600 to 3,100)	2,100 (1,600 to 2,600)
- over 5 years	5,900 (5,400 to 6,500)	7,100 (6,300 to 8,000)	10,000 (8,800 to 11,200)
QALYs			
- year 1	0.69 (0.64 to 0.73)	0.70 (0.67 to 0.74)	0.64 (0.59 to 0.68)
- year 2	0.82 (0.80 to 0.84)	0.78 (0.74 to 0.81)	0.82 (0.79 to 0.84)
- year 3	0.74 (0.69 to 0.78)	0.75 (0.71 to 0.78)	0.72 (0.68 to 0.77)
- year 4	0.72 (0.68 to 0.76)	0.70 (0.65 to 0.74)	0.74 (0.70 to 0.77)
- year 5	0.70 (0.66 to 0.74)	0.66 (0.61 to 0.70)	0.70 (0.66 to 0.74)
- over 2 years	1.51 (1.45 to 1.57)	1.48 (1.41 to 1.54)	1.45 (1.41 to 1.50)
- over 5 years	3.67 (3.55 to 3.79)	3.58 (3.45 to 3.70)	3.61 (3.50 to 3.72)

Outcomes based on single imputation nested in 10,000 bootstraps; costs expressed in euro's.
 TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; total medication costs= all RA medication costs; other medication costs= all RA medication costs, without bDMARDs; productivity loss costs= costs related to work loss or being less productive; #= using human capital approach; ^s= using friction cost approach, only counting costs for a period of absence up to 85 days; direct healthcare costs= all costs related to healthcare, also for other diseases (RA medication costs excluded); indirect non healthcare costs= costs related to patient and family (e.g., travel cost, buying stair lift, etc.); QALY= quality-adjusted life years [range 0-1, 1= full health].

Supplementary Table 4 Cost (€, rounded to the nearest hundreds) and QALYs over time per initial treatment strategy in means (2.5-97.5 percentiles).

	TCZ+MTX (n=106)	TCZ (n=103)	MTX (n=108)
Total medication costs			
- year 1	17,900 (17,000 to 18,800)	18,400 (17,500 to 19,200)	4,400 (3,500 to 5,300)
- year 2	10,400 (9,100 to 11,800)	11,100 (9,900 to 12,500)	9,100 (7,700 to 10,500)
- year 3	8,800 (7,700 to 9,900)	9,400 (8,300 to 10,500)	7,900 (6,700 to 9,100)
- year 4	5,400 (4,700 to 6,100)	6,000 (5,300 to 6,700)	5,400 (4,500 to 6,200)
- year 5	400 (200 to 600)	800 (400 to 1,300)	1,600 (900 to 2,400)
- over 5 years	42,900 (39,300 to 46,600)	45,700 (42,100 to 49,300)	28,400 (24,000 to 32,800)
- over 2 years	28,300 (26,500 to 30,300)	29,500 (27,600 to 31,400)	13,500 (11,400 to 15,600)
TCZ costs			
- year 1	16,900 (15,800 to 18,000)	17,900 (17,000 to 18,800)	2,600 (1,900 to 3,400)
- year 2	7,400 (6,000 to 8,900)	9,100 (7,600 to 10,500)	5,700 (4,300 to 7,200)
- year 3	6,200 (5,000 to 7,400)	7,700 (6,500 to 8,900)	5,000 (3,900 to 6,300)
- year 4	3,800 (3,100 to 4,500)	4,800 (4,100 to 5,600)	3,700 (2,800 to 4,500)
- year 5	200 (100 to 300)	600 (200 to 1,000)	1,600 (900 to 2,400)
- over 5 years	34,500 (30,300 to 38,700)	40,100 (35,9 to 44,300)	18,600 (14,400 to 23,000)
- over 2 years	24,300 (22,000 to 26,600)	27,000 (24,8 to 29,200)	8,400 (6,400 to 10,400)
Other bDMARD costs			
- year 1	900 (500 to 1,500)	400 (100 to 800)	1,600 (900 to 2,400)
- year 2	3,000 (1,900 to 4,100)	2,000 (1,100 to 3,000)	3,300 (2,300 to 4,500)
- year 3	2,500 (1,600 to 3,500)	1,700 (1,000 to 2,500)	2,800 (1,900 to 3,800)
- year 4	1,600 (1,000 to 2,200)	1,100 (600 to 1,600)	1,700 (1,100 to 2,300)
- year 5	200 (100 to 400)	200 (0 to 400)	0 (0 to 0)
- over 5 years	8,200 (5,300 to 11,300)	5,400 (3,100 to 7,900)	9,400 (6,300 to 12,800)
- over 2 years	3,900 (2,500 to 5,500)	2,400 (1,300 to 3,600)	5,000 (3,300 to 6,800)
Other medication costs			
- year 1	100 (100 to 100)	100 (100 to 100)	100 (100 to 200)
- year 2	0 (0 to 100)	0 (0 to 0)	100 (100 to 100)
- year 3	0 (0 to 0)	0 (0 to 0)	100 (0 to 100)
- year 4	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)
- year 5	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)
- over 5 years	200 (200 to 200)	200 (200 to 200)	300 (300 to 400)
- over 2 years	100 (100 to 200)	100 (100 to 100)	200 (200 to 200)
Total productivity loss costs*			
- year 1	6,700 (5,500 to 8,000)	5,600 (4,300 to 7,000)	6,500 (5,200 to 7,800)
- year 2	20,100 (16,400 to 24,000)	15,300 (11,700 to 19,000)	17,800 (14,200 to 21,500)
- year 3	16,900 (13,700 to 20,200)	12,900 (9,900 to 16,000)	15,000 (11,900 to 18,100)
- year 4	10,500 (8,500 to 12,500)	8,000 (6,100 to 10,000)	9,300 (7,400 to 11,400)
- year 5	800 (400 to 1,300)	700 (200 to 1,400)	900 (400 to 1,500)
- over 5 years	55,000 (44,700 to 65,600)	42,500 (32,800 to 52,800)	49,500 (39,500 to 59,800)
- over 2 years	26,800 (21,900 to 32,000)	20,900 (16,200 to 25,900)	24,200 (19,500 to 29,200)

Supplementary Table 4 (continued)

	TCZ+MTX (n=106)	TCZ (n=103)	MTX (n=108)
Total productivity loss costs^s			
- year 1	2,500 (2,100 to 3,000)	2,300 (1,800 to 2,800)	2,500 (2,100 to 3,000)
- year 2	3,700 (3,000 to 4,400)	3,200 (2,500 to 3,900)	3,300 (2,600 to 4,000)
- year 3	3,200 (2,500 to 3,800)	2,800 (2,100 to 3,400)	2,800 (2,200 to 3,400)
- year 4	2,100 (1,700 to 2,600)	1,900 (1,400 to 2,400)	1,900 (1,500 to 2,300)
- year 5	600 (300 to 1,100)	600 (1,100 to 1,300)	400 (200 to 700)
- over 5 years	12,100 (9,600 to 14,600)	10,700 (8,400 to 13,200)	10,900 (8,700 to 13,200)
- over 2 years	6,200 (5,100 to 7,300)	5,500 (4,400 to 6,600)	5,800 (4,700 to 6,900)
Direct healthcare costs			
- year 1	6,100 (5,500 to 6,700)	7,200 (6,200 to 8,500)	7,000 (6,200 to 7,900)
- year 2	18,700 (14,800 to 22,800)	7,000 (5,700 to 8,500)	8,700 (6,300 to 11,400)
- year 3	16,500 (13,200 to 19,800)	6,900 (5,800 to 8,200)	8,200 (6,300 to 10,400)
- year 4	12,000 (10,000 to 14,000)	6,700 (6,000 to 7,600)	7,300 (6,100 to 8,600)
- year 5	5,200 (4,600 to 5,900)	6,300 (5,700 to 6,900)	5,900 (5,400 to 6,300)
- over 5 years	58,400 (49,200 to 67,900)	34,200 (30,600 to 38,200)	37,100 (31,500 to 43,500)
- over 2 years	24,800 (20,900 to 31,400)	14,300 (12,500 to 16,200)	15,700 (13,200 to 18,600)
Indirect non-healthcare costs			
- year 1	1,100 (800 to 1,400)	1,600 (1,000 to 2,300)	1,500 (1,100 to 2,000)
- year 2	300 (200 to 500)	700 (500 to 900)	600 (400 to 700)
- year 3	700 (600 to 900)	1,000 (900 to 1,200)	1,300 (1,100 to 1,400)
- year 4	1,600 (1,400 to 1,700)	1,800 (1,600 to 1,900)	2,800 (2,500 to 3,100)
- year 5	2,800 (2,500 to 3,000)	2,800 (2,500 to 3,100)	5,000 (4,300 to 5,700)
- over 5 years	6,500 (6,000 to 7,100)	7,900 (7,000 to 8,800)	11,100 (9,800 to 1,500)
- over 2 years	1,500 (1,100 to 1,800)	2,300 (1,600 to 3,100)	2,100 (1,600 to 2,600)
QALYs			
- year 1	0.69 (0.64 to 0.73)	0.70 (0.67 to 0.74)	0.64 (0.59 to 0.68)
- year 2	0.84 (0.81 to 0.86)	0.79 (0.75 to 0.82)	0.83 (0.80 to 0.85)
- year 3	0.76 (0.71 to 0.81)	0.77 (0.73 to 0.80)	0.74 (0.70 to 0.79)
- year 4	0.76 (0.71 to 0.80)	0.73 (0.68 to 0.77)	0.77 (0.73 to 0.80)
- year 5	0.74 (0.70 to 0.78)	0.69 (0.65 to 0.74)	0.74 (0.70 to 0.79)
- over 5 years	3.78 (3.66 to 3.91)	3.69 (3.56 to 3.81)	3.72 (3.61 to 3.84)
- over 2 years	1.52 (1.46 to 1.58)	1.49 (1.43 to 1.56)	1.47 (1.42 to 1.51)

Outcomes based on single imputation nested in 10,000 bootstraps; costs expressed in euro's. Undiscounted values.

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; total medication costs= all RA medication costs; other medication costs= all RA medication costs, without bDMARDs; productivity loss costs= costs related to work loss or being less productive; ^= using human capital approach; \$= using friction cost approach, only counting costs for a period of absence up to 85 days; direct healthcare costs= all costs related to healthcare, also for other diseases (RA medication costs excluded); indirect non healthcare costs= costs related to patient and family (e.g., travel cost, buying stair lift, etc.); QALY= quality-adjusted life years [range 0-1, 1= full health].

Supplementary Table 5 Mean (2.5-97.5 percentile) difference in costs (€, rounded to the nearest hundreds) and QALYs, and ICER with distribution in cost effectiveness plane (%).

	TCZ+MTX vs. MTX	TCZ vs. MTX
Over 2 years		
Medication costs	14,800 (12,000 to 17,700)	16,000 (13,100 to 18,800)
Direct healthcare costs	9,000 (4,300 to 14,000)	-1,500 (-4,800 to 1,800)
Indirect non-healthcare costs	-600 (-1,700 to 0)	200 (-700 to 1,100)
Productivity loss costs [#]	2,600 (-4,300 to 9,600)	-3,300 (-10,200 to 3,500)
Productivity loss costs ^{\$}	0 (-1,100 to 2,000)	-300 (-1,900 to 1,200)
Total cost and effect		
Healthcare perspective	23,900 (18,300 to 29,600)	14,500 (9,900 to 19,000)
Societal perspective [#]	25,900 (17,400 to 34,100)	11,400 (3,300 to 19,400)
Societal perspective ^{\$}	23,700 (18,000 to 29,300)	14,400 (9,400 to 19,200)
QALYs	0.05 (-0.02 to 0.12)	0.03 (-0.06 to 0.11)
ICER ^{HC}	432.048	552.199
% (SE , SW, NW, NE)	(0,0,7,93)	(0,0,26,74)
ICER [#]	467.311	431.754
% (SE , SW, NW, NE)	(0,0,7,93)	(0,0,26,74)
ICER ^{\$}	427.920	546.065
% (SE , SW, NW, NE)	(0,0,7,93)	(0,0,26,74)
Over 5 years		
Medication costs	14,500 (8,800 to 20,200)	17,300 (11,500 to 22,900)
Direct healthcare costs	21,400 (10,400 to 32,400)	-2,800 (-10,100 to 4,100)
Indirect non-healthcare costs	-4,600 (-6,100 to -3,200)	-3,300 (-4,900 to -1,700)
Productivity loss costs [#]	5,500 (-8,900 to 20,000)	-7,000 (-21,200 to 7,200)
Productivity loss costs ^{\$}	1,200 (-2,000 to 4,500)	-200 (-3,500 to 3,100)
Total cost and effect		
Healthcare perspective	35,900 (22,700 to 48,800)	14,500 (4,400 to 24,200)
Societal perspective [#]	36,800 (18,700 to 54,600)	4,200 (-12,500 to 20,900)
Societal perspective ^{\$}	32,600 (19,800 to 45,500)	11,000 (600 to 21,100)
QALYs	0.06 (-0.11 to 0.23)	-0.04 (-0.21 to 0.14)
ICER ^{HC}	598,185	-377,128
% (SE , SW, NW, NE)	(0,0,24,76)	(0,0,67,33)
ICER [#]	613,487	-108,609
% (SE , SW, NW, NE)	(0,0,24,76)	(10,22,44,24)
ICER ^{\$}	542,200	-287,097
% (SE , SW, NW, NE)	(0,0,24,76)	(1,165,33)

Outcomes based on single imputation nested in 10,000 bootstraps; costs expressed in euro's. Undiscounted values.

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; Medication costs= all RA medication costs; Direct healthcare costs= all costs related to healthcare, also for other diseases (RA medication costs excluded); Indirect non healthcare costs= costs related to patient and family (e.g., travel cost, buying stair lift, etc.); Productivity loss costs= costs related to work loss or

Supplementary Table 5 (continued)

being less productive; # = using human capital approach; \$ = using friction cost approach, only counting costs for a period of absence up to 85 days; Healthcare perspective= direct healthcare costs + medication costs; Societal perspective= direct healthcare costs + indirect not healthcare costs + productivity loss costs + medication costs; QALY= quality-adjusted life years; ICER= incremental cost effectiveness ratio (using societal perspective according to human capital approach); HC= healthcare perspective; **SE**= south east: gain in QALY, less expensive (i.e., TCZ dominant); SW= south west: loss in QALY, less expensive; NW= north west: loss in QALY, more expensive (i.e., TCZ inferior); NE= north east: gain in QALY, more expensive.

Supplementary Table 6 Mean (2.5-97.5 percentile) difference in costs (€, rounded to the nearest hundreds) and QALYs with, and ICER with distribution in cost effectiveness plane (%).

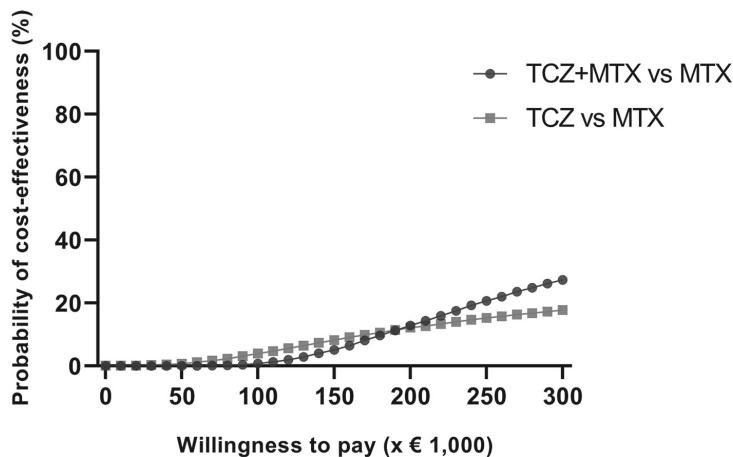
Analysis	Over 5 years	TCZ+MTX vs. MTX	TCZ vs. MTX
base case	Societal perspective# 36,800 (18,700 to 54,600) QALYs 0.06 (-0.11 to 0.23) ICER# 613,487 % (SE , SW, <u>NW</u> , NE) (0,0,24,76)	4,200 (-12,500 to 20,900) -0.04 (-0.21 to 0.14) -108,609 (10,22,44,24)	
Sensitivity analyses			
complete case	Societal perspective# 31,000 (8,200 to 53,200) QALYs 0.02 (-0.14 to 0.18) ICER# 2,127,794 % (SE , SW, <u>NW</u> , NE) (0,0,43,57)	9,400 (-11,900 to 30,400) -0.14 (-0.30 to 0.02) -62,967 (1,19,77,3)	
QALY +0.05 intervention	Societal perspective# 36,800 (18,700 to 54,600) QALYs 0.12 (-0.05 to 0.29) ICER# 306,841 % (SE , SW, <u>NW</u> , NE) (0,0,8,92)	4,200 (-12,500 to 20,900) 0.02 (-0.15 to 0.20) 192,750 (17,15,26,42)	
QALY -0.05 intervention	Societal perspective# 36,800 (18,700 to 54,600) QALYs 0.00 (-0.17 to 0.17) ICER# 961,430,116 % (SE , SW, <u>NW</u> , NE) (0,0,50,50)	4,200 (-12,500 to 20,900) -0.10 (-0.27 to 0.08) -42,368 (3,28,59,10)	
Scenario analyses			
TCZ-SC	Societal perspective# 32,900 (14,900 to 50,400) QALYs 0.06 (-0.11 to 0.23) ICER# 548,356 % (SE , SW, <u>NW</u> , NE) (0,0,24,76)	-1,100 (-17,300 to 15,200) -0.04 (-0.21 to 0.14) 29,373 (17,38,28,17)	
TCZ-SC -10%	Societal perspective# 31,700 (13,700 to 49,200) QALYs 0.06 (-0.11 to 0.23) ICER# 528,580 % (SE , SW, <u>NW</u> , NE) (0,0,24,76)	-2,700 (-48,700 to 13,400) -0.04 (-0.21 to 0.14) 71,271 (20,43,23,14)	

Supplementary Table 6 (continued)

Analysis	Over 5 years	TCZ+MTX vs. MTX	TCZ vs. MTX
TCZ-SC -30%	Societal perspective# 29,300 (11,400 to 46,800)	-6,000 (-21,800 to 9,900)	
	QALYs 0.06 (-0.11 to 0.23)	-0.04 (-0.21 to 0.14)	
	ICER# 488,763	155,624	
	% (SE , SW, <u>NW</u> , NE) (0,0,24,76)	(25,52,14,9)	
subgroup DAS28>5.1	Societal perspective# 35,500 (10,600 to 60,000)	200 (-23,800 to 23,900)	
	QALYs 0.07 (-0.17 to 0.31)	-0.01 (-0.26 to 0.25)	
	ICER# 513,598	-24,099	
	% (SE , SW, <u>NW</u> , NE) (0,0,29, 71)	(23,26,26,25)	

Outcomes based on single imputation nested in 10,000 bootstraps; costs expressed in euro's.

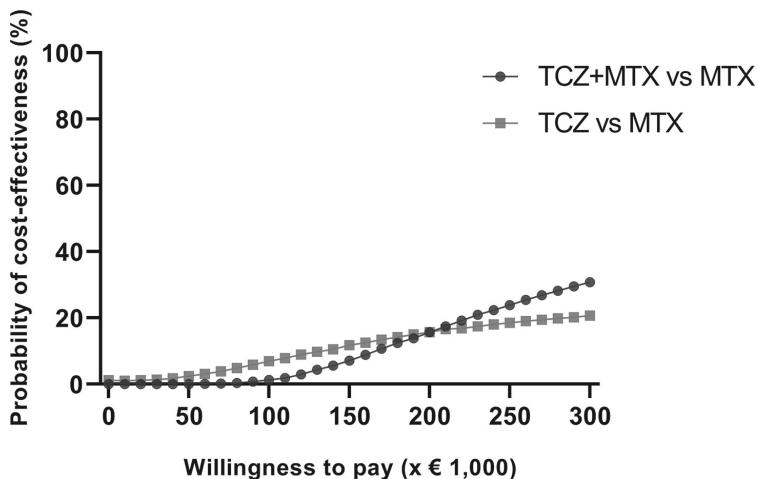
TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group; Societal perspective= direct healthcare costs + indirect not healthcare costs + productivity loss costs + medication costs; # = using human capital approach; QALY= quality-adjusted life years; ICER= incremental cost effectiveness ratio (using societal perspective according to human capital approach); **SE**= south east: gain in QALY, less expensive; SW= south west: Loss in QALY, less expensive; NW= north west: Loss in QALY, more expensive; NE= north east: gain in QALY, more expensive; SC= subcutaneous; IV= intravenous.



Supplementary Figure 1 Cost effectiveness acceptability curve for the TCZ based initiation treatment strategy groups versus the MTX initiation strategy group over 5 years, using the healthcare perspective.

TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group.

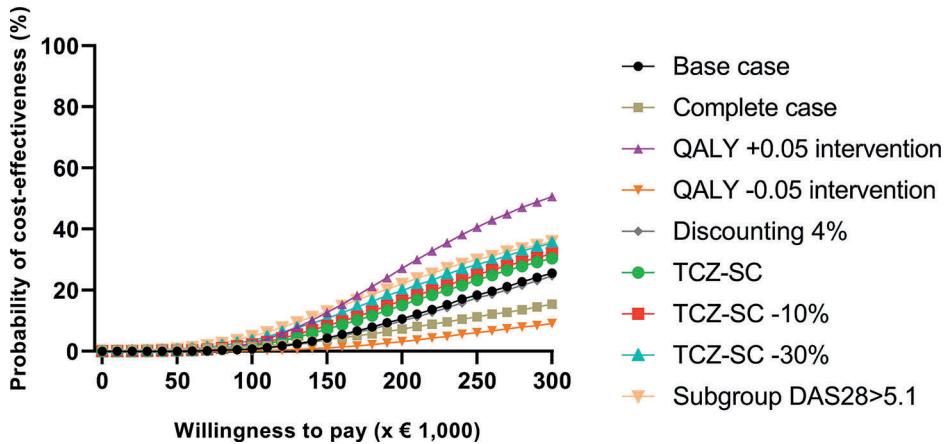
Example: Using a willingness to pay of €50,000, the probability that TCZ+MTX or TCZ is more cost-effective compared to MTX is 0 or 0%, respectively.



Supplementary Figure 2 Cost effectiveness acceptability curve for the TCZ based initiation treatment strategy groups versus the MTX initiation strategy group over 5 years, using the societal perspective (friction cost approach).

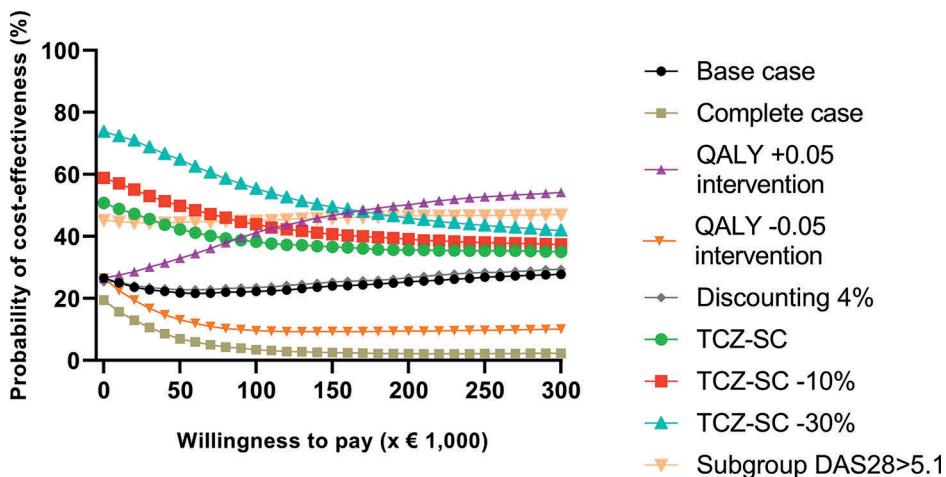
TCZ+MTX= initiation of tocilizumab + methotrexate strategy group; TCZ= initiation of tocilizumab + placebo-methotrexate strategy group; MTX= initiation of methotrexate + placebo-tocilizumab strategy group.

Example: Using a willingness to pay of €50,000, the probability that TCZ+MTX or TCZ is more cost-effective compared to MTX is 0 or 2%, respectively.



Supplementary Figure 3 Cost effectiveness acceptability curve, for base case and per sensitivity analysis, for the TCZ+MTX initiation treatment strategy group versus the MTX initiation strategy group over 5 years, using the societal perspective (human capital approach).

Base case= using data of all patients included in U-Act-Early; Complete case= using data of all patients also included in the post-trial follow-up; TCZ-SC= using TCZ subcutaneously; -10%= price reduction of 10%; -30%= price reduction of 30%.



Supplementary Figure 4 Cost effectiveness acceptability curve, for base case and per sensitivity analysis, for the TCZ initiation treatment strategy group versus the MTX initiation strategy group over 5 years, using the societal perspective (human capital approach).

Base case= using data of all patients included in U-Act-Early; Complete case= using data of all patients also included in the post-trial follow-up; TCZ-SC= using TCZ subcutaneously; -10%= price reduction of 10%; -30%= price reduction of 30%.

Chapter 8:

**Unravelling the cost of biological strategies
in rheumatoid arthritis: a kaleidoscope of
methodologies, interpretations and interests.**

J.W.G. Jacobs¹

M.M.A. Verhoeven¹

P.M.J. Welsing¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands.

In press, editorial in The Journal of Rheumatology December 2020

EDITORIAL

In this issue, Müskens et al. describes that after introduction on a Dutch rheumatology department of an etanercept biosimilar as substitute of the more expensive biological etanercept, the accumulated 3-monthly anti-rheumatic medication cost for patients in that hospital, mainly consisting of cost of biological DMARD (bDMARD), decreased, as expected. However, this financial advantage was lost within less than a year, due to an increase of the percentage of RA patients treated with a bDMARD. This means that the potential savings of using the biosimilar were spent on extra patients treated with a bDMARD, although the rheumatologists had not formally changed their bDMARD prescription policy. The brisk increase in percentage of patients treated with a bDMARD in this time period is not compatible with the general trend of slowly increasing bDMARD use over time.

Should the reader of the paper thus conclude that introduction of cheaper biosimilars is not effective in reducing medication cost on the longer term? Our answer would be that interpretations of this, and of any costing study, strongly depend on what we are looking at, how we are looking, and who is looking.

What we are looking at: treatment strategy

Müskens et al. found no statistically significant difference in disease activity in those starting a biological before the biosimilar introduction (mean disease activity score assessing 28 joints (DAS28) 4.7), versus in those starting a bDMARD after the biosimilar introduction (DAS28 4.5). After the biosimilar introduction, the percentage of bDMARD patients using concomitantly methotrexate (MTX) dropped from 68 to 54. Notably, the mean age of patients at start of bDMARD before the biosimilar introduction was statistically significantly lower than that after the biosimilar introduction, 52 versus 58 years, respectively. The most plausible explanation seems to be that after the introduction of the cheaper biosimilar, the rheumatologists in this centre felt more free to initiate bDMARD, also in elderly patients with adverse effects of conventional synthetic DMARD (csDMARD), to allow for stopping this csDMARD.

Then, the question arises whether this increased use of bDMARD and the resulting loss of financial benefit would also have occurred if stricter guidelines on bDMARD initiation and usage, e.g., on concomitant MTX usage and dose, had been applied. Perhaps the main conclusion of the paper is that persistent financial benefit in their study would have required more stable adherence to their prescription policy.

The study of Müskens can be regarded as a budget-impact analysis, i.e., an economic assessment that estimates the financial consequences of adopting a new intervention, but analysing only medication cost, not financial consequences e.g., of the intervention's impact on frequency of clinical visits and admissions. Moreover, not only

the (medication) cost perspective, but also the clinical perspective, could be relevant. What were the clinical effects of treating the extra patients with a bDMARD? Could the potential savings of using the biosimilar not have been more (cost-)effective spent on another treatment strategy? Take for instance a strategy with initiation of a bDMARD, preferably a biosimilar, next to MTX, in early RA, using the bDMARD for a limited period (thus limiting medication cost), targeting for drug free remission. This seems an attainable goal, given the results of the U-Act-Early trial.¹ In this study, approximately 30% of the early RA patients who started with a bDMARD (tocilizumab) as first treatment shortly after diagnosis achieved this target, whereas this was only 10% in the strategy group initiating MTX as first treatment.¹ The cost effectiveness analysis of the 2-year U-Act-Early trial and its 3-year post-trial follow-up period, during which treatment was according to the decision of the rheumatologist,² showed that the strategies initiating tocilizumab as first therapy in early RA were over these 5 years not cost-effective, compared to a strategy initiating MTX and later on adding tocilizumab, if indicated. A possible explanation, next to the finding that bDMARD use in the strategy arms during the study period became more similar as consequence of the strict treat-to-target strategy, is the lack of tapering during the non-protocolled follow-up period.³ Again, this underlines that for strategies aiming to save costs or improve cost effectiveness, clear recommendations on the use and strategy of expensive DMARD are needed, which have to be adhered to. In the study described above, selecting a prognostically unfavourable subgroup based on DAS28 and HAQ notably improved cost effectiveness. This suggests that it remains essential for rheumatology communities to continue developing novel strategies with bDMARD in RA, to reduce cost and improve long-term effectiveness.

How we are looking: methodology of costing studies

Müskens et al. performed a quasi-experimental (before after) study and data was collected in daily practice, appropriate to study real-world effects. However, only medication cost for RA was analysed, only in one hospital, a rather limited scope. A justification is that in the Netherlands, also for outpatients, bDMARD are paid for and delivered by the hospital. Discounts on the maximal governmentally set bDMARD purchase prices are negotiated by each individual hospital (or collaborative hospital group) with the pharmaceutical companies. Which percentage of the bDMARD cost is reimbursed to the hospital (groups) by the health insurance organisations is negotiated between these two parties, see Figure.⁴ This (intricate) system explains the interest of individual Dutch hospitals in number and cost of bDMARD prescribed in their center. As such, the results of the paper of Müskens are important. However, the negotiated prices of biologicals are not made available publically, and these, and not reference prices were used in the analyses of Müskens.⁵ This hampers generalizability of the results of this study in a single center. Interestingly, after introduction of biosimilars on the Dutch

market in 2015, the mean purchase price of etanercept for Dutch hospitals decreased almost 60%, reducing the price difference between etanercept and its biosimilar.⁴

As mentioned above, the study of Müskens et al. only analysing cost that are relevant for a certain budget, is classified as a budget impact analysis. For a more comprehensive evaluation of treatment strategies, health-economic studies are more appropriate. Depending on which costs and outcomes are taken into account in the analysis, an economic evaluation can be classified as a cost effectiveness (weighing costs against a disease specific health outcome), cost-utility (weighing costs against a general quality of life ‘utility’ measure, i.e., quality adjusted life years, QALYs), cost-minimization (only analyzing cost, assuming equal effectiveness and similar populations, which condition was not met in the study of Müskens) or cost-benefit (expressing health outcomes also in monetary value, leading to a net-monetary effect) analysis. Health-economic studies of medication strategies often are cost effectiveness or cost-utility analyses.

In all analyses, key decisions are on which costs specifically to include. First, costs within the health care sector, i.e., related to intervention, not only cost of medication, but also e.g., of intravenous administration,³ testing for latent tuberculosis, physician visits and consultations and admissions.⁶ Second, patient and family costs, occurring outside of health care system, like care provided by family members, and transportation.⁶ Third, costs in other sectors, e.g., related to loss of productivity, due to work disability, absenteeism (time missed from work due to health reasons) and presenteeism (impaired performance while at work due to health reasons resulting in productivity loss).^{6,7} This last cost category is often referred to as indirect cost. Cost of work disability can be estimated applying the human capital method, taking into account productivity loss over the whole period of the potential working life, and by the friction-cost method. Then, only the average time-span the employing organisation needs to restore the initial production level is taken into account.⁶

Next, the choice of the measure of effectiveness is of interest. This may be a unidimensional and disease specific health outcome, e.g., the percentage of patients achieving remission, but also a multidimensional, generic health outcome, such as QALY. Generic health outcomes enable comparisons of study outcomes of studies evaluating quite different interventions.⁶

If Müskens et al. would have performed a longer-term study with a wider scope, also looking at indirect cost, this study might have had a positive result, e.g., because a higher percentage of RA patients using a bDMARD could have translated into higher quality of life of patients and less productivity losses.

Who's looking: interests and interpretations of specific stakeholders

Interpretations of whether an expensive medication or novel therapeutic strategy is “worth it” depend not only on the methodology used, i.e., costs and outcomes assessed,

but also on the eye and interests of the beholder, i.e., stakeholder (see Figure). Of course, their interests are also determining choices in study design and methodology. For a Dutch hospital, financial data on prescribed bDMARD for outpatients, including biosimilars, are relevant. This would be different for hospitals in other countries, where bDMARDs are not paid for and delivered to outpatients by the hospital, or were bDMARD prices are set per region or the entire country, and all medication cost are reimbursed by the National Health Service system.^{8,9}

For RA patients, if they do not have to pay extra (i.e., beyond their health insurance) for the medications, quality of life and anti-disability effects of bDMARD strategy studies are the most relevant. For the Ministry of Health, control of the health budget, especially for the duration of its administration, is important, and probably most interesting will be results of budget impact analyses. For the government as a whole, as for the tax paying population, comprehensive cost-utility studies incorporating also indirect cost would be important. However, because reduction of expenditures for unemployment and for absenteeism, let alone for presenteeism, as a result of increased adequate use of bDMARD by RA patients is very difficult to estimate, and also because this would take an adequate period of many years to assess it fully, often not all indirect cost can be taken into account in these analyses.

In conclusion, Müskens et al. are to be complemented with their study on real-world medication cost with relevant results for a Dutch hospital. The authors report a negative study result. However, the study might have had positive results with a more stable bDMARD prescription policy, and especially with a broader scope, which we would prefer, next to health-economic evaluations of novel treatment strategies.

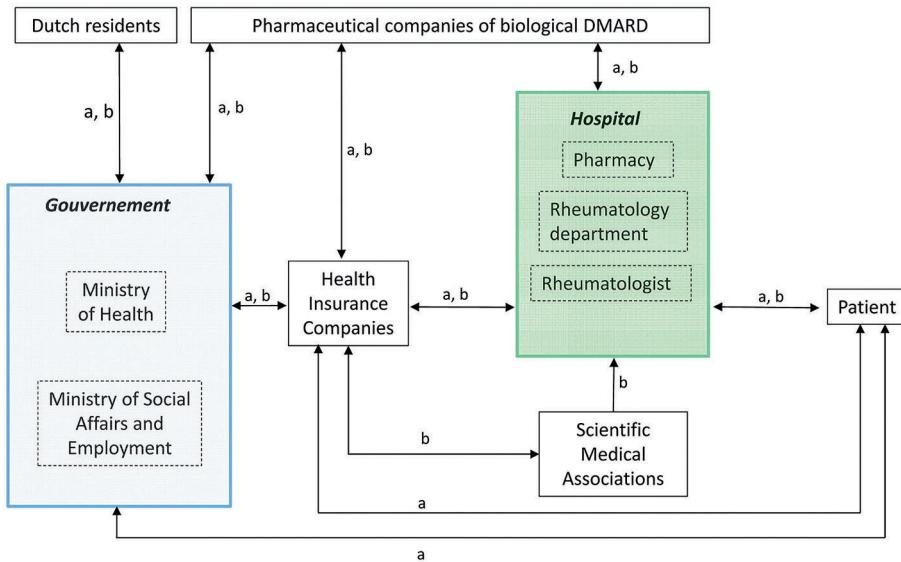


Figure The interplay of financial associations, guidelines and agreements between Dutch stakeholders of biological DMARD therapy.*

*This non-exhaustive model, based specifically on the Dutch situation serves only as example; for each country, situations will differ. Each stake-holder might be inclined to look at results of analyses specifically suited to their situation

a: financial interactions

b: agreements, guidelines, non-financial interactions

Within each colored rectangle, several interrelations of type a and b exist

REFERENCES

1. Bijlsma, J. *et al.* Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet* **388**, 343–355 (2016).
2. Verhoeven, M. M. A. *et al.* Effectiveness and safety over 3 years after the 2-year U-Act-Early trial of the strategies initiating tocilizumab and / or methotrexate. *Rheumatology* **59**, 2325–2333 (2020).
3. Verhoeven, M. M. A. *et al.* Effect on costs and quality-adjusted life-years of treat-to-target treatment strategies initiating methotrexate, or tocilizumab, or their combination in early rheumatoid arthritis. 5 year economic evaluation. *J. Rheumatol.* **2020 Aug (epub ahead of print)**.
4. Autoriteit Consument & Markt. Sectoronderzoek TNF-alfaremmers. <https://www.acm.nl/sites/default/files/documents/2019-09/sectoronderzoek-tnf-alfaremmers.pdf> (2019).
5. Zorginstituut Nederland. Medicijnkosten.nl.
6. Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. & Torrance, G. W. *Methods for the Economic Evaluation of Health Care Programmes*. (Oxford University Press, 2015).
7. Escorpizo, R. *et al.* Worker productivity outcome measures in arthritis. *J. Rheumatol.* **34**, 1372–1380 (2007).
8. WHO Regional Office for Europe. *Medicines reimbursement policies in Europe*. https://www.euro.who.int/_data/assets/pdf_file/0011/376625/pharmaceutical-reimbursement-eng.pdf (2018).
9. Putrik, P. *et al.* Variations in criteria regulating treatment with reimbursed biologic DMARDs across European countries. Are differences related to country's wealth? *Ann. Rheum. Dis.* **73**, 2010–2021 (2014).

PART II

**Monitoring disease activity in
treat-to-target strategies: exploring
usefulness of the HandScan.**

Chapter 9:

Utility of the HandScan in monitoring disease activity and prediction of clinical response in rheumatoid arthritis patients: an explorative study.

M.M.A. Verhoeven¹

J. Tekstra¹

A.C.A. Marijnissen¹

A.J.L. Meier²

A.A.A. Westgeest²

F.P.J.G. Lafeber¹

J.W.G. Jacobs¹

J.M. van Laar¹

P.M.J. Welsing¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands;

²Department of Rheumatology, Máxima Medical Center, Eindhoven, the Netherlands.

Epub ahead of print in Rheumatology Advances in Practice January 2021

ABSTRACT

Objective

To determine the ability of the HandScan (assessing inflammation in hand and wrist joints using optical spectral transmission, OST) to longitudinally measure RA disease activity compared to DAS28, and to determine whether short-term (i.e., 1 month) OST-score changes can predict treatment response at 3 or 6 months.

Methods

Participants visited the out-patient-clinic before start of (additional) RA-medication and 1, 3 and 6 months thereafter. Disease activity was monitored at each visit with the HandScan and DAS28 in parallel. A mixed effects model with DAS28 as outcome variable with a random intercept at patient level, visit month and DAS28 one visit earlier was used to evaluate whether OST-score changes are related to changes in DAS28. Binary logistic regression was used to test the predictive value of short-term OST-score changes together with baseline OST-score for achievement of treatment response (EULAR or ACR criteria). All models were adjusted for RA stage (early or established).

Results

In total 64 RA patients were included. One unit OST-score change was found to be related to an average DAS28 change of 0.03 (95%CI 0.01–0.06), $p=0.03$. When adding OST-score as variable in the longitudinal model, the model's ability to estimate DAS28 (i.e., explained variance) increased with 2%, to 59%. Neither baseline OST-score, nor short-term OST-score change was predictive for treatment response at 3 or at 6 months.

Conclusion

A longitudinal association of OST-score with DAS28 exists, although explained variance is low. The predictive ability of short-term changes in HandScan for treatment response is limited.

INTRODUCTION

The treatment of RA has significantly improved over the last decades due to earlier and more intensive treatment, with swift adjustment of treatment if the target is not achieved e.g., initiating biological (b) disease modifying anti-rheumatic drugs (DMARDs).^{1,2}

To treat RA patients effectively it is important to focus on achieving and maintaining remission (treat-to-target principle), thereby preventing or restricting joint damage. Therefore, patients visit the out-patient-clinic regularly to monitor disease activity (i.e., tight-control principle).^{3,4} The disease activity score assessing 28 joints (DAS28) is widely used to evaluate disease activity also in individual patients. Joint tenderness and swelling of 28 joints, together with an acute phase reactant (erythrocyte sedimentation rate (ESR) or C-reactive protein) and a visual analogue scale (VAS) for patient's experience of disease activity, are combined in the composite DAS28 measure. This method of evaluating disease activity, has a considerable inter and intra assessor variability, especially without formal training of assessors, is time consuming and partly subjective.⁵

The HandScan, based on the principle of optical spectral transmission (OST), is a new method which has been developed to measure RA inflammation in hand (i.e., MCP1-5, IP1 and PIP 2-5) and wrist joints. The RA patient places both hands in the HandScan, and by using red/near-infrared light, the grade of inflammation is assessed per joint (i.e., individual joint score), as well as a total score of all included joints (i.e., total OST-score). A HandScan measurement can be performed at any location, if the device is available, within 5 minutes, without taking much time of a healthcare professional.⁶ More detailed information is provided in Supplementary Data S1. In a cross-sectional study, the OST-score as provided by the HandScan (range 0-66=worst inflammation) was reproducible, and it correlated (coefficient=0.54) with the grade of inflammation of hand and wrist joints as assessed by ultrasonography.⁷ The outcome of the HandScan was more sensitive in detecting subclinical disease activity (as determined by ultrasonography) than physical exam, and its assessment is less time consuming than that of DAS28.⁸

In addition, the HandScan might facilitate early detection of response to treatment, typically assessed at 3 to 6 months after start of (added) therapy. This may be particularly relevant in (early) RA patients treated according to the tight-control principle, stepping up treatment to more intensive (biological) treatment modalities, for example, tumour necrosis factor inhibitors (TNFi).²

All previous research with the HandScan was cross-sectional. However, in light of the treat-to-target principle, it is specifically important to establish if changes in OST-score are associated with changes in DAS28 (as reference standard) in individual RA patients (i.e., if a *longitudinal* association of OST-score with disease activity exists). Also, for optimal treat-to-target strategies, it would be valuable if the OST-score could predict

clinical response to treatment early after treatment initiation. Furthermore, during the last decades patient reported outcomes became of more interest as measure for the impact of disease, and therefore the relation of OST-score with individual components of DAS28, functional disability and quality of life of patients is also of interest.⁹

The current explorative study aimed to determine the *longitudinal* association of the HandScan with DAS28 (i.e., are OST-score changes related to DAS28 changes) in individual RA patients, which, if present and strong enough, would provide a rationale for its use as a disease activity monitoring instrument like DAS28. In addition, the *longitudinal* association of OST-score with the swollen joint count (SJC), tender joint count (TJC), functional disability and quality of life of patients was determined.

Furthermore, the ability of short-term (i.e., baseline to 1 month) OST-score changes to predict clinical response to conventional synthetic (cs)DMARDs or TNFi treatment at 3 or 6 months was studied.

We hypothesized that a longitudinal association between OST-score and DAS28 exists. Furthermore, we hypothesized that short-term changes in OST-score can predict clinical response to treatment.

METHODS

This is an observational cohort study, among RA patients. The institutional review boards of the participating centres confirmed that the Medical Research Involving Human Subjects Act (WMO) was not applicable to this study, and all patients gave written informed consent.

Consecutive early and established RA patients visiting the out-patient-clinic of participating centres, from April 1, 2017, to May 31, 2019, and satisfying the inclusion criteria were all eligible for inclusion. Inclusion criteria were: meeting the 2010 ACR/EULAR criteria, and age above 18 years. Early RA patients were further required to be DMARD-naïve and started DMARD therapy, usually a csDMARD like methotrexate, according to the tight-controlled treat-to-target principle. Established RA patients started with or switched to another TNFi because of active disease, also in a tight-controlled manner, as additional therapy. Exclusion criteria for both cohorts were rheumatic autoimmune disease other than RA or a current inflammatory joint disease other than RA (e.g., gout). Other exclusion criteria were glucocorticoid use <6 weeks prior to baseline for early RA and previous use of the same TNFi (i.e., restarting treatment) for established RA.

All included patients visited the out-patient-clinic just before starting their (additional) treatment (baseline), and 1, 3 and 6 months thereafter (i.e., tight-controlled). In early RA patients the csDMARD dose (typically MTX start at 10 mg/week) was increased, if necessary, every month in steps of 5mg according to the treat-to-target principle. In established RA patients, the dose of the TNFi started was not modified during the study

period of 6 months. Disease activity was measured at each visit, first with the HandScan and shortly after with DAS28. The following baseline data was collected: age, gender, BMI, smoking status, alcohol use, rheumatoid factor (RF) status, and anti-cyclic citrullinated peptide antibody (anti-CCP) status. DAS28 (and its components) and OST-scores were collected at every visit, whereas the functional ability and quality of life were assessed at baseline and every 3 months thereafter, using respectively the Health Assessment Questionnaire (HAQ) and EuroQol five dimensions questionnaire (EQ5D-5L).

Statistical analysis

Baseline characteristics and treatment response were described for all patients, and stratified by RA stage (early or established; csDMARD therapy or TNFi therapy). Data of early and established RA were combined to obtain a more adequate sample size. The effect of RA stage was taken into account in all model based analyses (e.g., see explanation of the mixed effect models and the binary logistic regression models further in this section).¹⁰ Pearson or Spearman correlation coefficients, depending on distribution of the data, of DAS28, SJC, TJC, HAQ and EQ5D-5L, with OST-score were calculated for all patients, both concurrently, as well as with time-lags to explore the crude associations of OST-score over time with other frequently used outcome measures. To determine whether changes in OST-scores are related to changes in DAS28 in individual patients, an autoregressive mixed effects model with a random intercept at patient level was used.¹¹ The outcome variable was DAS28; independent variables were OST-score, visit month, RA stage and DAS28 at previous visit (i.e., autoregressor). The same analyses were performed for SJC (square root transformed), TJC (square root transformed), HAQ and EQ5D-5L as respective outcome variables. It was also explored whether RA stage (early vs. established; csDMARD vs. TNFi) modified the association between OST-score and the outcomes by adding the interaction term, e.g., OST-score*RA stage. Binary logistic regression was used to test the predictive value of short-term (i.e., 1 month) OST-score change together with baseline OST-score for the outcome EULAR good response (y/n), and ACR50 response (y/n) at 3 or 6 months. Baseline DAS28, and short-term (i.e., 1 month) DAS28 change were also evaluated in a similar separate analysis for comparison with the former model. This analysis was also adjusted for RA stage (early vs. established) as the initiated therapy differ (csDMARD vs. TNFi), and it was tested if RA stage modified the association between changes in OST-score and outcome (i.e., adding the interaction term OST-score*RA stage).

Because of the exploratory nature of this study, no power calculation was performed. The statistical analyses were performed in SAS version 9.4. All tested were two-sided and a p-value ≤ 0.05 was considered statistically significant. Seven of 64 patients had missing information on DAS28 and/or OST-score, but only at the 6 months visit. As mixed model analysis, using all longitudinally available data of the patients, is robust

against sporadically missing data, imputation was deemed to have no additional value in this situation and was not performed.¹²

RESULTS

In total 64 RA patients were included, n=32 with early RA (DMARD-naïve, starting MTX and prednisone) and n=32 with established RA (starting with first or consecutive TNFi as additional therapy). All early RA patients were treated according to EULAR guidelines, and remained on MTX treatment during the study. Regarding established RA patients, 26 of 32 were bDMARD-naïve and started treatment with a first TNFi, whereas the others started a consecutive TNFi. More detailed information about medication use is shown in Supplementary Table S1. In early RA patients, no treatment failures during the 6 months follow up were observed, whereas 5 patients in the established RA cohort discontinued TNFi therapy due to insufficient effectiveness. Three of them switched after 3 months to another bDMARD. One of 32 established RA patients experienced an adverse event (not related to TNFi therapy), and stopped therapy.

Table 1 provides an overview of the baseline characteristics and treatment response of all patients, and separately per cohort. Overall, similar outcomes were observed for early and established RA, except for SJC28 and number of alcohol users, both at baseline, and response to treatment during the study period, except for changes in HAQ, see Table 1. DAS28, OST-score, SJC, TJC, HAQ and EQ5D-5L values over time are shown in Figure 1.

Table 1 Baseline characteristics and treatment response.

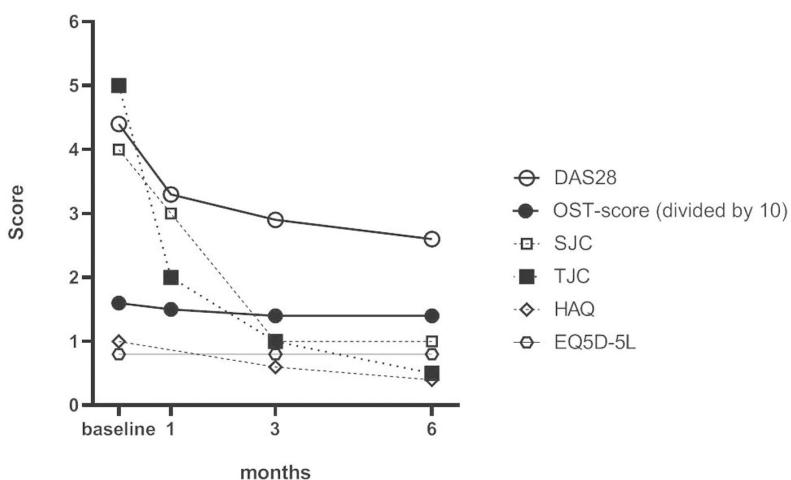
	All n= 64	Early n= 32	Established n= 32	p-value
Baseline characteristics				
Female, n (%)	42 (66)	22 (69)	20 (63)	0.60
Age (years), mean (SD)	57.6 (11.6)	55.8 (12.2)	59.6 (10.8)	0.26
BMI (kg/m ²), mean (SD)	26.4 (4.9)	25.7 (5.1)	27.2 (4.5)	0.18
Duration RA (years), median (IQR)	n.a	n.a.	8 (2 – 13)	n.a
Smoker, n (%)	13 (22)	7 (22)	6 (23)**	0.91
Alcohol user (\geq 1 unit/week), n (%)	36 (56)	22 (69)	14 (44)	0.04*
RF positivity, n (%)	48 (77)	23 (72)	25 (83)	0.51
Anti-CCP positivity, n (%)	50 (81)	29 (91)	21 (70)	0.11
DAS28, mean (SD)	4.4 (1.1)	4.5 (0.9)	4.3 (1.2)	0.39
SJC28, median (IQR)	4 (2 – 7)	6 (4 – 10)	3 (1 – 6)	0.01*
TJC28, median (IQR)	5 (2 – 8)	4.5 (3 – 7)	6 (2 – 11)	0.35
VAS global, median (IQR)	55 (38 – 70)	61 (42 – 75)	51 (30 – 63)	0.27
OST-score, mean (SD)	15.6 (5.3)	16.8 (5.8)	14.4 (4.3)	0.10
HAQ score, median (IQR)	1.0 (0.6 – 1.4)	0.9 (0.5 – 1.4)	1.1 (1.0 – 1.7)	0.20

Table 1 (continued)

	All n= 64	Early n= 32	Established n= 32	p-value
Treatment response				
ΔDAS28 month 3, mean (SD)	1.5 (1.3)	2.0 (1.0)	0.9 (1.3)	<0.01*
ΔDAS28 month 6, mean (SD)	1.9 (1.3)	2.5 (1.0)	1.1 (1.1)	<0.01*
ΔOST-score month 3, mean (SD)	1.6 (4.6)	2.6 (4.7)	0.5 (4.3)	0.04
ΔOST-score month 6, mean (SD)	1.7 (5.3)	2.6 (5.9)	0.7 (4.6)	0.10
EULAR good response month 3, n (%)	27 (42)	19 (59)	8 (25)	<0.01*
EULAR good response month 6, n (%)	34 (53)	28 (88)	6 (19)	<0.01*
ACR50 response month 3, n (%)	20 (31)	14 (44)	6 (19)	0.03*
ACR50 response month 6, n (%)	28 (44)	18 (56)	10 (31)	0.04*
ΔHAQ score month 3, median (IQR)	0.3 (0.0 – 0.8)	0.4 (0.1 – 0.8)	0.3 (0.0 – 0.8)	0.35
ΔHAQ score month 6, median (IQR)	0.5 (0.3 – 0.9)	0.6 (0.3 – 1.0)	0.4 (0.1 – 0.9)	0.24

All= early and established rheumatoid arthritis (RA) patients; Early= newly diagnosed RA patients starting conventional synthetic DMARD therapy; Established= established RA patients starting (first of new) TNF inhibitor as additional therapy.

BMI= body mass index; SD= standard deviation; RA= rheumatoid arthritis; RF= rheumatoid factor; CCP= cyclic citrullinated peptide; DAS28= disease activity score assessing 28 joints; SJC28= swollen joint count assessing 28 joints; IQR= interquartile range; TJC28= tender joint count assessing 28 joints; VAS= visual analogue scale, range 0-100=worst; OST= optical spectral transmission, range 0-66=worst; HAQ= Health Assessment Questionnaire, range 0-3=worst. *= statistically significant; **=6 missing.

**Figure 1** Disease activity measures, functional ability and quality of life over time.*

*means for DAS28 and OST-score and medians for SJC, TJC, HAQ and EQ5D-5L

DAS28= disease activity score assessing 28 joints; OST-score= optical spectral transmission score; SJC= swollen joint count; TJC= tender joint count; HAQ= health assessment questionnaire; EQ5D-5L= EuroQol five dimensions questionnaire.

The concurrent (i.e., at the same time-point) correlations between DAS28 and OST-score and between SJC and OST-score were moderate (correlation coefficients ranging from 0.18 to 0.39 and from 0.35 to 0.47, respectively) and statistically significant. Lower (often) non-statistically significant correlations of OST-score were found with TJC, HAQ and EQ5D-5L, see Table 2. Non-concurrent correlations were also generally lower and often not statistically significant.

Table 2 Correlation coefficients between OST-scores and DAS28/SJC28/TJC28/HAQ/EQ5D over time.

OST-score	baseline	month 1	month 3	month 6
DAS28, baseline	0.31*	0.25*	0.27*	0.11
DAS28, month 1	0.17	0.18	0.26*	0.11
DAS28, month 3	0.26*	0.25*	0.39*	0.39*
DAS28, month 6	0.17	0.20	0.31*	0.34*
SJC, baseline	0.41*	0.49*	0.42*	0.39*
SJC, month 1	0.33*	0.47*	0.51*	0.50*
SJC, month 3	0.23	0.30*	0.38*	0.47*
SJC, month 6	0.30*	0.19	0.16	0.35*
TJC, baseline	0.00	-0.02	-0.02	-0.02
TJC, month 1	-0.05	-0.04	0.07	0.08
TJC, month 3	0.16	0.17	0.31*	0.39*
TJC, month 6	0.07	0.16	0.29*	0.46*
HAQ, baseline	0.08		0.20	0.01
HAQ, month 3	-0.01		0.10	0.10
HAQ, month 6	-0.23		0.08	0.09
EQ5D, baseline	-0.05		-0.21	-0.26
EQ5D, month 3	0.08		-0.15	-0.21
EQ5D, month 6	-0.05		-0.14	-0.35*

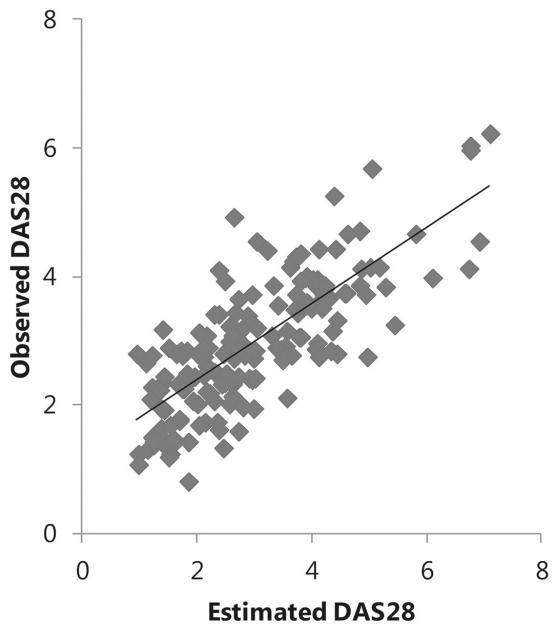
Outcomes based on all patients (early and established), n=64.

*= statistically significant

OST= optical spectral transmission; DAS28= disease activity score assessing 28 joints; SJC28= swollen joint count; TJC= tender joint count; HAQ= health assessment questionnaire; EQ5D= EuroQoL-5D.

The *longitudinal* analysis showed that one unit change in OST-score was associated with a change in DAS28 of on average 0.03 unit (95%CI 0.01–0.06). Using standardized values, this could be interpreted as a change of one SD unit in OST-score being related to a change in DAS28 of on average 0.13 SD unit (95%CI 0.03–0.23). Hence changes in DAS28 value can to some extent be estimated from the OST-score changes. This association was not modified by RA stage ($p=0.96$ for the interaction term). When adding OST-score to the model with only the previous DAS28 (autoregressor) and visit, the model's ability to estimate DAS28 over time (i.e., explained variance) increased, with 2%, to 59%, Figure

2. SJC and TJC changes of one SD unit were on average related to changes of 0.18 (95%CI 0.05–0.31) and 0.16 (95%CI 0.05–0.25) SD units of OST-score respectively. The explained variance increased with 4% and 3% respectively (to 32% and 43% respectively) when adding OST-score to the models. The association with SJC (but not TJC, $p=0.52$ for the interaction term) was found to be modified by RA stage ($p=0.03$ for the interaction term). Stratified analyses showed that one SD unit of OST-score change was on average related to 0.08 (95%CI -0.08–0.14) and 0.37 (95%CI 0.15–0.59) SD units in SJC, respectively, for early and established RA. No association of OST-score with HAQ nor EQ5D-5L was found (results not shown).



9

Figure 2 Observed DAS28 vs. estimated DAS28 (using full model with OST-score).
DAS28= disease activity score assessing 28 joints; Predicted DAS28= DAS28 as estimated by the model, with OST-score, visit month and DAS28 at the previous visit as variables.

Baseline OST-score (odds ratio (OR) 0.93, 95%CI 0.83–1.04; standardized OR 0.67, 95%CI 0.37–1.22), and the short-term OST-score change (OR 1.04, 95%CI 0.90–1.19; standardized OR 1.18, 95%CI 0.65–2.13) were both not statistically significantly predictive for EULAR response at 3 months. Baseline DAS28 neither was a significant predictor (OR 0.76, 95%CI 0.41–1.41; standardized OR 0.74, 95%CI 0.38–1.44), whereas short-term DAS28 change (OR 4.47, 95%CI 1.73–11.58; standardized OR 3.96, 95%CI 1.65–9.52) was. Results for EULAR response at 6 months were in line with the above, see Table 3. For ACR50 response at 3 months, none of the variables were significant predictors. Short-term DAS28 change (OR 3.92, 95%CI 1.57–9.28; standardized OR 3.69, 95%CI 1.65–9.52) was

a significant predictor for ACR50 response at 6 months, see Table 3. In all analyses the association of the short-term change in OST-score with treatment response was not modified by RA stage as tested in the models ($p=0.44/p=0.22$ and $p=0.20/p=0.30$ for ACR50 and EULAR good response at 3/6 months, respectively).

Table 3 Odds ratio's for predictive ability of variables on treatment response.

Parameter	OR (95%CI) multivariable	OR (95%CI) multivariable
EULAR good response at month 3		
Baseline OST-score	0.93 (0.83 – 1.04)	
ΔOST-score at month 1	1.04 (0.90 – 1.19)	
Baseline DAS28		0.76 (0.41 – 1.41)
ΔDAS28 at month 1		4.47 (1.73 – 11.58)*
RA stage	0.20 (0.07 – 0.62)*	0.41 (0.12 – 1.37)
EULAR good response at month 6		
Baseline OST-score	0.95 (0.93 – 1.10)	
ΔOST-score at month 1	1.07 (0.89 – 1.27)	
Baseline DAS28		0.71 (0.34 – 1.49)
ΔDAS28 at month 1		3.05 (1.05 – 8.90)*
RA stage	0.03 (0.01 – 0.13)*	0.04 (0.01 – 0.19)
ACR50 response at month 3		
Baseline OST-score	0.93 (0.82 – 1.05)	
ΔOST-score at month 1	1.10 (0.94 – 1.31)	
Baseline DAS28		0.96 (0.55 – 1.78)
ΔDAS28 at month 1		1.77 (0.81 – 3.88)
RA stage	0.22 (0.06 – 0.75)*	0.35 (0.10 – 1.22)
ACR50 response at month 6		
Baseline OST-score	0.94 (0.84 – 1.05)	
ΔOST-score at month 1	1.00 (0.88 – 1.14)	
Baseline DAS28		0.83 (0.47 – 1.49)
ΔDAS28 at month 1		3.92 (1.57 – 9.28)*
RA stage	0.27 (0.09 – 0.82)*	0.58 (0.18 – 1.87)

*= statistically significant (OR 95%CI not including 1).

OST= optical spectral transmission; DAS28= disease activity score assessing 28 joints; OR= Odds ratio; Δ= change; RA stage= early (starting conventional synthetic DMARD therapy) is reference.

DISCUSSION

In this first longitudinal study of the HandScan, the concurrent correlations of HandScan (expressed as OST-score) and DAS28 were in general low-to-moderate, consistent with data of a previous cross-sectional study.⁸ Although we established a longitudinal

association of the HandScan with DAS28, which would be a prerequisite for using such an instrument for monitoring disease activity over time, the added value (explained variance) was low. This limits the use of the HandScan as a comprehensive instrument for monitoring individual patients' disease activity.

A plausible explanation for the low ability to estimate DAS28 with OST-scores, may be that in addition to the number of tender and swollen joints (from 28), an acute phase reactant (i.e., ESR) and a VAS expressing the patients' assessment of disease activity are part of DAS28,¹³ whereas the OST-score measures only RA inflammation of hand and wrist joints (with a maximum of 22 joints). Therefore we evaluated also components of DAS28 separately. Indeed the association with TJC and, especially, SJC was (somewhat) stronger than with DAS28 as apparent from the higher standardized regression coefficients and increase in explained variance by adding OST-score to the longitudinal model.

We could not establish a predictive association of baseline OST-score or short-term changes in OST-score with later response to treatment. This lack of predictive ability may again, partly, be due the fact that OST-scores only reflects joint inflammation in a limited set of joints and response criteria are based on composite scores.¹⁴

As shown in Supplementary Table S1, GC therapy was used in early as well as in established RA. GC therapy diminishes disease activity, but this will likely have equally been the case for DAS28 and the OST-score. Therefore we think this has had no or only limited influence on the results of our main analysis, i.e., the longitudinal association between OST-score and DAS28.

Since inflammation of osteoarthritic joints is generally considerably less than in RA joints, and the DIP joints (mostly affected in osteoarthritis) are not assessed in the HandScan, we expect that the influence of concomitant osteoarthritis on OST-score results have been limited.

A limitation of this study is that the sample size is modest. The intention of the current study was to explore whether a longitudinal association of OST-score with DAS28 existed, which is a prerequisite for using OST-scores as a disease activity measurement in patients over time. Therefore, we aimed to include at least 30 early as well as 30 established RA. In the analysis we combined early (n=32) and established (n=32) RA patients in the analyses, correcting for RA stage. It turned out that RA stage did not influence the longitudinal association between OST-score and other outcomes, except for SJC. A possible explanation might be the fact that in early RA patients the SJC was often 0 at follow-up due to the strict treat-to-target treatment approach, possibly obscuring small changes over time, whereas SJC was generally higher in established RA patients.¹⁵ In addition, the predictive association between short-term changes in OST-score and longer-term response was also not influenced by RA stage. Furthermore, one may expect that type of treatment may influence the ability of OST-scores to detect

changes in disease activity as bDMARDs are known to more rapidly suppress tissue vascularity.^{16,17} We tested whether the relation between OST-scores and DAS28 (-based response) was different between early (i.e., csDMARD treated-) and established (i.e., bDMARD treated-) patients, but could not detect a significant effect. Of course, as the effect of RA stage and treatment modality are intertwined in our study, this may have muddled this effect. And, in the bDMARD treated group, patients could have started their next bDMARD, which could have diminished the potential change in joint vascularity (and thus the ability of the HandScan to detect it), as vascularity was already reduced by the previous bDMARD. It is known that in patients even with inadequate response to bDMARDs, progression of joint damage is inhibited¹⁸ and thus probably also joint vascularity. Lastly, given that the HandScan only measures hands and wrists joints, it might be mainly applicable for the subset of RA patients with involvement of the hand joints.

This first study assessing the *longitudinal* association of the HandScan with disease activity measures relevant in monitoring treatment response warrants future research focusing on the development of a composite measure to assess disease activity where a joint count assessment (i.e., SJC and TJC) is replaced by OST-scores. OST-scores can be obtained without visiting a physician, as a HandScan measurement can be performed easily by a non-healthcare professional, and at any location where the device can be placed, for example at the out-patient waiting room. By implementing a disease activity index (including only variables which are assessed without visiting a physician, i.a. OST-scores) into daily practice, the rheumatologists' and/or nurse practitioners' time might be saved in the busy out-patient-clinics, since only the patients with active disease would require an additional visit to the rheumatologist or healthcare professional for a more detailed assessment, including joint counts.

Conclusion

A longitudinal association of OST-score with DAS28 exists, although the relation is weak. As such, in this setting OST-score as a single measuring instrument is insufficient to comprehensively assess disease activity in RA patients. However, combining the OST-score with other (routinely used) disease activity parameters might result in an adequate composite disease activity measure.

REFERENCES

1. Smolen, J. S., Aletaha, D. & McInnes, I.B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
2. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
3. Bakker, M. F., Jacobs, J. W. G., Verstappen, S. M. M. & Bijlsma, J. W. J. Tight control in the treatment of rheumatoid arthritis: efficacy and feasibility. *Ann. Rheum. Dis.* **66**, 56–61 (2007).
4. Bakker, M. F. *et al.* Low-dose prednisone inclusion in a methotrexate-based, tight control strategy for early rheumatoid arthritis: a randomized trial. *Ann. Intern. Med.* **156**, 329–338 (2012).
5. Meier, A. J. L., Rensen, W. H. J., de Box, P. K. & de Nijs, R. N. J. Potential of optical spectral transmission measurements for joint inflammation measurements in rheumatoid arthritis patients. *J. Biomed. Opt.* **17**, 081420 (2012).
6. Besselink, N. J. *et al.* Novel optical spectral transmission (OST)-guided versus conventionally disease activity-guided treatment: study protocol of a randomized clinical trial on guidance of a treat-to-target strategy for early rheumatoid arthritis. *Trials* **20**, 226 (2019).
7. Besselink, N. J. *et al.* Optical spectral transmission to assess inflammation in hand and wrist joints of rheumatoid arthritis patients. *Rheumatology* **57**, 865–872 (2018).
8. van Onna, M. *et al.* Assessment of disease activity in patients with rheumatoid arthritis using optical spectral transmission measurements, a non-invasive imaging technique. *Ann. Rheum. Dis.* **75**, 511–518 (2016).
9. Hendrikx, J., de Jonge, M. J., Fransen, J., Kievit, W. & van Riel, P. L. C. M. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. *RMD Open* **2**, e000202 (2016).
10. Twisk, J. W. R. *Applied multilevel analysis*. (Cambridge University Press, 2006).
11. Funatogawa, I. I. & Funatogawa, T. Autoregressive linear mixed effects models. (Springer, Singapore, 2019).
12. Twisk, J., de Boer, M., de Vente, W. & Heymans, M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J. Clin. Epidemiol.* **66**, 1022–1028 (2013).
13. Felson, D. T. *et al.* American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum.* **63**, 573–586 (2011).
14. Fransen, J. & van Riel, P. L. C. M. The disease activity score and the EULAR response criteria. *Rheum. Dis. Clin. North Am.* **35**, 745–757 (2009).
15. Ramiro, S. *et al.* Is treat-to-target really working in rheumatoid arthritis? A longitudinal analysis of a cohort of patients treated in daily practice (RA BIODAM). *Ann. Rheum. Dis.* **79**, 453–459 (2020).
16. Fukae, J. *et al.* Radiographic prognosis of finger joint damage predicted by early alteration in synovial vascularity in patients with rheumatoid arthritis: potential utility of power doppler sonography in clinical practice. *Arthritis Care Res.* **63**, 1247–1253 (2011).
17. Akhavani, M. A., Larsen, H. & Paleolog, E. Circulating endothelial progenitor cells as a link between synovial vascularity and cardiovascular mortality in rheumatoid arthritis. *Scand. J. Rheumatol.* **36**, 83–90 (2007).
18. Smolen, J. S. *et al.* Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade. *Ann. Rheum. Dis.* **68**, 823–827 (2009).

SUPPLEMENTARY FILE

Supplementary Data S1: technical explanation of the HandScan

The HandScan is a (new) medical device, which is developed to assess disease activity in RA patients. The technology of the device is based on quantifying the inflammation related hemodynamic response to an applied stimulus.¹ The speed and magnitude of blood pooling differs between inflamed and healthy tissue due to the vascular changes, which are associated with inflammation. The process of blood pooling is measured by diffuse optical transmission. Both hands of a patient are placed in the HandScan and illuminated with red/near-infrared light; the transmitted light is detected by a camera on the other side of the patient's hands. Since blood is a strong absorber in the red/near-infrared region of the light spectrum, increases of pooling of blood as in arthritis will result in a decrease of the transmitted light, which is "translated" into a higher score. A non-healthcare professional (without medical background) can perform a HandScan measurement; within 5 minutes the OST-score (i.e., HandScan outcome) is produced, which provides a direct measure of inflammation of MCP1-5, IP1 and PIP2-5 and wrists bilaterally.

¹= Meier AJ, Rensen WH, de Bokx PK, de Nijs RN. Potential of optical spectral transmission measurements for joint inflammation measurements in rheumatoid arthritis patients. J Biomed Opt. 2012;17:081420.

Supplementary Table S1 Medication use.

	Early n= 32	Established n= 32
GC users, n (%)	32 (100)	10 (31)
- Dose in users mg/day, median (IQR)	10 (10 – 10)	5 (5 – 15)
MTX users, n (%)	32 (100)	19 (59)
- Dose in users mg/week, median (IQR)	20 (15 – 25)	20 (16 – 25)
Leflunomide users, n (%)	0 (0)	4 (13)
- Dose in users mg/day, median (IQR)		15 (10 – 20)
Sulfasalazine users, n (%)	0 (0)	2 (6)
- Dose in users mg/day, median (IQR)		1250 (1000 – 1500)
Hydroxychloroquine users, n(%)	0 (0)	6 (19)
- Dose in users mg/day, median (IQR)		300 (200 – 400)
TNFi users, n (%)	0 (0)	32 (100) ¹
- Infliximab, n (%)		- 1 (3)
- Adalimumab, n (%)		- 10 (31)
- Etanercept, n (%)		- 19 (60)
- Certolizumab pegol, n (%)		- 1 (3)
- Golimumab, n (%)		- 1 (3)

¹= 6 patients failed on TNFi. One of them experienced an adverse event (not related to TNFi therapy), but stopped therapy, and in 5 patients TNFi therapy was ineffective, 3 of them switched after 3 months to another bDMARD and the other 3 stopped their study participation.

Chapter 10:

**Development and validation of rheumatoid
arthritis disease activity indices including
HandScan (optical spectral transmission) scores.**

M.M.A. Verhoeven¹
A.A.A. Westgeest²
A. Schwarting³
J.W.G. Jacobs¹
C. Heller³
J.M. van Laar¹
F.P.J.G. Lafeber¹
J. Tekstra¹
K. Triantafyllias^{3}*
P.M.J. Welsing^{1}*

** the last two authors contributed equally (joined last author)*

¹Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, the Netherlands;

²Department of Rheumatology, Máxima MC, Eindhoven, the Netherlands;

³ACURA Rheumatology Center, Bad Kreuznach, Germany.

Submitted to Arthritis Care & Research

ABSTRACT

Objective

To develop and validate a composite RA disease activity index using optical spectral transmission (OST)-scores obtained with the HandScan.

Methods

RA patients from a single centre routinely undergoing HandScan measurements and at least one concurrent OST-score and DAS28 were included. Data was extracted from medical records. Linear regression analyses with DAS28 as outcome were performed to create a disease activity index (DAS-OST). OST-score, ESR and VAS, gender, age, disease duration and RF-status were evaluated as independent variables. Final models were derived, based on statistical significance of coefficients and model fit. Of the data, 2/3 was used for development and 1/3 for validation; external validation was performed in a cohort from another centre. Agreement between DAS-OST and DAS28 was assessed using the Bland-Altman plot method and intra-class correlation coefficient (ICC). Diagnostic value of DAS-OST was determined for established definitions of remission, and low (L), and high (H) disease activity (DA).

Results

Data of 3358 observations from 1505 unique RA patients were extracted. DAS-OST was defined as: $-0.44 + \text{OST} \times 0.03 + \text{male} \times -0.11 + \text{LN(ESR)} \times 0.77 + \text{VAS} \times 0.03$. The ICC between DAS-OST and DAS28 were 0.88 (95%CI 0.87–0.90) and 0.82 (95%CI 0.75–0.86) and measurement errors 0.58 and 0.87 in internal and external validation, respectively. Sensitivity for remission, LDA and HDA were 79%, 91%, 43%, and specificity 92%, 80%, 96% in external validation.

Conclusion

Using the HandScan, RA disease activity can be accurately estimated if combined with ESR, VAS and gender into a disease activity index (DAS-OST).

INTRODUCTION

Rheumatoid arthritis (RA) is a common chronic inflammatory disease, mainly affecting joints and surrounding tissues. The disease requires life-long treatment, preferably according to tight-control and treat-to-target principles. Such treatment strategies require that patients frequently (in early disease or with active disease every 1-3 months) visit their physician for evaluation of medication effects, adverse events and disease activity.¹

Disease activity is typically measured by a combination of parameters, including a swollen and tender joint count (e.g., assessing 28 joints, SJC28/TJC28), C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR) and an assessment of patient and/or physician of disease activity or general health typically using a visual analogue scale (VAS). These variables are often combined into an index like the disease activity score (e.g., DAS28).²

However, this method of assessing disease activity is time consuming, especially given the busy out-patient-clinics and limited time per patient that rheumatologists have nowadays. Also to assess the joints for swelling and tenderness as objectively as possible, training and standardisation of joint examinations would be needed.³⁻⁵ Therefore a tool assessing disease activity quickly, easily and objectively could be highly useful.

The HandScan is a tool which measures within 5 minutes inflammation of wrist and small hand joints (i.e., MCP, (P)IP), using optical spectral transmission (OST-score).⁶ Importantly a HandScan measurement can be performed by a non-health professional.⁷ This procedure is more objective, and less painful than joint count assessment.

The correlation between DAS28 and the OST-score (range 0-66=worst) is only moderate.⁸ This may not be surprising as RA disease activity is a multifaceted construct. This is reflected by the fact that patient VAS as well as CRP/ESR are part of validated disease activity indices like DAS28, while the OST-score is a substitute of the swollen joint count only. In line with this, the correlation coefficient (ρ) of OST-score with SJC28 was found to be slightly higher ($\rho=0.50$) than with DAS28 ($\rho=0.42$).⁸ Although the HandScan may be a substitution for joint count assessment only, this substitution may be beneficial, given its benefits, described above.

The current study aimed to develop and validate an index for assessing RA disease activity (states) using the OST-score and other disease activity parameters, and to determine the agreement of this index with DAS28 and its accuracy in estimating remission, low (L) and high (H) disease activity (DA).

METHODS

This study used 2 cohorts. For model development and internal validation, routinely collected data from electronic medical records of the rheumatology department of Máxima Medical Center (MMC) Eindhoven were used and for external validation, data of the ACURA Rheumatology Center Bad Kreuznach (ACURA). The institutional ethical review board of MMC indicated that the Medical Research Involving Human Subject Act (WMO) was not applicable as no interventions or extra measurements were performed and only pseudonymised routinely collected data was extracted from medical records. As such patients did not give informed consent. The local standing committee for ethical conduct of Rhineland-Palatinate, Germany, has approved the study of the ACURA cohort and patients gave written informed consent.

Development and internal validation cohort

Patients with RA, visiting the out-patient-clinic at MMC from April 2017 up to and including March 2019 were eligible for inclusion. Inclusion criteria were; (1) RA according to the American college of Rheumatology (ACR) 1987 or ACR/ European League against Rheumatism (EULAR) 2010 criteria, (2) no relevant visual deformations of hands or fingers (invalid HandScan measurement), (3) availability of a DAS28 measurement, with, thereafter, a HandScan measurement, performed during the same clinical visit, (4) age above 18 years, and (5) no participation in interventional studies.⁷ DAS28 was performed without knowledge of the HandScan score.

External validation cohort

OST was performed in eligible RA patients during their stay in the ACURA in-patient-clinic from September 2017 up to and including June 2018. All included patients met the 2010 ACR/EULAR classification criteria for RA.⁹ Patients with age<18 years, joint prostheses/implants, severe hand deformities, pronounced ulnar deviation, recent trauma or surgery and known photosensitivity were excluded.

Assessments

Data of DAS28 components (SJ28, TJC28, ESR and patient VAS), age, gender, disease duration, rheumatoid factor (RF) status, and anti-cyclic citrullinated peptide (anti-CCP) status were extracted, if available. OST-scores (i.e., total score and individual score of 22 joints) were obtained directly from the HandScan device. Data up to March 2019 was extracted. Data on the same variables was collected in the ACURA cohort. In both cohorts, DAS28 scores were assessed by a well-trained person, blinded with respect to OST-score. Inflammation markers, CRP and/or ESR, were routinely tested and used for DAS28 calculation. RF and anti-CCP status were assessed by ELISA.

Statistical analyses

Patient and disease characteristics were described using means with standard deviations (SD), medians with interquartile ranges (IQR) or frequencies with proportions, where appropriate.

Model development

A random sample of 2/3 of the data from the MMC cohort was used as development cohort.

Using linear regression, we developed disease activity scores including OST-scores (DAS-OST). In these analyses, DAS28-ESR was used as dependent variable (i.e., reference standard) and OST-score as independent variable. The association between OST-score and DAS28 was assessed for linearity graphically as well using categorical values (defined by quartiles) and using quadratic transformations of OST-score. Covariates (i.e., ESR and VAS, and age, gender, disease duration, RF and/or anti-CCP status, which were deemed to possibly influence OST-score) were added to the model and removed one-by-one, retaining all variables that showed added predicted value beyond the OST-score (as judged by p-value ≤ 0.20 or a decrease in adjusted R-square upon removal). Finally, modification of the association between OST-score and DAS28 by relevant covariates (i.e., those retained in the model) was evaluated, one-by-one, retaining interaction terms with a p-value ≤ 0.20 .

OST-scores could be used as a total sum or expressed as number of joints with inflammation (JC-OST: assessing 22 joints, MCP1-5, IP1, PIP2-5, wrist, all bilaterally) as typically used in disease activity indices. Therefore, we developed a DAS-OST based on JC-OST as well. To define JC-OST, first a mixed effect logistic regression analysis (mixed effect to account for clustering of joint scores within patients) using a random intercept was performed with joint swelling (yes/no, DAS28 component) as dependent variable and OST-score of the corresponding individual joints, side (left/right), and joint type (MCP, (P) IP, wrist) as independent variables. This was done to decide whether assuming one cut-off of OST-score for swelling is appropriate for every joint or whether specific cut-offs for a joint(-type) and/or side are more appropriate. Optimal cut-off(s) were defined thereafter using Youden's index. In line with the square root transformation of swollen joint count in DAS28, the square root of JC-OST was also used in DAS-OST development.

Further a DAS-OST formula without patient VAS was developed (i.e., 'objective index'), using the same methodology.¹⁰

Model validation

Using the formula (and applying the derived OST-score cut-offs for individual joints) as derived in the development cohort, DAS-OST was calculated in the internal validation

cohort (i.e., the 1/3 of the MMC data not used for model development). External validation was performed in the ACURA cohort by applying the derived formula.

Agreement of the different DAS-OST indices and DAS28 was determined in the internal- and external validation cohorts by using the Bland-Altman plot method, calculating the SD of the difference (i.e., measurement error), and by random two-way mixed effect intra-class correlation coefficient (ICC).¹¹

The area under the receiver operating characteristic curve (AU-ROC) was calculated to assess overall discrimination of DAS-OST for established definitions of remission, LDA and HDA, based on DAS28 and Boolean remission criteria (i.e., SJC≤1, TJC≤1, patient VAS≤10 and taking an CRP≤10mg/L, as estimated from the established ESR cut-off, see Supplementary Data S1). Furthermore, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of DAS-OST for these established definitions were determined using the same cut-offs for DAS-OST as used for DAS28 (i.e., remission= DAS-OST≤2.6, LDA= DAS-OST≤3.2, HDA= DAS-OST>5.1). Only for Boolean remission a cut-off was (pre-)defined in the development cohort using Youden's index.

All tests were two-sided, a p-value ≤0.05 was considered as statistically significant. No missing data was imputed and analyses were performed in SAS version 9.4.

RESULTS

Characteristics of the two patient cohorts

Data of 3358 observations were extracted, without missing values, from the medical records of the MMC, including 1505 unique RA patients. A random sample of 2/3 of the data (i.e., 2238 observations) was used as development cohort and the remaining data (i.e., 1120 observations) was used as internal validation cohort. The external validation cohort included 168 unique RA patients. Due to missing values of OST-score and/or DAS28, data of 151 RA patients was used for the analyses. Patients' demographic and clinical data are shown in Table 1. Overall, disease activity (i.e., DAS28, its components and OST-score) was statistically significantly higher in the ACURA cohort compared to the MMC cohorts, i.e., development and internal validation ($p<0.01$ and $p<0.01$, respectively, for all variables). Other statistical significant differences between the MMC cohorts and the ACURA cohort were shown for age, disease duration, and seropositivity ($p<0.01$, $p<0.01$ and $p<0.01$, respectively, for all variables).

Table 1 Patients' demographics and clinical data of the Máxima MC (MMC) and ACURA cohort.

	MMC cohort	MMC cohort <i>development</i>	MMC cohort <i>internal validation</i>	ACURA cohort <i>external validation</i>
Patients' demographics				
Number of patients,	1505	1272	817	151
Female, n (%)	979 (65)	831 (65)	541 (66)	100 (67)
Age (year), mean (SD)	65.1 (12.0)	64.9 (12.2)	65.6 (11.6)	60.5 (13.1)
duration of RA (year), mean (SD)	11.4 (8.3)	11.3 (8.4)	11.7 (8.2)	5.9 (8.0)
Seropositivity, n (%)	1068 (71)	901 (72)	588 (73)	116 (77)
Clinical data				
Number of observations	3358	2238	1120	151
DAS28, mean (SD)	2.5 (1.3)	2.5 (1.3)	2.5 (1.2)	3.8 (1.6)
ESR (mm/hr), median (IQR)	9 (5 – 21)	9 (5 – 21)	9 (5 – 21)	18 (10 – 34)
SJC28, median (IQR)	0 (0 – 2)	0 (0 – 2)	0 (0 – 2)	1 (0 – 4)
TJC28, median (IQR)	0 (0 – 2)	0 (0 – 2)	0 (0 – 2)	2 (0 – 8)
patient VAS, median (IQR)	30 (10 – 50)	30 (10 – 50)	28 (10 – 50)	40 (20 – 65)
OST-score, mean (SD)	12.6 (5.0)	12.7 (5.1)	12.6 (5.0)	15.0 (6.1)

SD= standard deviation; Seropositivity= presence of rheumatoid factor and/or anti-cyclic citrullinated peptide antibodies; IQR= interquartile range; DAS28= disease activity score assessing 28 joints; ESR= erythrocyte sedimentation rate; SJC28= swollen joint count assessing 28 joints; TJC28= tender joint count assessing 28 joints; VAS= visual analogue scale of general health, range 0-100=worst; OST = optical spectral transmission, score range 0-66=worst.

Model development

The derived formulae for the different DAS-OST indices are shown in Table 2. Next to ESR and VAS, gender was found to influence DAS28-ESR, independently from OST-score. None of the variables modified the association between OST-score and DAS28. The model without patient VAS had a markedly lower explained variance (i.e., 48% vs. 78% in models with VAS). The analysis on the association between joint specific OST-scores and presence of swelling showed that it was not possible to establish one cut-off for all joints, therefore cut-offs were established per individual joint to calculate a JC-OST, see Supplementary Table S1. This JC-OST was subsequently used to derive a DAS-OST(JC) formula, see Table 2.

Table 2 Developed DAS-OST formulae using DAS28-ESR as reference.*

Developed index	Formula	Explained variance of model (%)	Optimal cut-off for Boolean remission
DAS-OST	-0.44 + OST-score * 0.03 + male sex * -0.11 + LN(ESR) * 0.77 + VAS * 0.03	78%	2.2
DAS-OST(JC)	-0.34 + √JC-OST * 0.15 + male sex * -0.09 + LN(ESR) * 0.77 + VAS * 0.03	78%	2.0
DAS-OST without patient VAS	-0.11 + OST-score * 0.04 + male sex * -0.25 + LN(ESR) * 0.88	48%	2.6

*formulae derived from development cohort (i.e., 2238 observations from MMC cohort), DAS28-ESR was used as reference in all models.

OST= optical spectral transmission; ESR= erythrocyte sedimentation rate; VAS= patients' visual analogue scale of general health; JC-OST= joint count, assessing 22 joints for inflammation, based on optical spectral transmission.

Model validation

Agreement

Table 3 shows agreement of the different DAS-OST indices with DAS28. In line with above results DAS-OST including VAS showed higher agreement with the original DAS28-ESR compared to DAS-OST without patient VAS (ICC 0.88 and 0.82 in internal and external validation vs. 0.66 and 0.49) and measurement error was larger with DAS-OST without patient VAS (in external validation 0.87 vs. 1.28). Agreement and measurement error were similar and difference not statistically significant for models using total OST-score and JC-OST to derive DAS-OST, see Table 3 and Supplementary Figure S1-6.

Table 3 Agreement of DAS-OST indices with DAS28 and measurement error in internal and external validation cohort.

Validation cohort	Developed index	Measurement error	Agreement (ICC (95%CI))
Internal	DAS-OST	0.58	0.88 (0.87 – 0.90)
	DAS-OST(JC)	0.58	0.88 (0.87 – 0.89)
	DAS-OST without patient VAS	0.90	0.66 (0.62 – 0.69)
External	DAS-OST	0.87	0.82 (0.75 – 0.86)
	DAS-OST(JC)	0.87	0.81 (0.75 – 0.86)
	DAS-OST without patient VAS	1.28	0.49 (0.36 – 0.60)

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort. External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST consisted of total OST-score, gender, ESR and VAS; DAS-OST(JC) consisted of √JC-OST, gender, ESR and VAS; DAS-OST without patient VAS consisted of total OST-score, gender and ESR.

ICC= random two-way mixed effect intra-class correlation coefficient; CI= confidence interval.

Diagnostic value

Overall discrimination and diagnostic accuracy measures of DAS-OST indices for DAS28-based and Boolean remission, LDA and HDA are shown in Table 4. Discriminatory ability for DAS-OST value was generally high, ranging from 0.86 to 0.97 for indices including VAS in internal validation, and from 0.92 to 0.95 in external validation. Unfortunately, the AU-ROC for Boolean remission could not be calculated in the external validation cohort, since no patient was in this disease state. In line with agreement and measurement error results, AU-ROC and diagnostic accuracy measures of DAS-OST without patient VAS were lower.

Sensitivity and specificity for all disease states were in general good, but sensitivity for HDA was lower and 95% confidence interval wider, compared to DAS28-based remission and LDA, Table 4. Using the DAS-OST without patient VAS cut-off, no patient was classified into the disease state HDA.

Table 4 Diagnostic values of DAS-OST formula in internal and external validation cohort.

Validation cohort	Disease activity state	AU-ROC	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	accuracy
Internal	<i>DAS-OST</i>						
	Remission	0.93	0.84 (0.81 - 0.86)	0.86 (0.83 - 0.89)	0.89 (0.87 - 0.92)	0.79 (0.76 - 0.83)	0.85
	Boolean	0.87	0.83 (0.78 - 0.88)	0.70 (0.67 - 0.73)	0.41 (0.36 - 0.45)	0.94 (0.93 - 0.96)	0.73
	LDA	0.92	0.88 (0.86 - 0.90)	0.75 (0.70 - 0.80)	0.91 (0.89 - 0.93)	0.69 (0.64 - 0.74)	0.85
	HDA	0.97	0.49 (0.33 - 0.65)	0.99 (0.99 - 1.00)	0.75 (0.58 - 0.92)	0.98 (0.97 - 0.99)	0.98
	<i>DAS-OST(jc)</i>						
	Remission	0.94	0.84 (0.82 - 0.87)	0.86 (0.83 - 0.89)	0.89 (0.87 - 0.92)	0.80 (0.76 - 0.83)	0.85
	Boolean	0.86	0.75 (0.69 - 0.81)	0.77 (0.74 - 0.80)	0.45 (0.39 - 0.50)	0.93 (0.91 - 0.95)	0.77
	LDA	0.92	0.88 (0.86 - 0.90)	0.74 (0.69 - 0.79)	0.91 (0.89 - 0.93)	0.69 (0.64 - 0.74)	0.85
	HDA	0.97	0.49 (0.33 - 0.65)	1.00 (0.99 - 1.00)	0.78 (0.61 - 0.95)	0.98 (0.97 - 0.99)	0.98
	<i>DAS-OST without patient VAS</i>						
	Remission	0.82	0.83 (0.80 - 0.86)	0.65 (0.61 - 0.70)	0.77 (0.74 - 0.80)	0.73 (0.69 - 0.77)	0.75
	Boolean	0.64	0.78 (0.72 - 0.83)	0.41 (0.38 - 0.45)	0.25 (0.21 - 0.28)	0.88 (0.85 - 0.91)	0.49
	LDA	0.80	0.92 (0.91 - 0.94)	0.39 (0.33 - 0.45)	0.81 (0.79 - 0.84)	0.65 (0.58 - 0.72)	0.79
	HDA	0.88	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	n.a.	0.97 (0.96 - 0.98)	0.97

Table 4 (continued)

Validation cohort	Disease activity state	AU-ROC	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	accuracy
<i>DAS-OST</i>							
External	Remission	0.95	0.79 (0.67 - 0.91)	0.92 (0.86 - 0.97)	0.79 (0.67 - 0.91)	0.92 (0.86 - 0.97)	0.88
	Boolean	-	-	-	-	-	-
	LDA	0.93	0.91 (0.84 - 0.99)	0.80 (0.72 - 0.88)	0.75 (0.65 - 0.85)	0.94 (0.88 - 0.99)	0.85
	HDA	0.92	0.43 (0.26 - 0.59)	0.96 (0.93 - 1.00)	0.79 (0.61 - 0.97)	0.85 (0.78 - 0.91)	0.84
<i>DAS-OST(jC)</i>							
	Remission	0.95	0.84 (0.73 - 0.95)	0.91 (0.84 - 0.95)	0.77 (0.64 - 0.89)	0.93 (0.88 - 0.98)	0.89
	Boolean	-	-	-	-	-	-
	LDA	0.93	0.91 (0.84 - 0.99)	0.80 (0.72 - 0.89)	0.75 (0.65 - 0.85)	0.94 (0.88 - 0.99)	0.85
	HDA	0.92	0.42 (0.26 - 0.58)	0.96 (0.93 - 1.00)	0.79 (0.61 - 0.97)	0.84 (0.78 - 0.90)	0.84
<i>DAS-OST without patient VAS</i>							
	Remission	0.75	0.58 (0.43 - 0.73)	0.73 (0.64 - 0.81)	0.46 (0.33 - 0.60)	0.81 (0.73 - 0.89)	0.68
	Boolean	-	-	-	-	-	-
	LDA	0.74	0.93 (0.87 - 1.00)	0.47 (0.37 - 0.58)	0.53 (0.43 - 0.63)	0.91 (0.84 - 0.99)	0.65
	HDA	0.83	0.00 (0.00 - 0.00)	1.00 (1.00 - 1.00)	n.a.	0.77 (0.70 - 0.83)	0.77

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort. External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST consisted of total OST-score, gender, ESR and VAS; DAS-OST(jC) consisted of \sqrt{jC} -OST, gender, ESR and VAS; DAS-OST without patient VAS consisted of total OST-score, gender and ESR.

AU-ROC= area under the receiver operating characteristic curve; DAS28 Remission= DAS28 ≤ 2.6 ; LDA= low disease activity (DAS28 ≤ 3.2); HDA= high disease activity (DAS28 > 5.1); PPV= positive predictive value; NPV= negative predictive value; CI= confidence interval.

DISCUSSION

We developed and validated three disease activity indices using OST. DAS-OST(JC) performed similarly as DAS-OST, and therefore we prefer DAS-OST as this index has the most simple formula. DAS-OST without patient VAS performed significantly less well. In general, the (diagnostic) performance was found to be good and results remained well in the external validation (ACURA) cohort including patients with a more active disease compared to the out-patient-clinic patients of the MMC cohort. The measurement error of DAS28 is generally assumed to be around 0.6,^{12,13} in the validation cohorts similar and slightly higher measurement errors between DAS28 and DAS-OST were found.^{12,13} Misclassification occurred in both cohorts, but mainly regarding the low- and moderate disease activity categories and less regarding remission and high disease activity states, see Supplementary Table S2 and S3.

Diagnostic performance of DAS-OST for defining specific disease activity states was in general good, although specifically sensitivity for HDA was low, indicating that the performance in ruling out HDA using this index might be suboptimal (compared to DAS28, the reference). However, the primary use would be intensification of treatment strategy in absence of remission and/or LDA, and thus diagnostic accuracy for these outcomes are most important. Sensitivity and specificity for remission and LDA were adequate. Therefore we assume that our disease activity index, DAS-OST, might save the rheumatologists' time regarding evaluation of the disease activity of their patients in clinical practice.

Interestingly, in our analyses sensitivity and especially specificity were low when using Boolean remission as outcome for DAS-OST. Therefore, we calculated sensitivity and specificity for DAS28 itself to detect Boolean remission (data not shown). In this analysis sensitivity and specificity were also low(er) and similar with those of DAS-OST (87% and 68%, respectively).

A limitation of this study is the lack of information on smoking status as it is known that smoking decreases the blood flow of digits, which could influence the HandScan score.⁶ Furthermore, data on BMI and/or hand-size, which have been found to affect the HandScan outcome¹⁴ was absent. By including gender in our models, we may have partly corrected for the effect of hand size, which was found to influence the OST-score independently from the disease activity variables in DAS-OST, but it would be interesting to investigate whether hand size or BMI would have an additional effect.

Despite these limitations, the current study containing information of 1505 unique RA patients (with multiple observations) and an external validation cohort provided a unique opportunity to develop and validate a disease activity index using the HandScan. The fact that these were data of daily clinical practice of unselected patients enhances generalisability of results.

Conclusion

Using the HandScan, RA disease activity (states) can be accurately estimated if OST-scores are combined with ESR, VAS and gender into a disease activity index (DAS-OST), which results in a quick and more objective disease activity index compared to joint count based disease activity indices.

REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76**, 960–977 (2017).
2. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
3. Jacobs, J. W. G., ten Cate, D. F. & van Laar, J. M. Monitoring of rheumatoid arthritis disease activity in individual patients: still a hurdle when implementing the treat-to-target principle in daily clinical practice. *Rheumatology* **54**, 959–961 (2015).
4. Scott, D. L. & Houssien, D. A. Joint assessment in rheumatoid arthritis. *Br. J. Rheumatol.* **35**, 14–18 (1996).
5. Grunke, M. *et al.* Standardization of joint examination technique leads to a significant decrease in variability among different examiners. *J. Rheumatol.* **37**, 860–864 (2010).
6. Besselink, N. J. *et al.* Optical spectral transmission to assess inflammation in hand and wrist joints of rheumatoid arthritis patients. *Rheumatology* **57**, 865–872 (2018).
7. Besselink, N. J. *et al.* Novel optical spectral transmission (OST)-guided versus conventionally disease activity-guided treatment: study protocol of a randomized clinical trial on guidance of a treat-to-target strategy for early rheumatoid arthritis. *Trials* **20**, 226 (2019).
8. van Onna, M. *et al.* Assessment of disease activity in patients with rheumatoid arthritis using optical spectral transmission measurements, a non-invasive imaging technique. *Ann. Rheum. Dis.* **75**, 511–518 (2016).
9. Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).
10. Balsa, A. *et al.* Value of disease activity score 28 (DAS28) and DAS28-3 compared to American College of Rheumatology-defined remission in rheumatoid arthritis. *J. Rheumatol.* **31**, 40–46 (2004).
11. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
12. de Vet, H. C. W., Terwee, C. B., Mokkink, L. B. & Knol, D. L. *Measurement in Medicine*. (Cambridge University Press, 2011).
13. van Riel, P. L. C. M. The development of the disease activity score (DAS) and the disease activity score using 28 joint counts (DAS28). *Clin. Exp. Rheumatol.* **32**, S65–S74 (2014).
14. Triantafyllias, K., Heller, C., de Blasi, M., Galle, P. & Schwarting, A. Diagnostic value of optical spectral transmission in rheumatoid arthritis: associations with clinical characteristics and comparison with joint ultrasonography. *J. Rheumatol.* **47**, 1314–1322 (2020).

SUPPLEMENTARY FILE

Supplementary Data S1: Optimal cut-off estimation of CRP ≤10mg/L based on ESR

Since no data about CRP was available in the dataset of the MMC cohort, it was not possible to calculate directly remission using the Boolean criteria. Therefore, we used ESR to estimate CRP ≤10mg/L, and thereafter this estimate to calculate remission, according to the Boolean criteria. We obtained data including ESR and CRP (collected at the same moment) of 1646 (590 male; 1056 female) RA patients visiting the out-patient-clinic in the Máxima MC from April 2017 to including September 2019, but who were not part of our MMC cohort. We used a linear regression model with a natural log (LN) transformed CRP as outcome and a LN transformed ESR and gender as independent variables. The acute phase criterion of the Boolean remission criteria was defined as LN transformed, estimated CRP ≤10mg/L (i.e., ≤2.302585093).

$$\text{LN (estimated CRP)} = -1.08297 + 0.92 * \text{LN(BSE)} + 0.55 * \text{male}$$

Supplementary Table S1 Established OST-score cut-off per joint using Youden's index for swelling per joint with diagnostic values in the internal validation cohort.

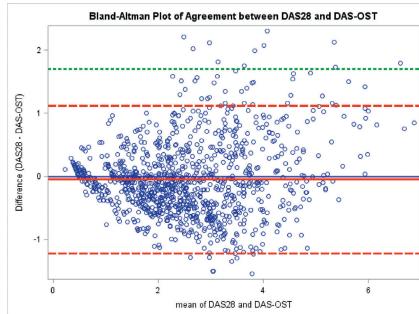
Joint	OST-score cut-off*	AU-ROC	Sensitivity	Specificity	PPV	NPV
wrist	1.0	0.55	0.37	0.72	0.13	0.92
MCP1	1.4	0.54	0.20	0.89	0.07	0.96
MCP2	1.0	0.65	0.42	0.87	0.27	0.93
MCP3	0.7	0.66	0.50	0.82	0.15	0.96
MPC4	1.3	0.58	0.23	0.96	0.07	0.99
MCP5	1.5	0.63	0.35	0.91	0.08	0.98
IP1	1.7	0.51	0.00	0.98	0.00	0.99
PIP2	0.9	0.65	0.35	0.90	0.19	0.96
PIP3	0.5	0.60	0.40	0.81	0.15	0.94
PIP4	1.0	0.63	0.39	0.88	0.09	0.98
PIP5	1.2	0.54	0.15	0.93	0.07	0.97

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort.

*if OST-score for specific joint ≥ cut-off, then that specific joint is estimated as inflamed.

AU-ROC= area under the receiver operating characteristic curve; PPV= positive predicted value;

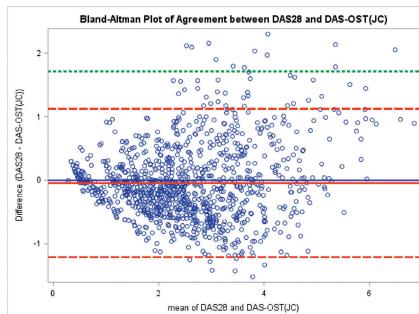
NPV= negative predicted value.



Supplementary Figure S1 Bland-Altman plot for difference (DAS28 – DAS-OST) using data of internal validation cohort.

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort. DAS-OST consisted of total OST-score, gender, ESR and VAS.

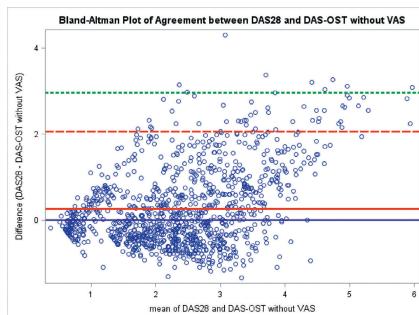
Blue line= 0; Red line= mean of the difference; Red lines (dashed)= +/- 2SD of mean of the difference; Green line (dashed)= + 3SD of mean of the difference.



Supplementary Figure S2 Bland-Altman plot for difference (DAS28 – DAS-OST(JC)) using data of internal validation cohort.

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort. DAS-OST(JC) consisted of √JC-OST, gender, ESR and VAS.

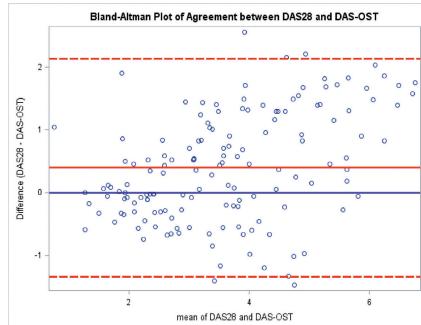
Blue line= 0; Red line= mean of the difference; Red lines (dashed)= +/- 2SD of mean of the difference; Green line (dashed)= + 3SD of mean of the difference.



Supplementary Figure S3 Bland-Altman plot for difference (DAS28 – DAS-OST without VAS) using data of internal validation cohort.

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort. DAS-OST without VAS consisted of total OST-score, gender and ESR.

Blue line= 0; Red line= mean of the difference; Red line (dashed)= + 2SD of mean of the difference; Green line (dashed)= + 3SD of mean of the difference.

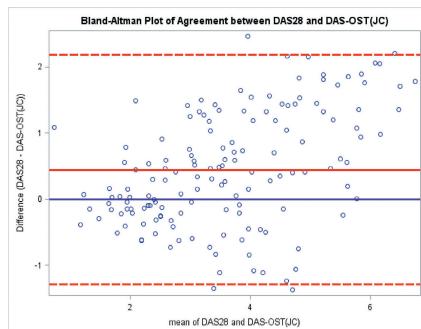


Supplementary Figure S4 Bland-Altman plot for difference (DAS28 – DAS-OST) using data of external validation cohort.

External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST consisted of total OST-score, gender, ESR and VAS.

Blue line= 0; Red line= mean of the difference; Red lines (dashed)= +/- 2SD of mean of the difference.

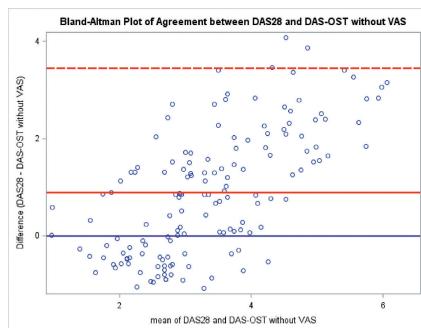


Supplementary Figure S5 Bland-Altman plot for difference (DAS28 – DAS-OST(JC)) using data of external validation cohort.

External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST(JC) consisted of \sqrt{JC} -OST, gender, ESR and VAS.

Blue line= 0; Red line= mean of the difference; Red lines (dashed)= +/- 2SD of mean of the difference.



Supplementary Figure S6 Bland-Altman plot for difference (DAS28 – DAS-OST without VAS) using data of external validation cohort.

External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST without VAS consisted of total OST-score, gender and ESR.

Blue line= 0; Red line= mean of the difference; Red line (dashed)= + 2SD of mean of the difference.

Supplementary Table S2 Number of patients classified per disease activity state based on DAS28 and DAS-OST in internal validation.

	Remission	LDA	MDA	HDA	misclassification per disease activity state
DAS-OST remission	<u>542</u>	43	24	0	11%
DAS-OST LDA	89	<u>52</u>	49	1	73%
DAS-OST MDA	17	83	<u>178</u>	18	40%
DAS-OST HDA	0	0	6	<u>18</u>	25%
overestimation		12%			
underestimation		17%			
total misclassification		29%			
weighted Cohen's Kappa*		0.65			

Internal validation cohort consisted of 1120 observations from MMC cohort, not used in development cohort.

DAS-OST consisted of total OST-score, gender, ESR and patient VAS.

*weights for the weighted Cohen's Kappa were: no difference between category: weight=0; difference is 1 category (next to reference category): weight=1; difference is 2 categories (next to reference category): weight=2; difference is 3 categories (next to reference category): weight=3.

LDA= low disease activity; MDA= moderate disease activity; HDA= high disease activity.

Supplementary Table S3 Number of patients classified per disease activity state based on DAS28 and DAS-OST in external validation.

	Remission	LDA	MDA	HDA	misclassification per disease activity state
DAS-OST remission	<u>34</u>	6	3	0	21%
DAS-OST LDA	9	<u>4</u>	14	1	86%
DAS-OST MDA	0	5	<u>35</u>	19	41%
DAS-OST HDA	0	0	4	<u>16</u>	20%
overestimation		29%			
underestimation		12%			
total misclassification		41%			
weighted Cohen's Kappa*		0.64			

External validation cohort consisted of 151 observations from the ACURA cohort.

DAS-OST consisted of total OST-score, gender, ESR and patient VAS.

*weights for the weighted Cohen's Kappa were: no difference between categories: weight=0; difference is 1 category (next to reference category): weight=1; difference is 2 categories (next to reference category): weight=2; difference is 3 categories (next to reference category): weight=3.

LDA= low disease activity; MDA= moderate disease activity; HDA= high disease activity.

Chapter 11:

Classification of RA into active or inactive disease, using a modified DAS with a HandScan score replacing joint counts.

M.M.A. Verhoeven¹

A.A.A. Westgeest²

J. Tekstra¹

J.M. van Laar¹

F.P.J.G. Lafeber¹

P.M.J. Welsing¹

J.W.G. Jacobs¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center

Utrecht, Utrecht University, the Netherlands;

²Department of Rheumatology, Máxima Medical Center, Eindhoven, the Netherlands.

Submitted for publication

ABSTRACT

Objective

To establish the value of a modified DAS (DAS-OST) without joint counts but with a HandScan score (OST), versus that of DAS28, to classify RA as active versus inactive, with as reference standard the rheumatologist's clinical classification.

Methods

RA patients with at least one HandScan and DAS28 measurement performed at the same visit were included. Data was extracted from medical records, as was the clinical interpretation as active or inactive RA by the rheumatologist. Logistic regression analyses were performed to calculate areas under the receiver operating characteristics (AU-ROC) curves. The clinical interpretation was used as reference standard in all analyses, and disease activity measures were used as predictor variables. The performance of predictor variables (AU-ROCs) was compared.

Results

Data of 1505 unique RA patients were used for analyses. The highest AU-ROC of 0.88 (95%CI 0.85 – 0.90) was shown for DAS28; AU-ROC of DAS-OST was 0.78 (95%CI 0.75 – 0.81), difference 0.10, $p < 0.01$. When removing OST-score from DAS-OST, AU-ROC decreased from 0.78 to 0.77.

Conclusion

Compared to DAS28, DAS-OST classified RA statistically significantly less well as active versus inactive, when using the clinical classification as reference standard. The contribution of the OST-score to DAS-OST was negligible in this setting.

INTRODUCTION

In rheumatoid arthritis (RA), life-long treatment according to the tight-control and treat-to-target principle is preferred, requiring frequent out-patient-clinic visits and contacts with rheumatologists.¹ Then, disease activity is usually assessed with DAS28,² a time consuming and partly subjective method. Moreover, training/standardisation of joint examinations, preferably yearly, is required to assess joints for swelling and tenderness as objectively as possible.³⁻⁶

The HandScan is a medical device measuring inflammation in both hand and wrists using optical spectral transmission (OST, score 0 to 66=worse). The OST-assessment takes 5 minutes without much involvement of a health professional.⁷ Some patients experience joint count assessments as painful, but the HandScan measurement is unpainful, an additional advantage.

Recently, to quickly assess disease activity in RA, using DAS28-ESR as reference, we developed and validated an index consisting of the OST-score, gender, erythrocyte sedimentation rate (ESR), and patients' general health on a visual analogue scale (VAS GH). This modification of DAS28, with OST-score replacing joint counts, was named DAS-OST.⁸ Its formula is $-0.44 + \text{OST-score} \times 0.03 + \text{male sex} \times -0.11 + \ln(\text{ESR}) \times 0.77 + \text{VAS GH} \times 0.03$, and its explained variance in the external validation cohort was 71%, using DAS28-ESR as reference. To define remission (i.e., DAS28≤2.6) DAS-OST had an area under the receiver operating characteristic curve (AU-ROC) of 0.95 (95%CI 0.91-0.98); sensitivity was 79%, specificity 92%, and accuracy 88%.

In the tight-control and treat-to-target principles, the decision whether RA is active or not is paramount. If there would be a tool to classify RA into active or inactive, the outpatients' contacts with the rheumatologist might be limited to patients with active RA and/or medical problems.¹

The aim of this study was establishing the value of DAS-OST, versus that of DAS28, to classify RA as active versus inactive, with the rheumatologist's clinical classification as active/inactive as reference standard, which is more comprehensive than disease activity indices alone.⁹

METHODS

This is a single center (Máxima MC Eindhoven; MMC) study, using routinely collected data from electronic medical records. The institutional ethical review board of MMC confirmed that the Medical Research Involving Human Subject Act (WMO) was not applicable to the protocol of this study (N19.002).

Patients

Detailed information on patients and obtained data has been reported elsewhere.⁸ For this study, a rheumatologist (AAAW) assessed the medical records for the treating rheumatologist's clinical impressions on disease activity. Methods and results of the assessments for this clinical classification are provided in the Supplementary file, Data S1. In short, 'active' was characterised by initiation of a new disease modifying anti-rheumatic drug (DMARD), or an increase of the dose of current DMARD, or intramuscular glucocorticoid injection, or the treating rheumatologist's notes indicating active disease, all at the consultation. 'Inactive' was characterised by tapering or stopping the current DMARD because of absence of disease activity (not because of adverse effects) or the treating rheumatologist's notes indicating inactive disease. A randomly drawn subset of 39 clinic visits assessed by AAAW was blindly re-assessed by two other rheumatologists (JT and JWGJ); agreement on the clinical classification was overall 92.5%.

Statistical analyses

Continuous variables were described using means with standard deviations (SD) or medians with interquartile ranges (IQR), where appropriate. Frequencies and proportions were calculated for categorical variables.

DAS28 was plotted against DAS-OST in a scatter plot, and agreement of DAS28 with DAS-OST was tested using a random one-way intra correlation coefficient (ICC). The mean of the differences between DAS28 and DAS-OST was illustrated in a Bland-Altman plot.

The value DAS-OST and DAS28 to classify RA as active or inactive was established applying logistic regression analyses and calculating AU-ROC, using the rheumatologist's clinical classification as reference standard. Similar analyses were performed with OST-scores only, to evaluate the additional value of OST in DAS-OST. AU-ROCs were tested for statistically significant differences. For remission (≤ 2.6) and low disease activity (LDA, ≤ 3.2), for both DAS28 and DAS-OST, sensitivity, specificity, positive- (PPV), negative predictive values (NPV) and accuracy were calculated. Youden's index was used to define optimal cut-offs for OST-scores, for inactive disease for males and females separately. Agreement between remission and LDA according to DAS28 and DAS-OST and the rheumatologist's clinical classification (reference standard) was tested using Cohen's kappa and Gwet's AC1.¹⁰

All tests were two-sided, a p-value ≤ 0.05 was considered as statistically significant. Analyses were performed in SAS version 9.4, NCSS version 12.0.12, R version 3.6.2 with the package pROC 1.16.1, and AgreeStat version 2015.6.1.

RESULTS

Patients

Data of 1505 unique RA patients was used. Patients' demographic and clinical data are shown in Table 1. Mean DAS28 and DAS-OST were both 2.5. Mean OST-score was somewhat higher in males than in females. According to clinical classification, RA of 79% of the patients was classified as inactive; 7 patients could not be clinically classified.

Table 1 Patients' demographics and clinical data.

	All	Female	Male
Patients' demographics			
Number of patients, n (%)	1505 (100)	979 (65)	526 (35)
Age (years), mean (SD)	65.5 (12.1)	64.1 (12.7)	68.0 (10.6)
Duration of RA (years), mean (SD)	11.5 (8.3)	11.8 (8.3)	11.0 (8.5)
Seropositivity, n (%)	1068 (71)	691 (71)	377 (72)
Clinical data			
DAS28, mean (SD)	2.5 (1.3)	2.6 (1.2)	2.3 (1.3)
ESR (mm in 1 st hour), median (IQR)	9.0 (5.0 – 21.0)	9.0 (5.0 – 21.0)	8.0 (2.0 – 19.0)
SJC28, median (IQR)	0 (0 – 2)	0 (0 – 2)	0 (0 – 2)
TJC28, median (IQR)	0 (0 – 1)	0 (0 – 2)	0 (0 – 1)
VAS global, median (IQR)	25.0 (10.0 – 50.0)	30.0 (10.0 – 50.0)	20.0 (10.0 – 43.0)
DAS-OST, mean (SD)	2.5 (1.2)	2.6 (1.2)	2.3 (1.2)
OST-score, mean (SD)	12.7 (5.3)	11.3 (4.7)	15.2 (5.3)
Clinical classification, n (%)			
- Active	295 (20)	195 (20)	100 (19)
- Inactive	1203 (79)	781 (79)	422 (80)
- Not interpretable	7 (1)	3 (1)	4 (1)

SD= standard deviation; Seropositivity= presence of rheumatoid factor and/or anti-cyclic citrullinated peptide antibodies; IQR= interquartile range; DAS28= disease activity score assessing 28 joints; ESR= erythrocyte sedimentation rate; SJC28= swollen joint count assessing 28 joints; TJC28= tender joint count assessing 28 joints; VAS= patients' visual analogue scale of general health, range 0-100=worst; DAS-OST= disease activity score using optical spectral transmission; OST= optical spectral transmission, range 0-66=worst.

Agreement

The relationship between DAS28 and DAS-OST is graphically shown in Supplementary Figure S1; the ICC was 0.88 (95%CI 0.87-0.89). The mean difference between DAS28 and DAS-OST is shown in Supplementary Figure S2.

Discriminatory values

AU-ROC to classify RA patients as active/inactive for DAS28 was 0.88 (95%CI 0.85-0.90) and for DAS-OST and single OST-score, 0.78 (95%CI 0.75-0.81) and 0.59 (95%CI

0.55-0.63), respectively, Table 2. The discriminative ability was statistically significantly higher for DAS28, compared to DAS-OST: $\Delta AU-ROC$ 0.10, 95%CI 0.08-0.12, whereas it was statistically significantly higher for DAS-OST, compared to the OST-score: $\Delta AU-ROC$ 0.19, 95%CI 0.14-0.23, see Table 2 and Figure 1.

Table 2 Area under the receiver operating characteristics curves for individual models and comparisons.

ROC model	AU-ROC (95%CI)	p-value
Individual model		
DAS28	0.88 (0.85 – 0.90)	<0.01
DAS-OST	0.78 (0.75 – 0.81)	<0.01
OST-score	0.59 (0.55 – 0.63)	<0.01
Comparisons		
Difference: DAS28 minus DAS-OST	0.10 (0.08 – 0.12)	<0.01
Difference: DAS-OST minus OST-score	0.19 (0.14 – 0.23)	<0.01

AU-ROC= Area under the receiver operating characteristics curve; CI= confidence interval; DAS28= disease activity score assessing 28 joints; DAS-OST= disease activity using optical spectral transmission; OST-score= optical spectral transmission score.

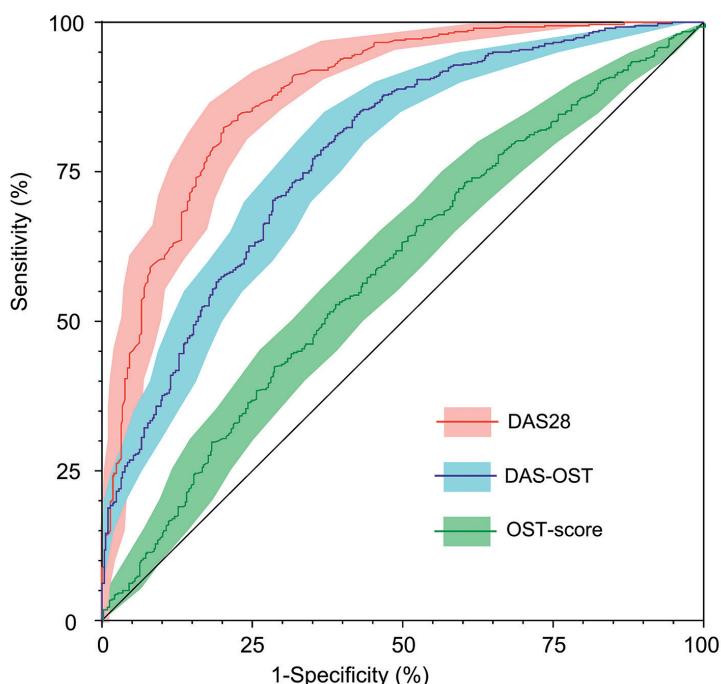


Figure 1 Receiver operating characteristics curves with 95% confidence areas of sensitivity. The classification as 'not active disease' according to the clinical interpretation was used as reference. DAS28= disease activity score assessing 28 joints; DAS-OST= disease activity using optical spectral transmission; OST-score= optical spectral transmission score.

For DAS28-based LDA classification, sensitivity and specificity were 87% and 72%, respectively and for DAS-OST-based LDA, these values were 81% and 59%, respectively, all versus the clinical reference. Lower values were obtained for both indices when using remission instead of LDA, see supplementary Table S1. Optimal cut-offs for inactive disease according OST-scores were ≤ 16.36 and ≤ 10.64 for males and females, respectively. Sensitivity and specificity for males were 66% and 50%, respectively, and 53% and 63%, respectively, for females, see supplementary Table S1. In line with the outcomes above, agreement (Cohen's kappa/Gwet's AC1) of the clinical classification with DAS28-based classification was higher compared with DAS-OST (0.41/0.55 and 0.54/0.76 vs. 0.24/0.37 and 0.36/0.64 for remission and LDA, respectively), see supplementary Table S1.

DISCUSSION

To our knowledge this is the first study that identifies the value of a disease activity measure with a HandScan score replacing joint counts to classify RA as active versus inactive, with as reference standard the rheumatologist's clinical classification. DAS-OST showed a moderate performance (AU-ROC 0.78), but statistically significantly less, compared to DAS28. The performance of OST-score alone was poor, in line with earlier published data,^{8,11} indicating that the OST-score has to be combined with other parameters into an index.

As expected, more favourable outcomes were obtained for DAS28-based LDA than for DAS-OST-based LDA. False negatives, as well as false positives were more often observed for DAS-OST than for DAS28 (4/20 vs. 2/20 and 2/20 vs. 1/20, respectively). Similarly, strength of agreement of the rheumatologist's classification with DAS28-based classification was moderate, and it was modest with DAS-OST. This limits the potential clinical applicability of DAS-OST. As a face-to-face visit with the rheumatologist is not required for DAS-OST, which saves time and cost, it would be efficient to select patients having no LDA based on DAS-OST for a face-to-face visit with the rheumatologist, assuming that most of those patients have active disease. In this setting, a higher specificity of DAS-OST (thus low number of patients with missed active RA) would preferable over a high sensitivity, because at the visit, false classifications of RA as active can be corrected. Cost might also be saved when using DAS-OST compared to DAS28, as healthcare workers/rheumatologists should be trained to perform joint count assessment as objective as possible.³ Of course, the HandScan machine has to be purchased.

The difference for males and females in optimal cut-off of OST-score for inactive disease is remarkable. Our previous study,⁸ as well as the current study show that OST-scores were higher in males than in females. A plausible explanation is the difference in size and volume of hands between men and women.¹²

A limitation of our study is that our reference, i.e., rheumatologist's clinical classification as active or inactive RA, probably is partly dependent of DAS28, as this

measurement is widely applied in the Netherlands at out-patient-clinic visits. This could have favoured our study results regarding DAS28, but not regarding DAS-OST. The results of the OST-score only apply utilisation of the HandScan in the setting of this study, e.g., categorizing disease as inactive or active.

Strength of our study is the relative large sample size of 1505 unique RA patients, with limited uninterpretable data.

Conclusion

Compared to DAS28, DAS-OST classified RA statistically significantly less well as active versus inactive, when using the clinical classification as reference standard, but it could be used in a strategy to limit outpatients' visits to the rheumatologist. The discriminative ability of OST-score alone was negligible in this setting.

REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
2. Felson, D. T. *et al.* American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis Rheum.* **63**, 573–586 (2011).
3. Jacobs, J. W. G., ten Cate, D. F. & van Laar, J. M. Monitoring of rheumatoid arthritis disease activity in individual patients: still a hurdle when implementing the treat-to-target principle in daily clinical practice. *Rheumatology* **54**, 959–961 (2015).
4. Grunke, M. *et al.* Standardization of joint examination technique leads to a significant decrease in variability among different examiners. *J. Rheumatol.* **37**, 860–864 (2010).
5. Scott, D. L. & Houssien, D. A. Joint assessment in rheumatoid arthritis. *Br. J. Rheumatol.* **35**, 14–18 (1996).
6. Porter, D. *et al.* DAS28 and rheumatoid arthritis: the need for standardization. *Musculoskeletal Care* **9**, 222–227 (2011).
7. Besseling, N. J. *et al.* Novel optical spectral transmission (OST)-guided versus conventionally disease activity-guided treatment: study protocol of a randomized clinical trial on guidance of a treat-to-target strategy for early rheumatoid arthritis. *Trials* **20**, 226 (2019).
8. Verhoeven, M. M. A. *et al.* Development and validation of rheumatoid arthritis disease activity indices including HandScan (optical spectral transmission) scores. *Revision submitted to Arthritis Care Res.*
9. Burmester, G. R. & Pope, J. E. Novel treatment strategies in rheumatoid arthritis. *Lancet* **389**, 2338–2348 (2017).
10. Wongpakaran, N., Wongpakaran, T., Wedding, D. & Gwet, K. L. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med. Res. Methodol.* **13**, 61 (2013).
11. Verhoeven, M. M. A. *et al.* Utility of the HandScan in monitoring disease activity and prediction of clinical response in rheumatoid arthritis patients. *In press, Rheumatology Advances in Practice January 2021.*
12. Triantafyllias, K., Heller, C., de Blasi, M., Galle, P. & Schwarting, A. Diagnostic value of optical spectral transmission in rheumatoid arthritis: associations with clinical characteristics and comparison with joint ultrasonography. *J. Rheumatol.* **47**, 1314–1322 (2020).

SUPPLEMENTARY FILE

Supplementary Data S1: Clinical classification

To determine the clinical classification of a physician (or physician assistant), we decided to make two categories; active disease or inactive disease.

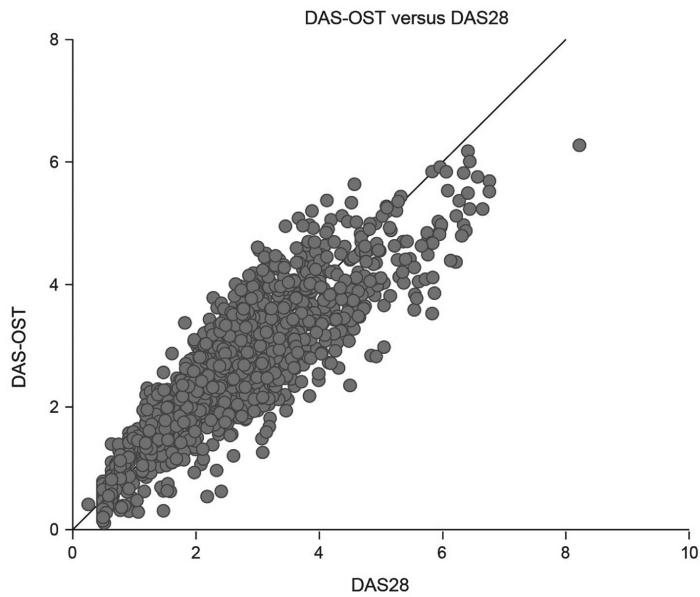
Active is defined as:

- The decision to start a new csDMARD / bDMARD / tsDMARD at present consultation
- Increasing dose of the used DMARD
- Intramuscular injection with glucocorticoid
- When the treating rheumatologist described "Active disease" or used synonym terms

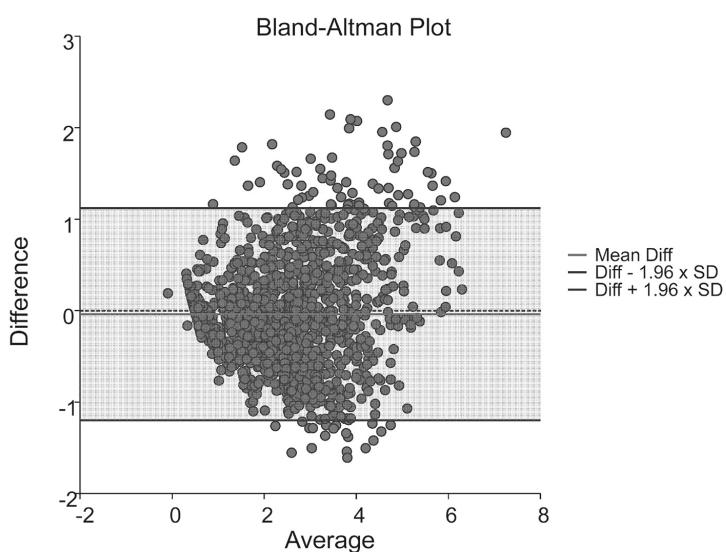
Inactive is defined as:

- Tapering the dose of csDMARD / bDMARD / tsDMARD, not because of adverse-effects
- Stopping csDMARD / bDMARD / tsDMARD, not because of adverse-effects
- When the treating rheumatologist described "remission" or used synonym terms

One rheumatologist, AAAW, went cross-sectionally through the medical records of the participating patients to assess disease activity as active/inactive based on his clinical interpretation on the records. One of the inclusion criteria for participation was the availability of a DAS28 measurement, with, thereafter, a HandScan measurement, performed during the same clinical visit. In addition, only one unique visit per patient was included. Of the included medical records, a randomly drawn subset of 39 clinic visits was blindly re-assessed by two other rheumatologists, JT and JWGJ for testing agreement on this clinical interpretation, which was overall 92.5% with classifications of AAAW.



Supplementary Figure S1 Scatter plot of the relationship between DAS-OST and DAS28 on patient level.
DAS-OST= disease activity using optical spectral transmission; DAS28= disease activity score assessing 28 joints



Supplementary Figure S2 Bland-Altman plot of the mean difference DAS28 – DAS-OST (y-axis) versus the mean of DAS28 and DAS-OST (x-axis).

DAS28= disease activity score assessing 28 joints; DAS-OST= disease activity using optical spectral transmission; Diff= difference; SD= standard deviation

Supplementary Table S1 Diagnostic values and agreement of clinical interpretation with (modified) disease activity outcomes.

Disease activity state	Diagnostic values					Agreement Gwet's AC1 (95%CI)
	Sensitivity	Specificity	PPV	NPV	Accuracy Cohen's Kappa (95%CI)	
DAS28[#]						
Remission	0.71	0.85	0.95	0.42	0.74	0.41 (0.37 - 0.46)
LDA	0.87	0.72	0.93	0.58	0.84	0.54 (0.49 - 0.59)
DAS-OST^{##}						
Remission	0.62	0.74	0.91	0.32	0.65	0.24 (0.20 - 0.29)
LDA	0.81	0.59	0.89	0.44	0.77	0.36 (0.30 - 0.41)
OST-score						
Optimal cut-off for inactive RA in males (OST-score ≤16.36)	0.66	0.50	0.85	0.26	0.63	0.11 (0.03 - 0.20)
Optimal cut-off for inactive RA in females (OST-score ≤10.64)	0.53	0.63	0.85	0.25	0.55	0.10 (0.05 - 0.15)
PPV= positive predictive value; NPV= negative predictive value; CI= confidence interval; DAS28= disease activity score assessing 28 joints; LDA= low disease activity; DAS-OST= disease activity using optical spectral transmission; OST-score= single measurement of optical spectral transmission.						

[#] Number of patients with; DAS28 remission and inactive= 858, DAS28 remission and active= 43, no DAS28 remission and inactive= 345, no DAS28 remission and active=252.

DAS28 LDA and inactive= 1048, DAS28 LDA and active= 84, no DAS28 LDA and inactive= 155, no DAS28 LDA and active=211.

^{##} Number of patients with; DAS-OST remission and inactive= 750, DAS-OST remission and active= 77, no DAS-OST remission and inactive= 453, no DAS-OST remission and active=218

DAS-OST LDA and inactive= 977, DAS-OST LDA and active= 120, no DAS-OST LDA and inactive= 226, no DAS-OST LDA and active=175.

Chapter 12:

A gender difference in HandScan scores in rheumatoid arthritis patients and controls?

M.M.A. Verhoeven¹

A.A.A. Westgeest²

J.W.G. Jacobs¹

¹Department of Rheumatology & Clinical Immunology, University Medical Center

Utrecht, Utrecht University, the Netherlands;

²Department of Rheumatology, Máxima Medical Center, Eindhoven, the Netherlands.

In press, accepted for publication in The Journal of Rheumatology December 2020

LETTER

To the Editor: A recent paper of *Triantafyllias et al.* described that optical spectral transmission (OST-) scores, obtained by the HandScan, were significantly higher in male compared to female rheumatoid arthritis (RA) patients and controls, and an association between OST-score and age, BMI, and hand surface area was found.¹

The difference in gender that *Triantafyllias et al.* showed is in line with results of studies performed at our department. We developed and validated a new disease activity index, DAS-OST, in which tender and swollen joint counts are replaced by the OST-score. In the construction of the DAS-OST formula, gender had to be taken into account, because also in this study, OST-scores were higher in males than in females.² In addition, by comparing single OST-scores to the rheumatologist's clinical classification into active or inactive RA as reference, also different optimal cut-offs for males and females were calculated, again indicating a gender difference. Based on these findings we investigated whether this difference in gender is also observed in controls (30 males vs. 45 females). This indeed turned out to be the case, $p=0.02$, in line with the results of *Triantafyllias et al.* The Figure shows OST-scores in RA patients and controls, separately for males and females, indicating the gender difference, and the difference between RA-patients (generally low disease activity) and controls, although there are considerable overlaps in scores between the four subgroups.

We further aimed to quantify the OST-score differences between males and females in RA patients as well as in controls, and to determine which factors influence OST-scores, including gender. In total 77 participants were included in this study. In this whole sample, adjusting for cohort (RA vs. controls), no statistically significant difference in gender was found ($p=0.12$). However, in the subgroup of controls (i.e., 22 females vs. 15 males) a statistically significant difference in OST-score was observed ($p=0.05$), in contrast to in the subgroup of RA patients ($p=0.90$, 20 females vs. 20 males). A possible explanation of the latter result, and of the absence of a statistically significant gender difference when analyzing the whole sample could be a difference in disease activity between male and female RA patients. Unfortunately, we were not able to adjust, for this, as disease activity data were not available. For OST-scores in the male subgroup, predictors were cohort, i.e., RA patients vs. controls (estimate, i.e., mean difference in OST-score, -3.08, 95%CI -6.52 to 0.36, indicating lower OST-scores in controls), age (estimate -0.15, 95%CI -0.29 to -0.02, indicating lower OST-scores in older persons), and hand surface area (estimate 0.24, 95%CI 0.03 to 0.45, indicating higher OST-scores with larger hand surface area). For OST-scores in the female subgroup, these were cohort (estimate -4.33, 95%CI -7.17 to -1.49, indicating lower OST-scores in controls), BMI (estimate -0.44, 95%CI -0.79 to -0.09, indicating lower OST-scores with higher BMI)

and hand volume (estimate 0.03, 95%CI 0.01 to 0.05, indicating higher OST-scores with more hand volume). In our opinion, not all of these results are easy to explain.

To exclude a system error of the HandScan device, although the manufacturer assured sex is no factor in the algorithm to calculate the OST-score, we aimed to perform an additional study in which a HandScan measurement is performed twice within the same participant; it is already known that the test-retest reliability, based on two measurements, is good on the patient-level.³ However, due to the COVID-19 pandemic, it was not feasible to perform this study yet.

In summary, OST-scores in controls and probably also in RA patients are statistically significantly higher in males compared to in females, although there is considerable overlap in OST-scores between subgroups. This suggests that the HandScan cannot be used for the diagnosis of RA, and has drawbacks when comparing disease activity between groups of female and male RA-patients in research. The gender difference, which has to be analysed further, plays no role however when repeated OST-scores, whether or not integrated in an index, are used in individual RA patients, to evaluate in clinical practice their disease activity over time.^{2,4}

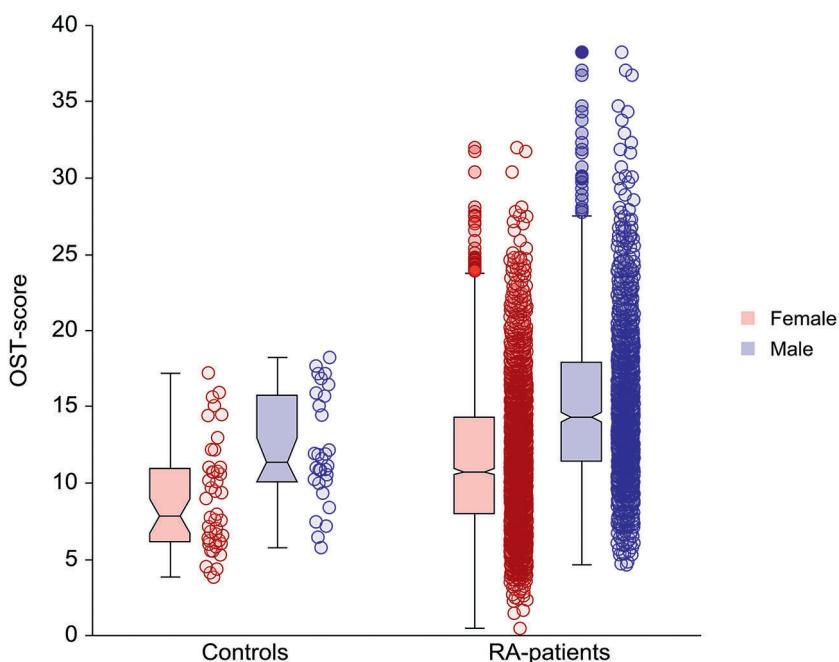


Figure Distribution of OST-scores in subgroups, a single score for each individual. Control females, n= 45; Control males, n= 30; RA females, n= 2200; RA males, n= 1157. RA= rheumatoid arthritis; OST= optical spectral transmission, score range 0 – 66 (=worst).

REFERENCES

1. Triantafyllias, K., Heller, C., de Blasi, M., Galle, P. & Schwarting, A. Diagnostic value of optical spectral transmission in rheumatoid arthritis: associations with clinical characteristics and comparison with joint ultrasonography. *J. Rheumatol.* **47**, 1314–1322 (2020).
2. Verhoeven, M. M. A. *et al.* Development of a disease activity index for RA patients using HandScan. *Ann. Rheum. Dis.* **79**, 879 (2020).
3. Besselink, N. J. *et al.* Optical spectral transmission to assess inflammation in hand and wrist joints of rheumatoid arthritis patients. *Rheumatology* **57**, 865–872 (2018).
4. Verhoeven, M. M. A. *et al.* Utility of the HandScan in monitoring disease activity and prediction of clinical response in rheumatoid arthritis patients. *In press, Rheumatology Advances in Practice January 2020.*

Chapter 13:

Summary and general discussion

SUMMARY

The last decades, rheumatoid arthritis (RA) patients are treated according to the tight-control and treat-to-target principles with the target of remission or low disease activity (LDA).^{1,2} Regarding disease modifying anti-rheumatic drug (DMARD) therapy, initial therapy with a single conventional synthetic (cs)DMARD, preferably methotrexate (MTX), with or without short-term glucocorticoids (GCs; 'bridging therapy') is advised. Biological (b)DMARDs, regarded as more effective DMARDs and often combined with a csDMARD, are not advised as initial therapy. This seems counterintuitive given that a potentially larger impact might be expected on patients outcome when a more intensive treatment is given earlier in the disease within the 'window of opportunity' period.^{1,3} The benefits of intensive treatment strategies (with bDMARDs) should be carefully weighed against adverse effects, costs related to the disease including its treatment, and the risk of potential overtreatment. Aspects of treatment strategies with the aim to further improve efficiency and effectiveness of care for RA patients are subjects of study in this thesis. The first part evaluates DMARD therapy approaches in treat-to-target treatment strategies in (early) RA, including effectiveness, cost-effectiveness and longer-term outcomes. The second part of this thesis explores whether monitoring of disease activity can be performed accurately using objective reading of inflammation with the HandScan, sparing deployment time of a healthcare professional, with the ultimate aim of better implementation of the treat-to-target principle in daily care.

Summary of PART I of the thesis

Several clinical trials have studied more intensive initial DMARD strategies with the initial aim of rapidly inducing clinical remission, compared to starting a csDMARD, with/without GC 'bridging', the latter strategy according to current recommendations in early RA. We performed a systematic literature review on this subject, see **chapter 2**. From our results it could be concluded that more intensive strategies, specifically those initiating a bDMARD, are more effective in achieving remission when compared to strategies initiating single csDMARD therapy. However, when compared to csDMARD strategies including GC bridging, the more intensive strategies with bDMARD were no longer more effective.

These results confirmed that GCs are a highly effective therapy in (early) RA (and can be regarded as DMARD). In **chapter 3**, we aimed to compare a treatment strategy initiating GC in low-median dose as DMARD (i.e., not short-term as bridging therapy) concomitantly to MTX, versus 2 strategies initiating a bDMARD (i.e., TCZ in both strategies) in early DMARD-naïve patients. We found that TCZ-based strategies compared to MTX with prednisone resulted in lower disease activity using disease activity score assessing 28 joints (DAS28)-based outcomes, whereas using a modification

of the Clinical Disease Activity Index (CDAI) as outcome, no difference in efficacy between TCZ-based strategies and MTX with prednisone could be detected. An explanation could be the known direct acute phase suppression by TCZ, extra lowering DAS28.

As it is known that, approximately 30% of the patients does not benefit from MTX¹ or does not tolerate it, identifying these patients up-front would be useful for a more patient-tailored treatment strategy. A prediction model (including baseline predictors; DAS28, HAQ, erythrocyte folate, genetic markers (SNPs; ABCB1 and ABCC3), smoking and BMI) for insufficient clinical response to MTX at three months of treatment (i.e., DAS28>3.2) was previously developed.⁴ We performed an (external) validation of this prediction model using data from the trial arm initiating MTX-monotherapy in U-Act-Early (**chapter 4**). Results showed a similar predictive value in this validation cohort as in the development cohort. Investigation of the specific predictive effect of individual components in the model showed baseline DAS28>5.1 and HAQ>0.6 to be the strongest predictors. Lifestyle parameters (BMI/smoking) and erythrocyte folate further increased the predictive value. Whether this prediction model is strong enough at patient level remains to be established.

An important outcome of RA is joint damage, as typically scored on X-rays ('radiographic progression'). Although it is generally found for disease activity outcomes that TCZ-monotherapy and TCZ combination therapy with MTX are equally effective,⁵ their relative effect on preventing radiographic progression is not clear. **Chapter 5** aimed to address this by combining data of several clinical trials. We explored the effect of TCZ strategies on different components of radiographic progression, and whether the level of disease activity, joint damage present at baseline and disease duration modified this relative effect. Results indicated that combination therapy with MTX (TCZ+MTX) is more effective in preventing radiographic progression compared to TCZ-monotherapy, but that the joint-sparing effectiveness of TCZ-monotherapy might approximate that of TCZ+MTX in specific subgroups. Specifically, in early RA patients with more joint damage or a lower DAS28 at baseline, and in established RA patients with longer disease duration.

As the impact of a treatment strategy for a chronic disease like RA becomes more apparent over a longer-term, participants of U-Act-Early¹ were followed for three additional years in an extension study. During this period, treatment was according to standard of care. The aim of **chapter 6** was to determine the longer-term effectiveness in daily clinical practice of the initial treatment strategies used in U-Act-Early. Although

¹ U-Act-Early was a 2-year, randomized placebo controlled, double-blind trial, in which DMARD-naïve early RA patients were treated to the target of sustained remission. Patients were randomly assigned to a treatment strategy initiating TCZ, MTX or their combination. If the treatment target was not achieved, MTX (in TCZ-monotherapy group) or TCZ (in MTX group) was added. When patients achieved sustained remission (i.e., remission for ≥24 weeks) medication was tapered, and finally stopped.

outcomes over the first 2 years showed that the strategies initiating TCZ yielded the most clinical benefit, over the longer-term differences, in disease activity, physical function and radiographic progression between the strategies faded.

Next to longer-term effectiveness of treatment, the cost-effectiveness is of relevance to implementation of treatment strategies. The research question of **chapter 7** was to evaluate the cost effectiveness over 5 years of treat-to-target treatment strategies in early DMARD-naïve RA patients initiating TCZ with or without MTX versus initiating MTX as monotherapy. Our analyses indicated that early initiation of TCZ or TCZ+MTX was not cost-effective compared with MTX initiation in a treat-to-target treatment strategy over 5 years in early RA patients. Sensitivity and scenario analyses showed that lower-priced TCZ improved cost effectiveness for the TCZ-based strategies, as well as reserving the TCZ-based strategies for patients anticipated (predicted) not to respond sufficiently to MTX. Although, this did not lead to the TCZ-based strategies being cost-effective.

The final chapter, **chapter 8**, of part I is an editorial entitled “Unravelling the cost of biological strategies in rheumatoid arthritis: a kaleidoscope of methodologies, interpretations and interests”, and is a discussion of, and response to a paper on the budget impact of introducing an etanercept biosimilar in a single rheumatology centre.⁶ The paper showed that costs of bDMARDs decreased but this advantage was lost within less than a year, due to an increase of the percentage of the patients with RA treated with a bDMARD. We concluded that with a different, explicit strategy for prescribing bDMARD or a broader scope of the cost(-effectiveness) study, a different conclusion might have been reached. We stressed the importance to continue conducting health-economic evaluations of current and novel DMARD strategies, including those investigating in whom, and when to initiate, taper and discontinue bDMARD treatment. Proper implementation of and adherence to these strategies in clinical practice are important for improvement of patients’ health and control of heath care costs.

Summary of PART II of the thesis

A good measuring instrument to monitor disease activity should be able to detect changes in disease activity within individual patients, the study described in **chapter 9** was performed to determine the ability of the HandScan (an objective surrogate for measuring joint inflammation) to longitudinally measure RA disease activity using DAS28 as reference (being the current standard method for measuring disease activity within patients). Based on the results we concluded that a longitudinal association of optical spectral transmission (OST)-score (obtained by the HandScan) with DAS28 exists, although it is weak.

Results indicate that the HandScan may have potential as monitoring instrument for RA, but that it likely performs better when combined with other disease activity parameters for a comprehensive view of RA disease activity. For that reason, we aimed

to develop and validate a composite disease activity measure for assessing RA patients using OST values obtained by the HandScan instead of the clinical joint assessment (SJC/TJC; **chapter 10**). The developed formula for the new disease activity index, DAS-OST, was: $-0.44 + \text{OST} \times 0.03 + \text{male} \times -0.11 + \text{LN(ESR)} \times 0.77 + \text{VAS} \times 0.03$. The intra-class correlation coefficient, evaluating agreement between DAS-OST and DAS28, was high in the internal and external validation cohort. Diagnostic performance of DAS-OST for discriminating specific disease activity states was in general good as well.

As there exists no gold standard for measuring disease activity in RA patients, we used DAS28 as reference standard for the development of DAS-OST, described above. Another reference standard relevant for clinical practice, because it would indicate the need for intensifying of the therapeutic strategy or not, would be the assessment whether RA is active or inactive, according to the rheumatologist. Therefore, **chapter 11** aimed to establish the value of DAS-OST, versus that of DAS28, to classify RA as active versus inactive, with the rheumatologist's clinical classification as reference standard. In our analyses, DAS-OST showed a clearly better performance than a single OST-score, but statistically significantly less compared to DAS28.

The final chapter, **chapter 12**, of part II of this thesis is a letter in response to a paper of *Triantafyllias et al.*⁷ Our letter aimed to explain the gender difference in OST-score observed by *Triantafyllias et al.* as observed also in our studies. We concluded that the OST-score in controls and probably also in RA patients is statistically significantly higher in males compared to in females, although there is considerable overlap in OST-scores between subgroups.

GENERAL DISCUSSION

What is the optimal DMARD treatment strategy for DMARD-naïve early RA patients? The finding that the additional effectiveness of more intensive treatment strategies (i.e., using bDMARDs as initial treatment) is limited compared to single csDMARD strategies when combined with GCs is relevant for clinical practice as bDMARDs are not available in all countries, mainly due to their high costs.⁸ It is reassuring that csDMARD therapy with GC as bridging therapy and as additional DMARD when used over a longer-term might be effective and safe,⁹ and probably a more cost-effective alternative. Also the longer-term effectiveness (of main importance in a chronic disease like RA) of initiating more intensive treatment strategies compared to a strategy starting with MTX-monotherapy in early RA does not seem to do better.¹⁰ This is likely due to the treat-to-target design with swift treatment intensification in case of insufficient effect of initial therapy.² In line with this it is plausible that the specific initial DMARD therapy used is not the most important, but treating patients within the 'window of opportunity' to rapidly suppress disease activity is.¹¹

In the current era of scarce resources in healthcare, also the cost effectiveness of a treatment strategy is important.¹² Overall, in treat-to-target treatment strategies aiming at remission, initiation of bDMARDs, specifically TCZ, are not cost-effective compared to initiation of a single csDMARD (**chapter 7**). Based on our result in **chapter 3**, we hypothesize that initiating as first therapy in early RA a strategy with a csDMARD combined with GC is cost-effective or even cost-saving compared to a more intensive strategy with a bDMARD, due to lower drug costs and similar effectiveness in treat-to-target treatment strategies. This is in line with a previous cost-utility analysis study showing that initial combination therapy with prednisone is preferred over initial combination therapy with a bDMARD, both in a treat-to-target strategy aiming at remission in newly diagnosed RA patients over 2 years.¹³ Surprisingly, in our own cost effectiveness analyses, we observed that productivity loss costs were saved in the groups initiating TCZ, particularly in the group starting with TCZ-monotherapy (**chapter 7**), and less so in the initial TCZ+MTX group, compared to the MTX-monotherapy group, suggesting that MTX toxicity may be partially responsible for the productivity loss.¹⁴ However, this does not offset the cost of therapy.

We expect that initiation of bDMARDs could become cost-effective as initial treatment strategy in early RA under certain circumstances. We hypothesize that initial treatment with bDMARDs combined with active tapering might enable drug free remission, an even sharper target than remission only. If this target is indeed achievable in a considerable proportion of patients over time, we expect that this increases quality adjusted life years, not only by preventing irreversible joint damage, but also adverse effects of treatment (specifically of MTX), while also saving medication costs. Further, decrease of bDMARD prices in the future driven by introduction of biosimilars will add to this; after introduction of biosimilars on the Dutch market in 2015, the mean purchase price of etanercept for Dutch hospitals decreased almost 60%.^{6,15} In addition, tailoring initial bDMARD treatment strategy to specific subpopulations may be promising to further improve cost effectiveness. For example, focussing on patients who are anticipated to have an insufficient response to the anchor treatment of MTX,¹ and for those implementing prediction models in clinical practice make sense.¹⁶ In **chapter 7** we observed that, in a specific early RA subgroup (i.e., baseline DAS28 >5.1 based on prediction model as described in **chapter 4**), TCZ-based initial treatment might become more cost-effective, although the strategy did not reach generally accepted levels of cost effectiveness in the Netherlands,^{17,18} probably due to the lack of active tapering in the follow-up period. Notwithstanding, analyses simulating specific stratified treatment strategies, using existing or future prediction models, should be performed more frequently and continuously to develop effective as well as cost-effective treatment strategies, specifically in light of achieving drug free remission.

In conclusion, based on this thesis and literature, we assume that a treat-to-target treatment strategy initiating a csDMARD in combination with GC is the most cost-effective strategy nowadays, which is widely available as well. However, with bDMARDs becoming less costly, in specific subgroups, e.g., MTX non-responders, initial treatment with a bDMARD may possibly be more effective as well as cost-effective when combined with an active tapering protocol. Whatever the initial strategy, many RA patients eventually will (concomitantly) need a bDMARD.

What is the optimal instrument to monitor disease activity in RA patients?

Measurement instruments used for monitoring disease activity such as DAS28 currently are not always easy to implement in daily practice, as they are time consuming, and rather subjective.^{8,19,20} A DAS28 measurement will take approximately 5-10 minutes of rheumatologists' or healthcare professionals' time, which is quite a proportion of the total time generally available for an out-patient-clinic visit. Further, as most of these measurements (except CDAI) include an acute phase reactant (APR), rheumatologists have to be aware that (at least) one of the current medical therapies (i.e., TCZ) strongly directly affects this factor, which in theory may result in underestimation of disease activity score and/or overestimation of treatment effect.²¹

The HandScan could provide a more objective and faster measurement, and might be an alternative method for assessing disease activity. However, the correlation between DAS28 and OST-scores at one time point is moderate only,²² and the model's ability to estimate DAS28 over time (i.e., monitor disease activity) by OST-scores was low. Besides, OST-scores seem to be influenced by gender (**chapter 10, 11 and 12**): OST-scores were higher in males compared to females. This could be in part a system/device error; future research is warranted to exclude this. The OST-score can be considered an imaging tool; other studies, mostly investigating a simple correlation between an imaging tool (MRI or ultrasound) and a clinical disease activity tool, show similar results, i.e., a moderate correlation.²³⁻²⁵ As a consequence, some clinical trials applied a treatment target based on a combination of clinical remission AND remission defined on imaging. However, in these studies outcomes for RA patients didn't improve. On the contrary, it rather seemed to result in additional DMARD therapy, which might be considered as overtreatment.^{26,27}

We argue that replacement of the time consuming joint count assessments, by the OST-score in a composite index like DAS28 could be favourable as OST quantifies inflammation faster. Replacement would make more sense than adding the OST-score to the composite index, given that the advantage of saving time would be lost, and that adding an OST-based target would probably not result in better patient outcomes, as described above for other imaging modalities. The concept, DAS-OST, wherein joint counts have been replaced by OST, is tested against DAS28 and the rheumatologist's

clinical classification (i.e., active or inactive disease based on (adjustments of) DMARD therapy). In general, the performance of DAS-OST was good compared to DAS28 and the rheumatologist's clinical interpretation, **chapter 10 and 11**. In addition, the gender effect, which is seen in OST-scores, plays no role when using DAS-OST in which gender is taken into account. However, the clinical applicability of DAS-OST may be limited as more patients with active disease according to the rheumatologist were incorrectly classified as having LDA, compared to when using DAS28. This indicates that we should optimize the cut-offs of DAS-OST in future research, as we used the established DAS28 cut-offs, since DAS-OST was modelled to reflect DAS28 as reference standard. Perhaps such a new cut-off could be validated in a currently ongoing trial at our department. Before results of the current thesis were present, we started a clinical non-inferiority trial in 2017, to investigate HandScan-guided treatment aimed at 'HandScan remission' versus clinically guided treatment aimed at ACR/EULAR 2011 Boolean remission.²⁸ HandScan remission in this trial is based on a single OST-score, but in case of a continued discrepancy (at three consecutive time points) between HandScan-guided decisions and decisions that would have been taken based on the Boolean criteria, as safety measure in this trial, the Boolean-guided decision will be followed. It would be interesting to perform post-hoc analyses in this trial by calculating DAS-OST and simulate treatment decisions. In this analysis, we can investigate whether other/more appropriate decisions would have been made by applying the established as well as optimised cut-offs for DAS-OST. Moreover, it would be interesting to replace the joint count assessment in more recently developed disease activity indices, e.g., CDAI, by the OST-score. Especially CDAI is of interests as in this measurement an APR is not included, and since CDAI (as well as simplified disease activity index(SDAI)) uses equal weighting for all components.²⁹

Thus, if a suitable cut-off for DAS-OST can be defined in future research, i.e., if DAS-OST can distinguish between patients having LDA or not, DAS-OST could be implemented in daily practice. Then it may be a faster and more objective alternative measurement compared to the current indices, as a DAS-OST measurement can be performed by a non-healthcare professional. This allows the rheumatologists to focus on patients with active disease (treat-to-target).

We conclude that a more promising approach regarding the HandScan than using only the resulting OST-scores seems to be replacement of (time consuming) joint counts by the OST-score in the composite DAS28 measurement, generating a new measurement instrument; DAS-OST. Although this measurement seems to perform reasonably, several challenges have to be overcome before implementation in clinical practice. Until this time, we recommend to (at least in parallel) use the current composite measurements to monitor disease activity.

REFERENCES

1. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
2. Ramiro, S. *et al.* Is treat-to-target really working in rheumatoid arthritis? A longitudinal analysis of a cohort of patients treated in daily practice (RA BIODAM). *Ann. Rheum. Dis.* **79**, 453–459 (2020).
3. Raza, K. & Filer, A. The therapeutic window of opportunity in rheumatoid arthritis: does it ever close? *Ann. Rheum. Dis.* **74**, 793–794 (2015).
4. De Rotte, M. C. *et al.* Development and validation of a prognostic multivariable model to predict insufficient clinical response to methotrexate in rheumatoid arthritis. *PLoS One* **13**, e0208534 (2018).
5. Teitsma, X. M., Marijnissen, A. C. A., Bijlsma, F. W. J., Lafeber, F. P. J. & Jacobs, J. W. G. Tocilizumab as monotherapy or combination therapy for treating active rheumatoid arthritis: a meta-analysis of efficacy and safety reported in randomized controlled trials. *Arthritis Res. Ther.* **18**, 211 (2016).
6. Müskens, W. D., Rongen-van Dartel, S. A. A., van Riel, P. L. C. M. & Adang, E. M. M. Does Etanercept biosimilar prescription in a rheumatology center bend the medication cost curve? *In press, J. Rheumatol.* November 2020.
7. Triantafyllias, K., Heller, C., de Blasi, M., Galle, P. & Schwarting, A. Diagnostic value of optical spectral transmission in rheumatoid arthritis: associations with clinical characteristics and comparison with joint ultrasonography. *J. Rheumatol.* **47**, 1314–1322 (2020).
8. Pincus, T. & Segurado, O. G. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. *Ann. Rheum. Dis.* **65**, 820–822 (2006).
9. Safy, M. *et al.* Long-term outcome is better when a methotrexate-based treatment strategy is combined with 10 mg prednisone daily: follow-up after the second computer-assisted management in early rheumatoid arthritis trial. *Ann. Rheum. Dis.* **76**, 1432–1435 (2017).
10. Goekoop-Ruiterman, Y. P. M. *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.* **52**, 3381–3390 (2005).
11. O'Dell, J. R. Treating rheumatoid arthritis early: a window of opportunity? *Arthritis Rheum.* **46**, 283–285 (2002).
12. Hsieh, P. H. *et al.* Economic burden of rheumatoid arthritis: a systematic review of literature in biologic era. *Ann. Rheum. Dis.* **79**, S771–S777 (2020).
13. van der Hout, W. B. *et al.* Cost-utility analysis of treatment strategies in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum.* **61**, 291–299 (2009).
14. Salliot, C. & van der Heijde, D. Long-term safety of methotrexate monotherapy in patients with rheumatoid arthritis: a systematic literature research. *Ann. Rheum. Dis.* **68**, 1100–1104 (2009).
15. Autoriteit Consument & Markt. Sectoronderzoek TNF-alfaremmers. <https://www.acm.nl/sites/default/files/documents/2019-09/sectoronderzoek-tnf-alfaremmers.pdf> (2019).
16. van Nies, J. A. B., Tsionaka, R., Fautrel, B. & van Mil-van der Helm, A. H. M. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann. Rheum. Dis.* **74**, 806–812 (2015).
17. Zorginstituut Nederland. *Ziektelest in de praktijk.* (2018).
18. Zorginstituut Nederland. *Kosteneffectiviteit in de praktijk.* (2015).
19. Scott, D. L. & Houssien, D. A. Joint assessment in rheumatoid arthritis. *Br. J. Rheumatol.* **35**, 14–18 (1996).
20. Jacobs, J. W. G., ten Cate, D. F. & van Laar, J. M. Monitoring of rheumatoid arthritis disease activity in individual patients: still a hurdle when implementing the treat-to-target principle in daily clinical practice. *Rheumatology* **54**, 959–961 (2015).

21. Smolen, J. S. & Aletaha, D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. *Arthritis Rheum.* **63**, 43–52 (2011).
22. van Onna, M. *et al.* Assessment of disease activity in patients with rheumatoid arthritis using optical spectral transmission measurements, a non-invasive imaging technique. *Ann. Rheum. Dis.* **75**, 511–518 (2016).
23. Haavardsholm, E. A. *et al.* Monitoring anti-TNF α treatment in rheumatoid arthritis: Responsiveness of magnetic resonance imaging and ultrasonography of the dominant wrist joint compared with conventional measures of disease activity and structural damage. *Ann. Rheum. Dis.* **68**, 1572–1579 (2009).
24. Xu, H., Zhang, Y., Zhang, H., Wang, C. & Mao, P. Comparison of the clinical effectiveness of US grading scoring system vs MRI in the diagnosis of early rheumatoid arthritis (RA). *J. Orthop. Surg. Res.* **12**, 152 (2017).
25. Hameed, B., Pilcher, J., Heron, C. & Kiely, P. D. W. The relation between composite ultrasound measures and the DAS28 score, its components and acute phase markers in adult RA. *Rheumatology* **47**, 476–480 (2008).
26. Haavardsholm, E. A. *et al.* Ultrasound in management of rheumatoid arthritis: ARCTIC randomised controlled strategy trial. *BMJ* **354**, i4205 (2016).
27. Mandl, P. & Aletaha, D. The role of ultrasound and magnetic resonance imaging for treat to target in rheumatoid arthritis and psoriatic arthritis. *Rheumatology* **58**, 2091–2098 (2019).
28. Besselink, N. J. *et al.* Novel optical spectral transmission (OST)-guided versus conventionally disease activity-guided treatment: study protocol of a randomized clinical trial on guidance of a treat-to-target strategy for early rheumatoid arthritis. *Trials* **20**, 226 (2019).
29. Aletaha, D. & Smolen, J. The simplified disease activity index (SDAI) and the clinical disease activity index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin. Exp. Rheumatol.* **23**, s100–s108 (2005).

Addendum

NEDERLANDSE SAMENVATTING

Introductie

Reumatoïde artritis (RA) is een relatief veelvoorkomende ontstekingsziekte, die met name gekenmerkt wordt door ontsteking van gewrichten, vooral die van handen, polsen en enkels. De ontstoken gewrichten zijn vaak pijnlijk, dik en stijf; de ontsteking kan leiden tot gewrichtsschade. Ongeveer 1-2% van de Nederlandse volwassenen heeft RA en de ziekte treft meer vrouwen dan mannen. De klachten zoals pijn, stijfheid en ook vermoeidheid beperken RA-patiënten in hun dagelijkse bezigheden, wat kan leiden tot ziekteverzuim en vermindering van kwaliteit van leven.

In de afgelopen jaren hebben verschillende onderzoeken aangetoond dat de gevlogen van RA op lange termijn veel minder erg zijn, als zo snel mogelijk na de diagnose met ziekte-onderdrukkende therapie, met zogeheten *disease modifying anti-rheumatic drugs (DMARDs)* wordt gestart, bij voorkeur binnen 3-6 maanden. Deze periode wordt 'window of opportunity' genoemd, omdat later starten minder effect heeft op langere termijn. Daarnaast worden lange termijn gevlogen van RA beperkt door vaak meten (monitoren) van ziekteactiviteit ('tight-control') met, indien nodig, het direct aanpassen van de DMARD-therapie (bijvoorbeeld ophogen van de dosis als de RA actief is). Het behandelen volgens dit 'tight-control'-principe wordt in internationale richtlijnen geadviseerd, naast het behandelen volgens het 'treat-to-target'-principe. Het 'treat-to-target'-principe houdt in dat het doel ('target') van behandeling, meestal volledige onderdrukking van ziekteactiviteit ('remissie') of tenminste lage ziekteactiviteit, wordt nastreefd. Richtlijnen bevelen aan om als eerste therapie bij pas gediagnosticeerde RA te starten met één van de conventionele DMARDs, bij voorkeur methotrexaat (MTX), al dan niet samen met een glucocorticoïd (GC), zoals prednison. De conventionele DMARD werkt pas optimaal na enkele weken en prednison binnen een dag, dus prednison wordt vaak voorgeschreven ter overbrugging van die onwerkzame periode van de DMARD (overbruggingstherapie).

Op dit moment wordt nog niet aanbevolen te starten met een mogelijk effectievere, maar duurdere behandeling, namelijk met een biologische (b)DMARD. Biologisch betekent dat het middel een eiwit is, met als consequentie dat het niet in tabletvorm gegeven kan worden, omdat eiwitten worden afgebroken in de maag. Om deze reden moet het toegediend worden per infuus of injectie. Dit in tegenstelling tot de conventionele DMARDs, die 'synthetisch' zijn, d.w.z. een chemische, niet eiwitachtige samenstelling hebben. Dat starten met een bDMARD niet wordt aanbevolen lijkt onlogisch, omdat een krachtige (intensieve) behandeling, meteen na het stellen van de diagnose, dus binnen de 'window of opportunity', waarschijnlijk het meest bepalend is voor een gunstig ziektebeloop.

De belangrijkste redenen om een bDMARD niet als starttherapie aan te bevelen zijn (1) de kosten, en (2) het risico dat sommige patiënten dan behandeld worden met een sterk werkend middel, met meer kans op bijwerkingen, terwijl ze met een minder krachtig middel ook goed behandeld hadden kunnen worden. In het laatste geval spreken we van overbehandeling. Dat bij een deel van patiënten in de eerste fase van de ziekte het starten van een bDMARD in feite overbehandeling is, en dat starten met een conventionele DMARD zoals MTX (met of zonder prednison als overbruggingstherapie) bij hen ook voldoende effectief zou zijn geweest, en het feit dat bDMARDs veel duurder zijn, maken dat de (kosten)effectiviteit van het starten met een bDMARD beperkt is, in vergelijking met die van het starten met een conventionele DMARD, eventueel met tijdelijk prednison. Er moet dan wel sprake zijn van behandeling volgens de '*tight-control*'- en '*treat-to-target*'-principes.

Gezien de hogere kosten en het risico op overbehandeling (en dus het gevaar van onnodige bijwerkingen) moeten de voordelen van intensievere behandelstrategieën zorgvuldig worden afgewogen.

Verbeteringen van de kosteneffectiviteit en risico-batenverhouding van bDMARDs kunnen worden verwacht van nieuwe behandelstrategieën, bijvoorbeeld strategieën die snel en fors RA onderdrukken met een bDMARD en conventionele DMARD, en vervolgens de duurdere bDMARD afbouwen en stoppen, waarbij, in de eerste fase, de behandeling met de conventionele DMARD wordt voortgezet. Als het dan heel goed blijft gaan kan de conventionele DMARD ook afgebouwd worden. Daarnaast is verbetering van kosteneffectiviteit te verwachten, indien het mogelijk zou worden, te voorspellen voor welke RA-patiënt het starten met een conventionele DMARD als eerste therapie voldoende effectief zou zijn, en voor wie niet. Dat zou een meer op het individu afgestemde ('gepersonaliseerde') behandelstrategie mogelijk maken.

Om de behandeling van RA volgens de tijdsintensieve '*tight-control*'- en '*treat-to-target*'-principes meer haalbaar te maken, kan gekeken worden naar andere manieren van monitoren van ziekteactiviteit. De ziekteactiviteit wordt in het algemeen vastgesteld met een ziekteactiviteit-score, waarbij onder andere het aantal ontstoken en pijnlijke gewrichten van een set van 28 gewrichten wordt geteld. Deze 'disease activity score' genaamd DAS28 omvat ook de waarde van de bloedbezinking en een score door de patiënt van de ziekteactiviteit. Het gewrichtsonderzoek voor het beoordelen van het aantal pijnlijke en gezwollen gewrichten kost relatief veel tijd. Daarnaast moeten de zorgmedewerkers worden getraind in hoe dat onderzoek betrouwbaar dient te worden uitgevoerd. Desondanks kan er ongewenste variatie zijn in de resultaten tussen zorgmedewerkers, als zij dezelfde gewrichten scoren. De zogeheten HandScan is een nieuw apparaat, dat ontsteking van hand- en polsgewrichten geautomatiseerd en objectief kan meten. De HandScan maakt gebruik van optische spectrale transmissie (OST). Dit houdt het meten van de hoeveelheid van bepaald rood-bijna infrarood licht

in, dat de gewrichtsregio's kan passeren (transmissie); is het gewicht ontstoken, dan passeert er minder licht en is de OST-score hoger. Binnen 5 minuten kan een niet medisch geschoold medewerker een meting uitvoeren.

Dit proefschrift, waarvan **hoofdstuk 1** de inleiding is, bestaat uit twee delen. Het eerste deel beschrijft het evalueren van de effectiviteit en kosteneffectiviteit van verschillende behandelstrategieën van RA. Het tweede deel beschrijft onderzoek van monitoren van ziekteactiviteit van RA met behulp van de HandScan.

Deel I: Evaluatie van effectiviteit & kosteneffectiviteit van verschillende 'treat-to-target'- DMARD-strategieën bij patiënten met (pas gediagnosticeerde) RA.

Verschillende onderzoeken hebben bij pas gediagnosticeerde RA-patiënten het effect van het starten met een intensieve '*treat-to-target*'-behandelstrategie vergeleken met dat van het starten van de standaard '*treat-to-target*'-behandelstrategie volgens de huidige richtlijnen (conventionele DMARD met of zonder GC-overbruggingstherapie).

Hoofdstuk 2 beschrijft ons systematische literatuuronderzoek over dit onderwerp. Het laat zien dat intensievere strategieën (bijvoorbeeld met een bDMARD) effectiever zijn in vergelijking met een standaardstrategie met conventionele DMARD *zonder* GC overbruggingstherapie. Maar de intensievere strategieën zijn niet effectiever dan een standaardstrategie met een conventionele DMARD *met* een GC als overbruggingstherapie.

De resultaten beschreven in hoofdstuk 2 bevestigen eerdere resultaten dat een GC zeer effectief is bij pas gediagnosticeerde RA-patiënten als kortdurende overbruggingstherapie. Maar prednison wordt niet alleen als overbruggingstherapie, maar ook wel langdurig gebruikt. Bij langdurig gebruik van prednison is gebleken dat het ontstaan van gewrichtsschade afgeremd kan worden, en het GC dus dezelfde werking heeft als een DMARD. Het doel van het onderzoek, beschreven in **hoofdstuk 3** is om de effectiviteit van langduriger gebruik van een relatief lage dosering van prednison, gecombineerd met MTX (MTX+GC), te vergelijken met die van twee behandelstrategieën met de bDMARD tocilizumab (TCZ), één strategie met en één zonder MTX. Dit als eerste therapieën bij patiënten met pas gediagnosticeerde RA. De TCZ-strategieën resulteerden in betere DAS28(-gebaseerde) resultaten in vergelijking met MTX+GC. Een deel van dit effect is echter waarschijnlijk toe te schrijven aan het aspecifieke, extra verlagende effect van TCZ op de bloedbezinking, die een van de componenten van de DAS28 is. Er blijkt namelijk geen beter resultaat van de TCZ-strategieën, als evaluatie van het effect wordt verricht met ziekteactiviteit-scores, waarvan de bezinking geen deel uitmaakt.

Ongeveer 30% van alle RA-patiënten heeft of te weinig effect van behandeling met MTX, of ervaart bijwerkingen ervan. Als we deze groep patiënten voor de start van de behandeling zouden kunnen identificeren, zouden we hen kunnen behandelen

met een andere (b)DMARD, of met een lagere dosering MTX. Eerder was er al een voorspellend rekenmodel (predictiemodel) ontwikkeld voor te weinig klinisch effect na 3 maanden behandeling met MTX. In het onderzoek, beschreven in **hoofdstuk 4** is dit predictiemodel op zijn waarde nagekeken in een andere groep RA-patiënten (=extern gevalideerd). Voor deze andere groep (validatiecohort) werden gegevens gebruikt van pas gediagnosticeerde RA-patiënten, die aan het zogeheten U-Act-Early¹ onderzoek hadden deelgenomen, en gestart waren met MTX. De voorspellende waarde van het predictiemodel in dit validatiecohort blijkt overeen te komen met die in de oorspronkelijke groep, waarin het model werd ontwikkeld (het ontwikkelingscohort). Daarmee kan het predictiemodel als deugdelijk (valide) worden beschouwd. Van alle componenten van het predictiemodel blijken, voor het voorspellen, voor de start van de behandeling met MTX, van te weinig klinisch effect na 3 maanden behandeling, scores van de DAS28 boven de 5,1, wijzend op fors actieve RA, en scores van de health assessment questionnaire (HAQ) boven 0,6, wijzend op functionele beperkingen door RA, het meest voorspellend te zijn.

Een andere belangrijke uitkomstmaat van de ziekte RA is beschadiging van gewrichten, die te zien is op röntgenfoto's (radiografische schade). In **hoofdstuk 5** zijn gegevens van meerdere, reeds verrichte gerandomiseerde onderzoeken gebruikt, om het effect van TCZ-therapie vergeleken met dat van TCZ samen met MTX (TCZ+MTX) te bepalen op het ontstaan van radiografische schade. Het blijkt dat TCZ+MTX effectiever het ontstaan van radiografische schade afremt dan TCZ-therapie alleen. Dit verschil is echter niet aanwezig in twee subgroepen van patiënten: patiënten met pas gediagnosticeerde RA & een lagere score van de DAS28, en patiënten met pas gediagnosticeerde RA & meer radiografische schade.

Het effect van een behandelstrategie voor de chronische ziekte RA wordt pas helemaal duidelijk op de langere termijn. Daarom werden deelnemers aan het 2 jaar durende U-Act-Early¹ onderzoek na afloop van het onderzoek nog 3 jaar gevolgd (3 jaar follow-up). In deze laatste periode was de behandeling volgens standaardzorg. **Hoofdstuk 6** beschrijft deze langere termijn uitkomsten. Alhoewel op korte termijn de twee initiële TCZ-strategieën (starten met TCZ of met TCZ+MTX) klinisch meer voordeel opleverden dan starten met MTX, is er op de langere termijn geen verschil in effectiviteit meer tussen de drie initiële behandelstrategieën. Dit is waarschijnlijk het gevolg van

¹ U-Act-Early was een 2 jaar durend, zogeheten gerandomiseerd geblindeerd placebogecontroleerd onderzoek, waarin pas gediagnosticeerde RA-patiënten met medicijnen startten met als doel persistente volledige onderdrukking (remissie) van RA. Patiënten lootten voor een behandeling startend met TCZ of met MTX, of met TCZ+MTX, maar wisten niet welke medicijnen ze kregen, door gebruik van placebo's. Indien het behandelingsdoel niet werd behaald, werd in de TCZ-groep MTX en in MTX-groep TCZ toegevoegd. Wanneer het doel werd behaald, en steeds behaald bleef, werd de medicatie afgebouwd en gestopt.

het toegepaste ‘*treat-to-target*’-principe, ook gedurende de 3 jaar follow-up, waarbij de behandeling steeds wordt aangepast totdat de ‘target’ (remissie) is behaald.

Naast effectiviteit op langere termijn van deze strategieën is ook de kosteneffectiviteit van belang. Zowel tijdens het U-Act-Early¹ onderzoek als tijdens de 3 jaar follow-up zijn gegevens verzameld over zorg-gerelateerde kosten en productiviteitsverlies. In **hoofdstuk 7** worden de resultaten van de analyses daarvan (kosteneffectiviteitsanalyses) beschreven. Het blijkt dat starten met TCZ of met TCZ+MTX niet kosteneffectief is, in vergelijking met starten met MTX bij deze pas gediagnosticeerde RA-patiënten. De kosteneffectiviteit van de twee TCZ-strategieën is wel aanzienlijk beter voor de subgroep van patiënten met een baseline score van de DAS28 hoger dan 5,1, die wijst op zeer actieve RA. Uiteraard wordt de kosteneffectiviteit van de twee TCZ-strategieën ook beter als TCZ minder duur zou zijn. Maar zelfs in deze twee scenario’s van betere kosteneffectiviteit is het starten met één van de twee TCZ-strategieën niet kosten-effectiever dan starten met MTX.

Het laatste hoofdstuk van dit deel van het proefschrift, **hoofdstuk 8**, is een beschouwing (editorial) van een artikel van andere auteurs over de impact op het budget van de introductie van een biosimilar (een goedkopere versie van een bDMARD) in een reumatologiecentrum in Nederland. Daardoor daalden de bDMARD-kosten, maar binnen een jaar was het voordeel verdwenen door toename van het percentage RA-patiënten dat werd behandeld met een bDMARD. Het resultaat van dit onderzoek werd door de auteurs zelf als negatief gezien. We beargumenteren dat met een andere visie op, en analyse van, dit onderzoek het resultaat ook positief had kunnen uitvallen. Want meer patiënten met bDMARDs behandelen betekent niet alleen meer medicatiekosten, maar mogelijk ook financiële voordelen, zoals minder kosten gerelateerd aan actieve RA. We benadrukken het belang van goed opgezet onderzoek naar kosteneffectiviteit van behandelstrategieën in de dagelijkse praktijk met (b)DMARDs.

Deel II: Monitoren van ziekteactiviteit tijdens ‘*treat-to-target*’-behandelstrategieën: het toepassen van de HandScan.

Een methode om ziekteactiviteit te monitoren is pas bruikbaar als die betrouwbaar veranderingen over de tijd van ziekteactiviteit van één en dezelfde patiënt kan detecteren, want zo wordt die methode in de dagelijkse praktijk gebruikt. Het onderzoek beschreven in **hoofdstuk 9** is overeenkomstig uitgevoerd: het gaat binnen een groep RA-patiënten na of de HandScan hun ziekteactiviteit over de tijd (longitudinaal) goed kan meten, waarbij de ziekteactiviteit gemeten met de DAS28 de referentie is. In deze longitudinale analyse wordt slechts een zwakke relatie tussen de HandScan OST-score en DAS28 vastgesteld.

Dat betekent dat de OST-score gecombineerd moet worden met andere parameters van ziekteactiviteit, bijvoorbeeld in een index, om de ziekteactiviteit bruikbaar te

kunnen monitoren. We hebben daarom een index ontwikkeld en gevalideerd voor het monitoren van ziekteactiviteit van RA-patiënten, gebruikmakend van de OST-score (**hoofdstuk 10**). In deze index genaamd DAS-OST wordt de OST-score gecombineerd met de waarde van de bezinking, een score door de patiënt van de ziekteactiviteit en met geslacht. De overeenstemming tussen deze DAS-OST en de DAS28 was relatief hoog. We concluderen dat met de DAS-OST ziekteactiviteit bij RA-patiënten voldoende nauwkeurig kan worden gemonitord. DAS-OST geeft sneller en objectiever een indruk van de ziekteactiviteit dan DAS28.

Bij het ontwikkelen van DAS-OST hebben we de DAS28 als referentiestandaard gebruikt; er is geen echte gouden standaard voor het meten van ziekteactiviteit van RA-patiënten. Een andere referentiestandaard die relevant is voor de klinische praktijk, is het oordeel van de behandelende reumatoloog of de ziekte actief of rustig is. In het eerste geval moet de DMARD-therapie worden aangepast. In **hoofdstuk 11** is beschreven in hoeverre de OST-score, de DAS-OST en de DAS28 de RA-activiteit kunnen classificeren als actief of rustig, waarbij het oordeel van de behandelende reumatoloog de referentiestandaard is. Het blijkt dat de DAS-OST dat duidelijk beter kan dan de OST-score alleen, maar wel minder goed dan de DAS28.

Het voorlaatste hoofdstuk (**hoofdstuk 12**) betreft een door ons ingezonden brief ('letter'), waarin ons onderzoek wordt beschreven dat probeert een ook al eerder vastgesteld systematisch verschil in OST-scores tussen mannen en vrouwen te verklaren: mannen scoren over het algemeen hoger dan vrouwen. We vinden bij dit onderzoek ook een forse overlap van OST-scores van RA-patiënten en die van vrijwilligers zonder RA, maar vinden geen afdoende verklaring voor het geslachtsverschil. De overlap en het geslachtsverschil spelen echter geen belangrijke rol als herhaalde OST-scores over de tijd worden gebruikt binnen dezelfde patiënt en evenmin als de DAS-OST wordt gebruikt, waarin geslacht een parameter is.

Conclusie

Het onderzoek, beschreven in dit proefschrift beoogt bouwstenen aan te dragen met als uiteindelijk doel de behandeling van (pas gediagnosticeerde) RA-patiënten te verbeteren.

Het *eerste deel* van dit proefschrift laat zien dat bij pas gediagnosticeerde RA-patiënten het starten met intensievere '*treat-to-target*'- DMARD-strategieën effectiever is dan volgens huidige richtlijnen starten met de '*treat-to-target*'-behandelstrategie met MTX met zo nodig ophogen van de dosis, d.w.z. bij onvoldoende effect. Maar starten met MTX met de GC prednison (kortdurend als overbruggingstherapie of langer durend, waarbij de prednison dan ook als DMARD werkt) is mogelijk even effectief als starten met andere intensievere strategieën met een bDMARD, die wel duurder zijn. Starten met MTX met prednison bij pas gediagnosticeerde RA-patiënten is het meest kosteneffectief en deze goedkope strategie is ook wereldwijd beschikbaar

en betaalbaar. Onderzoek naar het voorspellen van het effect van een behandeling met MTX voordat dit middel gestart is laat veelbelovende resultaten zien, maar het personaliseren van de behandelstrategie voor individuele RA-patiënten is nu nog niet goed mogelijk. Wel lijken specifieke subgroepen van pas gediagnosticeerde RA-patiënten te kunnen worden geselecteerd met verhoogd risico op te weinig effect van MTX, als dat gestart zou worden als eerste therapie. Deze subgroepen zouden beter kunnen starten met een intensiere strategie, bijvoorbeeld MTX met prednison, of met een bDMARD, zoals TCZ. Het is dan wel een uitdaging de te lage kosteneffectiviteit van behandelstrategieën met bDMARDs te verbeteren, bijvoorbeeld door het toepassen van specifieke, gestandaardiseerde afbouwprotocollen. Gunstig daarbij is dat bDMARDs langzaam in de tijd in prijs lijken te dalen, dat goedkopere versies van bDMARDs (biosimilars) beschikbaar zijn en komen, en dat steeds ook nieuwe, andere vormen van therapie worden ontwikkeld.

Het *tweede deel* van dit proefschrift laat zien dat de HandScan kan worden gebruikt om ziekteactiviteit bij RA-patiënten te monitoren, indien de OST-score gecombineerd wordt met andere (routinematisch verzamelde) parameters van ziekteactiviteit. De zodanig ontwikkelde en gevalideerde index, DAS-OST, lijkt een objectieve, snelle en relatief goedkope manier om de ziekteactiviteit te monitoren bij RA-patiënten, hoewel dat wat minder betrouwbaar gaat dan met de DAS28. Maar DAS-OST zou in de dagelijkse, drukke praktijk het implementeren van de tijdsintensieve ‘*tight-control*’- en ‘*treat-to-target*’-principes van de behandeling van RA-patiënten kunnen vereenvoudigen. Dat moet dan wel eerst verder onderzocht worden.

DANKWOORD

Samen sta je sterk! Zonder de betrokkenheid van meerdere personen was dit proefschrift niet tot stand gekomen. Sommige mensen wil ik graag in het bijzonder bedanken.

Op de eerste plaats gaat mijn dank uit naar alle patiënten die deelgenomen hebben aan de onderzoeken, die beschreven worden in dit proefschrift en de betrokken medewerkers bij deze onderzoeken. Zonder jullie waren er geen gegevens en was er geen proefschrift.

Daarnaast wil ik graag mijn promotieteam bedanken voor de prettige en deskundige begeleiding in de afgelopen jaren:

Prof. dr. Lafeber, beste Floris, als eerste promotor heb je altijd nauwlettend de grote lijnen van het promotietraject in de gaten gehouden. We hadden maandelijks een afspraak met het gehele promotieteam om de voortgang te bespreken. Dit waren efficiënte overleggen, en meestal een bevestiging voor mij dat alles volgens plan verliep. Ik bewonder de hoeveelheid kennis die je bezit en altijd paraat hebt. Ik vond het prettig dat de deur altijd openstond voor een korte vraag, ook ten tijde van het thuiswerkbeleid. Ik heb veel van je geleerd.

Prof. dr. van Laar, beste Jaap, als tweede promotor ben je nauw betrokken geweest bij de projecten. Je kennis en ervaring in het uitvoeren van klinische onderzoeken heeft mij vooral geholpen bij het uitvoeren van de HandScan-trial, die nog loopt. Ik heb bewondering voor de manier waarop jij discussies (op de juiste momenten) kunt beëindigen, waarmee alle betrokken partijen tevreden zijn. Hiervan, en van je relaxte “helicopterview” heb ik veel geleerd.

Dr. Welsing, beste Paco, als copromotor en dagelijkse begeleider gaat veel dank uit naar jou. Ook jij had je deur op H3 altijd openstaan voor een korte vraag, of om te sparren over het uitvoeren van analyses. Jij hebt me geleerd dat het soms noodzakelijk is om een beetje van je oorspronkelijk plan af te wijken om een project tot een goed einde te brengen. Mede door jouw optimisme en relaxte persoonlijkheid heb ik geleerd dat het met minder stress ook wel goed kan komen. Bedankt voor je bijdrage aan het proces en aan het proefschrift.

Dr. Tekstra, beste Janneke, je hebt jezelf altijd ook 100% ingezet voor de projecten, ondanks dat je iets later bent ingestapt als copromotor. Jij bent de clinicus die in artikelen de klinische boodschap goed naar voren wist te brengen als het dreigde te epidemiologisch te worden. Ik heb je leren kennen als een “hardwerker” en doorzetter met oog voor de patiënt. Bij jou gaat kwaliteit boven kwantiteit. Hier heb ik veel bewondering voor, en ik heb er uiteraard veel van geleerd. Bedankt voor je bijdrage aan het proces en aan het proefschrift.

Dr. Jacobs, beste Hans, ook jou wil ik bedanken voor je begeleiding, grotendeels van afstand, ten tijde van het promotietraject. Jij bent altijd bereid om ergens ‘even voor te gaan zitten’ of, als hedendaags, ‘even te ZOOMen’. De meeste manuscripten (en overige teksten) zijn door jou tot in detail nagekeken, met uitgebreide commentaren en onderbouwingen. Hiervan heb ik veel geleerd. Het was beoogd om jouw naam en functie te benoemen op het titelblad, echter is dit afgekeurd. Zelf ben ik overtuigd dat jij, naast de promotoren en copromotoren, hier een plek verdient. Bedankt voor jouw bijdrage bij de totstandkoming van dit proefschrift, de leerzame overleggen en wetenschappelijke discussies.

Ook mensen buiten het promotieteam hebben een groot aandeel geleverd.

Allereerst Dr. Westgeest, beste Toon, ondanks dat jij geen direct begeleider was, is het tweede deel van het proefschrift grotendeels dankzij jou tot stand gekomen. Ook al waren niet alle resultaten altijd even goed te verklaren, jij bent steeds door blijven gaan met het scannen van handen en het verzamelen van gegevens. Bedankt voor jouw bijdrage en niet te vergeten, je enthousiasme en doorzettingsvermogen.

Dr. de Hair, beste Marjolein, ik wil je bedanken voor de eerste anderhalf jaar van mijn promotietraject. Samen met Paco hebben wij de eerste lijnen uitgezet van het onderzoek voor dit proefschrift. Daarnaast heb je mij geïntroduceerd op de afdeling en ervoor gezorgd dat ik mij hier thuis voel(de).

Verder wil ik alle (ex-)collega’s bedanken van de afdeling Reumatologie & Klinische Immunologie van het UMC Utrecht, in het bijzonder de mensen van H3 en het lab. Door het thuiswerkbeleid hebben we elkaar het laatste jaar weinig gesproken; ik denk vaak terug aan de koffiemomentjes en gezellige werksfeer. Hopelijk kunnen jullie dit snel hervatten! Speciaal wil ik Anne Karien bedanken voor de hulp en adviezen binnen de HandScan-onderzoeken, en natuurlijk de prettige samenwerking.

Daarnaast gaat mijn dank uit naar alle artsen en andere medewerkers van de afdeling Reumatologie & Klinische Immunologie. Met name Prof. dr. Hans Bijlsma, als coauteur op vijf manuscripten heeft u een belangrijk aandeel gehad. Vaak kwam u even binnen wandelen om bij te praten over de lopende projecten, ook een gemis gedurende het afgelopen jaar (thuiswerkbeleid). De researchverpleegkundigen wil ik bedanken voor alle werkzaamheden binnen de HandScan-onderzoeken.

Ook gaat mijn dank uit naar mensen buiten het UMC Utrecht met wie ik de afgelopen jaren heb samengewerkt en die een bijdrage hebben geleverd aan de totstandkoming van dit proefschrift. In het bijzonder dr. Michelle Borm (Roche Nederland B.V.) en dr. Attila Pethö-Schramm (F. Hoffman-La Roche) die nauw betrokken zijn geweest bij de

projecten gedurende het promotietraject. Bedankt voor de fijne samenwerkingen en deskundige input die jullie hebben geleverd bij de totstandkoming van de manuscripten. *Thank you for your pleasant collaboration and the input you provided, highly appreciated.* De medewerkers van Demcon Hemics B.V. wil ik bedanken voor de prettige samenwerking binnen de HandScan-onderzoeken. Verder wil ik Prof. dr. Robert de Jonge, Amsterdam UMC, en Helen Gosselt, Amsterdam UMC en Erasmus MC, bedanken voor de prettige samenwerking bij verschillende projecten. Prof. Dr. Schwarting, Dr. Triantafyllias en Dr. Heller, ACURA klinik Rheinland-Pfalz, wil ik bedanken voor de samenwerking binnen het HandScan-onderzoek, zonder jullie data was DAS-OST niet extern gevalideerd. *Wir danken Ihnen sehr herzlich für Ihre Mitarbeit, ohne Ihre Daten wäre DAS-OST nicht extern validiert worden!* Verder wil ik hier dr. Evert-Jan ter Borg, dr. Reinhard Bos, dr. George Bruyn, dr. Ed Griep, dr. Ruth Klaasen, dr. Suzanne Linn-Rasker en dr. Louise Meier, allen coauteurs, ook bedanken voor de prettige samenwerking.

Andere mensen met wie ik veel heb samengewerkt bij het uitvoeren van de HandScan-trial, die nog loopt, wil ik ook bedanken; alle reumatologen, reumaconsulten en overige betrokken medewerkers van Gelre Ziekenhuizen Apeldoorn, Máxima Medische Centrum Eindhoven, Meander Medische Centrum Amersfoort en Noordwest Ziekenhuisgroep Alkmaar. Bedankt voor jullie inzet! Daarnaast wens ik Matthijs van der Leeuw als nieuwe onderzoeker op dit project veel succes met het uitvoeren van dit onderzoek en het analyseren en publiceren van alle verkregen gegevens.

Mijn dankwoord wil ik afsluiten met het bedanken van mijn partner, familie en vrienden voor de support en afleidingen. Deze hebben een grote rol gespeeld tijdens mijn promotietijd. Vier personen wil ik speciaal bedanken:

Allereerste mijn ouders, Jos en Irma, lieve papa en mama, ik wil jullie bedanken voor de persoon die ik dankzij jullie geworden ben. Al 28 jaar lang steunen en motiveren jullie mij waar nodig. Zonder jullie had ik dit nooit kunnen bereiken, erg bedankt!

Mijn zus, lieve Justine, ik ken weinig mensen die zo positief in het leven staan en relaxed zijn als jij. Ook jou wil ik bedanken voor de goede gesprekken en de input die je hebt gegeven op de Nederlandse stukken van het proefschrift. Ik ben ontzettend trots op jou! Daarnaast wil ik je ook bedanken voor de kleding(adviezen).

Als laatste de persoon die het promotietraject van zeer nabij aan de zijlijn heeft gevuld. Lieve Tom, we zijn al enige tijd samen, en sedert kort ook samenwonend. De lockdown periodes en het afronden van het promotietraject waren voor onze relatie direct een goede test, die natuurlijk helemaal is geslaagd. Bedankt dat jij er altijd voor mij bent!

CURRICULUM VITAE

Maxime Marie Antoinette Verhoeven was born on the 12th of March 1993 in Oldenzaal, the Netherlands. In 2010, she finished secondary school at the Twents Carmel College (TCC) Lyceumstraat. In September 2010, she started studying Nutrition and Dietetics at HAN University of Applied Science. After a 1-year internship at the Radboud UMC Nijmegen, she obtained her Bachelor of Science (BSc) degree in July 2014. Afterwards she started with the research master Nutrition and Health at Wageningen University of Research. For her master thesis, she performed research on the association between visceral fat area and survival in colorectal cancer patients. To conduct this research, she had to assessed patients' body composition using preoperative CT images at the level of the 3rd lumbar vertebrae. During the summer holidays, she continued this work at the Wageningen University of Research. During her internship at a dairy company, "FrieslandCampina", she performed a systematic literature review and meta-analysis with the aim to assess the association between intake of different dairy sources of vitamin B12 and vitamin B12 status in healthy humans. These results have been published in 2019. After finishing her master thesis and internship, she obtained her Master of Science (MSc) degree in February 2017. In May 2017, she started as a PhD candidate at the department of Rheumatology & Clinical Immunology at the University Medical Center Utrecht under supervision of dr. J. Tekstra, dr. P.M.J. Welsing, Prof. dr. J.M. van Laar, and Prof. dr. F.P.J.G. Lafeber. Other members of the promotion team were dr. M.J.H. de Hair, who changed jobs at the end of 2018, and dr. J.W.G. Jacobs. During her PhD journey she performed research on the effectiveness, safety and cost effectiveness of different (medical) treatment strategies in (early) rheumatoid arthritis patients. Furthermore, she established the additional value of a (new) medical device, and implementing this device in clinical practice by developing and validating a new disease activity measurement instrument for rheumatoid arthritis patients. In addition, she coordinated a clinical trial, which is currently being performed in 5 centres in the Netherlands. Almost all her research has been published the recent years in peer-reviewed international journals. Currently she works at the "Kennisinstituut" as an advisor.

LIST OF PUBLICATIONS

Verhoeven MMA, Westgeest AAA, Tekstra J, van Laar JM, Lafeber FPJG, Welsing PMJ, Jacobs JWG. Classification of RA into active or inactive disease, using a modified DAS with a HandScan score replacing joint counts. *Submitted for publication*

Gosselt HR, **Verhoeven MMA**, Bulatovic-Calasan M, Welsing PMJ, de Rotte MCFJ, Hazes JMW, Lafeber FPJG, Hoogendoorn M, de Jonge R. Complex machine learning algorithms and multivariable logistic regression on par in prediction of insufficient clinical response to methotrexate in rheumatoid arthritis. *J. Pers. Med.* 2021;11:44

Verhoeven MMA, Westgeest AAA, Schwarting A, Jacobs JWG, Heller C, van Laar JM, Lafeber FPJG, Tekstra J, Triantafyllias K, Welsing PMJ. Development and validation of rheumatoid arthritis disease activity indices including HandScan (optical spectral transmission) scores. *Revision submitted to Arthritis Care Res.*

Verhoeven MMA, Tekstra J, Marijnissen ACA, Meier AJL, Westgeest AAA, Lafeber FPJG, Jacobs JWG, van Laar JM, Welsing PMJ. Utility of the HandScan in monitoring disease activity and prediction of clinical response in rheumatoid arthritis patients: an explorative study. *Rheumatology Advances in Practice*. 2021 January [Epub ahead of print]

Verhoeven MMA, Westgeest AAA, Jacobs JWG. A gender difference in HandScan scores in rheumatoid arthritis patients and controls? *J Rheumatol.* 2020 December [in press]

Verhoeven MMA, Tekstra J, Jacobs JWG, Bijlsma JWJ, van Laar JM, Pethö-Schramm A, Borm MEA, Lafeber FPJG, Welsing PMJ. Is tocilizumab monotherapy as effective in preventing radiographic progression in rheumatoid arthritis as its combination with methotrexate? *Arthritis Care Res.* 2020 November [Epub ahead of print]

Jacobs JWG, **Verhoeven MMA**, Welsing PMJ. Unravelling the cost of biological strategies in rheumatoid arthritis: a kaleidoscope of methodologies, interpretations and interests. *J Rheumatol.* 2020 November [in press]

Gosselt HR, **Verhoeven MMA**, de Rotte MCFJ, Pluijm SMF, Muller IB, Jansen G, Tekstra J, Bulatovic-Calasan M, Heil SG, Lafeber FPJG, Hazes JMW, de Jonge R. Validation of a prognostic multivariable prediction model for insufficient clinical response to methotrexate in early rheumatoid arthritis and its clinical application in Evidencio. *Rheumatol Ther.* 2020;7:837-850

Verhoeven MMA, Tekstra J, van Laar JM, Pethö-Schramm A, Borm MEA, Bijlsma JWJ, Jacobs JWG, Lafeber FPJG, Welsing PMJ. Effect on costs and quality-adjusted life-years of treat-to-target treatment strategies initiating methotrexate, or tocilizumab, or their combination in early rheumatoid arthritis. *J Rheumatol.* 2020 August [Epub ahead of print]

Verhoeven MMA, Tekstra J, Welsing PMJ, Pethö-Schramm A, Borm MEA, Bruyn GAW, Bos R, Griep EN, Klaasen R, van Laar JM, Lafeber FPJG, Bijlsma JWJ, de Hair MJH, Jacobs JWG. Effectiveness and safety over 3 years after the 2-year U-Act-Early trial of the strategies initiating tocilizumab and/or methotrexate. *Rheumatology.* 2020;59:2325-2333

Verhoeven MMA, de Hair MJH, Tekstra J, Bijlsma JWJ, van Laar JM, Pethö-Schramm A, Borm MEA, ter Borg EJ, Linn-Rasker SP, Teitsma XM, Lafeber FPJG, Jacobs JWG, Welsing PMJ. Initiating tocilizumab, with or without methotrexate, compared with starting methotrexate with prednisone within step-up treatment strategies in early rheumatoid arthritis: an indirect comparison of effectiveness and safety of the U-Act-Early and CAMERA-II treat-to-target trials. *Ann Rheum Dis.* 2019;78:1333-1338

Besselink NJ, Westgeest AAA, Klaasen R, Gamala M, van Woerkom JM, Tekstra J, **Verhoeven MMA**, van Spil WE, Lafeber FPJG, Marijnissen ACA, van Laar JM, Jacobs JWG. Novel optical spectral (OST)-guided versus conventionally disease activity-guided treatment: study protocol of a randomized clinical trial on guidance of a treat-to-target strategy for early RA. *Trials.* 2019;20:226

Verhoeven MMA, Welsing PMJ, Bijlsma JWJ, van Laar JM, Lafeber FPJG, Tekstra J, Jacobs JWG. Effectiveness of remission induction strategies for early rheumatoid arthritis: a systematic literature review. *Curr Rheumatol Rep.* 2019;21:24

Obeid R, Heil SG, **Verhoeven MMA**, van den Heuvel EGHM, de Groot LCPGM, Eussen SJPM. Vitamine B12 intake from animal foods, biomakers, and health aspects. *Front. Nutr.* 2019;6:93

LIST OF CONFERENCE ABSTRACTS

Development of a disease activity index for rheumatoid arthritis patients using HandScan. Poster presentation at the ACR congress 2020, November 6-11, e-Congress.

Development of a disease activity index for rheumatoid arthritis patients using HandScan. Poster presentation at the EULAR congress 2020, June 3-6, e-Congress.

Impact on costs and quality of life over 5 years of treat-to-target treatment strategies initiating tocilizumab, methotrexate or their combination in early rheumatoid arthritis. Economic evaluation of the U-Act-Early trial. Poster presentation at the ACR congress 2019, November 8-13, Atlanta, GA, USA.

Validation of a prognostic multivariable prediction model for insufficient clinical response to methotrexate in early rheumatoid arthritis. Poster presentation at the ACR congress 2019, November 8-13, Atlanta, GA, USA.

Individual patient data meta-analysis of the effectiveness of tocilizumab on inhibiting radiographic progression in rheumatoid arthritis. Poster presentation at the NVR congress 2019, September 25-27, Arnhem, the Netherlands.

Individual patient data meta-analysis of the effectiveness of tocilizumab on inhibiting radiographic progression in rheumatoid arthritis. Poster presentation at the EULAR congress 2019, June 12-15, Madrid, Spain.

Does initiating tocilizumab lead to better disease control compared to initiating MTX with low-moderate dose prednisone in early rheumatoid arthritis; an indirect comparison of U-Act-Early and CAMERA-II treat-to-target trials. Poster presentation at the EULAR congress 2019, June 12-15, Madrid, Spain.

Effectiveness of remission-induction strategies for early rheumatoid arthritis: a systematic literature review. Poster presentation at the EULAR congress 2019, June 12-15, Madrid, Spain.

U-Act-Early trial 3 years follow-up: Radiographic joint damage and use of bDMARDs over 5 years in early RA patients treated-to-target with strategies initiating tocilizumab, methotrexate or their combination. Poster presentation at the ACR congress 2018, October 19-24, Chicago, IL, USA.

U-Act-Early trial 3 years follow-up. The longer effectiveness of treat-to-target strategies in early rheumatoid arthritis with tocilizumab, methotrexate, or their combination. Oral presentation at the NVR congress 2018, September 27-28, Arnhem, the Netherlands.

An indirect comparison of effectiveness and safety of three treat-to-target treatment strategies: Tocilizumab with or without MTX and MTX with prednisone. Poster presentation at the NVR congress 2018, September 27-28, Arnhem, the Netherlands.

U-Act-Early trial 3 years follow-up. The longer effectiveness of treat-to-target strategies in early rheumatoid arthritis with tocilizumab, methotrexate, or their combination. Poster presentation at the EULAR congress 2018, June 13-16, Amsterdam, the Netherlands.

