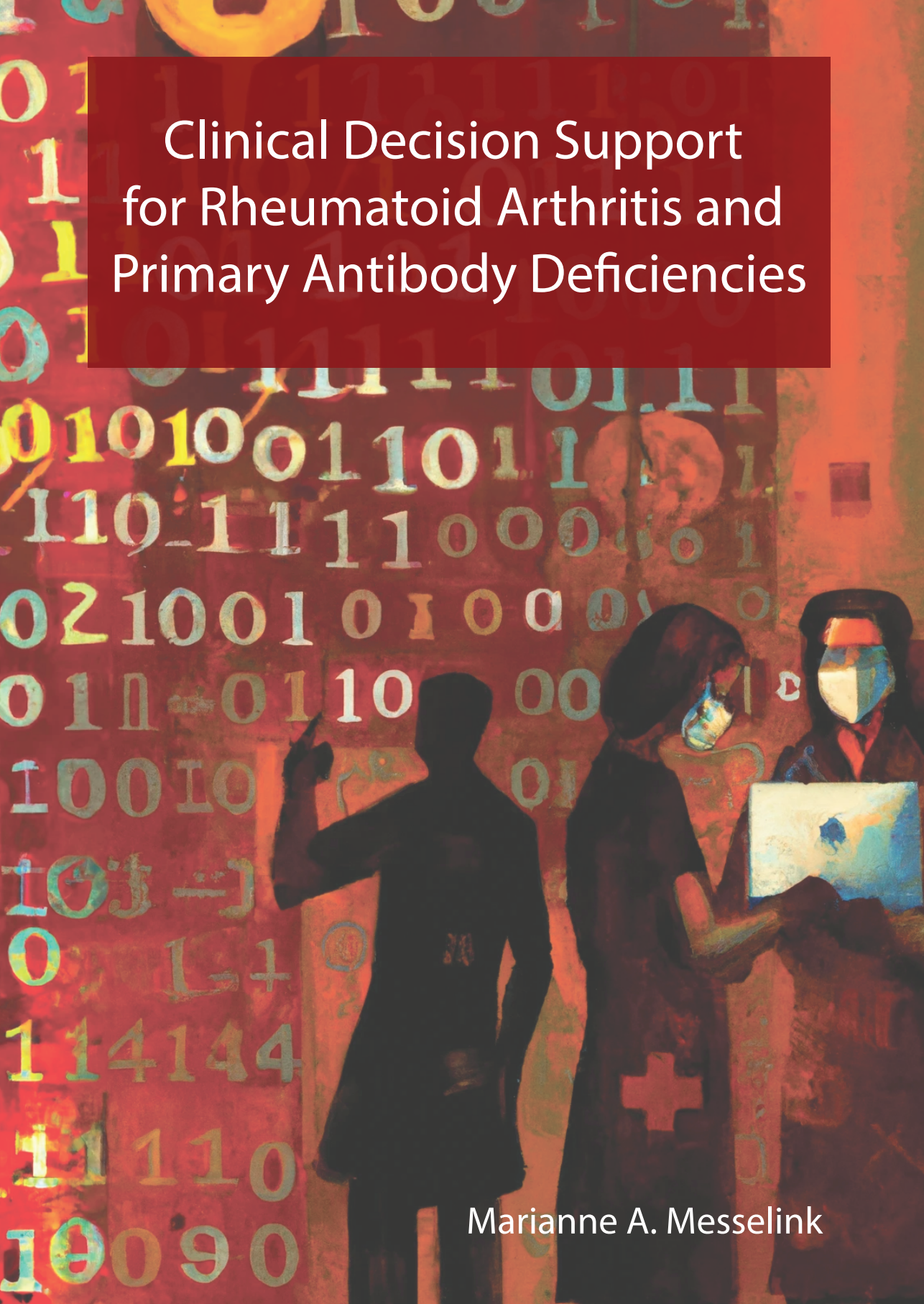
Clinical Decision Support for Rheumatoid Arthritis and Primary Antibody Deficiencies



Marianne A. Messelink

Clinical Decision Support for Rheumatoid Arthritis & Primary Antibody Deficiencies

Marianne Alice Messelink

DOI: https://doi.org/10.33540/2263
Cover designed using DALL-E || openai.com/dall-e-2
Printed by: Proefschriftmaken || Proefschriftmaken.nl

Printing of this thesis was financially supported by CSL Behring B.V., Galapagos N.V. and Takeda Nederland B.V.

Clinical Decision Support for Rheumatoid Arthritis & Primary Antibody Deficiencies

Klinische beslisondersteuning voor reumatoïde artritis en primaire antistofdeficiënties
(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

dinsdag 4 juni 2024 des ochtends te 10:15

door

Marianne Alice Messelink

geboren op 2 december 1992
te Margraten

Promotoren
Prof. dr. F.P.J.G. Lafeber
Dr. A.A. den Broeder

Copromotoren
Dr. P.M.J. Welsing
Dr. H.L. Leavis

Beoordelingscommissie
Prof. dr. M.S. de Bruin-Weller
Prof. dr. J. Frenkel (voorzitter)
Prof. dr. van der Helm
Prof. dr. D.L. Oberski
Dr. R.P. Venekamp

Table of contents

Chapter 1

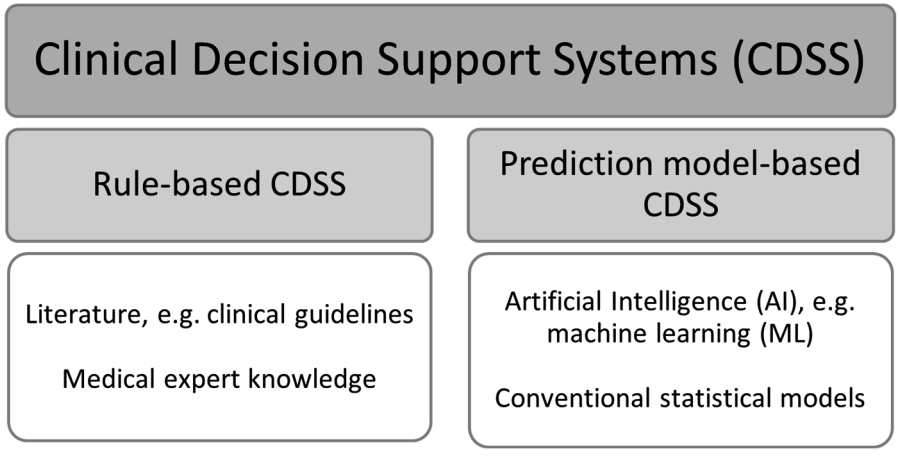
General introduction

Chapter 1

Clinical Decision Support

The surge in the digitalization of healthcare services has laid the ground for the rapid evolution of clinical decision support systems (CDSSs) since the 1990s.[1,2] A CDSS is an electronic system designed to aid in clinical decision-making, in which individual patient-, disease or treatment characteristics are used to generate patient-specific assessments and recommendations, that are then presented to clinicians for consideration.[3] In this thesis we will make a distinction between 'rule-based' and 'prediction model-based' CDSSs (Figure 1).[5] Rule-based CDSSs use 'if-then' statements, which can be derived from clinical guidelines, other published literature or medical expert knowledge. An example of a rule-based CDSS is a pop-up warning for medication interactions. Prediction model-based CDSSs use conventional statistical- or artificial intelligence (AI) methods such as machine learning (ML). These models are developed using individual patient data. There is an expanding number of publications about the development of prediction models within the fields of rheumatology and clinical immunology.[6-10] Examples include prediction models for early diagnosis, disease progression, prognosis or treatment response.[11-14]

Figure 1



The prediction models used in a model-based CDSS predict the probability of a certain outcome, e.g. the risk of a flare of disease activity. This predicted risk is usually accompanied by a cut-off value. For example, a predicted risk of a flare of ≥ 20% could be classified as a 'high-risk', and everything < 20% as 'low-risk'. Dichotomizing the risk (e.g. high-risk and low-risk), rather than presenting it as a continuum (e.g. 13% or 36%), allows to convert the risk into a clear treatment advice. For example, in case of a high predicted risk of flare, consider increasing the medication dose. The

dichotomization with a cut-off value also enables the calculation of certain measures of predictive performance (e.g. sensitivity), which will be discussed below.

Prediction model performance

The performance of prediction models can be measured in many ways. Measures of performance that require a cut-off value include sensitivity, specificity, or positive- and negative predictive values. Sensitivity and specificity can also be plotted in a receiver operator curve (ROC), from which the area under the curve (AUC) can be derived as an overall measure of predictive performance irrespective of a specific cut-off value. Validation of prediction models comprises multiple steps. During internal validation, the performance of the model is assessed within the same dataset from which the model was developed. This can be done using cross-validation, in which the development data is split into a training-set and a testing-set. Multiple rounds (or 'folds') of cross-validation can be performed using different partitions. However, internal validation can give an optimistic estimate of performance due to overfitting. Therefore, an important subsequent step is to evaluate performance in a different dataset (e.g. from a different hospital), called external validation. Model performance should be monitored continuously after clinical implementation, as over time a 'temporal drift' may occur. This can be due to e.g. availability of new treatments, changes in patient population, new disease classifications, or changes in coding habits. Guidelines such as the TRIPOD statement (Transparent Reporting a of multivariable prediction model for Individual Prognosis or Diagnosis) and the STARD guideline (Standards for Reporting Diagnostic accuracy studies) provide a framework for reporting the development and validation of prediction models.[15]

Clinical impact

The effectiveness of a CDSS does not just depend on the predictive accuracy, but on the clinical impact it has. Clinical impact can initially be estimated based on metrics such as the 'number needed to treat' or the 'number needed to harm'. Alternatively, the effect of prediction-aided decisions on outcomes (e.g. disease flares, medication dose) can be simulated. Such estimates should subsequently be studied in a test-treatment trial (when feasible), where treatments with and without the CDSS are compared and health outcomes are measured. Several reviews have described trials on the effects of CDSSs on health outcomes, including both rule-based and (more recently) model-based CDSSs. Reviews of CDSSs in diabetes have shown improvements in glucose levels and blood pressure, and in oncology improvements in guideline adherence and prescriptions errors.[16,17] However, other reviews conclude that the impact of CDSSs on patient outcomes was insufficiently reported or lacking.[18,19] Several factors may contribute to the clinical impact of the CDSS, such as the clarity of the advice, the acceptance and

adherence by patients and physicians, and whether the recommended treatment strategy is more effective than usual care.

Implementation

A CDSS which has been shown to improve health outcomes must meet several further requirements before implementation. For example, the necessary patient data should be accessible and the CDSS must be compatible with the clinical workflow. Furthermore, the patient burden of implementing the CDSS (e.g. additional tests or overdiagnosis) should be taken into account. A treatment strategy based on a CDSS should also be cost-effective in comparison to the current standard of care. Moreover, the CDSS must comply with legal frameworks such as the medical device regulation (MDR) and general data protection regulation (GDPR).[20,21] These and other challenges of implementation will be further addressed in the discussion of this thesis.

Routine care data

In order for a CDSS (rule-based or model-based) to provide a treatment advice, individual patient data is required as input. Model-based CDSSs also require individual patient data for model development. This data can consist of patient-, disease- and treatment characteristics. Examples include demographics, medical history, laboratory results, imaging, results from physical exams and medication prescriptions. In this thesis we will mainly focus on CDSSs that use routine care data. This is data that is already routinely collected in usual clinical care, rather than e.g. complex molecular biomarkers. The data can be either structured (in a predefined format, e.g. laboratory results) or unstructured (e.g. clinical notes). The use of routine care data has several important advantages.[22] First, it is available for a vast number of patients. This is especially of interest for rare diseases, where it can be difficult to obtain sufficient sample sizes in clinical studies. Second, it is often available over many years of follow-up. This allows to study the development of a disease over time, and can increase patient numbers by including historic patient data. Third, analyses based on routine care data reduce the risk of selection bias compared to clinical trials. Several studies have shown that the patients included in clinical trials often do not represent the target population regarding e.g. sex, ethnicity, age, pregnancy/lactation or comorbidities.[23-25] This selection may hamper the generalizability of the resulting prediction models to the population of interest. Fourth, the wide variety of available variables allows to study many different risk factors. Lastly, CDSSs that use routine care data can be implemented without the need for additional (expensive) exams that disrupt the usual workflow. The use of routine care data also poses several challenges, such as missing- and unstructured data and privacy concerns. These will be further addressed in the general discussion.

Rheumatoid arthritis

This thesis will cover several studies focused on clinical decision support for rheumatoid arthritis (RA). RA is a chronic, systemic, autoimmune inflammatory disorder that primarily involves synovial joints.[26] Extra-articular organ involvement such as cardiovascular- or pulmonary disease can however also manifest.[27] Although several risk factors for developing RA have been identified, the etiology is not fully known.[28] RA treatment is aimed at limiting and controlling disease activity, as prolonged high levels of activity result in patient burden and increase the risk of progressive joint damage and mortality.[29-31]

In the proactive treat-to-target (T2T) strategy, RA disease activity is frequently and systematically assessed using a validated measure, which is then compared to a pre-specified treatment target. If the target is not reached within a particular timeframe, treatment is intensified.[32] Targets are commonly based on a composite disease activity score with a cut-off value for remission or low disease activity (LDA).[33] Examples of composite scores include the Disease Activity Score (DAS, including 28 or 44 joints in the DAS28/DAS44 respectively), the Simplified- and the Clinical Disease Activity Indices (SDAI/CDAI).[34]

The recommended first line treatment for RA consists of conventional synthetic disease modifying antirheumatic drugs (csDMARDs), which may be accompanied by short-term glucocorticoids.[35] In case of inadequate response, biological (b) or targeted synthetic (ts) DMARDs should be considered. Within this thesis, we specifically focus on bDMARDs. bDMARDs are complex molecules produced by cells that can inhibit proinflammatory cells, cellular interactions, or cytokines.[36] Examples include TNF-α blockers (e.g. adalimumab, etanercept), B or T cell inhibitors (rituximab, abatacept) and IL-6 receptor blockers (tocilizumab, sarilumab). Although bDMARDs are effective in the treatment of RA, they may also lead to adverse events, require self-injections or hospital visits and are relatively expensive (although net prices have dropped considerably in many countries).[37-39] Thus, once a stable treatment target has been reached, tapering bDMARDs to the lowest effective dose is of great clinical and societal interest. Tapering bDMARDs however also increases the risk of flares of disease activity.[40,41] It is thus relevant to explore methods to reduce the risk of disease flares during bDMARD tapering.

Although b/tsDMARDs are highly effective, there is a subgroup of patients that remain symptomatic after treatment with several b/tsDMARDs in accordance with EULAR guidelines.[35] These patients are referred to as having 'difficult-to-treat' RA.[42,43] Identification of D2T RA in routine care data can allow to study the development of this

disease state over time. In a clinical setting, identification may serve as a warning system to facilitate optimization of treatment according to published recommendations (instead of just 'cycling' to a next b/tsDMARD).[44] Furthermore, predicting the risk of future D2T RA development early in the RA disease course could allow for close monitoring of these patients for factors contributing to D2T RA, ultimately aiming to prevent this disease state.

Primary antibody deficiencies

This thesis also described the development of a CDSS for primary antibody deficiencies (PAD). PAD are a heterogeneous group of immunodeficiencies that are characterized by an inability to produce a clinically effective antibody response.[45,46] Examples of clinically relevant PAD include common variable immunodeficiency (CVID), IgG subclass deficiency and specific antibody deficiency (SpAD).[47] The clinical presentation encompasses a wide range of symptoms including increased susceptibility to respiratory- and gastro-intestinal tract infections, autoimmunity, and an increased risk of certain malignancies.[46,48] Due to the heterogeneous presentation and low prevalence, diagnosis of PAD can be challenging. This is evident from the reported median delay in diagnosis between 2–10 years.[49-54] Numerous studies have shown that diagnostic delay is associated with increased infections, hospitalizations and mortality.[48,49,51,55-57] As effective therapies are available, reducing the diagnostic delay of PAD is of clinical interest.[54]

Thesis outline

Part 1: Patient perspective

In **chapter 2** the RA-patient perspective about the use of prediction models in medical decision-making is explored using semi-structured interviews and thematic analysis. The results describe factors to consider during the development and implementation of prediction model based CDSSs in order to enhance patient acceptability.

Part 2: Diagnostic decision support

In **chapter 3**, various machine learning techniques are applied for the identification and prediction of difficult-to-treat RA (D2T RA) using structured and unstructured routine care data. The developed models provide the first step towards a CDSS for the early detection of D2T RA to facilitate optimization of therapy.

Chapter 4 and **chapter 5** focus on the development and validation of a clinical decision support for the early detection of PAD in primary care. In chapter 4, a rule-based CDSS

is developed with components (e.g. pneumonia) based on literature and clinical expertise. The weights of individual components of this CDSS are based on analysis of aggregated data from primary care of PAD patients and controls. In chapter 5, the CDSS is clinically validated, and further refined by adding a prediction model to improve performance within a subset of high-risk patients using individual patient data.

Part 3: Treatment decision support

As mentioned, T2T is the recommended treatment strategy for RA. However, there is no consensus about the optimal target, e.g. DAS28 remission or SDAI low disease activity. In **chapter 6**, the effect of different targets on clinical and radiographical outcomes is compared in a systematic review and meta-regression analysis. Determining the optimal target can support the decision to start, stop, or taper (b)DMARD treatment.

In the PATIO study, the use of a model-based CDSS for tapering bDMARDs in RA patients is compared to tapering without this prediction support within a pragmatic test-treatment RCT. The model predicts the risk of a flare occurring at each step of the tapering process. In case of a high predicted risk, it is advised to stop tapering, aiming to prevent a flare from occurring. The main aim of this CDSS is to minimize the number of flares during bDMARD tapering whilst still achieving a considerable dose-reduction. The development, validation and simulated impact of the flare prediction model is described in **chapter 7**, and the PATIO study protocol in **chapter 8.**

Finally, **chapter 9** summarizes the results of the different chapters, discussing them in a broader context and providing suggestions for future research.

References

1. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform 2016;Suppl 1(Suppl 1):S103-16.
2. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. Ann Intern Med 1994;120(2):135-42.
3. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998;280(15):1339-46.
4. Augstein P, Heinke P, Vogt L, Kohnert KD, Salzsieder E. Patient-Tailored Decision Support System Improves Short- and Long-Term Glycemic Control in Type 2 Diabetes. J Diabetes Sci Technol 2022;16(5):1159-66.
5. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17.
6. Arbeeva L, Minnig MC, Yates KA, Nelson AE. Machine Learning Approaches to the Prediction of Osteoarthritis Phenotypes and Outcomes. Curr Rheumatol Rep 2023;25(11):213-25.
7. Madrid-García A, Merino-Barbancho B, Rodríguez-González A, Fernández-Gutiérrez B, Rodríguez-Rodríguez L, Menasalvas-Ruiz E. Understanding the role and adoption of artificial intelligence techniques in rheumatology research: An in-depth review of the literature. Semin Arthritis Rheum 2023;61:152213.
8. Wang J, Tian Y, Zhou T, Tong D, Ma J, Li J. A survey of artificial intelligence in rheumatoid arthritis. Rheumatol Immunol Res 2023;4(2):69-77.
9. Galozzi P, Basso D, Plebani M, Padoan A. Artificial intelligence and laboratory data in rheumatic diseases. Clin Chim Acta 2023;546:117388.
10. Rider NL, Srinivasan R, Khoury P. Artificial intelligence and the hunt for immunological disorders. Curr Opin Allergy Clin Immunol 2020;20(6):565-73.
11. Liu J, Zhu Q, Han J, Zhang H, Li Y, Ma Y, et al. IgG Galactosylation status combined with MYOM2-rs2294066 precisely predicts anti-TNF response in ankylosing spondylitis. Mol Med 2019;25(1):25.
12. Stoel BC. Artificial intelligence in detecting early RA. Semin Arthritis Rheum 2019;49(3S):S25-S8.
13. B N, B.S G, L T, D L, M G, B O, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA network open 2019.
14. Myasoedova E, Athreya AP, Crowson CS, Davis JM, 3rd, Warrington KJ, Walchak RC, et al. Toward Individualized Prediction of Response to Methotrexate in Early Rheumatoid Arthritis: A Pharmacogenomics-Driven Machine Learning Approach. Arthritis Care Res (Hoboken) 2022;74(6):879-88.
15. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ (Online) 2015;350(January):1-9.
16. Huang S, Liang Y, Li J, Li X. Applications of Clinical Decision Support Systems in Diabetes Care: Scoping Review. J Med Internet Res 2023;25:e51024.
17. Nafees A, Khan M, Chow R, Fazelzad R, Hope A, Liu G, et al. Evaluation of clinical decision support systems in oncology: An updated systematic review. Crit Rev Oncol Hematol 2023;192:104143.
18. Jia P, Zhang L, Chen J, Zhao P, Zhang M. The Effects of Clinical Decision Support Systems on Medication Safety: An Overview. PLoS One 2016;11(12):e0167683.
19. Susanto AP, Lyell D, Widyantoro B, Berkovsky S, Magrabi F. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. J Am Med Inform Assoc 2023;30(12):2050-63.
20. Medical Device Regulation: European Union;Accessed on 2023Dec 7. Available from: https://www.medical-device-regulation.eu/download-mdr/.
21. General Data Protection Regulation: European Union;Accessed on 2023Dec 7. Available from: https://gdpr-info.eu/.

22. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. Ann Rheum Dis 2023;82(3):306-11.
23. Strait A, Castillo F, Choden S, Li J, Whitaker E, Falasinnu T, et al. Demographic Characteristics of Participants in Rheumatoid Arthritis Randomized Clinical Trials: A Systematic Review. JAMA Netw Open 2019;2(11):e1914745.
24. Ramirez FD, Motazedian P, Jung RG, Di Santo P, MacDonald Z, Simard T, et al. Sex Bias Is Increasingly Prevalent in Preclinical Cardiovascular Research: Implications for Translational Medicine and Health Equity for Women: A Systematic Assessment of Leading Cardiovascular Journals Over a 10-Year Period. Circulation 2017;135(6):625-6.
25. Hoel AW, Kayssi A, Brahmanandam S, Belkin M, Conte MS, Nguyen LL. Under-representation of women and ethnic minorities in vascular surgery randomized controlled trials. J Vasc Surg 2009;50(2):349-54.
26. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. Bone Res 2018;6:15.
27. Figus FA, Piga M, Azzolin I, McConnell R, Iagnocco A. Rheumatoid arthritis: Extra-articular manifestations and comorbidities. Autoimmun Rev 2021;20(4):102776.
28. van der Woude D, van der Helm-van Mil AHM. Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis. Best Pract Res Clin Rheumatol 2018;32(2):174-87.
29. Book C, Saxne T, Jacobsson LT. Prediction of mortality in rheumatoid arthritis based on disease activity markers. J Rheumatol 2005;32(3):430-4.
30. Mikuls TR, Fay BT, Michaud K, Sayles H, Thiele GM, Caplan L, et al. Associations of disease activity and treatments with mortality in men with rheumatoid arthritis: results from the VARA registry. Rheumatology (Oxford) 2011;50(1):101-9.
31. Listing J, Kekow J, Manger B, Burmester GR, Pattloch D, Zink A, Strangfeld A. Mortality in rheumatoid arthritis: the impact of disease activity, treatment with glucocorticoids, TNFα inhibitors and rituximab. Ann Rheum Dis 2015;74(2):415-21.
32. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2016;75(1):3-15.
33. Fleischmann RM, Van Der Heijde D, Gardiner PV, Szumski A, Marshall L, Bananis E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open 2017;3(1):e000382.
34. Aletaha D, Smolen JS. The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract Res Clin Rheumatol 2007;21(4):663-75.
35. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82(1):3-18.
36. Rein P, Mueller RB. Treatment with Biologicals in Rheumatoid Arthritis: An Overview. Rheumatol Ther 2017;4(2):247-61.
37. Joensuu JT, Huoponen S, Aaltonen KJ, Konttinen YT, Nordström D, Blom M. The cost-effectiveness of biologics for the treatment of rheumatoid arthritis: a systematic review. PLoS One 2015;10(3):e0119683-e.
38. Bittner B, Richter W, Schmidt J. Subcutaneous Administration of Biotherapeutics: An Overview of Current Challenges and Opportunities. Biodrugs 2018;32(5):425-40.
39. Ramiro S, Sepriano A, Chatzidionysiou K, Nam JL, Smolen JS, Van Der Heijde D, et al. Safety of synthetic and biological DMARDs: A systematic literature review informing the 2016 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis 2017;76(6):1093-101.
40. Van Herwaarden N, Van Maas AD, Minten MJM, Van Den Hoogen FHJ, Kievit W, Van Vollenhoven RF, et al. Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis:Open label, randomised controlled, non-inferiority trial. BMJ (Online) 2015;350.

41. Fautrel B, Pham T, Alfaiate T, Gandjbakhch F, Foltz V, Morel J, et al. Step-down strategy of spacing TNF-blocker injections for established rheumatoid arthritis in remission: Results of the multicentre non-inferiority randomised open-label controlled trial (STRASS: Spacing of TNF-blocker injections in Rheumatoid ArthritiS St. Ann Rheum Dis 2016;75:59-67.
42. Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, van der Goes MC, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2021;80(1):31-5.
43. de Hair MJH, Jacobs JWG, Schoneveld JLM, Van Laar JM. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. Rheumatology (Oxford) 2017;57(7):1135-44.
44. Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, van der Goes MC, et al. EULAR points to consider for the management of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2022;81(1):20-33.
45. Wood PM. Primary antibody deficiency syndromes. Curr Opin Hematol 2010;17(4):356-61.
46. Durandy A, Kracker S, Fischer A. Primary antibody deficiencies. Nat Rev Immunol 2013;13(7):519-33.
47. Fried AJ, Bonilla FA. Pathogenesis, diagnosis, and management of primary antibody deficiencies and infections. Clin Microbiol Rev 2009;22(3):396-414.
48. Wood P, Stanworth S, Burton J, Jones A, Peckham DG, Green T, et al. Recognition, clinical diagnosis and management of patients with primary antibody deficiencies: a systematic review. 2007;149(3):410-23.
49. Seymour B, Miles J, Haeney M. Primary antibody deficiency and diagnostic delay. J Clin Pathol 2005;58(5):546-7.
50. Slade CA, Bosco JJ, Giang TB, Kruse E, Stirling RG, Cameron PU, et al. Delayed diagnosis and complications of predominantly antibody deficiencies in a cohort of Australian adults. Front Immunol 2018;14:9:694.
51. Aghamohammadi A, Bahrami A, Mamishi S, Mohammadi B, Abolhassani H, Parvaneh N, Rezaei N. Impact of delayed diagnosis in children with primary antibody deficiencies. J Microbiol Immunol Infect 2011;44(3):229-34.
52. Maarschalk-Ellerbroek LJ, Hoepelman AIM, Van Montfrans JM, Ellerbroek PM. The spectrum of disease manifestations in patients with common variable immunodeficiency disorders and partial antibody deficiency in a university hospital. J Clin Immunol 2012;32(5):901-21.
53. El-Helou SM, Biegner AK, Bode S, Ehl SR, Heeg M, Maccari ME, et al. The German National Registry of Primary Immunodeficiencies (2012-2017). Front Immunol 2019;10:1272.
54. Odnoletkova I, Kindle G, Quinti I, Grimbacher B, Knerr V, Gathmann B, et al. The burden of common variable immunodeficiency disorders: a retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. Orphanet J Rare Dis 2018;13(1):201.
55. Graziano V, Pecoraro A, Mormile I, Quaremba G, Genovese A, Buccelli C, et al. Delay in diagnosis affects the clinical outcome in a cohort of cvid patients with marked reduction of iga serum levels. Clin Immunol 2017;180:1-4.
56. Busse PJ, Razvi S, Cunningham-Rundles C. Efficacy of intravenous immunoglobulin in the prevention of pneumonia in patients with common variable immunodeficiency. J Allergy Clin Immunol 2002;109(6):1001-4.
57. Chan HY, Yang YH, Yu HH, Chien YH, Chiang LL, Chiang BL. Clinical characteristics and outcomes of primary antibody deficiency: A 20-year follow-up study. J Formos Med Assoc 2014;113(6):340-8.

PART I
Patient perspective

Chapter 2

Rheumatoid arthritis patients' perspective on the use of prediction models in clinical decision-making

Marianne A. Messelink*
Sina Fadaei*
Lise M. Verhoef
Paco M.J. Welsing
Nienke C. Nijhof
Heleen Westland

**these authors contributed equally*

Accepted for publication in Rheumatology

Abstract

Objectives A rapidly expanding number of prediction models is being developed aiming to improve rheumatoid arthritis (RA) diagnosis and treatment. However, few are actually implemented in clinical practice. This study explores factors influencing the acceptance of prediction models in clinical decision-making by RA patients.

Methods A qualitative study design was used with thematic analysis of semi-structured interviews. Purposive sampling was applied to capture a complete overview of influencing factors. The interview topic list was based on pilot data.

Results Data saturation was reached after 12 interviews. Patients were generally positive about the use of prediction models in clinical decision-making. Six key themes were identified from the interviews. First, patients have the need for information on prediction models. Second, factors influencing trust in model-supported treatment are described. Third, patients envision the model to have a supportive role in clinical decision-making. Fourth, patients hope to personally benefit from model-supported treatment in various ways. Fifth, patients are willing to contribute time and effort to contribute to model input. And lastly, we discuss the theme on effects of the relationship with the caregiver in model-supported treatment

Conclusion In this study, RA patients were generally positive about the use of prediction models in their treatment given certain conditions were met and concerns addressed. The results of this study can be used during the development and implementation in RA care of prediction models in order to enhance patient acceptability.

Introduction

An expanding number of publications highlight the potential of prediction models to decrease diagnostic delay and facilitate more effective treatment in healthcare, including in the field of rheumatoid arthritis (RA).[1-3] These models are based on computational techniques, such as machine learning, that identify patterns in healthcare data in order to make predictions for individual patients. Healthcare data can include medical history, radiological examinations, patient-reported outcomes and laboratory values. Prediction models embedded in user-friendly software can be used to support clinical decision-making. As an example, the PATIO (Prediction Aided Tapering In rheumatoid arthritis patients treated with biOlogicals) trial investigates the use of a flare prediction model during medication tapering, aiming to minimize flares whilst still achieving a considerable dose reduction.[4,5] Other examples of RA prediction models include early detection of RA, identification of RA subtypes for specific treatment, and prediction of treatment responses.[6-10]

Despite the abundance of developed models, very few are used in clinical practice. Key challenges for implementation include uncertainties about the clinical impact, the integration with electronic health records, legal concerns and acceptance by physicians and patients.[11,12] Previous studies have shown that patients wish to be involved in the integration of prediction models in healthcare, and that their support is necessary to enable meaningful implementation.[13,14] Furthermore, the EULAR recommendations for the use of big data in rheumatology stress the importance of early consideration of implementation and engagement of patients.[15] Thus, in order to allow the field of prediction model aided rheumatoid arthritis (RA) care to move forward, it is essential to get insight into the perspective of patients. Under which conditions would they allow a prediction model in their medical treatment? What are their hopes and concerns? The current study aims to explore factors influencing the acceptance of prediction models in clinical decision-making by RA patients.

Methods

An exploratory qualitative study design with semi-structured interviews was used. The need for ethical approval was waived by the Medical Research Ethics Committee Utrecht (file number 21/340). Informed consent was obtained verbally at the start of the interview.

Sampling and recruitment

We recruited patients from the outpatient rheumatology clinic of the University Medical Centre Utrecht in the Netherlands. Patients aged ≥18 years with a clinical RA diagnosis

and sufficient proficiency in the Dutch language were eligible to participate. Purposive sampling was applied in which we strived for maximum variation in the sample regarding gender, age, and technological proficiency, as these factors were expected to influence patient acceptance of prediction models. Technological proficiency was initially judged by the caregiver and subsequently verbally verified during the interviews using the six questions of the Pharos Quickscan for digital health literacy (S1 Appendix).[16] Sampling was continued until data saturation was achieved, i.e. when no new themes arose from the last three interviews. Eligible patients were invited to participate by their treating health care professional. Patients who were interested to participate received written study information from the research team.

Semi-structured interviews

Data were collected through semi-structured interviews using a topic list. The topic list was developed based on exploratory pilot data from three data collections conducted by a medical doctor and patient research partner (MM and NN): 1. an online open questionnaire (37 complete responses, 32 incomplete, patients with a rheumatic disorder); 2. a focus-group (3 RA patients); and 3. an individual interview (1 RA patient). The research team used the insights from these data to develop the topic list (MM, SF, LV, NN, HW), see S2 Appendix. At the start of the interview, the concept of prediction models was explained using examples of potential applications, e.g. using a prediction model during medication tapering. It was verified if the concept of prediction models as a clinical decision aid was clear for patients. All interviews started with the question: 'What are your first thoughts about using a predictive computer model in your medical care?'. Subsequent questions focused on factors influencing the acceptance of prediction models in clinical decision-making. Interviews were performed either in person, by telephone or online using Microsoft Teams, according to the preference of the patient. Most interviews [11/12] were conducted by a male and female interviewer together (SF and MM), one interview was conducted by SF only. Both interviewers were medical doctors and PhD-candidates and had research experience with prediction models and clinical RA care, which was communicated with patients at the start of the interview. Interviewers and patients did not know each other beforehand. The interview guide and techniques were evaluated and refined continuously (MM, SF, LV, HW). All interviews were audio recorded. Interviews lasted 41-76 minutes (mean 54 minutes). Transcriptions of the interviews were sent to the interviewees for verification.

Data-analysis

We used the six phases of thematic analysis as described by Braun and Clarke.[17] Data-analysis started after the first four interviews and continued after every subsequent interview. Interviews were, transcribed verbatim, checked for accuracy and initial ideas

for coding were made (MM, SF). The first four interviews were coded independently by both MM and SF, after which codes were discussed and an initial code list was agreed upon (MM, SF, LV, HW). Subsequent interviews were coded by either MM or SF while the other provided a cross-check, making additions or changes where necessary. The topic list was continuously adjusted to further explore new codes that arose from the interviews. Initial codes were collated into potential themes (MM, SF), and their preliminary descriptions were discussed (MM, SF, LV, PW, NN, HW). The transcripts were re-read by SF and MM at multiple timepoints to ensure the validity of the themes with the primary data. Potential themes were further refined (MM, SF, LV, PW, NN, HW), and specific content of each theme was worked out and themes were defined. Themes were shared with the participants to provide feedback. The report was drafted and representative quotes were selected to illustrate the themes (MM, SF, LV, HW) and were reviewed by all authors. Data-analysis was supported by NVivo software (QSR International Pty Ltd, Version 12.0, 2017).

Credibility and dependability of the data was increased by researcher triangulation and regular review within the team throughout all phases of the study.[18] Two experts on qualitative research (LV, HW) were involved in the data collection and analysis to strengthen the accuracy and dependability of the process.[18]. Member check (verification of data and results by participants), memo writing and the involvement of a patient partner (NN) in the pilot studies and beyond supported the study reliability. The consolidated criteria for reporting qualitative studies (COREQ) checklist was used for reporting of the results.[19]

Results

Data saturation was reached after 12 interviews. Characteristics of the patients are shown in Table 1. Most were female, with ages ranging from 34-84 years, and a disease duration varying from 6 to 52 years. Regarding digital proficiency, all patients had access to internet, but two patients did not use the internet to search for information. Five patients were not able to use Microsoft Teams, and therefore four patients were interviewed in person and one by phone. Almost all patients (11/12) regarded themselves to be, in general, positive about the use of prediction models in clinical decision-making.

Six key themes emerged from the interviews covering factors that influence the acceptance of prediction models in clinical decision making: 1. Need for information about the model, 2. Trust in model-supported treatment, 3. Role of the model in medical decision-making, 4. Personal benefit from model-supported treatment, 5. Willingness

to contribute to model input, and 6. Relationship with the caregiver. In the following section the themes will be discussed and illustrated with quotes (Table 2).

Table 1. Patient characteristics

General	**Total (n=12)**
Age, median (min-max)	61 (34-84)
Female, n (%)	9 (75%)
Disease duration in years, median (min-max)	23 (6-52)
Number of csdmards used, median (min-max)	3 (1-7)
Number of bdmards used, median (min-max)	1 (0-5)
Number of tsdmards used, median (min-max)	0 (0-0)
Digital proficiency	
Has access to internet, n (%)	12 (100%)
Uses the internet to search for information, n (%)	10 (83%)
Has an e-mail address, n (%)	12 (100%)
Uses applications on phone or tablet, n (%)	10 (83%)
Can download an application on a phone or tablet, n (%)	10 (83%)
Uses digid (an online identification app from the Dutch government), n (%)	10 (83%)

B: biological, cs: conventional synthetic, DMARD: disease modifying antirheumatic drug, ts: targeted synthetic.
**As measured by the Quickscan questionnaire[16]*

1. Need for information on the model

Before using a predictive computer model in clinical practice, patients wanted to receive information. Patients emphasized they wanted to know the possible consequences for their personal treatment, for example, how their medication could change or risks of side effects. Some patients indicated that they would want information about the number of patients involved in model development, the model performance and the kind of variables that are used in the model. Most patients were *not* interested in the technical specifics about model development, because they thought this was too difficult or because they thought it was not their responsibility to verify this. After implementation in clinical practice, several patients mentioned that they expected their caregiver to explain the model output to them.

2. Trust in model-supported treatment

Patients felt that a model-derived advice was more objective, science-based, neutral and consistent than human advice. Nevertheless, they were reluctant to trust the model-derived advised without the supervision of a doctor. Several factors were mentioned that would enhance trust in the ability of the model to make a correct prediction for patients' personal situation. For example, some patients indicated that they would want to verify the data entered into the model. Furthermore, it was believed that the involvement of a high number of patients in model development and a high number of included variables would improve model performance. Variables that increased trust were e.g. stress, the weather, diet and menopause, as in patients'

personal experience these variables triggered their RA symptoms and would therefore adequately predict their individual disease course.

Table 2. Quotes

Theme	Quote
Need for information	*"I think a lot of it comes down to: what are the possible consequences if steps are taken on the basis of, be it the computer model, but I would also like to know that from the doctor if there is a change[sic]. What relapse or which risks, I would like to know that in the same way from a computer model." [resp.7]*
Trust in model-supported treatment	*"Yes, I think that it would be good if you can kind of check the most important data that is entered [into the model]… I do think it would increase the feeling of trust."[resp.11]* *"The more patients that are involved, the more I believe that it [the prediction] is correct."[resp.6]* *"…if I look at myself and my past, then I am very sure that a computer model really could not have predicted if I would get a flare. Because you know, personally, I react very strongly to stress. And yeah of course a computer model does not know when I'm stressed."[resp. 9]*
Supportive role of the model in medical decision-making	*"But I want his [caregiver] human advice. But not that he walks away and says, well, a note will come out of that thing and that's what you should do. No, then I want to be able to discuss like, do you agree with what mister computer is telling us? "Two know more than one. And those two are the computer and the doctor."[resp.5]* *"…And the final decision is made by the patient and rheumatologist together."[resp.4]*
Personal benefit from model-supported treatment	*"I think that maybe a computer model could be nice because it could independently suggest any medication that exists, so to speak. And maybe a rheumatologist always uses methotrexate because that just has their preference or they are familiar with it. Maybe it [the model] gives a more, like, independent advice for the correct medication."[resp.1]* *"I think that maybe it could be linked to your healthcare record, if you can access that anyway. But an app could also be possible. If you say like a month in advance, before a conversation or a consultation with your specialist, you could have a look like hey, what are the possibilities for the next conversation? That you can prepare for that. So also for a specialist, so then the time of the consultation is maybe also more focused."[resp.3]*
Willing to contribute to model input	*"I'm certainly willing to participate to develop and improve that [the model]. There is one 'but': because if we're talking about radiographs then I think, well, you should only get a certain amount of radiation in your life,. So then I think, well, if they need a couple of extra X-rays once I wouldn't mind, but if it has a certain frequency, then I would say: hey, hold on there– Look, I totally don't mind blood withdrawals. That is something completely different. That doesn't harm me."[resp.6]*
Relationship with the caregiver	*"R:No, I don't see that it [the relationship] would be [different] because of a computer model. I mean, I often feel that there's a kind of, yes, call it a connection or something with your rheumatologist. That you say, well, I find that a pleasant person to talk to. I think that's important. But I can't imagine that would be influenced by that computer model……* *I: If the rheumatologist and the model do not agree with each other, would you be more inclined to follow the advice of the rheumatologist than the model?* *R: I find that quite difficult. It depends on the situation. If it would be a doctor that has known me for 15 years or more, then I would be more inclined to trust that. But in the situation where I am now, with a new doctor, then I would be inclined to say: try that research, that result [prediction], because we don't know each other that well yet."[resp.1]*

In case of an incorrect prediction, most patients envisioned they would still have sufficient trust to continue to use the model. They felt that the caregiver can also make a wrong judgement. In order to retain trust, patients expressed a need for elaborate consultation with the caregiver and emphasized that the model should be able to learn from mistakes. Several patients mentioned that their response to an incorrect prediction would depend on the severity of the consequences for their health. Patients expected they would lose their trust after a second incorrect prediction.

Patients also described their doubts. Some were concerned that the model could not adequately predict their individual disease course. They thought factors such as emotions and subtleties from the physical exam could not be captured by prediction models. Some believed the disease RA to be too complex to capture in a model due to their own experience with the unpredictability of the disease. A few patients mentioned that model-supported treatment might not be suitable for all patients, such as when there is a complex disease-course or for patients that are socially vulnerable. Patients felt that the caregiver should therefore assess whether model-supported treatment is suitable for each individual patient. Lastly, some patients were concerned about the involvement of pharmaceutical- or insurance companies in model development. They were concerned that involvement of companies would influence patients' access to certain medications. For example, an insurer might limit their access to expensive medication, while a pharmaceutical company might promote their own medications.

3. Supportive role of the model in medical decision-making

Patients stated that the model should have a supportive role in medical decision-making. They believed that the caregiver should always be able to give a treatment advice independent of the model. The envisioned role of the model was to give an extra justification for treatment decisions. The final decision about treatment should be made by the patient and caregiver together. Patients stressed that a caregiver should always help them if they experience symptoms, irrespective of the model-derived advice.

In case of a discrepancy between the advice derived from the prediction model and the caregiver, patients judged that the caregiver may deviate from the model-derived advice. However, he/she should be able to give an adequate explanation for this deviation. Patients mentioned they would take the initiative to suggest using a predictive computer model if they knew it was available, but not yet by their own caregiver.

4. Personal benefit from model-supported treatment

Patients expressed a wide range of hopes about how they could personally benefit from model-supported treatment. For example, patients suggested that the model could help

to reduce disease progression, side effects of medication and the time to find the right medication. Furthermore, patients thought the model could help to anticipate the disease course, and thus plan ahead regarding e.g. medication changes. Patients indicated that remote access to model outcomes, could help them to prepare for outpatient visits. There were also hopes that the model could provide new ideas. For example by suggesting medications that the caregiver had not considered themselves, or by identifying relationships between symptoms and the weather or certain foods. It was mentioned that the model could provide consistency in care, e.g. when there are changes in caregivers or when multiple specialists are involved. In the latter case, there was a hope the model could provide the "overall picture", by integrating multidisciplinary data. Furthermore, patients expected that they would feel more confident to start tapering medication if this was in line with the model-derived advice. When patients were forced to choose between a more transparent model where all the variables and their interactions are known, versus a better performing model with unknown variables, patients consistently chose the better performing model. Although patients had preferences about the included variables, ultimately patients wanted the best possible prediction for their health.

5. Willing to contribute to model input

Patients were willing to invest a reasonable amount of time and effort to contribute to model-supported treatment. For example, they were willing to undergo additional blood tests if this would improve the predictions made by the model. Patients were also willing to undergo additional radiography, although there were fears about accumulating exposure to radiation. In addition, they were willing to monitor their lifestyle, such as diet or exercise. However, if this registration would require a great amount of time or effort, they would not do it. Most patients were willing to share all of their data from their electronic health record (preferably anonymous). Such a contribution to the model was often motivated by hopes to personally benefit from model-supported treatment. Some were also motivated to contribute to science, because they had also benefited from scientific breakthroughs themselves.

6. Relationship with the caregiver

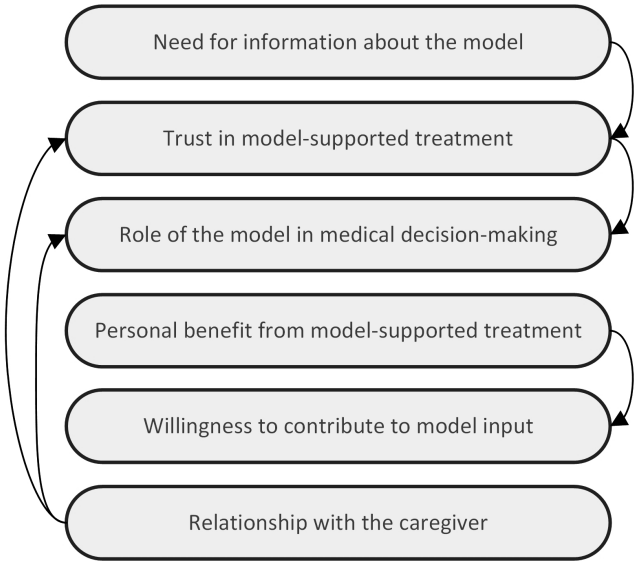
Patients expressed that their acceptance of the model was dependent on their relationship with their caregiver. Some patients who had a good connection with their caregiver were certain that their relationship would not change. Some patients feared that with the use of a model they would lose personal contact with their caregiver, although patients who described themselves as being assertive were confident they would stand up for themselves to prevent this from happening. Patients who highly valued personal contact tended to be negative about the possibility of prediction models to promote remote care. Furthermore, several patients voiced their concern

that (especially young) caregivers could become too dependent on the model. However, patients with a strong trust in their current caregiver did not share this concern. Lastly, a few patients mentioned that they would rely more heavily on the model in case of a bad relationship with the physician.

Relationships between the themes

We identified relationships between the themes (Figure 1). First, trust in model-supported treatment was influenced by the received information on said models. For example, interviewers observed that patients tended to be more positive about predictive models after receiving more information. Some patients also described, e.g.: *"To be honest, when [caregiver] spoke about this [prediction models] I thought: but how, where is the personal aspect? Then you delve a bit into the information I received from you, and it quickly became clear to me: it is just a kind of tool to look more broadly." [resp. 8]* Second, patients envisioned a supportive role for the model due to a lack of trust in the model as a stand-alone system. Third, the willingness to contribute to model input was often motivated by hopes to personally benefit from model-supported treatment. Fourth, the current relationship with the caregiver influenced the trust in model-supported treatment. For example, if patients had a good relationship, they were not concerned that their caregiver would rely too heavily on the model. Lastly, the relationship with the caregiver also influenced the envisioned role of the model, e.g. in case of a bad relationship patients envisioned a more prominent role of the model.

Figure 1: Themes and their relationships



Discussion

In this study we explored factors influencing the acceptance of prediction models in clinical decision-making by RA patients. We identified six themes: need for information, trust in model-supported treatment, supportive role of the model in medical decision-making, personally benefitting from model-supported treatment, the willingness to contribute to model input, and the effects on the relationship with the caregiver. To the best of our knowledge, this is the first study to focus on the patient perspective about the use of model-based clinical decision aids in RA.

Previous studies have described the perceptions about the use of artificial intelligence (AI) from the perspective of the public or hospital patients in general.[13,14,20-30] Of note, AI is a broad term, covering not only clinical decision aids, but also automated reporting, image pattern recognition, wearables and chatbots.[13] Similar to our study, many publications report that patients are generally positive about the use of prediction models in healthcare[13,20-25,31], and that patients are willing to keep track of their health-related information.[32,33] Patients also envisioned prediction models to have a supportive role in addition to the caregiver, rather than being a replacement.[13,20,23,26,28,31] In contrast, other studies report concerns about cybersecurity[13,21,31,34], legal concerns[13,14,35,36], and discrimination[13,21,23], which did not arise in our study. The latter may be influenced by our population being Caucasian, and not including ethnic minorities. In addition, our population has a chronic disease, and many of them expressed a strong trust in the hospital where they have been treated for years. This might have mitigated worries about cyber security, as well as discrimination. Lastly, patients with a chronic disease may also be more willing to try new technologies, as their potential personal benefit is more evident.[37]

Within the field of rheumatology, patients' perspective on the use of predictive software for identifying osteoporosis based on radiography have been described.[34] In line with our results, patients were generally not interested in understanding the (technical) details of the predictive software. In contrast to our study, these patients were concerned about negative results (e.g. bad prognosis), which has also been described in a publication on predictive RA biomarkers.[37] Other rheumatology publications describe the patient perspective on the use of chatbots[38], telemedicine[39], digital referral decision-trees[40] and the emotional impact of knowing your predicted risk of developing RA[41]. Our study focuses explicitly on the use of model-based clinical decision aids based on prediction models, which is important given the rapidly expanding number of developed prediction models in rheumatology and their potential to influence care.

Strengths of the current study include that our exploratory pilot data allowed us to gain insights that facilitated in-depth interviews. Moreover, in-person interviews enabled the participation of several patients with low digital literacy. Although technological proficiency according to the Quickscan questionnaire appeared to be quite high, 5/12 patients were not able to use Microsoft Teams. We therefore believe that this study also represents patients with limited digital literacy. Our study also has several limitations. Because patients were approached for participation by their caregiver, reasons not to participate were not registered due to feasibility. This study primarily included middle-aged Caucasians, which is in line with the demographics of RA. However, results may be less transferable to patients with other characteristics, especially ethnic minorities.[42]

The results of this study have practical implications for the development and implementation of model based clinical decision aids. First of all, patients tended to be more positive about prediction models after being informed, stressing the importance of information provision to patients. This information should focus on the possible personal treatment consequences, rather than the technical details. Furthermore, patients expressed their trust in models that included variables about diet, stress, exercise levels and the weather. As these variables are not commonly included in prediction models, it could be of interest to explore their predictive capacity in order to improve patient acceptability. Moreover, patients were concerned that the involvement of insurers or pharmaceutical companies could influence their access to certain medications. The role of these third parties should thus be carefully considered and, where possible, guaranties should be given for access to treatments. Regarding model selection and maintenance, patients expressed a strong preference for models that are able to learn from mistakes and are thus regularly updated. When designing the user-interface for prediction models, patients' wish to verify model input and to prepare for clinical visits based on the model output should be taken into consideration. As there is an overlap between RA and other rheumatologic or chronic conditions, the results of this study could possibly transferable to other patients groups.

In conclusion, RA patients are generally positive about the potential of prediction models to aid clinical decision-making. Key findings include the need for information, trust in model supported treatment, a supportive role for the model in decision making, personal benefits for patients, the willingness of patients to contribute to model input, and the effect on the relationship with the caregiver. The results of this study can be used to facilitate the development and implementation of effective RA prediction models by enhancing patient acceptability.

References

1. Madrid-García A, Merino-Barbancho B, Rodríguez-González A, Fernández-Gutiérrez B, Rodríguez-Rodríguez L, Menasalvas-Ruiz E. Understanding the role and adoption of artificial intelligence techniques in rheumatology research: An in-depth review of the literature. Semin Arthritis Rheum 2023;61:152213.
2. Wang J, Tian Y, Zhou T, Tong D, Ma J, Li J. A survey of artificial intelligence in rheumatoid arthritis. Rheumatol Immunol Res 2023;4(2):69-77.
3. Nelson AE, Arbeeva L. Narrative Review of Machine Learning in Rheumatic and Musculoskeletal Diseases for Clinicians and Researchers: Biases, Goals, and Future Directions. J Rheumatol 2022;49(11):1191-200.
4. Van der Leeuw MS, Messelink MA, Tekstra J, Medina O, van Laar JM, Haitjema S, et al. Using real-world data to dynamically predict flares during tapering of biological dmards in rheumatoid arthritis: development, validation, and potential impact of prediction-aided decisions. Arthritis Research and Therapy 2022;24(1).
5. Messelink MA, van der Leeuw MS, den Broeder AA, Tekstra J, van der Goes MC, Heijstek MW, et al. Prediction Aided Tapering In rheumatoid arthritis patients treated with biologicals (PATIO): protocol for a randomized controlled trial. Trials 2022;23(1):494.
6. Lim AJW, Tyniana CT, Lim LJ, Tan JWL, Koh ET, Chong SS, et al. Robust SNP-based prediction of rheumatoid arthritis through machine-learning-optimized polygenic risk score. J Transl Med 2023;21(1):92.
7. Guan Y, Zhang H, Quang D, Wang Z, Parker SCJ, Pappas DA, et al. Machine Learning to Predict Anti-Tumor Necrosis Factor Drug Responses of Rheumatoid Arthritis Patients by Integrating Clinical and Genetic Markers. Arthritis Rheumatol 2019;71(12):1987-96.
8. Souto-Carneiro M, Tóth L, Behnisch R, Urbach K, Klika KD, Carvalho RA, Lorenz HM. Differences in the serum metabolome and lipidome identify potential biomarkers for seronegative rheumatoid arthritis versus psoriatic arthritis. Ann Rheum Dis 2020;79(4):499-506.
9. Orange DE, Agius P, dicarlo EF, Robine N, Geiger H, Szymonifka J, et al. Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data. Arthritis Rheumatol 2018;70(5):690-701.
10. Gosselt HR, Verhoeven MMA, Bulatović-Ćalasan M, Welsing PM, de Rotte M, Hazes JMW, et al. Complex Machine-Learning Algorithms and Multivariable Logistic Regression on Par in the Prediction of Insufficient Clinical Response to Methotrexate in Rheumatoid Arthritis. J Pers Med 2021;11(1):44.
11. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019;21(7):e13659.
12. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30-6.
13. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599-e611.
14. Lammons W, Silkens M, Hunter J, Shah S, Stavropoulou C. Centering Public Perceptions on Translating AI Into Clinical Practice: Patient and Public Involvement and Engagement Consultation Focus Group Study. J Med Internet Res 2023;25:e49303.
15. Gossec L, Kedra J, Servy H, Pandit A, Stones S, Berenbaum F, et al. EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann Rheum Dis 2020;79(1):69-76.
16. Quickscan digitale vaardigheden: Pharos;Accessed on 12-10-2023. Available from: https://www.pharos.nl/kennisbank/quickscan-digitale-vaardigheden-van-uw-patienten/.
17. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative research in psychology 2006;3(2):77-101.
18. Guba EG, Lincoln YS. Competing paradigms in qualitative research. Handbook of qualitative research 1994;2(163-194):105.
19. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-57.
20. Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. Digit Health 2022;8:20552076221116772.
21. Musbahi O, Syed L, Le Feuvre P, Cobb J, Jones G. Public patient views of artificial intelligence in healthcare: A nominal group technique study. Digit Health 2021;7:20552076211063682.
22. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient Perceptions on Data Sharing and Applying Artificial Intelligence to Health Care Data: Cross-sectional Survey. J Med Internet Res 2021;23(8):e26162.
23. Ploug T, Sundby A, Moeslund TB, Holm S. Population Preferences for Performance and Explainability of Artificial Intelligence in Health Care: Choice-Based Conjoint Survey. J Med Internet Res 2021;23(12):e26611.

24. Spänig S, Emberger-Klein A, Sowa JP, Canbay A, Menrad K, Heider D. The virtual doctor: An interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. Artif Intell Med 2019;100:101706.
25. Gao S, He L, Chen Y, Li D, Lai K. Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media. J Med Internet Res 2020;22(7):e16649.
26. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. JAMA Dermatol 2020;156(5):501-12.
27. Ongena YP, Haan M, Yakar D, Kwee TC. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. Eur Radiol 2020;30(2):1033-40.
28. Mccradden MD, Sarker T, Paprica PA. Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research. BMJ Open 2020;10(10):e039798.
29. Stai B, Heller N, mcsweeney S, Rickman J, Blake P, Vasdev R, et al. Public Perceptions of Artificial Intelligence and Robotics in Medicine. J Endourol 2020;34(10):1041-8.
30. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, macisaac RJ, He M. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018;8(1):4330.
31. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. Front Med (Lausanne) 2020;7:233.
32. Solomon D, Lu F, Xu C, Colls J, Suh DH, Murray M, et al. THU0155 Patient Adherence with a Smartphone App for Disease Monitoring in Rheumatoid Arthritis. BMJ Publishing Group Ltd; 2019.
33. Navarro-Millán I, Zinski A, Shurbaji S, Johnson B, Fraenkel L, Willig J, et al. Perspectives of Rheumatoid Arthritis Patients on Electronic Communication and Patient-Reported Outcome Data Collection: A Qualitative Study. Arthritis Care Res (Hoboken) 2019;71(1):80-7.
34. Manning F, Mahmoud A, Meertens R. Understanding patient views and acceptability of predictive software in osteoporosis identification. Radiography (Lond) 2023;29(6):1046-53.
35. Yang L, Ene IC, Arabi Belaghi R, Koff D, Stein N, Santaguida PL. Stakeholders' perspectives on the future of artificial intelligence in radiology: a scoping review. Eur Radiol 2022;32(3):1477-95.
36. Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A Qualitative Study to Understand Patient Perspective on the Use of Artificial Intelligence in Radiology. J Am Coll Radiol 2019;16(10):1416-9.
37. Kumar K, Peters S, Barton A. Rheumatoid arthritis patient perceptions on the value of predictive testing for treatments: a qualitative study. BMC Musculoskelet Disord 2016;17(1):460.
38. Ye C, Zweck E, Ma Z, Smith J, Katz S. Doctor versus AI: Patient and physician evaluation of large language model responses to rheumatology patient questions, a cross sectional study. Arthritis Rheumatol 2023.
39. Sloan M, Lever E, Harwood R, Gordon C, Wincup C, Blane M, et al. Telemedicine in rheumatology: a mixed methods study exploring acceptability, preferences and experiences among patients and clinicians. Rheumatology (Oxford) 2022;61(6):2262-74.
40. Van Delft E, Bos R, Pennings P, Hazes M, Lopes Barreto D, Weel-Koenders A. Barriers and facilitators for implementation of a digital referral algorithm for inflammatory arthritis - a qualitative assessment in patients and caregivers. BMC Prim Care 2022;23(1):248.
41. Stack RJ, Stoffer M, Englbrecht M, Mosor E, Falahee M, Simons G, et al. Perceptions of risk and predictive testing held by the first-degree relatives of patients with rheumatoid arthritis in England, Austria and Germany: a qualitative study. BMJ Open 2016;6(6):e010555.
42. Yip K, Navarro-Millán I. Racial, ethnic, and healthcare disparities in rheumatoid arthritis. Curr Opin Rheumatol 2021;33(2):117-21.

S1 Appendix Pharos Quickscan

Questionnaire to assess digital proficiency of patients[16]

1. Do you have a computer, telephone or tablet with internet?
2. Do you sometimes search for information (about health and disease) on the internet?
3. Do you use e-mail, so that I can send you a link?
4. Do you use apps?
5. Are you able to download an app by yourself?
6. Do you use digid (i.e. Online identification tool from the Dutch government), for example to access your health care record?

S2 Appendix Interview topic list

Topics	Questions (examples)
Opening question	(after explanation about predictive computer models) what are your first thoughts about using a predictive computer model in your medical care?
Information	Imagine you would receive a flyer about the predictive computer model. What would you like this flyer to say?
Advantages	Can you think of any (other) advantages of using a predictive computer model in your medical treatment?
Disadvantages	Can you think of any (other) disadvantages of using a predictive computer model in your medical treatment?
Privacy	In order to make a prediction, the computer model needs information from your health care file. For which information from your health care record would you give permission?
Incorrect predictions	Imagine you and your caregiver use a predictive computer model when choosing a new medication. You follow the advice of the model, but after some time it appears that the medication does not help you. The prediction of the model was incorrect. How would you respond?
Discrepancy caregiver and model	Imagine you and your caregiver use a predictive computer model when tapering medication. The model predicts that you can lower your medication dose, without increasing your complaints. However, your caregiver decides to deviate from this advice, and recommends you to maintain the same dose. How would you respond?
Differences caregiver and computer model	What are your intuitions about the differences between medical advices from a caregiver and from a computer model? Do you trust one more than the other? Why?
Additional exams	Assume you would be asked to perform additional blood tests or x-rays, in order to improve the performance of the model. How would you respond?
Remote care	Predictive computer models could increase the opportunities for remote care. How would you feel about that?
Relationship caregiver	Do you think the relationship with your caregiver would change if you used a predictive computer model in your clinical care? If so, how?
Applications model	In which situations would you be interested to use a predictive computer model in your medical care? Why?
Acceptance	Would you allow a predictive computer model in your clinical care? If so, under which conditions?
Societal impact	Do you think the use of prediction models in healthcare will have a societal impact? If so, how do you feel about that?
Additional information	Is there anything you would like to add to the topics we discussed?

PART II
Diagnostic decision support

Chapter 3

Identification and prediction of difficult-to-treat rheumatoid arthritis patients in structured and unstructured routine care data: results from a hackathon

Marianne A. Messelink*
Nadia M.T. Roodenrijs*
Bram van Es
Cornelia A.R. Hulsbergen-Veelken
Sebastiaan Jong
L. Malin Overmars
Leon C. Reteig
Sander C. Tan
Tjebbe Tauber
Jacob M. van Laar
Paco M.J. Welsing
Saskia Haitjema

**these authors contributed equally*

Arthritis Research & Therapy. 2021;23(1):184

Abstract

Background The new concept of difficult-to-treat rheumatoid arthritis (D2T RA) refers to RA patients who remain symptomatic after several lines of treatment, resulting in a high patient and economic burden. During a hackathon, we aimed to identify and predict D2T RA patients in structured and unstructured routine care data.

Methods Routine care data of 1873 RA patients were extracted from the Utrecht Patient Oriented Database. Data from a previous cross-sectional study, in which 152 RA patients were clinically classified as either D2T or non-D2T, served as a validation set. Machine learning techniques, text-mining and feature importance analyses were performed to identify and predict D2T RA patients based on structured and unstructured routine care data.

Results We identified 123 potentially new D2T RA patients by applying the D2T RA definition in structured and unstructured routine care data. Additionally, we developed a D2T RA identification model derived from a feature importance analysis of all available structured data (AUC-ROC 0.88 (95%CI 0.82-0.94)), and we demonstrated the potential of longitudinal hematological data to differentiate D2T from non-D2T RA patients using supervised dimension reduction. Lastly, using data up to the time of starting the first biological treatment, we predicted future development of D2TRA (accuracy 79% and AUC-ROC 0.73 (95%CI 0.71-0.75)).

Conclusions During this hackathon, we have demonstrated the potential of different techniques for the identification and prediction of D2T RA patients in structured as well as unstructured routine care data. The results are promising and should be optimized and validated in future research.

Introduction

The treatment for rheumatoid arthritis (RA) has substantially improved over the past decades, enabling many patients to reach and maintain a state of low disease activity or even remission.[1] However, even when following current management recommendations, there is still a subgroup of patients that remains symptomatic after treatment with several (biological and/or targeted synthetic) disease-modifying antirheumatic drugs ((b/ts)DMARDs).[1–3] These patients are referred to as having "difficult-to-treat (D2T)" RA. Depending on the definition used, this disease state is estimated to affect 5% to 20% of all RA patients.[2–4] D2T RA is likely the subgroup of RA patients with the highest medical need.[5–7] Identifying and optimizing treatment could thus have great clinical impact for individual patients as well as for the sustainability of the healthcare system as a whole.

The importance of focusing on this subgroup of RA patients was previously acknowledged by an international survey among rheumatologists.[5] This survey indicated that several topics that are considered important for the management of D2T RA are not addressed in current RA management recommendations, reflecting an unmet clinical need. Additionally, results showed a wide variety in the existing concepts of D2T RA. Consequently, a European League Against Rheumatism (EULAR, from 2021 European Alliance of Associations for Rheumatology) Task Force recently defined D2T RA (Supplemental table 1)[8] and specific management recommendations for this patient population are under development.[8–11]

In the process of developing these recommendations, it became clear that evidence regarding this patient population is scarce and that further research is urgently needed.[9–11] This is however complicated by the difficulty of identifying D2T RA patients both retrospectively in cohorts and prospectively in clinical practice, due to the multidimensionality of the D2T RA definition and the presumed fluctuation of the disease state over time. Additionally, D2T RA comprises a heterogeneous group of patients with potential differences in contributing factors and underlying pathology.[6,8,12] Identifying D2T RA patients in routine care data enhances research opportunities, as it allows to retrospectively study the development of RA into D2T RA and the progression of the D2T RA state over time. Clear identification of these patients in retrospective data could also enable the development of models that can predict the development of D2T RA early on in the disease course, ultimately aiding in preventing D2T RA by a timely adjustment of therapy.

We previously conducted a cross-sectional study at the department of Rheumatology & Clinical Immunology of the University Medical Center Utrecht (UMC Utrecht), the

Netherlands, in which RA patients meeting the D2T RA definition[8] and a control group of RA patients not fulfilling all three criteria of the definition were enrolled.[6] This resulted in a valuable dataset with elaborate information on clinically classified D2T and non-D2T RA patients. This data served as a validation set during a hackathon (November 2020), in which data scientists and clinicians collaborated to identify and predict the development of D2T RA in structured and unstructured routine care data of all RA patients at the UMC Utrecht.

Methods

Routine care data

Structured and unstructured routine care data were extracted from the Utrecht Patient Oriented Database (UPOD) and pseudonymized. The organization and content of the UPOD have been described in more detail elsewhere.[13] In brief, the UPOD is an infrastructure of relational databases comprising electronic health record data of all patients treated at the UMC Utrecht and was established in 2004. UPOD data acquisition and management are in accordance with current regulations concerning privacy and ethics. For this hackathon, first, we identified the RA population according to the 10th revision of International Classification of Diseases (ICD-10) codes. We included patients with classification M05.X (seropositive rheumatoid arthritis) and M06.X (other rheumatoid arthritis) and subsequently excluded patients with M06.1 (adult-onset Still disease). Subsequently, the following structured data were extracted from the UPOD:

- Age (at time of RA diagnosis) and sex.
- Medication prescriptions: We included relevant medication based on Anatomical Therapeutic Chemical (ATC) codes (Supplemental table 2). All inpatient and outpatient prescriptions, including ATC codes and start dates, were extracted. As medication stop dates are prone to administrative errors, we only used start dates in our analyses. The b/tsDMARDs were labelled according to their mechanism of action (MoA). Medication prescriptions dated back to 2007.
- Laboratory analyses: We extracted laboratory measurements deemed clinically relevant (Supplemental table 3). In addition, we included all hematological parameters, as these are available in the UPOD for all patients for whom one or more components of the complete blood count (CBC) have been requested (e.g. hemoglobin). These parameters include the entire CBC, as well as research only values and raw scatter pattern measurements from the Abbott Celldyn Sapphire machines (Abbott hematology, Santa Clara, CA, USA). This data was available from 2003.
- Clinical measurements: Clinical measurements including 28 joint counts for swelling and for tenderness (SJC28/TJC28), length, weight, blood pressure and general health related to RA according to the patient as scored on a Visual

Analogue Scale (VAS-GH) were extracted for all patients. This data was available since 2002.

- Hospital visits: Visits to the outpatient rheumatology clinic (since 1995) as well as hospitalizations on the rheumatology ward (since 1987) were extracted for all patients.

In addition, clinical correspondence was extracted as unstructured data from the UPOD. This included all clinical letters from the rheumatology department as available since 1988.

Clinically classified patients

In a previous cross-sectional study,[6] 52 D2T and 100 non-D2T RA patients were clinically classified according to the EULAR definition (Supplemental table 1) in 2019-2020.[8] See Supplemental table 4 for an overview of the clinical characteristics of these patients. Both the structured and unstructured UPOD data as well as the study data were extracted. Study data included patient and disease characteristics as well as factors potentially contributing to D2T RA (e.g. treatment non-adherence, fibromyalgia), which were collected during a single study visit including a physical examination, laboratory analyses and by a subsequent questionnaire set. The data from these clinically classified patients served as a validation set, used to define the ability of the identification and prediction models to correctly classify D2T RA patients.

Identification of D2T RA patients

Four different techniques were employed to identify D2T RA patients in routine care data. The first two were based on the application of the criteria of the D2T RA definition in structured and unstructured data, respectively. Both methods focused on the first two criteria of the D2T RA definition (failing ≥2 b/tsDMARDS with different MoA and signs of active/progressive disease, see Supplemental table 1 for details).[8] The third criterion (problematic management) was deemed too subjective to be extracted from the available data. The third method explored the ability of other variables available in the structured data to differentiate D2T from non-D2T RA patients using a feature importance analysis. The fourth method entailed an exploratory dimension reduction of longitudinal hematological data.

Classification in structured data

In this approach, the structured data of medication prescriptions, laboratory analyses, clinical measurements, diagnostic codes and hospital visits were analyzed for all RA patients in the UPOD. Patients were classified as D2T or non-D2T RA using these data (Supplemental table 1).[8] Patients with registered medication prescriptions of at least two b/tsDMARDs with different MoA were deemed eligible to meet the first criterion of the D2T RA definition.[8] To define “active disease” (second criterion) we aimed to

calculate the disease activity score assessing 28 joints (DAS28) from SJC28/TJC28 and VAS-GH combined with erythrocyte sedimentation rate (ESR) or C-reactive protein) CRP where available. However, as these were missing for many patient visits in the database, a model was developed that approximated the DAS28-ESR. This model was based on laboratory values, number of hospital visits, patient characteristics and swiftness of cycling through b/tsDMARDs with a different MoA (see Supplemental table 5 for a brief description of the model and an overview of included parameters). The model had a mean absolute error of 0.8 (for reference: the DAS28 itself has a measurement error of 0.6).[14] Patients who had a mean approximated DAS28-ESR ≥3.2 in the period from 3 to 12 months after starting a b/tsDMARD of a second MoA were deemed to have failed their treatment due to active disease, thus fulfilled the first and second criterion of the D2T RA definition.[8] Patients who started a third b/tsDMARD with a different MoA were also deemed to have failed the b/tsDMARD of a second MoA and thus also met the first and second criterion of the D2T RA definition. This way, the RA patients in the UPOD dataset could be classified as being either D2T or non-D2T based on the available structured data.

Classification in unstructured data

In this approach, text mining techniques were applied to analyze clinical letters of RA patients in the UPOD to classify patients as D2T or non-D2T RA (Supplemental table 1).[8] Medication prescriptions were extracted from the headings "medication" and "DMARD history". Patients who had a history of a prescription of at least 2 b/tsDMARDs with different MoA were deemed to meet the first criterion of the D2T RA definition. To meet the second criterion, relevant subheadings were screened for synonyms of active disease, such as "flare". Negations such as "no flare" were excluded. This way, the RA patients in the UPOD dataset could be classified as being either D2T or non-D2T based on the available unstructured data.

Feature importance analysis

To gain insight in the importance of structured data variables regarding their ability to differentiate D2T from non-D2T RA patients, we performed an exploratory feature importance analysis using logistic regression. We included all available structured data variables from the UPOD of the 152 clinically classified patients, including those used for the application of the EULAR definition. (8) We determined the importance of different variables with multivariable logistic regression with L1 regularization (based on 1000 bootstrapped cross-validations with a 140/12 split). L1 regularization limits the number of coefficients by eliminating uninformative coefficients. This was preceded by standard scaling and multiple imputation using Bayesian Ridge regression and univariate feature filtering using a false discovery rate with alpha 0.05. The repeated measured variables were time-aggregated using the mean, median, standard deviation,

mean difference and mean minus the median. The resulting variables were univariately filtered based on their ability to differentiate between D2T and non-D2T RA patients. An identification model was derived using XGBoost, of which we present the receiver operating characteristic (ROC) curve based on ten-fold cross validation. XGBoost is a machine learning model which uses gradient boosting. In gradient boosting, multiple decision tree models are combined together into an ensemble. Each sequential model is trained to correct for the errors of the previous model. An important advantage of XGBoost is that it can handle missing data without imputation, which makes it a suitable model for real-life EHR data. We also considered multivariate logistic regression and a dense neural network, but the XGBoost model had a better performance in terms of AUC.

Dimension reduction of longitudinal hematological data

To explore the possibility to differentiate D2T- from non D2T RA patients solely based on longitudinal hematological data, a non-linear dimensionality reduction was performed. In dimension reduction, all available hematological parameters are reduced to two parameters, which allows for this information to be plotted on a 2-dimensional x-y graph. Dimension reduction was performed using uniform manifold approximation and projection (UMAP).[15] UMAP is a non-linear alternative to principal component analysis, which explicitly aims to preserve the Euclidean distance between samples.
This method was applied to all hematological data of the 152 clinically classified patients for training purposes using supervised techniques. Subsequently, this method was applied to the hematological data of all RA patients from the UPOD, to assess its ability to differentiate D2T from non-D2T RA patients. A Y-score was calculated for each patient, indicating the likelihood of having D2T RA. This was based on the combined outcomes of the classifications in structured and unstructured data (as described above), and the clinical classification (if available).

The results of these analyses are visualized for each individual patient using the median of the reduced dimensions (d1 and d2) of the hematological data over time. This was done both for the clinically classified patients as well as all RA patients from the UPOD. The aim of this method is to investigate if distinct clusters can be distinguished to separate D2T from non-D2T RA patients based on longitudinal hematological data.

Prediction model

In an effort to predict D2T RA patients early in the disease course (i.e. before satisfying the D2T RA definition) we developed a prediction model based on machine learning techniques using XGBoost. All available structured UPOD data from before the start of the first b/tsDMARD of the clinically classified D2T and non-D2T RA patients were used. The longitudinal data were regularized to a one-month time interval using forward fill-

in. This implies that missing values are imputed based on the last known values. The XGBoost classifier was used as the predictive model because of its robustness regarding data preprocessing. We used 10-fold cross-validation and the area under the ROC (AUC) statistic to determine model performance.

Results

Data extraction from the UPOD

Based on the ICD-10 codes, 1873 RA patients were identified in the UPOD.

Identification

Classification in structured data

Of the 1873 RA patients in the UPOD, 122 patients met the first criterion of the D2T RA definition (7%) as determined in structured UPOD data. For 100 of these patients, sufficient data was available to determine the fulfilment of the second criterion. Patients for whom insufficient data was available were classified as non-D2T. Twenty-five (25) of 52 patients clinically classified as D2T RA patients were correctly classified based on the structured data (sensitivity 48%, see Table 1). Two (2) of the 100 patients clinically classified as non-D2T RA were incorrectly classified (specificity 98%, Table 1). Using this approach, 43 additional (potential) D2T RA patients were identified.

Table 1. Classification of D2T and non-D2T patients in structured routine care data

Validation Classification in structured data	Clinically classified D2T RA	Clinically classified non-D2T RA	Newly classified patients in the UPOD	Total
D2T RA	25	2	43	70
Non-D2T RA	27	98	1678	1803
Total	52	100	1721	1873

*Patients were classified by applying the D2T RA definition[8] in structured routine care data from the UPOD. D2T: difficult-to-treat; DAS28-ESR: disease activity score based on 28-joint count and erythrocyte sedimentation rate; RA: rheumatoid arthritis; UPOD: Utrecht Patient Oriented Database; *: Clinical classification of D2T and non-D2T RA patients as performed in the cross-sectional study.[6]*

Classification in unstructured data

In the UPOD, 16,780 clinical letters of 1873 patients were available and extracted as unstructured data. Two-hundred thirty-nine (239) of all RA patients from the UPOD (13%) met the first D2T RA criterion, based on the unstructured data. This included all 52 clinically classified D2T RA patients from the cross-sectional study. One hundred sixty-one (161) patients also met the second criterion of the definition. Thirty-six (36) of 52 patients clinically classified as D2T RA patients were correctly classified using the

unstructured data (sensitivity 69%, see Table 2). Eight (8) of the 100 patients clinically classified as non-D2T RA were incorrectly classified (specificity 92%, Table 2). 117 additional (potential) D2T RA patients were identified. When comparing these patients with the 43 identified additional (potential) D2T RA patients using the structured data approach, 123 unique, additional (potential) D2T RA patients were found.

Table 2. Classification of D2T and non-D2T patients in unstructured routine care data.

Validation Classification in unstructured data	Clinically classified D2T RA	Clinically classified non-D2T RA	Newly classified patients in the UPOD	Total
D2T RA	36	8	117	161
Non-D2T RA	16	92	1604	1712
Total	52	100	1721	1873

*Patients were classified by applying the D2T RA definition[8] in unstructured routine care data from the UPOD. D2T: difficult-to-treat; RA: rheumatoid arthritis; UPOD: Utrecht Patient Oriented Database; *: Clinical classification of D2T and non-D2T RA patients as performed in the cross-sectional study.[6]*

Feature importance analysis

The most important structured data variables (features) to identify D2T and non-D2T RA patients and their logistic regression coefficients are shown in Tables 3A and 3B. Among others, this included the number of different medication prescriptions, the time period since RA diagnosis and the mean DAS28-ESR. Based on these features, an identification model was derived with an AUC-ROC of 0.88 (95%CI 0.82-0.94), Figure 1.

Table 3A. The most important features to identify D2T RA patients based on logistic regression coefficients

Feature	**Logistic Regression coefficient**
Number of different medication prescriptions, based on the extracted medication in Supplemental table 2	1.05
Mean DAS28-ESR score over time	0.76
Median DAS28-ESR score over time	0.70
Median non-invasively measured blood pressure over time	0.64
Standard deviation of the creatinine laboratory measurements over time	0.63
Time since RA diagnosis	0.52
Median of banded neutrophils over time	0.37
Ratio of segmented neutrophils by percentage of immature granulocytes over time	0.30
Standard deviation of percentage of reticulocytes over time	0.30
Median of the delta over time of banded neutrophils over time	0.29

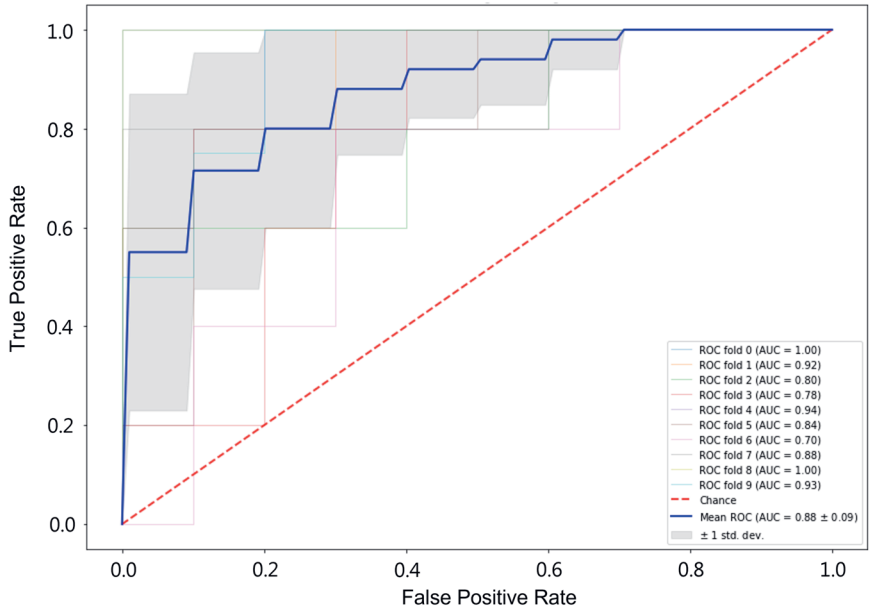
Features are noted in order of importance. A higher value of a feature corresponds to a higher likelihood of having D2T RA. DAS28: disease activity score based on 28-joint count; ESR: erythrocyte sedimentation rate; RA: rheumatoid arthritis.

Table 3B. The most important features to identify non-D2T RA patients based on logistic regression coefficients

Feature	Logistic Regression coefficient
Maximum ESR over time	0.84
Standard deviation of ESR values over time	0.78
Mean minus median of intermediate angle scatter of platelets over time	0.63
White blood cell count divided by lymphocyte concentration over time	0.62
Median length	0.58
Minimum potassium value over time	0.56
Female sex	0.56
Median neutrophils over time	0.46
Median percentage of reticulocytes over time	0.43
Standard deviation of DAS28-ESR score over time	0.43

Features are noted in order of importance. A higher value of a feature corresponds to a higher likelihood of having non-D2T RA. DAS28: disease activity score based on 28-joint count; ESR: erythrocyte sedimentation rate; IAS: intermediate angle scatter of platelet.

Figure 1. ROC-curve of the D2T RA identification model based on a feature importance analysis

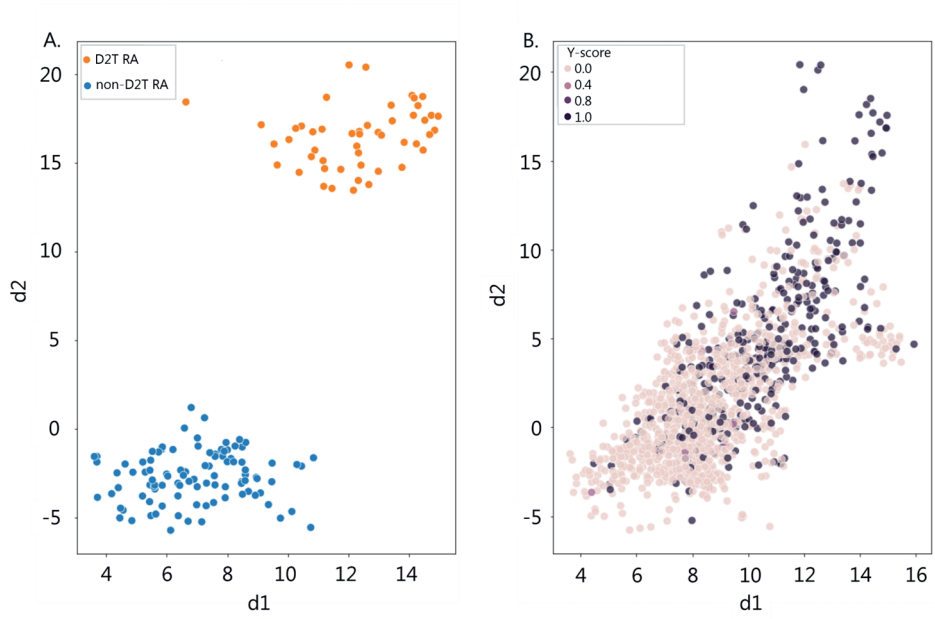


AUC-ROC for an identification model to identify D2T and non-D2T RA patients based on structured UPOD data. The model is based on the most important features derived with logistic regression techniques from the available structured data from the UPOD. D2T: difficult-to-treat; RA: rheumatoid arthritis; AUC: area under the curve, ROC: receiver-operator curve; UPOD: Utrecht Patient Oriented Database.

Dimension reduction of longitudinal hematological data

Figure 2A depicts the medians of the reduced dimensions of the longitudinal hematological data of the clinically classified D2T and non-D2T RA patients. Each point represents a single patient, and the axes represent the two reduced dimensions d1 and d2. Two distinct clusters are visible, which are strictly separated due to the supervised techniques. Figure 2B depicts the medians of the reduced dimensions of the hematological data of all 1873 RA patients in the UPOD. A tendency towards two separate clusters is visible based on the likelihood of having D2T RA, although these are not strictly separated.

Figure 2. Reduced dimensions of longitudinal hematological data



A: Medians of the reduced dimensions of the longitudinal hematological data of all 52 clinically classified D2T and 100 clinically classified non-D2T RA patients. B: Medians of the reduced dimensions of the longitudinal hematological data of all 1873 RA patients in the UPOD-database, where a higher Y-score indicates a higher estimated probability of having D2T RA according to the classifications in structured and unstructured data, and the clinical classification (if available). All available hematological parameters were reduced to two dimensions (d1 and d2). For each patient, the median of these reduced dimensions over time is visualized. d: reduced dimension; D2T: difficult-to-treat; RA: rheumatoid arthritis; UPOD: Utrecht Patient Oriented Database.

Prediction model

The machine learning prediction model was trained on the data of the clinically classified RA patients for whom data was available before prescribing the first b/tsDMARD (28 D2T and 88 non-D2T RA patients). The most important features mainly included hematological parameters, e.g. white blood cell count, percentage of

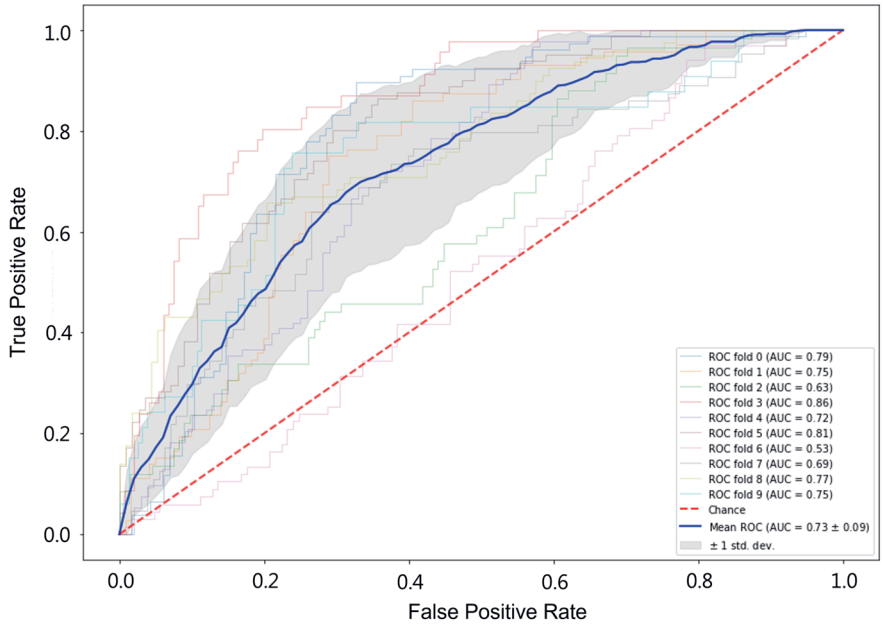
neutrophils, segmented neutrophils and hemoglobin(see Supplemental Table 6 for further details). With this XGBoost model, we were able to correctly predict 22 of the clinically classified D2T RA patients and 44 of the clinically classified non-D2T RA patients (sensitivity 79%, specificity 50%, Table 4). The average AUC-ROC over the 10-fold cross-validation was 0.73 (95%CI 0.71-0.75), Figure 3).

Table 4. The number of predicted D2T and non-D2T RA patients

Validation Prediction	Clinically classified D2T RA	Clinically classified non-D2T RA	Total
D2T RA	22	44	66
Non-D2T RA	6	44	50
Total	28	88	116

*Predictions are based on data from before the start of the first b/tsDMARD. B/tsDMARD: biological or targeted synthetic disease-modifying antirheumatic drug; D2T: difficult-to-treat; RA: rheumatoid arthritis; *: Clinical classification of D2T and non-D2T RA patients as performed in the cross-sectional study.[6] A decision threshold of 0.15 was applied.*

Figure 3. ROC-curve of the D2T RA machine learning prediction model



AUC-ROC of the D2T RA prediction model based on data from before the start of the first b/tsDMARD. AUC: area under the curve; b/tsDMARD: biological or targeted synthetic disease-modifying antirheumatic drug; csDMARD: conventional synthetic disease-modifying antirheumatic drug; D2T: difficult-to-treat; RA: rheumatoid arthritis; ROC: receiver-operator characteristic; std dev: standard deviation.

Discussion

The current study presents the results of a hackathon aimed at the identification and prediction of D2T RA patients in structured and unstructured routine care data. We were able to identify 123 potentially new D2T RA patients by applying the criteria of the D2T RA definition in structured and unstructured data. Additionally, we developed an identification model based on a feature importance analysis with high diagnostic performance (AUC-ROC 0.88), and we have shown the potential of longitudinal hematological parameters to differentiate D2T from non-D2T RA patients using supervised dimension reduction. To predict the risk of developing D2T RA, we developed a machine learning model based on structured data that correctly predicted 79% of clinically classified D2T RA patients using data available from before the time of prescribing the first b/tsDMARD (AUC-ROC 0.73). To our knowledge, there is no previous literature using these techniques in the context of (D2T) RA.

Routine care data is a valuable source of information, as it comprises a vast amount of "real world" patient data that is ample available. Unfortunately, this data often remains unutilized, due to technical challenges in their analysis. Yet routine care data could play a crucial role in the developing field of personalized medicine. A major strength of this study is that we have shown various data-analytical techniques to utilize this valuable source of information in the identification and prediction of D2T RA. Identifying D2T RA patients from routine care data enhances research possibilities, as it allows for retrospective analysis of the development of RA into D2T RA and the progression of the D2T RA state over time. Moreover, in clinical practice it creates an opportunity to optimize the treatment of D2T RA patients according to current and emerging guidelines. Correct identification of patients in longitudinal routine care data may also enhance the performance of models that can predict D2T RA early in the disease course. When patients at risk can be identified at an early stage, they may be monitored more intensively for the presence or development of factors contributing to D2T RA (e.g. treatment non-adherence or depression).[6] When these contributing factors develop and are adequately addressed, the risk of acquiring D2T RA could potentially be diminished.

Interestingly, our feature importance analysis, our machine learning prediction model and our exploratory dimension reduction all show an important role for hematological data in the identification and prediction of D2T RA patients. This is in line with previous research that has shown the potential role of the neutrophil-lymphocyte and platelet-lymphocyte ratios as biomarkers of disease activity in RA patients, although the underlying pathophysiology is not well-understood.[16–18] Of note, the large contribution of hematological parameters in our analyses is likely influenced by the

ample availability of these structured data, as this is a key feature of the UPOD. Nevertheless, as hematological parameters are low in costs, often readily available and require a minimal effort of the treating physician, they could be valuable potential markers in the evaluation of RA disease progression.

The performance of our identification strategies based on structured and unstructured data have been estimated conservatively. Patients for whom insufficient data were available to apply the D2T RA definition were now classified as "non-D2T", which may have contributed to the relatively low sensitivity that was observed. The D2T RA patients that were not identified by our models could especially include the D2T RA patients who were referred to the UMC Utrecht from other hospitals as a "second opinion", as data transfers between hospitals are often incomplete and electronic health record data from different hospitals, general practitioners and pharmacies are (unfortunately) not synchronized in the Netherlands. Improving the availability of these data could thus potentially improve the performance of our identification and, subsequently, prediction models.

Although the results of this study are promising regarding the accuracy of identification of D2T RA patients as well as predicting the development of D2T RA, this preliminary study also has several limitations. For example, not all components of the D2T RA definition (Supplemental table 1)[8] were incorporated in the structured and unstructured data approaches. This was done for several reasons. First of all, the subjective character of criterion 3 *"the management of the signs and/or symptoms is perceived as problematic by the rheumatologist and/or the patient"* was deemed too subjective to extract from the available data. Additionally, whether the management of patients is perceived as problematic will most often not be routinely noted in health records. This issue will therefore remain a challenge in further research on D2T RA. Second, for criterion 2c "*inability to taper glucocorticoid treatment below 7.5mg/day prednisone or equivalent"*, the stop dates of the medication that are available in the digital prescriptions system were deemed too unreliable. For example, additional medication prescriptions may be requested from the general practitioner instead of the rheumatologist (which are noted in separate systems), resulting in missing data in the prescription system and incorrect stop dates. Inclusion of these criteria in future identification and/or prediction models could further improve their performance.

Furthermore, an inherent limitation of working with routine care data is the dependency on the availability of certain data parameters. Several factors that have previously been reported in association with more severe RA disease activity, such as smoking status and radiographic progression, were not readily available in the UPOD[19]. Improvement of registration of these parameters and the optimization of

free text mining techniques could allow for future inclusion of these parameters in model development resulting in still better performing prediction models.

In future studies, the possibility of combining the different techniques presented in this paper for the identification of D2T RA patients in structured and unstructured routine care data should be addressed. In addition, other data sources could be utilized to explore other known contributing and risk factors for D2T RA, such a low socio-economic status based on e.g. postal codes[6,20]. Furthermore, the performance of the presented identification and prediction models should be evaluated in external data.

Conclusions

In conclusion, during this hackathon we have demonstrated potential techniques (including text mining, feature importance analysis and machine learning) for the identification and prediction of D2T RA patients in structured and unstructured routine care data. The results are promising to fuel research in this emerging field and should be optimized in further research.

References

1. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020;79:685–99.
2. de Hair MJH, Jacobs JWG, Schoneveld JLM, Van Laar JM. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. Rheumatology (Oxford). 2017;57:1135–44.
3. Buch MH. Defining refractory rheumatoid arthritis. Ann Rheum Dis. BMJ Publishing Group Ltd; 2018;77:966–9.
4. Kearsley-Fleet L, Davies R, De Cock D, Watson KD, Lunt M, Buch MH, et al. Biologic refractory disease in rheumatoid arthritis: Results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. Ann Rheum Dis. 2018;77:1405–12.
5. Roodenrijs NMT, de Hair MJH, van der Goes MC, Jacobs JWG, Welsing PMJ, van der Heijde D, et al. Characteristics of difficult-to-treat rheumatoid arthritis: results of an international survey. Ann Rheum Dis. BMJ Publishing Group Ltd; 2018;77:1705–9.
6. Roodenrijs NMT, van der Goes MC, Welsing PMJ, Tekstra J, Lafeber FPJG, Jacobs JWG, et al. Difficult-to-treat rheumatoid arthritis: contributing factors and burden of disease. Rheumatology. 2020;keaa860.
7. Roodenrijs NMT, Welsing PMJ, Goes MC van der, Tekstra J, Lafeber FPJG, Jacobs JWG, et al. Health care utilisation and economic burden of difficult-to-treat rheumatoid arthritis: a cost-of-illness study. Rheumatology. 2021;keab078.
8. Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, van der Goes MC, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis. BMJ; 2021;80:31–5.
9. Roodenrijs NMT, Kedves M, Hamar A, Nagy G, van Laar JM, van der Heijde D, et al. Diagnostic issues in difficult-to-treat rheumatoid arthritis: a systematic literature review informing the 2020 EULAR recommendations for the management of difficult-to-treat rheumatoid arthritis. RMD open. 2021;7:e001511.
10. Roodenrijs NMT, Hamar A, Kedves M, Nagy G, van Laar JM, van der Heijde D, et al. Pharmacological and non-pharmacological therapeutic strategies in difficult-to-treat rheumatoid arthritis: a systematic literature review informing the EULAR recommendations for the management of difficult-to-treat rheumatoid arthritis. RMD open. 2021;7:e001512.
11. Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, Goes MC van der. EULAR points to consider for the management of difficult-to-treat rheumatoid arthritis. :Submitted.
12. Buch MH, Eyre S, McGonagle D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. Nat Rev Rheumatol. Nature Research; 2021;17:17–33.
13. Ten Berg MJ, Huisman A, Van Den Bemt PMLA, Schobben AFAM, Egberts ACG, Van Solinge WW. Linking laboratory and medication data: New opportunities for pharmacoepidemiological research. Clin Chem Lab Med. 2007;45:13–9.
14. Fransen J, Stucki G, van Riel PLCM. Rheumatoid arthritis measures: Disease Activity Score (DAS), Disease Activity Score-28 (DAS28), Rapid Assessment of Disease Activity in Rheumatology (RADAR), and Rheumatoid Arthritis Disease Activity Index (RADAI). Arthritis Rheum. Wiley; 2003;49:S214–24.
15. McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv. 2018;1802.03426.
16. Chen Q, Chen DY, Xu XZ, Liu YY, Yin TT, Li D. Platelet/lymphocyte, lymphocyte/monocyte, and neutrophil/lymphocyte ratios as biomarkers in patients with rheumatoid arthritis and rheumatoid arthritis-associated interstitial lung disease. Med Sci Monit. 2019;25:6474–81.
17. Boulos D, Proudman SM, Metcalf RG, McWilliams L, Hall C, Wicks IP. The neutrophil-lymphocyte ratio in early rheumatoid arthritis and its ability to predict subsequent failure of triple therapy. Semin Arthritis Rheum. United States; 2019;49:373–6.
18. Sargin G, Senturk T, Yavasoglu I, Kose R. Relationship between neutrophil-lymphocyte, platelet-lymphocyte ratio and disease activity in rheumatoid arthritis treated with rituximab. Int J Rheum Dis. 2018;21:2122–7.

19. Rojas-Serrano J, Pérez LL, García CG, Moctezuma F, Álvarez-Hernández E, Vázquez-Mellado J, et al. Current smoking status is associated to a non-ACR 50 response in early rheumatoid arthritis. A cohort study. Clin Rheumatol. Springer-Verlag; 2011;30:1589–93.
20. Verstappen SMM. The impact of socio-economic status in rheumatoid arthritis. Rheumatology. 2016;56:1051–2.
21. Fautrel B, Granger B, Combe B, Saraux A, Guillemin F, Le Loet X. Matrix to predict rapid radiographic progression of early rheumatoid arthritis patients from the community treated with methotrexate or leflunomide: results from the ESPOIR cohort. Arthritis Res Ther. 2012;14:R249.
22. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016. p. 1603.02754.

Supplemental table 1. EULAR definition of D2T RA[8]

1. Treatment according to EULAR recommendations and failure of ≥2 b/tsDMARDs (with different mechanisms of action)[1] after failing csDMARD therapy (unless contraindicated)[2]
2. Signs suggestive of active/progressive disease, defined as ≥1 of:
    a. At least moderate disease activity (according to validated composite measures including joint counts e.g. DAS28-ESR > 3.2 or CDAI >10)
    b. Signs (including acute phase reactants and imaging) and/or symptoms suggestive of active disease (joint related or other)
    c. Inability to taper glucocorticoid treatment (below 7.5 mg/day prednisone or equivalent)
    d. Rapid radiographic progression (with or without signs of active disease)[3]
    e. Well-controlled disease according to above standards, but still having RA symptoms that are causing a reduction in quality of life
3. The management of signs and/or symptoms is perceived as problematic by the rheumatologist and/or the patient

All three criteria need to be present in D2T RA. b: biological; CDAI: clinical disease activity index; cs: conventional synthetic; DAS28-ESR: disease activity score assessing 28 joints using erythrocyte sedimentation rate; DMARD: disease-modifying antirheumatic drug; EULAR: European League Against Rheumatism, from 2021 European Alliance of Associations for Rheumatology; mg: milligram; RA: rheumatoid arthritis; ts: targeted synthetic. 1) Unless restricted by access to treatment due to socioeconomic factors. 2) If csDMARD treatment is contraindicated, failure of ≥2 b/tsDMARDs with different mechanisms of action is sufficient. 3) Rapid radiographic progression: change in van der Heijde-modified Sharp score ≥ 5 points at 1 year[21]

Supplemental table 2. Selected medication and ATC codes for extraction from the Utrecht Patient Oriented Database (UPOD)

Category	ATC code	Description
Glucocorticoid	H02AB01	Betamethasone
Glucocorticoid	H02AB02	Dexamethasone
Glucocorticoid	H02AB04	Methylprednisolone
Glucocorticoid	H02AB06	Prednisolone
Glucocorticoid	H02AB07	Prednisone
Glucocorticoid	H02AB08	Triamcinolon
Glucocorticoid	H02AB09	Hydrocortison
csDMARD	A07EC01	Sulfasalazine
csDMARD	L01AA01	Cyclophosphamide
csDMARD	L01BA01 / L04AX03	Methotrexate
csDMARD	L04AX01	Azathioprine
csDMARD	L04AA13	Leflunomide
csDMARD	L04AD01	Ciclosporin
csDMARD	P01BA01	Chloroquine
csDMARD	P01BA02	Hydroxychloroquine
bDMARD – CD20 inhibitor	L01XC02	Rituximab
bDMARD - TNFi	L04AB01	Etanercept
bDMARD – TNFi	L04AB02	Infliximab
bDMARD – TNFi	L04AB04	Adalimumab
bDMARD – TNFi	L04AB05	Certolizumab pegol
bDMARD – TNFi	L04AB06	Golimumab
bDMARD – CTLA4 inhibitor	L04AA24	Abatacept
bDMARD – IL-1 inhibitor	L04AC03	Anakinra

bDMARD – IL-6 receptor antagonist	L04AC07	Tocilizumab
bDMARD – IL-6 inhibitor	L04AC14	Sarilumab
tsDMARD – JAK inhibitor	L04AA29	Tofacitinib
tsDMARD – JAK inhibitor	L04AA37	Baricitinib

ATC: Anatomical Therapeutic Chemical; b: biological; cs: conventional synthetic; DMARD: disease-modifying antirheumatic drug; IL: interleukin; JAK: janus kinase; TNFi: tumor necrosis factor inhibitor; ts: targeted synthetic.

Supplemental table 3. Selected laboratory measurements for extraction from the Utrecht Patient Oriented Database (UPOD)

25-OH-Vitamin D	Low-density lipoprotein cholesterol
Anti-citrullinated protein antibodies	Lactate dehydrogenase
Albumin	Lipase
Alanine transaminase	M-protein screening
Alkaline phosphatase	Magnesium
Amylase	P-antineutrophil cytoplasmic antibodies
Antinuclear antibodies	Parathyroid hormone
Activated partial thromboplastin time	Phosphate
Aspartate transaminase	Potassium
Bicarbonate	Prothrombin time
Total bilirubin	Prothrombin time and international normalized ratio
Brain natriuretic peptide	Rheumatoid factor
C-antineutrophil cytoplasmic antibodies	Sodium
CA 15.3	Soluble interleukin-2 receptor
Calcium	Total protein level
Cholesterol	Transferrin
Creatinine kinase	Transferrin saturation
Creatinine kinase-myocardial band mass	Triglycerides
Creatinine	Troponin
CRP	Thyroid-stimulating hormone
Erythrocyte sedimentation rate	Uric acid
Ferritin	Vitamin B1
Folic acid	Vitamin B12
Free T4	Vitamin B6
GGT	Sediment (urine)
Glucose	Creatinine (urine)
Hemoglobulin A1c	Uric acid (urine)
High-density lipoprotein cholesterol	Creatinine (collected urine)
Immunoglobulin G total	Uric acid (collected urine)
Iron	

Supplemental table 4. Patient characteristics of clinically classified D2T and non-D2T patients

	Clinically classified D2T RA (n=52)	Clinically classified non-D2T RA (n=100)
Age (mean, SD)	60.2 (11.4)	64.5 (10.9)
Female, %	73	72
Disease duration, years (median, IQR)	17.0 (9.0-25.0)	14.0 (8.0-24.0)
Age at onset, years (mean, SD)	41.9 (12.4)	47.8 (15.6)
RF positivity, %	75	65
ACPA positivity, %	73	65
Joint erosions, %	63	49
DAS28-ESR (median, IQR)	4.1 (3.5-6.1)	2.5 (1.8-3.3)
csDMARD(s), %	71	86
bDMARD, %	52	39
tsDMARD, %	23	0

Data is based on the cross-sectional study of Roodenrijs et al.[6] ACPA: anti-citrullinated protein antibody; b: biological; cs: conventional synthetic; DAS28: disease activity score based on 28 joint count; DMARD: disease modifying anti-rheumatic drug; D2T: difficult-to-treat; ESR: erythrocyte sedimentation rate; IQR: interquartile range; RF: rheumatoid factor; SD: standard deviation; ts: targeted synthetic.

Supplemental table 5. Most important features of machine learning model to predict the DAS28-ESR*

Feature	Shapley value (order of importance)	Direction of the effect
Minimum hemoglobin over time	0.151	-
Higher age	0.127	-
Maximum platelet count as measured by impedance over time	0.100	+
Maximum fraction of immature reticulocytes over time	0.078	-
Minimum platelet crit over time	0.063	-
Female gender	0.063	-
Maximum platelet crit over time	0.056	+
Mean size of neutrophils over time, measured by the mean axial light loss	0.055	+
Number of visits to the outpatient clinic in the past 6 months	0.054	+
Maximum red cell distribution width over time	0.053	+

*DAS28: disease activity score based on 28-joint count; ESR: erythrocyte sedimentation rate *:The DAS28-ESR prediction model was trained on all available DAS28-ESR scores in the Utrecht Patient Oriented Database. The model was developed using the machine learning model XGBoost,[22] which uses gradient boosting. In gradient boosting, multiple decision tree models are combined together into an ensemble. Each sequential model is trained to correct for the errors of the previous model. An important advantage of XGBoost is that it can handle missing data without imputation, which makes it a suitable model for real-life EHR data. Feature importance was determined with Shapley values, which determines the contribution of a feature to each tree in the model. Note that this method is non-linear and there are feature interactions. The third column indicates the direction of the effect on the outcome variable. The DAS28-ESR prediction model had a mean absolute error of 0.8.*

Supplemental table 6. Most important features of the machine learning model to predict the development of D2T RA before the start of the first b/tsDMARD

Feature	Shapley value (order of importance)	Direction of the effect
Percentage macrocytic erythrocytes, difference between maximum and minimum values over time	1.016	+
White blood cell count, difference between maximum and minimum values over time	0.664	+
Variance of polarized side scatter of platelets, difference between maximum and minimum values over time	0.620	-
Minimum non-invasively measured blood pressure over time	0.568	+
White cell viability, difference between maximum and minimum values over time	0.511	+
Minimum weight of patient	0.495	+
Mean polarized side scatter of neutrophils, difference between maximum and minimum values over time	0.411	+
Hemoglobin, difference between maximum and minimum values over time	0.410	+
Percentage of neutrophils, difference between maximum and minimum values over time	0.379	+
Segmented neutrophils, difference between maximum and minimum values over time	0.378	+

b/tsDMARD: biological or targeted synthetic disease-modifying antirheumatic drug; D2T: difficult-to-treat; RA: rheumatoid arthritis. The model was developed using the machine learning model XGBoost,[22] which uses gradient boosting. In gradient boosting, multiple decision tree models are combined together into an ensemble. Each sequential model is trained to correct for the errors of the previous model. An important advantage of XGBoost is that it can handle missing data without imputation, which makes it a suitable model for real-life EHR data. The order of importance was determined with Shapley values, which determines the contribution of a feature to each tree in the model. Note that this method is non-linear and there are feature interactions. The third column indicates the direction of the effect on the outcome variable.

Chapter 4

Development of a primary care screening algorithm for the early detection of patients at risk of primary antibody deficiency

Marianne A. Messelink
Roos M. Berbers
Joris M. van Montfrans
Pauline M. Ellerbroek
André Gladiator
Paco M.J. Welsing
Helen L. Leavis

Allergy, Asthma & Clinical Immunology, 2023;19(1):44

Abstract

Background Primary antibody deficiencies (PAD) are characterized by a heterogeneous clinical presentation and low prevalence, contributing to a median diagnostic delay of 3–10 years. This increases the risk of morbidity and mortality from undiagnosed PAD, which may be prevented with adequate therapy. To reduce the diagnostic delay of PAD, we developed a screening algorithm using primary care electronic health record (EHR) data to identify patients at risk of PAD. This screening algorithm can be used as an aid to notify general practitioners when further laboratory evaluation of immunoglobulins should be considered, thereby facilitating a timely diagnosis of PAD.

Methods Candidate components for the algorithm were based on a broad range of presenting signs and symptoms of PAD that are available in primary care EHRs. The decision on inclusion and weight of the components in the algorithm was based on the prevalence of these components among PAD patients and control groups, as well as clinical rationale.

Results We analyzed the primary care EHRs of 30 PAD patients, 26 primary care immunodeficiency patients and 58,223 control patients. The median diagnostic delay of PAD patients was 9.5 years. Several candidate components showed a clear difference in prevalence between PAD patients and controls, most notably the mean number of antibiotic prescriptions in the 4 years prior to diagnosis (5.14 vs. 0.48). The final algorithm included antibiotic prescriptions, diagnostic codes for respiratory tract- and other infections, gastro-intestinal complaints, auto-immune symptoms, malignancies and lymphoproliferative symptoms, as well as laboratory values and visits to the general practitioner.

Conclusions In this study, we developed a screening algorithm based on a broad range of presenting signs and symptoms of PAD, which is suitable to implement in primary care. It has the potential to considerably reduce diagnostic delay in PAD, and will be validated in a prospective study.

Background

Primary antibody deficiencies (PAD) are characterized by an inability to produce clinically effective immunoglobulin responses and represent the majority of all primary immunodeficiency (PID) disorders [1-3]. PAD encompasses a heterogeneous group of diseases such as Common Variable Immunodeficiency (CVID), X-linked agammaglobulinemia (XLA), IgG subclass deficiency and specific antibody deficiency (SpAD) [4]. The estimated prevalence of PAD varies widely from 1:700 to 1:25.000, in part due to the suspected large number of undiagnosed patients [5-8].

The onset of symptoms of PAD is most commonly in the second to fourth decade of life [9-12]. The clinical presentation is heterogeneous, including increased susceptibility for respiratory tract- and gastro-intestinal infections, auto-immune symptoms, lymphoproliferative disease and an increased risk of certain malignancies [4, 13-15]. Due to the wide variety in presenting symptoms and the rarity of PAD, diagnosis can be challenging. This is evident by the reported mean diagnostic delay of 6-12 years (median 3-10 years), which has not significantly improved over the past 5 decades [13, 16-18]. This diagnostic delay may increase the risk of morbidity and mortality, as effective therapies and prophylaxes are available [18-21]. A timely diagnosis may also lead to substantial health care cost savings, even when correcting for the cost of treatment [22]. Thus, reducing the diagnostic delay of PAD is of key interest [23].

To this end, several 'Early warning signs' have previously been developed. However, these do not include the full spectrum of presenting symptoms of PAD, show suboptimal diagnostic performance, and are often used in a non-automated manner. [24-29]. As manual screening systems depend on the awareness of an individual physician, these are suboptimal for rare diseases with heterogeneous presentations. Therefore, we aimed to develop a screening algorithm that encompasses an extended spectrum of PAD signs and symptoms based on electronic health record (EHR) data, that can easily be automated. Such an algorithm may be used as an aid to notify general practitioners (GPs) when further laboratory evaluation of immunoglobulins should be considered, thereby facilitating a timely diagnosis of PAD.

Primary care may be the optimal setting for such an algorithm for several reasons. Firstly, a patient will initially present with their symptoms in primary care, especially in countries where the general practitioner has a gatekeeper function to secondary care. Therefore, screening in primary care could allow to detect PAD patients in an earlier phase when compared to screening in secondary care. Secondly, primary care EHRs encompass a comprehensive overview of the complaints for which a patient has sought medical care. In contrast, in secondary care usually only the symptoms for which

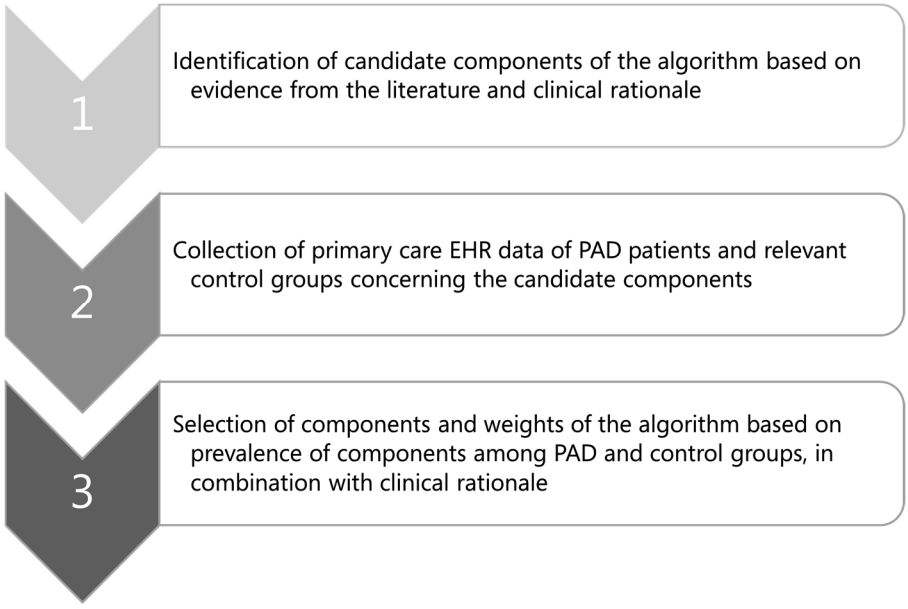
a patient has been referred are registered structurally (i.e. in diagnostic codes). For example, if a patient is only referred for respiratory tract infections, relevant gastro-intestinal or auto-immune symptoms could be missed. Thus, focusing on primary care allows to screen for a broad range of presenting signs and symptoms of PAD.

The aim of the current study is to develop a screening algorithm to identify patients with an increased risk of PAD in a primary care setting. This algorithm could be applied to notify GPs of patients at risk of PAD, for whom further laboratory investigation of immunoglobulins and/or consultation of an immunologist is indicated.

Methods

The algorithm for the early detection of PAD was developed based on primary care EHR data of PAD patients and control groups. We focused on data that are available as structured data in the EHRs of primary care facilities in the Netherlands. Almost all inhabitants of the Netherlands have an EHR at a public primary care physician, and the primary care physician has a gatekeeper function to secondary care, signifying that in non-urgent cases referral from the general practitioner is necessary to access hospital care.[30] Available structured data included age, diagnostic ICPC-codes (International Classification of Primary Care), ATC-codes (Anatomical Therapeutic Chemical) for medication prescriptions and (requests for) laboratory assessments.
The algorithm was developed in 3 steps, see Figure 1.

Figure 1. Steps in the development of the screening algorithm for primary antibody deficiencies



EHR: electronic health record, PAD: primary antibody deficiency.

1. Identification of Candidate Components of the Algorithm

The hallmark clinical features of PAD are recurrent respiratory tract infections (RTI), including otitis media, sinusitis and pneumonia [12, 21, 31]. Gastro-intestinal complaints, such as chronic diarrhea and parasitic infections, also occur frequently in certain PADs [12, 32-35]. In addition, it is estimated that around 30% of PAD patients develop auto-immune disorders, e.g. idiopathic thrombocytopenic purpura, auto-immune hemolytic anemia or rheumatoid arthritis [36-38]. Finally, certain PADs can also present with lymphoproliferative symptoms and are associated with an increased risk of non-Hodgkin lymphoma and gastric cancer [39-42]. All diagnostic ICPC codes related to these presenting symptoms of PAD were considered as candidate components of the algorithm.

Since PAD patients often present with recurrent infections that require antibiotic treatment, we also considered the type and number of antibiotic prescriptions for inclusion in the algorithm. We considered antibiotics registered in the Netherlands for the treatment of respiratory tract- or parasitic intestinal infections [43]. As antibiotics for upper respiratory tract infections (RTIs) are frequently prescribed in the early years of childhood, we only took prescriptions from the age of 6 years into account [44].

As PAD, by definition, can be characterized by a reduced level of one or more (subtypes of) immunoglobulins, we considered these laboratory results for inclusion in the algorithm. Since immunoglobulin diagnostics may not be requested commonly in primary care, we also considered calculated globulin. Calculated globulin is a measure based on total protein and albumin levels, which can indicate hypogammaglobulinemia [45, 46].

A previous study reported that a higher number of consultations at the GP, as well as a higher number of requests for lung function tests and blood tests for infection, were statistically significantly associated with an increased odds of CVID diagnosis [47]. Therefore, we also considered these factors for inclusion in the algorithm.

In contrast to primary antibody deficiencies, secondary antibody deficiencies are the consequence of an underlying disease, such as hematological malignancy. Diagnoses known to be causes of secondary antibody deficiencies (e.g. leukemia, multiple myeloma, HIV) were added as exclusion criteria [48-50]. In the case that a diagnosis could be both a cause of secondary antibody deficiency as well a complication of PAD (e.g. non-Hodgkin lymphoma), it was labelled as ‘‘ambiguous diagnosis’’. For these diagnoses, the EHRs were screened up to the moment of the ambiguous diagnosis. This approach allows to detect patients for whom the ambiguous diagnosis is the consequence of an underlying PAD, whilst still excluding patients with a secondary

antibody deficiency. Immunosuppressant medication was not added as an exclusion criterion, as these may also be prescribed for auto-immune symptoms that are caused by an underlying PAD. Table 1 shows an overview of the candidate components of the algorithm.

2. Collection of EHR Data of PAD Patients and Control Groups

To gain insight in the prevalence of the candidate algorithm components among PAD patients and control groups, we used two data sources. First, we collected the primary care EHR data of 30 patients from different age groups with a variety of PAD diagnoses from an academic hospital in the Netherlands (the University Medical Center Utrecht). Patients signed informed consent and research agreements were signed by their GPs. Second, we collected primary care EHR data from the pseudonymized database of the Julius General Practitioner Network (JHN), a research collaboration of GPs in the region of Utrecht, the Netherlands [51]. Due to privacy regulations from the JHN database we only had access to the metadata (e.g. means, medians and proportions) concerning the candidate components of the algorithm for several patient groups of interest, defined based on registered diagnostic ICPC codes. The groups of interest included PAD patients as well as several control groups. Because no specific ICPC code for PAD exists, we selected the first group based on the code for immunodeficiency (T99.01). This group approximates PAD patients, as causes of secondary immunodeficiencies were excluded based on the exclusion criteria, and because PAD represents the majority of the remaining PIDs. Therefore, this group of 'primary care immunodeficiency' patients was considered to be of interest, in addition to the group of confirmed PAD-patients from the academic hospital.

Table 1. Overview of candidate components from primary care electronic health records

Type of electronic health record data	Description
ICPC-codes	Diagnostic codes related to presenting signs and symptoms of PAD For exclusion criteria: diagnostic codes related to causes of secondary antibody deficiencies
ATC-codes	Antibiotic prescriptions for the treatment of respiratory tract- or parasitic infections
Laboratory results	IgA, IgM, IgG (total and subclasses), calculated globulin
Number of visits	Number of physical or telephone appointments with the general practitioner
Number of requests for additional tests	Number of requests for CRP, leukocytes and lung function tests

Overview of candidate components for the algorithm as extracted from structured primary care EHR data. ATC Anatomical Therapeutic Chemical, CRP C-reactive protein, EHR electronic health record, ICPC International Classification of Primary Care, Ig immunoglobulin, PAD primary antibody deficiency.

The selected control groups were 'patients from the general GP population', 'patients with upper RTIs', 'patients with chronic obstructive pulmonary disease (COPD) or asthma', 'patients with inflammatory bowel disease (IBD)' and 'patients with a malignancy'. These groups were selected as they were expected to have overlapping presenting symptoms with PAD patients, but should be distinguished from potential PAD patients by the algorithm in order to prevent false-positives. For all control groups we applied the exclusion criteria, and excluded patients that had an ICPC code for immunodeficiency (T99.01). The meta-data of all available patients in the JHN-database that met the criteria for the control groups were utilized.

From both the academic hospital and the primary care-database we selected patients aged 12-70 years, as PAD most commonly presents in the second to fourth decade of life [9-12]. In addition, as recurrent infections and antibiotic prescriptions are common in younger children, other cut-off values are to be expected for this age group. Lastly, our algorithm was developed in line with the study design of a subsequent external validation study, where high-risk patients will be invited for laboratory analysis of immunoglobulins. Considering the medical ethical guidelines, this validation study is only feasible in patients aged ≥12 years.
A distinct 'censoring date' was selected for different patient groups. All available EHR data before the censoring date was extracted. The censoring date indicates the date up to which time the EHR data was extracted. For most patients, this was the date of data-extraction, November 18, 2021. For the PAD and primary care immunodeficiency patients, we extracted the EHR data both pre- and post-diagnosis. The censoring date could therefore be the date of PAD diagnosis (pre-diagnosis), or the date of data-extraction (post-diagnosis). When deciding which components were to be maintained in the algorithm, pre-diagnosis data was used, since our aim is to identify patients before diagnosis. For patients with an ambiguous diagnosis, the date of this specific diagnosis was defined as their censoring date.

To determine the validity of the diagnostic ICPC codes, we compared the presenting symptoms registered as diagnostic ICPC codes in the primary care EHR of the PAD patients with the presenting symptoms registered as free text in the secondary care EHR of the academic hospital and determined their concordance.

3. Selection of Components and Weights of the Algorithm
The decision to include a candidate component in the algorithm and the corresponding weight was based on the results of analyses of the corresponding EHR data in combination with clinical rationale. To estimate the ability of the presence of individual diagnostic ICPC codes to differentiate PAD patients from the control group, we calculated the Youden's index. Youden's index is a summary measure of diagnostic

quality based on sensitivity and specificity, where a higher index indicates that the component has a better discriminatory ability [52]. The sensitivity of the presence of a component was calculated for the PAD patients and primary care immunodeficiency patients, and the mean of the sensitivities the was used to determine Youden's index. Specificity was calculated from the general GP population control group. The weight of the diagnostic ICPC codes was based on the discriminative value as expressed by Youden's index in combination with clinical rationale.

The diagnostic ICPC codes included in the algorithm were grouped according to the following categories: 'respiratory tract infections', 'gastro-intestinal complaints', 'other infections', 'auto-immune symptoms' and 'malignancies, lymphoproliferative- and other symptoms'. For each of these groups, we determined whether the entire EHR before the censoring date should be screened for these diagnostic codes, or only the 10 years before the censoring date. This decision was based on clinical rationale and on the discriminative value as expressed by Youden's index.

For the ATC-codes, we calculated the mean number of antibiotic prescriptions per year in the 10 years before the censoring date. The decision on the inclusion and weight of individual antibiotics was based on both clinical rationale and the difference in the mean number of prescriptions between the control groups and immunodeficiency patients.

For immunoglobulin levels and calculated globulin, we determined the mean laboratory values in the EHR in the 10 years before the censoring date. Lastly, regarding the number of consultations at the general practitioner and the number of requests for lung function tests/blood tests for infection, we calculated the mean and median number of visits and requests in the year before the censoring date for both the PAD/primary care immunodeficiency groups and the control groups. To determine the optimal cut-off point for the number of lung function tests, blood tests for infection and number of visits to the GP, we calculated the Youden's index for several cut-off points spread around the median values in PAD patients.
Ethical approval for this study was received from the Medical Research Ethics Committee Utrecht, protocol number 19-748. Statistical analysis was done with the base package in R version 4.0.2.

Results

Patient Characteristics
Of the 30 included PAD patients from the academic hospital, 12 were diagnosed with CVID, 8 with IgG subclass deficiency, 4 with unclassified antibody deficiency, 3 with IgA-

and IgG subclass deficiency, 2 with selective IgA deficiency and 1 patient was diagnosed with SpAD. The mean age at the time of data-extraction was 29.8 years, with a large variation (standard deviation of 14). The PAD-patients thus represented a wide range of diagnoses and ages. The mean delay from symptom onset to diagnosis of these patients was 12.4 years (SD 12.2, median 9.5 (3.0-19.5)). The patient characteristics of the PAD patients from the academic hospital, of the primary care immunodeficiency patients and the control groups (all aged 12-70 years) are shown in Table 2.

Table 2. Patient characteristics

Group	No. of patients	% Female	Mean age (SD)
General population	58,223	52.9%	39.8 (13.9)
Upper RTI	13,133	57.4%	38.4 (16.3)
COPD/Asthma	4,427	53.2%	42.3 (15.7)
IBD	403	54.6%	43.8 (12.9)
Malignancy	1,526	60.5%	55.2 (12.0)
Primary care immunodeficiency patients	26	53.8%	at diagnosis: 33.3 (15.4)
			at data-extraction date: 40 (14.8)
PAD patients	30	40%	onset symptoms: 6.4 (7.4)
			at diagnosis: 19.4 (16.7)
			at data-extraction date: 29.8 (14)

Patient characteristics of the different control groups, the immunodeficiency patients registered in primary care and the PAD patients registered in an academic hospital. If not stated otherwise, the values were captured on the date of data-extraction (November 2021). COPD: chronic obstructive pulmonary disease, IBD: inflammatory bowel disease, JHN: Julius General Practitioner Network, a general practitioner network in Utrecht the Netherlands, No. number, RTI: respiratory tract infection, SD: standard deviation.

Diagnostic ICPC Codes

Table 3 shows the concordance between the presenting symptoms registered as diagnostic ICPC codes in the primary care EHRs and the symptoms registered in free text in the secondary care EHRs. Concordance was assessed by calculating the percentage of patients for whom the presence or non-presence of a symptom was equal between the diagnostic codes in the primary care EHR and the free text in the secondary care EHR. For almost all presenting symptoms, apart from gastro-intestinal complaints, concordance between the diagnostic primary care ICPC codes and the free text in secondary care EHRs was high. The selection of diagnostic codes and their attributed weight was based on the Youden's index (Supplemental Table S1) in combination with evidence from literature and clinical expertise of PAD specialists in the University Medical Center Utrecht. The diagnostic codes included in the final algorithm are shown in Table 5.

Table 3. Concordance between PAD symptoms registered as diagnostic ICPC codes and free text

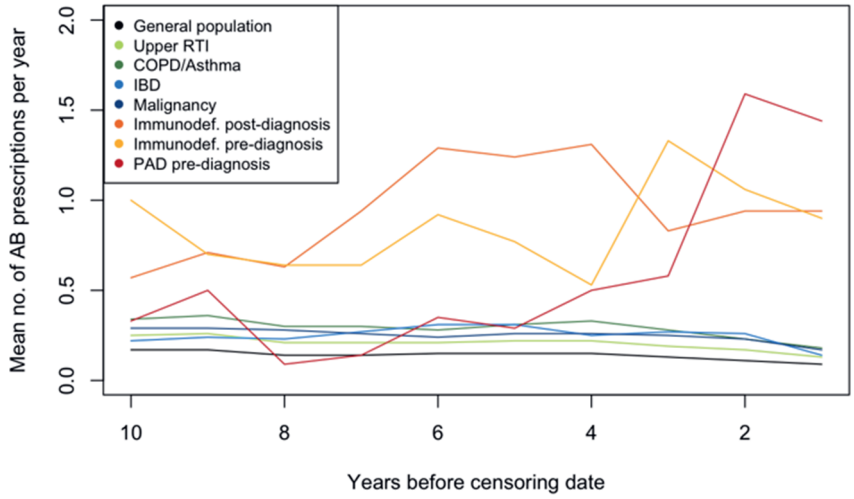
Presenting symptom	Registered in primary care EHR (diagnostic codes)	Registered in secondary care EHR (free text)	Concordance
Upper RTI	30 (100%)	28 (93%)	93%
Gastro-intestinal symptoms	17 (57%)	12 (40%)	57%
Pneumonia	11 (37%)	11 (37%)	87%
Auto-immune symptoms	11 (37%)	9 (30%)	87%
Bronchiectasis	6 (20%)	6 (20%)	87%
Arthritis/arthralgia	0 (0%)	4 (13%)	87%
Meningitis	4 (13%)	4 (13%)	100%

Concordance between symptoms registered as diagnostic codes in the primary care EHR and symptoms registered as free text in the secondary EHR. Data of 30 patients of an academic hospital before their PAD diagnosis. Data in columns 2 and 3 are presented as number of patients (%). EHR: Electronic Health Record, ICPC: International Classification of Primary Care, PAD: primary antibody deficiency, RTI: respiratory tract infection.

Antibiotic Prescriptions

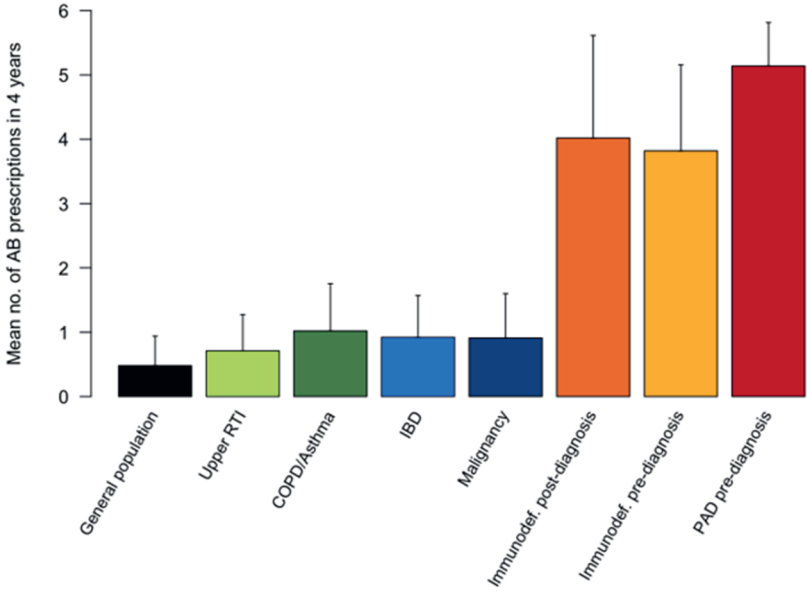
Figure 2A shows the mean number of antibiotic prescriptions per year in the 10 years before the censoring date for the different patient groups. PAD-patients and primary care immunodeficiency patients had a higher mean number of antibiotic prescriptions than all control groups. In the pre-diagnosis groups this effect appeared to be most pronounced from 4 years before the diagnosis. Figure 2B shows the mean total number of antibiotic prescriptions in the 4 years before the censoring date, where the immunodeficiency and PAD groups had a higher mean compared to the control groups (5.14 in PAD vs. 0.48 in the general population). It was therefore decided to include antibiotic prescriptions from the past 4 years in the algorithm, see Table 5. The mean number of antibiotic prescriptions per individual antibiotic is shown in Supplemental Table S2. The weight of the individual antibiotics was based on the difference in means between immunodeficiency- and control groups in combination with clinical rationale. For example, if the difference in means was < 0.05, we generally attributed a weight of 1, rather than a weight of 2, for each antibiotic prescription.

Figure 2A. Mean number of antibiotic prescriptions per year in the 10 years before the censoring date



AB: antibiotic, COPD: chronic obstructive pulmonary disease, IBD: inflammatory bowel disease, immunodef. immunodeficiency patients from primary care, no.: number, PAD: primary antibody deficiency patients from an academic hospital, RTI: respiratory tract infection.

Figure 2B. Mean total number of antibiotic prescriptions in 4 years before the censoring date



AB: antibiotic, COPD: chronic obstructive pulmonary disease, IBD: inflammatory bowel disease, Immunodef: immunodeficiency patients from primary care, no.: number, PAD: primary antibody deficiency patients from an academic hospital, RTI: respiratory tract infection.

Laboratory Results and Number of Requests

Table 4 shows the mean laboratory values of the immunoglobulins and calculated globulin in the EHR in the 10 years before the censoring date. Immunoglobulins were only requested for a very small proportion of patients from the control groups. Calculated globulin was requested more often than immunoglobulins in these control groups, although still for a very small proportion. For example, calculated globulin was requested only for 0.38% of the general GP population. For the PAD- and primary care immunodeficiency patients, immunoglobulins and calculated globulin were not requested pre-diagnosis. Although these parameters are thus not likely to apply to many patients, a reduced level of immunoglobulins or calculated globulin was still added to the algorithm based on the strong clinical relevance. Although it could be expected that a PAD diagnosis is quickly made after immunoglobulin testing, patients may be missed by the GP, thus making this a relevant component to add to the screening algorithm.

Table 4. Mean laboratory values of immunoglobulins in the 10 years before the censoring date

	General population (N=58,223)	Upper RTI (N=13,133)	COPD/ Asthma (N=4,427)	IBD (N=403)	Malignancy (N=1,526)	Immunodef. pre-diagnosis (N=26)	PAD pre-diagnosis (N=30)
IgA (g/L)	2.01 (1.06) (n = 541)	2.09 (1.11) (n=240)	2.32 (1.44) (n=63)	2.51 (1.00) (n=10)	2.31 (0.98) (n=24)	-	-
IgG1 (g/L)	3.8 (0) (n = 1)	3.8 (0) (n = 1)	-	-	-	-	-
IgG2 (g/L)	3.34 (0) (n = 1)	3.34 (0) (n = 1)	-	-	-	-	-
IgG3 (g/L)	0.67 (0) (n = 1)	0.67 (0) (n = 1)	-	-	-	-	-
IgG4 (g/L)	0.06 (0) (n = 1)	0.06 (0) (n = 1)	-	-	-	-	-
IgG total (g/L)	10.6 (2.87) (n = 65)	11.28 (2.09) (n=34)	11.08 (1.55) (n=11)	-	11.37 (1.96) (n=9)	-	-
IgM (g/L)	2.63 (4.73) (n = 67)	1.79 (1.97) (n=34)	2.44 (9.53) (n=11)	-	2.49 (6.42) (n=10)	-	-
Calculated globulin (g/L)	26.34 (12.12) (n=221)	30.66 (9.25) (n=89)	26.79 (13.53) (n=38)	33.70 (0.71) (n=3)	28.61 (15.83) (n=23)	-	-

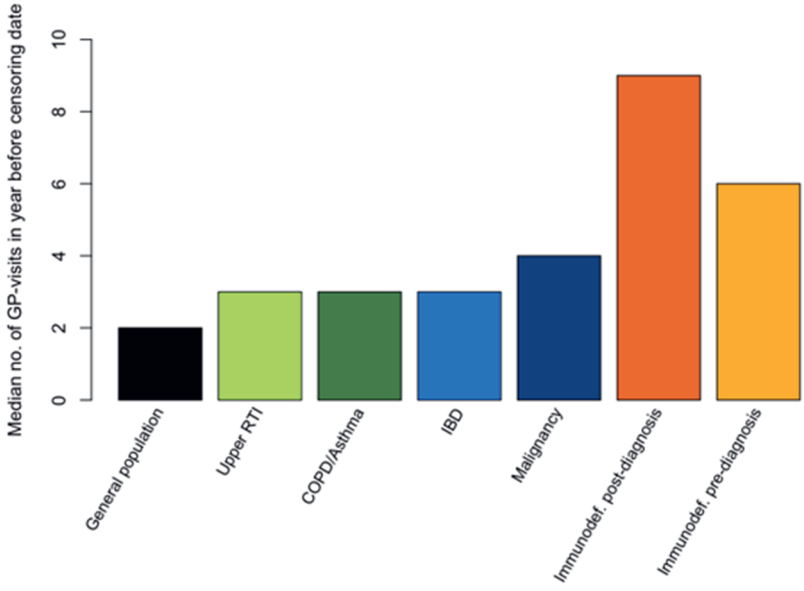
Values are shown as mean(standard deviation). EHR data from the 10 years before the censoring date. A '–' indicates that this laboratory value has not been requested for this patient group. COPD: chronic obstructive pulmonary disease, IBD: inflammatory bowel disease, Ig: immunoglobulin, immunodef: immunodeficiency patients from primary care, PAD: primary antibody deficiency, RTI: respiratory tract infection.

Based on previous research by Ilkjær et al. [47] we also considered the number of requests of leukocytes, CRP and lung function tests by the general practitioner. These however did not appear to have a discriminatory value when comparing the general population with immunodeficiency patients, see Supplemental Figure S1. These components were thus not added to the algorithm.

Visits to the General Practitioner

Figure 3 shows the median number of visits to the general practitioner in the year before the censoring date. The immunodeficiency groups showed a higher median number of visits to the general practitioner compared to the control groups. For example, PAD patients pre-diagnosis had a median of 6 visits per year, compared to 2 visits per year in the general population. This data was not available for the PAD patients from the academic hospital. Based on Youden's index the cut-off value for the algorithm was set to ≥ 6 visits to the general practitioner in a year.

Figure. 3 Median number of visits to the general practitioner in the year before the censoring date



Both physical and electronic visits are included. COPD: chronic obstructive pulmonary disease, GP: general practitioner, IBD: inflammatory bowel disease, Immunodef: immunodeficiency patients from primary care, no.: number, RTI: respiratory tract infection.

Algorithm

The components and corresponding weights of the final PAD screening algorithm are shown in Table 5. The categories of the algorithm include 'Antibiotics', 'Respiratory tract infections', 'Gastro-intestinal complaints', 'Other infections', 'Auto-immune

symptoms', 'Malignancies, lymphoproliferative- and other symptoms', 'Laboratory values' and 'Visits to general practitioner'.

Table 5: Screening algorithm for early detection of primary antibody deficiencies in primary care

ANTIBIOTICS

Total score in the 4 years before the censoring date is calculated. If a patient is enrolled in the GP clinic less than 4 years, this is corrected for with the formula: (total number of prescriptions*4)/(number of days enrolled/365). Prescriptions before the age of 6 are not taken into account.

ATC-code	Description	Score per prescription
J01AA02	Doxycycline	2
J01CA04	Amoxicillin	2
J01CF05	Flucloxacillin	1
J01CR02	Amoxicilline / clavulanic acid	2
J01EE01	Cotrimoxazole	2
J01FA01	Erythromycin	2
J01FA09	Clarithromycin	2
J01FA10	Azithromycin	2
J01MA02	Ciprofloxacin	1
J01MA12	Levofloxacin	2
J01MA14	Moxifloxacin	2
P01AB01	Metronidazole	1
J01CE05	Pheneticillin	2
J01CE02	Phenoxymethylpenicillin	2
J01DB01	Cefalexin	2
J01DC04	Cefaclor	2
J01DD14	Ceftibuten	2
J01DC02	Cefuroximaxetil	2
S02CA03	Hydrocortisone / colistin / bacitracin ear suspension	0.5
S02AA16	Ofloxacin ear suspension	0.5

RESPIRATORY TRACT INFECTIONS

Score is attributed if the ICPC code is registered in the 10 years before the censoring date.

ICPC code	Description	Score for presence of code
H01	Ear pain	1
H04	Discharge from ear	1
H71	Acute otitis media / myringitis	2
H72	Otitis media with effusion	2
H74	Chronic otitis media / other ear infections	1
H74.01	Chronic otitis media	1
H74.02	Mastoiditis	2
R05	Coughing	1
R07	Sneezing / nasal congestion / running nose	0.5
R09	Symptoms/complaints sinuses	1
R73	Furuncle / abscess nose	2

R74	Acute upper respiratory tract infection	2
R74.01	Common cold	1
R96	Asthma	2
R90	Hypertrophy / chronic infection tonsils / adenoid	2
R72	Streptococcal pharyngitis / red spark	1
R72.01	Streptococcal pharyngitis	1
R72.02	Red spark	1
R74.02	Acute pharyngitis	1
R75	Acute / chronic rhinosinusitis	2
R75.01	Acute rhinosinusitis	2
R75.02	Chronic rhinosinusitis	2
R76	Acute tonsillitis / peritonsillar abscess	2
R76.01	Acute tonsillitis	2
R76.02	Peritonsillar abscess	1
R77	Acute laryngitis / tracheitis	2
R77.01	Subglottic laryngitis / pseudo croup	1
R77.02	Acute epiglottitis	1
R78	Acute bronchitis / bronchiolitis	2
R81	Pneumonia	3
R91	Chronic bronchitis / bronchiectasis	1
R91.01	Chronic bronchitis	1
R91.02	Bronchiectasis	4

GASTRO-INTESTINAL COMPLAINTS

Score is attributed if the ICPC code is registered in the 10 years before the censoring date. Auto-immune symptoms are registered below.

ICPC code	Description	Score for presence of code
D11	Diarrhea	2
D70	Infectious diarrhea, dysentery	1
D70.01	Salmonella	1
D70.02	Shigella-/Yersinia-/Campylobacter intestinal infection	2
D70.03	Giardia	2
D73	Presumed gastro-intestinal infection	2
D86	Other peptic ulcer	1
D93	Inflammatory Bowel Syndrome	1
D94	Ulcerative colitis / Chronic enteritis	1
D94.01	Ulcerative colitis	1

OTHER INFECTIONS

Score is attributed if the ICPC code is registered in the EHR at any time-point before the censoring date.

ICPC code	Description	Score for presence of code
L70.01	Osteomyelitis	1
L70.02	Septic arthritis	1
N71	Meningitis / Encephalitis	2
N71.01	Bacterial meningitis	2
N71.02	Viral meningitis	2
N71.03	Encephalitis	2
N71.04	Myelitis	2

AUTO-IMMUNE SYMPTOMS

Score is attributed if the ICPC code is registered in the EHR at any time-point before the censoring date.

ICPC code	Description	Score for presence of code
B04	Symptoms / complaints blood / blood forming organs	1
B81	Pernicious / Folic acid anemia	1
B82	Other / Non specified anemia	1
B83	Purpura / coagulation disorder / aberrant thrombocytes	1
B83.02	Idiopathic thrombocytopenic purpura (ITP)	2
L88	Rheumatoid arthritis / related diseases	1
L88.01	Rheumatoid arthritis	2
R83.02	Sarcoidosis	1
S23.01	Alopecia areata	1
S99.04	Vitiligo	1
T86	Hypothyroidism	1
T99.02	Thyroiditis	1
T99.12	Adrenal insufficiency	1
D94.02	Crohn's Disease	2
D99.06	Coeliac disease	1
N99	Myasthenia Gravis	1

MALIGNANCIES, LYMPHOPROLIFERATIVE- AND OTHER SYMPTOMS

Score is attributed if the ICPC code is registered in the EHR at any time-point before the censoring date.

ICPC code	Description	Score
D74	Gastric cancer	1
B72	Hodgkin's disease	1
B72.01	Hodgkin's disease	1
B72.02	Non-Hodgkin lymphoma	2
T08	Weight loss	1
B87	Splenomegaly	2
B02	Lymphadenopathy	1
D96	Hepatomegaly	1
T10	Failure to thrive	2
A04	Fatigue / weakness	1
B84	Aberrant leukocytes	0.5

N94	Other peripheral neuritis / neuropathy	1

LABORATORY VALUES

Score is attributed if a reduced lab value is registered in the EHR at any time-point before the censoring date.

Laboratory measure	Aberrant value	Score if aberrant value present
IgG - total	< 7 g/L	8
IgG1	< 4.9 g/L	8
IgG2	< 1.5 g/L	8
IgG3	< 0.2 g/L	8
IgG4	< 0.08 g/L	8
IgM-total	< 0.4 g/L	8
IgA-total	< 0.7 g/L	4
Calculated globulin (total protein – albumin)	< 18g/L	6

VISITS TO GENERAL PRACTITIONER

Both physical and electronic visits are taken into account.

Description	Cut-off value	Score if visits >= cut-off value
Visits to the general practitioner clinic in the year before the censoring date.	>=6 visits	3

AMBIGUOUS CODES

For these codes, EHRs were screened up to the moment of the ambiguous diagnosis

ICPC code	Description
B72	Hodgkin's lymphoma
B72.01	Hodgkin's lymphoma
B72.02	Non-Hodgkin lymphoma
A87.02	Post-transplantation
D74	Gastric cancer
D74	Colon or rectal cancer

EXCLUSION CRITERIA

ICPC code / other	Description
B73	Leukemia
B74.01	Multiple Myeloma
B90	HIV-infection
B90.01	HIV seropositive without symptoms
B90.02	AIDS / AIDS-related complex
P15.01	Alcoholism
P15.02	Delirium tremens
P15.03	Wernicke – Korsakoff
P19.03	Addiction to hard drugs
T06	Anorexia nervosa / bulimia
T06.01	Anorexia nervosa
T06.02	Bulimia

T99.01	Immunodeficiency
T99.10	Cystic fibrosis
Age	< 12 years or > 70 years

AIDS: acquired immune deficiency syndrome, ATC: anatomic therapeutic chemical, EHR: electronic health record, HIV: human immunodeficiency virus, ICPC: International Classification of Primary Care, Ig: immunoglobulin

Discussion

In the current study we have developed a screening algorithm to identify patients at risk of PAD in a primary care setting, based on EHR data. The components of the algorithm encompass a broad range of presenting signs and symptoms of PAD, including antibiotic prescriptions, diagnostic ICPC codes, laboratory values and the number of visits to the GP. The presented algorithm can be used as an aid to notify general practitioners when further laboratory evaluation of immunoglobulins should be considered, and thus has the potential to considerably reduce the diagnostic delay of PAD.

Our results show a clear distinction in the number of antibiotic prescriptions when comparing PAD patients with relevant control groups. This effect was especially pronounced in the 4 years before PAD-diagnosis. Furthermore, our results show a distinction regarding the number of GP-consultations per year. This is in line with the main results from a previous study that compared primary care date of CVID patients with controls [47]. This study also found an increased odds ratio regarding the number requests for lung function tests and blood tests for infection, a finding which could not be reproduced in the current study. However, the authors reported that a low effectiveness was expected of these measures as a screening tool, due to either a low procedure occurrence among cases or a relatively high occurrence of these procedures amongst controls. Lastly, our results show that serum immunoglobulin levels were very rarely requested by GPs. This could indicate that GPs do not often consider the PAD diagnosis, or that they are not yet aware of the added value of assessing immunoglobulin levels. The implementation of the algorithm presented in this study could aid general practitioners when to consider requesting serum immunoglobulin levels, or calculated globulin as an alternative.

Since reducing the diagnostic delay of PAD is of great clinical interest, several other studies have investigated screening possibilities for PAD and other primary immunodeficiencies. Several sets of early warning signs (EWS) have been developed by for example the Jeffrey Modell Foundation (JMF) and the European Society of Immunodeficiencies. However, these have been shown to have a poor performance, especially in adults [24-27, 29]. This could partially be because these warning signs

focus almost exclusively on infectious complications, rather than for instance auto-immune symptoms. In addition, these manual screening systems are suboptimal for the detection of rare and heterogeneous diseases, because they depend on the awareness of an individual treating physician. For this reason, efforts have been made to develop automated algorithms for the early detection of PID such as the "SPIRIT software" of the JMF, and the "PI Prob", and a third which is currently in development[53-55]. Of these, only the "PI Prob" has been internally validated. These efforts are of great importance, however these algorithms have been developed mainly for secondary care, and mostly based on pediatric data. This may be less suitable for the recognition of PAD, as most patients present in adulthood. [9-12] Therefore, these pediatric- and secondary care algorithms could be of complementary value to the primary care algorithm presented in this paper.

In this study, we focused specifically on the early detection of PAD in primary care. As mentioned, the advantage of primary care EHRs is that they include a comprehensive overview of medical complaints. In addition, patients often primarily present at the GP, thus allowing for an early recognition of high risk patients. Focusing on PAD, rather than the full spectrum of PIDs, allows to use this algorithm in combination with serologic testing for immunoglobulins. Serologic testing for the full spectrum of PIDs would entail more expensive tests which are difficult to request and interpret by GPs, such as vaccination responses. By using the current algorithm to identify patients at risk for PAD for whom immunoglobulin analysis is indicated, it is both feasible and affordable to implement in primary care.

Because PADs are rare, our study was limited by a small sample size of PAD patients. In an effort to increase the sample size, we also analyzed primary care patients with a diagnostic code for immunodeficiency (T99.01). This group is expected to approximate PAD patients, as causes of secondary immunodeficiencies were excluded and because PAD represents the majority of the PIDs. Therefore, this group was considered to be of interest, in addition to the group of confirmed PAD-patients. A second limitation was that we only had access to metadata (e.g. means/medians) from the Julius General Practitioner's Network, due to privacy regulations. We could therefore only evaluate the diagnostic accuracy of individual components, rather than combinations of different components.

As we aimed to design an algorithm that can easily be automated, we focused on structured EHR data. The components of the algorithm are therefore inherently dependent on the available structured EHR data, such as diagnostic codes and medication prescriptions. Certain factors of possible interest, such as the number of hospitalizations, were unfortunately not structurally available in primary care.

Furthermore, we did not have access to the data regarding the number of visits to the general practitioner of the PAD patients from the academic hospital. This data was available for the primary care immunodeficiency patients. Although this is a limitation, it is important to note that the number of visits to the GP is only 1 of the 106 components of the algorithm.

Our algorithm is designed to detect adolescent and adult patients aged 12-70 years at risk of an undiagnosed PAD in a primary care setting. We focused on patients aged ≥12 years, as other cut-off points are to be expected for younger patients (e.g. due to frequent infections and antibiotic prescriptions), and because most PADs present in adulthood. [9-12] Consequently, our algorithm may be less suitable to detect XLA patients, as these usually present with symptoms during the first few years of life. However, our main aim is to reduce the diagnostic delay of PAD, which is considerably lower for XLA patients (e.g. reported median of 1 year for XLA vs. 7.5 years for PADs in general[12]), and almost all cases of XLA are diagnosed before 5 years of age. [56] In addition, it has recently been suggested to add XLA to the newborn screening, which is likely a more effective way to reduce diagnostic delay for this particular PAD. [57] Lastly, the diagnostic delay in our cohort was relatively high (median 9.5 years), which may be a consequence of our focus on a population aged 12-70 years, although the delay is within the ranges described in previous cohorts of PAD patients.[12, 13]

A major strength of the current study is that our algorithm is based completely on structured EHR data, making it suitable to implement in an automated manner. As ICPC- and ATC-codes are used internationally, this algorithm could potentially also be applied in other countries. In a consecutive study, we will prospectively validate the algorithm by applying it to 60,000 primary care EHRs in the Netherlands. The highest scoring patients will undergo laboratory assessment of serum immunoglobulin levels, and referral to an immunologist if necessary. This prospective study will allow to analyze patient-level data and determine the optimal cut-off value for high-risk patients.

Conclusions

In conclusion, in this study we developed a screening algorithm based on presenting signs and symptoms of PAD which is suitable to implement in primary care. It has the potential to considerably reduce the diagnostic delay of PAD, and will be evaluated in a prospective study.

References

1. Grimbacher BERWP. The European Society for Immunodeficiencies (ESID) registry 2014. Clin Exp Immunol. 2014;178:18-20.
2. Wood PM. Primary antibody deficiency syndromes. Curr Opin Hematol. 2010;17:356-61.
3. Chan HY, Yang YH, Yu HH, Chien YH, Chiang LL, Chiang BL. Clinical characteristics and outcomes of primary antibody deficiency: A 20-year follow-up study. J. Formos. Med. Assoc. 2014;113:340-8.
4. Fried AJ, Bonilla FA. Pathogenesis, diagnosis, and management of primary antibody deficiencies and infections. Clin. Microbiol. Rev. 2009;22:396-414.
5. Casanova JL, Abel L. Primary immunodeficiencies: a field in its infancy. Science. 2007;317:617-9.
6. Boyle JM, Buckley RH. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. J. Clin. Immunol. 2007;27:497-502.
7. Bousfiha AA, Jeddane L, Ailal F, Benhsaien I, Mahlaoui N, Casanova JL, et al. Primary immunodeficiency diseases worldwide: more common than generally thought. J. Clin. Immunol. 2013;33:1-7.
8. Wood P, Stanworth S, Burton J, Jones A, Peckham DG, Green T, et al. Recognition, clinical diagnosis and management of patients with primary antibody deficiencies: A systematic review. 2007.
9. Cunningham-Rundles C. The many faces of common variable immunodeficiency. Hematology Am. Soc. Hematol. Educ. Program. 2012;2012:301-5.
10. Wiesik-Szewczyk E, Jahnz-Rozyk K. From infections to autoimmunity: Diagnostic challenges in common variable immunodeficiency. World J Clin Cases. 2020;8:3942-55.
11. Song J, Lleo A, Yang GX, Zhang W, Bowlus CL, Gershwin ME, et al. Common Variable Immunodeficiency and Liver Involvement. Clin. Rev. Allergy Immunol. 2018;55:340-51.
12. Slade CA, Bosco JJ, Giang TB, Kruse E, Stirling RG, Cameron PU, et al. Delayed diagnosis and complications of predominantly antibody deficiencies in a cohort of Australian adults. Front. Immunol. 2018;14:9:694.
13. Maarschalk-Ellerbroek LJ, Hoepelman AIM, Van Montfrans JM, Ellerbroek PM. The spectrum of disease manifestations in patients with common variable immunodeficiency disorders and partial antibody deficiency in a university hospital. J. Clin. Immunol. 2012;32:901-21.
14. Aghamohammadi A, Pouladi N, Parvaneh N, Yeganeh M, Movahedi M, Gharagolou M, et al. Mortality and morbidity in common variable immunodeficiency. J. Trop. Pediatr. 2007;53:32-8.
15. Jolles S. The variable in common variable immunodeficiency: a disease of complex phenotypes. J Allergy Clin Immunol Pract. 2013;1:545-56.
16. Dong J, Liang H, Wen D, Wang J. Adult Common Variable Immunodeficiency. Am. J. Med. Sci. 2016;351:239-43.
17. El-Helou SM, Biegner AK, Bode S, Ehl SR, Heeg M, Maccari ME, et al. The German National Registry of Primary Immunodeficiencies (2012-2017). Front. Immunol. 2019;10:1272.
18. Odnoletkova I, Kindle G, Quinti I, Grimbacher B, Knerr V, Gathmann B, et al. The burden of common variable immunodeficiency disorders : A retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. Orphanet J. Rare Dis. 2018;13:201.
19. Gathmann B, Mahlaoui N, Gérard L, Oksenhendler E, Warnatz K, Schulze I, et al. Clinical picture and treatment of 2212 patients with common variable immunodeficiency. J. Allergy Clin. Immunol. 2014.
20. Blore J, Haeney MR. Primary antibody deficiency and diagnostic delay. BMJ. 1989;298:516-7.
21. Seymour B, Miles J, Haeney M. Primary antibody deficiency and diagnostic delay. J. Clin. Pathol. 2005;58:546-7.
22. Modell V, Orange JS, Quinn J, Modell F. Global report on primary immunodeficiencies: 2018 update from the Jeffrey Modell Centers Network on disease classification, regional trends, treatment modalities, and physician reported outcomes. Immunol. Res. 2018.
23. Odnoletkova I, Kindle G, Quinti I, Grimbacher B, Knerr V, Gathmann B, et al. The burden of common variable immunodeficiency disorders: a retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. Orphanet J. Rare Dis. 2018;13:201.
24. Reda SM, El-Ghoneimy DH, Afifi HM. Clinical predictors of primary immunodeficiency diseases in children. Allergy, Asthma and Immunology Research. 2013;5:88-95.
25. Arkwright PD, Gennery AR. Ten warning signs of primary immunodeficiency: A new paradigm is needed for the 21st century. Ann. N. Y. Acad. Sci. 2011;1238:7-14.
26. O'Sullivan MD, Cant AJ. The 10 warning signs: a time for a change? Curr. Opin. Allergy Clin. Immunol. 2012;12:588-94.
27. Arslan S, Ucar R, Caliskaner AZ, Reisli I, Guner SN, Sayar EH, et al. How effective are the 6 European Society of Immunodeficiency warning signs for primary immunodeficiency disease? Annals of Allergy, Asthma and Immunology. 2016;116:151-5.e1.
28. Lankisch P, Schiffner J, Ghosh S, Babor F, Borkhardt A, Laws HJ. The Duesseldorf Warning Signs for Primary Immunodeficiency: Is it Time to Change the Rules? J. Clin. Immunol. 2015;35:273-9.

29. Bjelac JA, Yonkof JR, Fernandez J. Differing Performance of the Warning Signs for Immunodeficiency in the Diagnosis of Pediatric Versus Adult Patients in a Two-Center Tertiary Referral Population. J. Clin. Immunol. 2019;39:90-8.
30. Netherlands Institute for Health Services Research, Nivel.nl Netherlands [updated 08-11-2022; cited 2022 28-11-2022]. Available from: https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/cijfers-over-zorgverlening/cijfers-zorgverlening-huisartsen.
31. Thickett KM, Kumararatne DS, Banerjee AK, Dudley R, Stableforth DE. Common variable immune deficiency: respiratory manifestations, pulmonary function and high-resolution CT scan findings. QJM. 2002;95:655-62.
32. Heyworth MF. Immunological aspects of Giardia infections. Parasite. 2014;21:55.
33. Khodadad A, Aghamohammadi A, Parvaneh N, Rezaei N, Mahjoob F, Bashashati M, et al. Gastrointestinal manifestations in patients with common variable immunodeficiency. Dig. Dis. Sci. 2007;52:2977-83.
34. Cunningham-Rundles C, Bodian C. Common variable immunodeficiency: Clinical and immunological features of 248 patients. Clin. Immunol. 1999;92:34-48.
35. Lederman HM, Winkelstein JA. X-linked agammaglobulinemia: an analysis of 96 patients. Medicine (Baltimore). 1985;64:145-56.
36. Knight AK, Cunningham-Rundles C. Inflammatory and autoimmune complications of common variable immune deficiency. Autoimmun Rev. 2006;5:156-9.
37. Agarwal S, Cunningham-Rundles C. Autoimmunity in common variable immunodeficiency. Ann. Allergy. Asthma. Immunol. 2019;123:454-60.
38. Azizi G, Bagheri Y, Tavakol M, Askarimoghaddam F, Porrostami K, Rafiemanesh H, et al. The Clinical and Immunological Features of Patients with Primary Antibody Deficiencies. Endocr. Metab. Immune Disord. Drug Targets. 2018;18:537-45.
39. Mellemkjaer L, Hammarstrom L, Andersen V, Yuen J, Heilmann C, Barington T, et al. Cancer risk among patients with IgA deficiency or common variable immunodeficiency and their relatives: a combined Danish and Swedish study. Clin. Exp. Immunol. 2002;130:495-500.
40. Kinlen LJ, Webster AD, Bird AG, Haile R, Peto J, Soothill JF, et al. Prospective study of cancer in patients with hypogammaglobulinaemia. Lancet. 1985;1:263-6.
41. Quinti I, Soresina A, Spadaro G, Martino S, Donnanno S, Agostini C, et al. Long-term follow-up and outcome of a large cohort of patients with common variable immunodeficiency. J. Clin. Immunol. 2007;27:308-16.
42. Graziano V, Pecoraro A, Mormile I, Quaremba G, Genovese A, Buccelli C, et al. Delay in diagnosis affects the clinical outcome in a cohort of cvid patients with marked reduction of iga serum levels. Clin. Immunol. 2017;180:1-4.
43. Zorginstituut Nederland. Farmacotherapeutisch Kompas. [Available from: www.farmacotherapeutischkompas.nl.
44. Hobbs MR, Grant CC, Ritchie SR, Chelimo C, Morton SMB, Berry S, et al. Antibiotic consumption by New Zealand children: exposure is near universal by the age of 5 years. J. Antimicrob. Chemother. 2017;72:1832-40.
45. Holding S, Khan S, Sewell WA, Jolles S, Dore PC. Using calculated globulin fraction to reduce diagnostic delay in primary and secondary hypogammaglobulinaemias: results of a demonstration project. Ann. Clin. Biochem. 2015;52:319-26.
46. Holding S, Jolles S. Current screening approaches for antibody deficiency. Curr. Opin. Allergy Clin. Immunol. 2015;15:547-55.
47. Ilkjær FV, Rasmussen LD, Martin-Iguacel R, Westh L, Katzenstein TL, Hansen ABE, et al. How to Identify Common Variable Immunodeficiency Patients Earlier: General Practice Patterns. J. Clin. Immunol. 2019;39:641-52.
48. Raje N, Dinakar C. Overview of Immunodeficiency Disorders. Immunol. Allergy Clin. North Am. 2015;35:599-623.
49. Srivastava S, Wood P. Secondary antibody deficiency - causes and approach to diagnosis. Clin Med (Lond). 2016;16:571-6.
50. Cunningham-Rundles C, Sidi P, Estrella L, Doucette J. Identifying undiagnosed primary immunodeficiency diseases in minority subjects by using computer sorting of diagnosis codes. J. Allergy Clin. Immunol. 2004;113:747-55.
51. Smeets HM, Kortekaas MF, Rutten FH, Bots ML, van der Kraan W, Daggelders G, et al. Routine primary care data for scientific research, quality of care programs and educational purposes: the Julius General Practitioners' Network (JGPN). BMC Health Serv. Res. 2018;18:735.
52. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32-5.

53. Modell V, Gee B, Lewis DB, Orange JS, Roifman CM, Routes JM, et al. Global study of primary immunodeficiency diseases (PI)--diagnosis, treatment, and economic impact: an updated report from the Jeffrey Modell Foundation. Immunol. Res. 2011;51:61-70.
54. Rider NL, Cahill G, Motazedi T, Wei L, Kurian A, Noroski LM, et al. PI Prob: A risk prediction and clinical guidance system for evaluating patients with recurrent infections. PLoS One. 2021;16:e0237285.
55. Kumar B, Zetumer S, Swee M, Endelman ELK, Suneja M, Davis B. Reducing Delays in Diagnosing Primary Immunodeficiency Through the Development and Implementation of a Clinical Decision Support Tool: Protocol for a Quality Improvement Project. JMIR Res Protoc. 2022;11:e32635.
56. Lackey AE, Ahmad F. X-linked Agammaglobulinemia. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022.
57. Blom M, Bredius RGM, van der Burg M. Future Perspectives of Newborn Screening for Inborn Errors of Immunity. Int J Neonatal Screen. 2021;7.

Supplemental material

Supplemental Table S1. Youden's index per ICPC code
Youden's index was calculated based on the presence of each code in the patient groups with primary antibody deficiency versus the control group. A higher index indicates a higher discriminative value of the ICPC code. ICPC: International Classification of Primary Care.

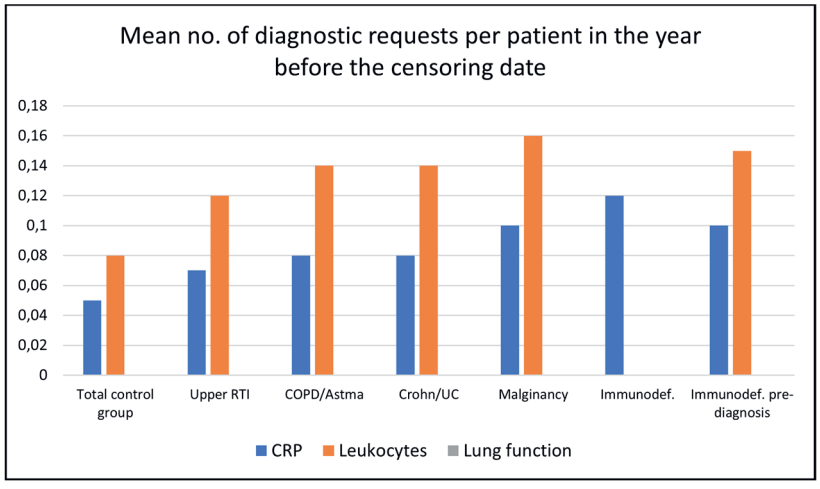
ICPC code	Description	Youden's index
A04	Tiredness/weakness	10
A91.07	Subclinical hypothyroidism	1.5
B02	Enlarged lymph nodes	16
B04	Symptoms / complaints blood / blood forming organs	2
B72	Hodgkin's disease	0
B72.01	Hodgkin's disease	0
B72.02	Non-hodgkin lymphoma	3.5
B81	Pernicious / folic acid anemia	0
B82	Other / non specified anemia	4
B83	Purpura / coagulation disorder / aberrant thrombocytes	0
B83.02	Idiopathic thrombocytopenic purpura (ITP)	3.5
B84	Aberrant leukocyte count	0
B87	Splenomegaly	15
D11	Diarrhea	17.5
D70	Infectious diarrhea, dysentery	-1
D70.01	Salmonella	0
D70.02	Shigella-/Yersinia-/Campylobacter-intestinal infection	1.5
D70.03	Giardia	1.5
D73	Presumed gastro-intestinal infection	2
D74	Gastric cancer	0
D77.01	Malignancy esophagus	0
D86	Other peptic ulcer	2
D86.01	Ventricular ulcer	0
D93	Inflammatory Bowel Syndrome	4
D94	Ulcerative colitis / chronic enteritis	0
D94.01	Ulcerative colitis	0
D94.02	Crohn's disease	7
D99.06	Celiac disease	0
H01	Pain in ear	5.5
H04	Discharge from ear	2.5
H70	External otitis	0.5
H71	Acute otitis media / myringitis	16
H72	Otitis media with effusion	7
H74	Chronic otitis media / other ear infection	8.5
H74.01	Chronic otitis media	1.5
H74.02	Mastoiditis	0
L70	Infectious disease of the musculoskeletal system	0
L70.01	Osteomyelitis	0

L70.02	Septic arthritis	0
L88	Rheumatoid arthritis / related diagnoses	4
L88.01	Rheumatoid arthritis	0
N71	Meningitis / encephalitis	8.5
N71.01	Bacterial meningitis	0
N71.02	Viral meningitis	0
N71.03	Encephalitis	0
N71.04	Myelitis	0
N94	Other peripheral neuritis / neuropathy	1
N99	Myasthenia gravis	0
R05	Coughing	39.5
R07	Sneezing / congested nose / running nose	0
R09	Symptoms / complaints sinuses	1.5
R72	Streptococcal pharyngitis / red spark	-1
R72.01	Streptococcal pharyngitis	0
R72.02	Red spark	0
R73	Furuncle / abscess nose	2
R74	Acute upper respiratory tract infection	44.5
R74.01	Common cold	5.5
R74.02	Acute pharyngitis	-4
R75	Acute / chronic rhinosinusitis	12.5
R75.01	Acute rhinosinusitis	11
R75.02	Chronic rhinosinusitis	20
R76	Acute tonsillitis / Peritonsillar abscess	2.5
R76.01	Acute tonsillitis	4
R76.02	Peritonsillar abscess	0
R77	Acute laryngitis / tracheitis	3
R77.01	Subglottic laryngitis / pseudo croup	0
R77.02	Acute epiglottitis	0
R78	Acute bronchitis/bronchiolitis	8
R81	Pneumonia	20.5
R83.02	Sarcoidosis	3.5
R90	Hypertrophy / chronic infection tonsils / adenoid	10
R91	Chronic bronchitis / bronchiectasis	0
R91.01	Chronic bronchitis	0
R91.02	Bronchiectasis	10
R95	COPD (chronic obstructive pulmonary disease)	-1
R96	Asthma	9
S23.01	Alopecia areata	2.5
S99.04	Vitiligo	3.5
T08	Weight loss	5.5
T10	Failure to thrive	4
T86	Hypothyroidism	2.5
T90.01	Diabetes mellitus type 1	2
T99.02	Thyroiditis	0
T99.12	Adrenal gland insufficiency	0

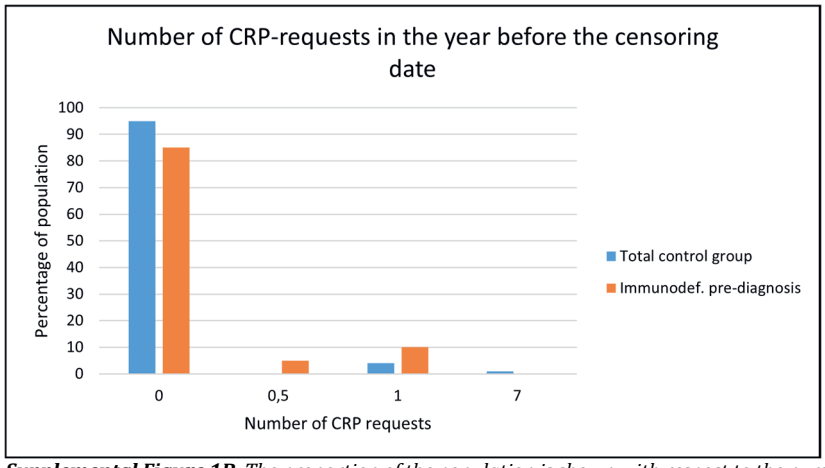
Supplemental Table S2. Mean number of antibiotic prescriptions per ATC (Anatomical Therapeutic Chemical) code in the 4 years before the extraction date

ATC-code	Description	Control group (mean)	Immunodeficiency pre-diagnosis (mean)	Difference in means
J01AA02	Doxycycline	0.02	0.05	0.03
J01CA04	Amoxicillin	0.03	0.15	0.12
J01CE02	Phenoxymethylpenicillin	0	NA	NA
J01CE05	Pheneticillin	0	0.08	0.08
J01CF05	Flucloxacillin	0.01	0.03	0.02
J01CR02	Amoxicillin / Clavulanic acid	0.01	0.12	0.11
J01DB01	Cefalexin	0	NA	NA
J01DC02	Cefuroximaxetil	0	NA	NA
J01DC04	Cefaclor	0	NA	NA
J01DD14	Ceftibuten	0	NA	NA
J01EE01	Cotrimoxazole	0	0.19	0.19
J01FA01	Erythromycin	0	NA	NA
J01FA09	Clarithromycin	0	0.02	0.02
J01FA10	Azithromycin	0.01	0.21	0.2
J01MA02	Ciprofloxacin	0.01	0.03	0.02
J01MA12	Levofloxacin	0	NA	NA
J01MA14	Moxifloxacin	0	0.03	0.03
J01XD01	Metronidazole	NA	NA	NA
S02AA16	Ofloxacin ear suspension	0	NA	NA
S02CA03	Hydrocortisone / colistin / bacitracin ear suspension	0.01	0	-0.01

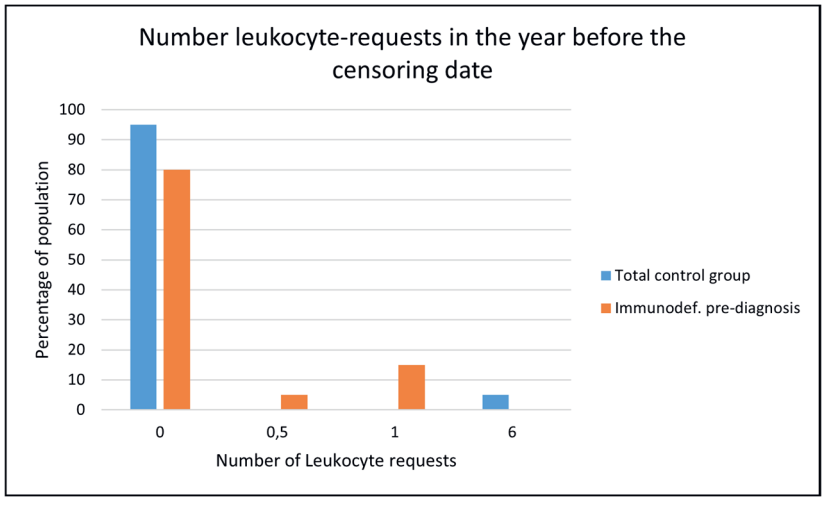
Supplemental Figure S1. Data from the primary care electronic health record (EHR) on the number of diagnostic requests for diagnostic requests for leukocytes, C-reactive protein (CRP) and lung function tests. The censoring date is the date before which the electronic health care record (EHR) was screened. For most patients this is the date of data-extraction, November 2021. For immunodeficiency patients pre-diagnosis, the censoring date is the diagnosis date. For details on the censoring date, see the Methods section.



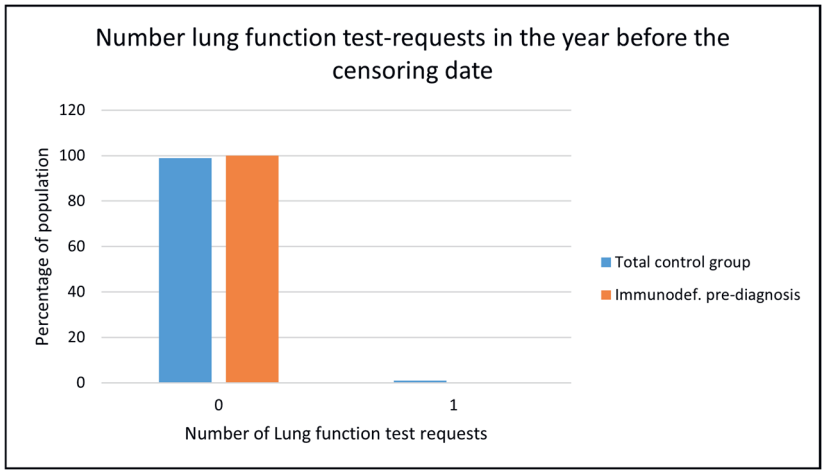
Supplemental Figure 1A. *Mean number of diagnostic requests per year for different patient populations. The mean number of requested lung function tests per patient in the year before the censoring date was too low to be visible in this figure. There is no clear distinction between the non-specified immunodeficiency patients pre-diagnosis and the other control groups, indicating that these parameters are not of additional value to the algorithm. COPD: chronic obstructive pulmonary disease, CRP: C-reactive protein, immunodef: immunodeficiency patients from primary care, no.: number, RTI: respiratory tract infection.*



Supplemental Figure 1B. *The proportion of the population is shown with respect to the number of requests for CRP in the year before the censoring date. No clear distinction is visible between the general population and the immunodeficiency patients pre-diagnosis, indicating that adding this parameter to the algorithm is of no additional value. CRP: C-reactive protein, immunodef: immunodeficiency patient from primary care.*



Supplemental Figure 1C. *The proportion of the population is shown with respect to the number of requests for leukocytes in the year before the censoring date. No clear distinction is visible between the general population and the immunodeficiency patients pre-diagnosis, indicating that adding this parameter to the algorithm is of no additional value. Immunodef: immunodeficiency patients from primary care.*



Supplemental Figure 1D. *The proportion of the population is shown with respect to the number of requests for lung function tests in the year before the censoring date. No clear distinction is visible between the general population and the immunodeficiency patients pre-diagnosis, indicating that adding this parameter to the algorithm is of no additional value. Immunodef: immunodeficiency patients from primary care.*

Chapter 5

Clinical validation of a primary antibody deficiency screening algorithm for primary care

Marianne A. Messelink
Paco M.J. Welsing
Giovanna Devercelli
Jan Willem N. Marsden
Helen L. Leavis

Journal of Clinical Immunology, 2023;43(8):2022-2032

Abstract

Background The diagnostic delay of primary antibody deficiencies (PADs) is associated with increased morbidity, mortality, and health care costs. We have therefore previously developed a screening algorithm for the early detection of patients at risk of PAD in primary care.

Objective To clinically validate and optimize the PAD screening algorithm by applying it to a primary care database in the Netherlands.

Methods The algorithm was applied to a dataset of 61 172 electronic health records (EHRs). Four hundred high scoring EHRs were screened for exclusion criteria, after which remaining patients were invited for analysis of serum immunoglobulins and referred if clinically necessary.

Results Of the 104 patients included in this study, 16 were referred by their general practitioner (GP) for suspected PAD, of whom 10 had a PAD diagnosis. In patients selected by the screening algorithm and included for laboratory analysis, prevalence of PAD was ~1:10 compared to 1:1700–1:25 000 in the general population. To optimize efficiency of the screening process, we refitted the algorithm with the subset of high-risk patients, which improved the AUC-ROC to 0.80 (95% confidence interval 0.63–0.97). We propose a two-step screening process, first applying the original algorithm to distinguish high-risk from low-risk patients, and subsequently the optimized algorithm to select high-risk patients for analysis of serum immunoglobulins.

Conclusion With the screening algorithm we were able to identify 10 new PAD patients from a primary care population, thus reducing diagnostic delay. Future studies should address further validation in other populations and full cost-effectiveness analyses.

Introduction

Primary antibody deficiencies (PADs) form the majority of primary immunodeficiencies (PIDs) and are characterized by an inability to produce a clinically effective antibody response.[1,2] PADs represent a heterogeneous group of disorders such as common variable immunodeficiency (CVID), IgG subclass deficiency, and specific antibody deficiency (SpAD).[3] The reported prevalence of PAD varies considerably from 1:1700 to 1:25 000, partly due to the suspected large number of undiagnosed patients.[4-6] The clinical presentation encompasses a wide range of symptoms including increased susceptibility to respiratory- and gastro-intestinal tract infections, autoimmunity, and an increased risk of certain malignancies.[2,6]

Due to the heterogeneous presentation and low prevalence, diagnosis of PAD can be challenging. This is evident from the reported median delay in diagnosis between 2–10 years, which has not improved substantially over the past 5 decades.[7-12] This diagnostic delay is associated with increased morbidity and mortality, as effective therapies are available.[12-14] A timely diagnosis may also result in substantial health care cost savings, even when taking the cost of treatment into consideration.[15] Reducing the diagnostic delay of PAD is thus of key importance.[12]

To this end, we have developed an algorithm that can be used to detect patients with a high risk of PAD in a primary care setting.[16] An advantage of focusing on primary care is that most patients initially present their complaints at a general practitioner (GP), especially in countries where the GP has a gatekeeper function to secondary care. This allows to detect PAD patients in an early phase. In addition, primary care electronic health records (EHRs) encompass a comprehensive overview of the symptoms for which a patient has sought medical care. In contrast, in secondary care usually only the symptoms for which a patient has been referred are documented structurally. For example, if a patient is referred for suspected inflammatory bowel disease, the secondary care EHR might not include recurrent respiratory tract infections for which a patient has visited the GP. Focusing on primary care thus allows screening for a broad range of PAD symptoms at an early stage.

The algorithm is based on structured EHR-data including diagnostic codes, antibiotic prescriptions, laboratory results, and the number of visits to the general practitioner. Focusing on structured EHR-data enables the application of the algorithm in an automated manner to large databases. The aim of this study was to clinically validate and optimize the algorithm, by applying it to a primary care database. Patients identified by the algorithm as being at increased risk of PAD were invited for laboratory

evaluation of immunoglobulin levels and referred to an immunologist if deemed clinically necessary.

Methods

Details on the algorithm have been reported previously and in Table E1.[16] In short, the algorithm was developed using EHR-data from PAD-patients (University Medical Centre Utrecht), aggregated subgroup data from control groups (Julius General Practitioner Network [JHN] Utrecht), literature and clinical expertise.[17] The algorithm encompasses 107 items within 8 categories: 'Antibiotic prescriptions,' 'Respiratory tract infections' (RTI), 'Gastro-intestinal (GI) complaints,' 'Other infections,' 'Auto-immune symptoms,' 'Malignancies, lymphoproliferative- and other symptoms,' 'Laboratory values,' and 'Number of visits to the GP'.

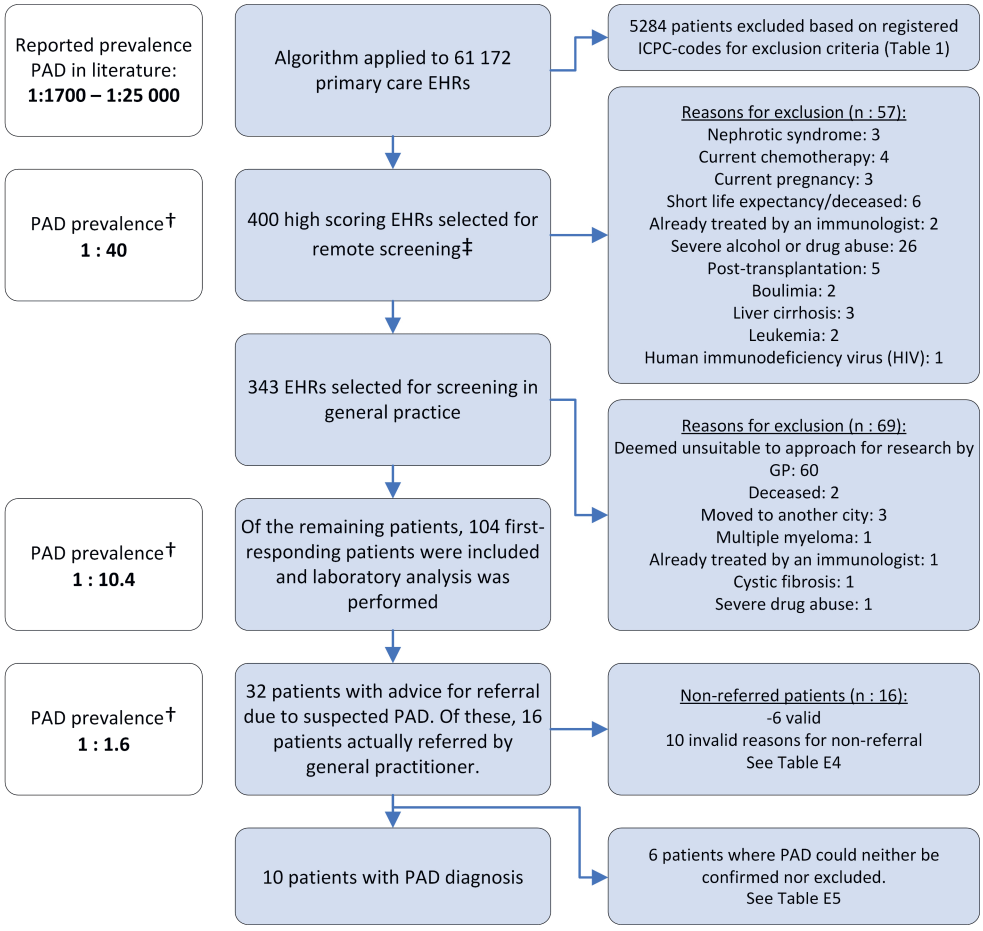
In the current study, the algorithm was applied to a JHN dataset containing 61 172 EHRs from 13 general practices. All patients in the JHN database were offered an opt-out prior to registration. EHR-data was extracted for a certain period before the 'censoring date' (e.g. 4 years for antibiotics, Table E1). Usually, this was the date of application of the algorithm to the database (February 8th, 2022). For certain uncommon diagnoses that can be both a complication of PAD, but also the cause of a secondary antibody deficiency (SAD, e.g. non-Hodgkin lymphoma), the censoring date was the registration date of this ambiguous diagnosis (Table E1).

As the estimated prevalence of PAD is 1:1700–1:25 000, 2-36 PAD cases were a priori expected to be present in the dataset of 61 172 patients.[4-6] Previous PID-screening studies selected 0.1%-0.4% of their population for further analysis.[18,19] Based on the above, expert opinion, and feasibility, we aimed to screen the 400 highest scoring EHRs to confirm eligibility. Of the remaining patients, we aimed to include at least 100 patients (0.2% of the dataset) for laboratory analysis. See Figure 1 for an overview of the study flow. We focussed on patients aged 12-70 years as PAD usually presents in the second to fourth decade of life, because of restrictions regarding study participation of patients < 12 years, and because differing clinical presentations have been described for children vs. adults.[8,20-23]

Exclusion criteria (Table I) for which an ICPC-code (International Classification of Primary Care) was available were applied by the algorithm to the population of 61 172 patients, and include previously diagnosed (secondary) immunodeficiencies. Subsequently, exclusion criteria were verified manually in the 400 EHRs selected for screening. This was performed in a 2-step manner due to COVID-19 restrictions. First,

pseudonymized EHRs were screened remotely from a secured server, and subsequently remaining patients were discussed with the GP on location.

Figure 1. Overview of study flow.



EHR, electronic health record; GP, general practitioner; ICPC, International Classification of Primary Care; JHN, Julius General Practitioner Network; PAD, primary antibody deficiency. †Prevalence was determined based on the 10 PAD patients in this dataset. ‡This selection included patients with a rank number < 400 due to an error in data-extraction, see results section and Table E6.

Patients that remained eligible after screening were invited for participation through a letter from their GP. Participation consisted of a single visit with analysis of serum immunoglobulins and calculated globulin, and the Early Warning Signs (EWS) questionnaire.[24,25] GPs were advised to refer patients with reduced immunoglobulins for further evaluation of PAD to an immunologist or infectious disease specialist, except if it concerned solitary reduced IgG4 as this has little clinical relevance.[26] In addition, GPs were advised to refer the 10% highest scoring patients, to account for SpAD which presents without concomitant reduced immunoglobulins.[27] Lastly, GPs were advised to

consult an internist in case of incidental findings of elevated immunoglobulins that were not suspect for PAD. Six months after inclusion, GPs were contacted to verify referral outcomes. PAD was classified according to the IUIS criteria by a clinical immunologist.[28] This study was approved by the Medical Research Ethics Committee NedMec under protocol number NL74 944.041.20. All patients included for laboratory analysis provided written informed consent. For the TRIPOD checklist, see Table E2.

Table I. Exclusion criteria

Exclusion criteria
Leukaemia (B73)
Multiple myeloma (B74.01)
HIV (B90, B90.01, B90.02)
Anorexia Nervosa / Bulimia (T06, T06.01, T06.02)
Cystic Fibrosis (T99.10)
Severe alcohol addiction (P15.01, P15.02, P15.03)
Addiction to hard drugs (P19.03)
Previously diagnosed immunodeficiency (T99.01)
Nephrotic syndrome
Stadium 3-4 liver cirrhosis
Current systemic chemotherapy
Current pregnancy
Short life expectancy / deceased
Patient is already treated by an immunologist
Patient is not deemed suitable for participation by GP

EHR, electronic health record; GP, general practitioner; HIV, human immunodeficiency virus; ICPC, International Classification of Primary Care. If an ICPC code was available, this is represented in the table in brackets. Exclusion criteria were applied by the algorithm if an ICPC code was available, and subsequently verified by manually screening the EHR of the 400 highest scoring patients.

Descriptive statistics are presented as means with standard deviations, medians with interquartile ranges, or frequencies with percentages. To compare continuous data between PAD and non-PAD patients t-tests or non-parametric Wilcoxon rank sum tests (for 2 groups) were performed, or for ≥ 2 groups ANOVA or Kruskall-Wallis (non-parametric) tests. For categorical characteristics Chi-square tests were used, or Fisher's exact test if small cell frequencies were expected (< 5).

In addition to the original algorithm (version 1), we explored three alternative algorithms (versions 2–4) to optimize predictive performance within the subset of high-risk patients with a confirmed PAD/non-PAD diagnosis. We performed penalized logistic regression analyses, with the presence of PAD as dependent variable. In the original algorithm (version 1), the 8 categories were weighted equally. In version 2, category weights were adjusted based on Ridge regression coefficients (λ=1SE), with the category scores as independent variables. In version 3, the items (e.g. 'pneumonia')

per category (e.g. 'RTI') were first grouped using a principal component (PC) analysis to prevent overfitting due to the large number of items compared to the number of patients. The number of PCs was based on an eigenvalue of ≥ 1. Items were grouped in the PC where they had the highest contribution, or based on clinical rationale if they were not present in this dataset. Version 3 was derived by first determining the weight per item group, and subsequently the weight per category using Ridge regression. In version 4, we explored the addition of new variables which were not available during algorithm development or have an ambiguous relationship to PAD, i.e. use of immunosuppressant medication in the past 4 years, presence of an ICPC-code for chronic obstructive pulmonary disease or malignancy, ≥ 6 GP-visits in the past 2–4 years, and the EWS score. These variables, and the total score of the optimal algorithm (from versions 1–3), were combined in a Lasso regression; Algorithm 4 consisted of the retained variables. The predictive performance of all algorithms was determined with the AUC-ROC, sensitivity, and specificity using optimal cut-offs based on Youden's index or 100% sensitivity. Statistics were performed in R version 4.2.0.

Results

The algorithm was applied to 61 172 EHRs; 5284 patients were excluded based on ICPC-codes (see Table I) of which 8 had a previous PAD-diagnosis. From the remaining 55 888 patients, 400 high ranking patients were selected for EHR-screening. Of these 400 patients, 104 were included for immunoglobulin assessment. From these, 16 were referred, of whom 10 were diagnosed with a PAD (Table E3). Sixteen patients were not referred despite a referral advice, of which 6 were deemed valid and 10 invalid (Table E4). For example, a sufficient explanation for frequent antibiotic use was deemed a valid reason, while 'referral was too much effort' was deemed invalid, as PAD cannot be excluded in this case. The valid non-referred patients(n=6) and the patients without a referral advice(n=72) were labelled as 'Unlikely PAD-diagnosis'(n=78). For 6 referred patients PAD could not be confirmed nor excluded (Table E5), these were labelled as 'Inconclusive'(n=16) together with the invalid non-referred patients(n=10). The prevalence of PAD in the subpopulations selected with each step of the study is shown in Figure 1. In the general population, the prevalence of PAD is estimated to be 1:1700–1:25 000.[4-6] In the 400 patients selected for EHR-screening prevalence was estimated at 1:40, in those where immunoglobulin analyses was performed at ~1:10, and in those referred for suspected PAD at ~1:2. This can be translated to a number needed to screen of 40, a number needed to test of 10, and a number needed to refer of 2 to identify one case of PAD.

Initially, we aimed to select the 400 highest scoring EHRs for screening. After termination of the study, it appeared however that lower ranks were also screened due

to data-extraction error concerning antibiotic prescriptions, of which the ethical committee was informed. See Table E6 for details. The included patients were still within the highest scoring 2% of the total population of 61 172 patients (Figure E1). An unintended benefit of this occurrence is that it allowed us to study patients with a wider range of ranks. Three of the newly identified PAD-patients had a rank lower than 400 (528, 657, and 791), as well as one patient with an inconclusive diagnosis (803). It thus appears that the initially estimated cut-off point of 21.5 (based on 400 highest ranks) was too strict, a finding which would have remained unnoticed if we had only screened the top 400 EHRs. A cut-off of ≥ 17 (corresponding to a rank of 1000) may be more suitable, as all confirmed and inconclusive PAD-cases are well within this range.

The baseline characteristics are shown in Table II; there were no statistically significant differences between groups of included patients (i.e. 'PAD-diagnosis,' 'unlikely PAD,' and 'inconclusive diagnosis'). The algorithm scores are shown in Table III. Statistically significant differences were present for the total score and for the categories 'Antibiotic prescriptions' and 'RTIs,' but not for other categories. Most points were scored in the categories 'Antibiotic prescriptions,' 'RTIs,' and 'Visits to the GP', while points were rarely scored for 'Other infections' (e.g. meningitis, osteomyelitis, Table E1) and 'Auto-immune symptoms'. There were no previously registered reduced immunoglobulin levels in the EHRs of included patients, most likely because these were not requested by GPs: total IgG and IgM were determined in only one patient, and IgG subclasses in none.

Table II. Baseline characteristics of included and screened patients

	Included patients (n=104)			
	PAD-diagnosis	**Unlikely PAD †**	**Inconclusive ‡**	**Screened patients**
n	10	78	16	400
Age (y) mean (SD)	58.4 (6.4)	51.9 (14.8)	55.8 (11.8)	52.3 (14.9)
Patients aged 12-18 years n(%)	0 (0%)	4 (5.1%)	0 (0.0%)	18 (4.5%)
Patients aged 60-70 years n(%)	4 (40.0%)	31 (39.7%)	8 (50.0%)	168 (42.0%)
Female n(%)	8 (80.0%)	67 (85.9%)	14 (87.5%)	294 (73.5%)
Caucasian n(%)	9 (90.0%)	62 (79.5%)	13 (18.3%)	NA
Pack years median (IQR)	22.5 (22.5–24.0)	11.3 (2.9 – 24.0)	22.0 (12.0-24.0)	NA
JMF 10 Early Warning Signs, median (IQR)	3.5 (2.3-4.0)	2.0 (1.0-4.0)	3.0 (2.0-4.0)	NA

IQR, interquartile range; JMF, Jeffrey Modell Foundation; n, number; PAD, primary antibody deficiency; SD, standard deviation. There were no statistical significant differences between groups of included patients.

† Included patients who did not receive an advice for referral for suspected PAD, or who were not referred by their GP based on valid reasons (Table E4). Referral was advised based on reduced immunoglobulin results or if patients were within the top 10% of highest scoring patients on the algorithm.
‡ Diagnosis was inconclusive if 1) referral was advised, but the patient was not referred based on invalid reasons (Table E4) or 2) the patient was referred, but PAD could not be confirmed nor excluded (see Table E5).

Table III. Algorithm score, total and per category

	Included patients (n=104)				
	PAD-diagnosis (n=10)	**Unlikely PAD † (n=78)**	**Inconclusive‡ (n=16)**	**Sig.**	**Screened patients (n=400)**
Total score on algorithm median (IQR)	24.8 (20.8-31.5)	22.0 (19.0-29.0)	29.8 (25.1-44.3)	*	22.0 (19.0-29.0)
Antibiotic prescriptions					
Number of prescriptions in 4 years, median (IQR) (range)	5.0 (2.0-13.8) (2.0-25.0)	4.0 (2.0-7.0) (0.0-17.0)	9.5 (6.0-15.0) (4.0-31.0)	*	4.0 (2.0-7.0) (0.0-45.0)
Score, median (IQR)	7.3 (3.3-17.0)	6.8 (4.0-14.0)	16.9 (10.4-28.0)	*	7.0 (3.0-14.0)
RTIs					
Number of ICPC-codes, median (IQR) (range)	5.5 (5.0-6.0) (3.0-10.0)	5.0 (4.0-6.0) (0.0-15.0)	4.0 (2.8-5.3) (1.0-6.0)	*	5.0 (4.0-6.0) (0.0-15.0)
Score, median (IQR)	9.5 (7.5-11.8)	9.0 (7.0-11.8)	6.0 (5.0-9.3)	*	9.0 (6.8-11.0)
GI-complaints					
Number of ICPC-codes, median (IQR) (range)	1.0 (0.3-1.8) (0.0-2.0)	1.0 (0.0-2.0) (0.0-3.0)	0.5 (0.0-2.0) (0.0-2.0)		1.0 (0.0-1.0) (0.0-4.0)
Score, median (IQR)	2.0 (0.3-2.8)	2.0 (0.0-3.0)	0.5 (0.0-2.3)		2.0 (0.0-2.0)
Other infections ‡‡					
Number of ICPC-codes, median (IQR) (range)	0.0 (0.0-0.0) (0.0-0.0)	0.0 (0.0-0.0) (0.0-0.0)	0.0 (0.0-0.0) (0.0-1.0)		0.0 (0.0-0.0) (0.0-2.0)
Score, median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)		0.0 (0.0-0.0)
Auto-immune symptoms					
Number of ICPC-codes, median (IQR) (range)	0.0 (0.0-0.0) (0.0-2.0)	0.0 (0.0-0.0) (0.0-3.0)	0.0 (0.0-0.0) (0.0-3.0)		0.0 (0.0-1.0) (0.0-3.0)
Score, median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)		0.0 (0.0-1.0)
Malignancy/lymphoproliferative/other symptoms					
Number of ICPC-codes, median (IQR) (range)	1.0 (1.0-1.0) (0.0-1.0)	1.0 (0.0-1.0) (0.0-3.0)	1.0 (1.0-1.0) (0.0-2.0)		1.0 (0.0-1.0) (0.0-5.0)
Score, median (IQR)	1.0 (1.0-1.0)	1.0 (0.0-1.0)	1.0 (1.0-1.3)		1.0 (0.0-1.1)
Previously registered reduced immunoglobulin levels in EHR					
Number of patients (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		2.0 (0.5%)
Score, median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)		0.0 (0.0-0.0)

≥ 6 visits to the general practitioner per year					
Total number of visits, median (IQR) (range)	15.5 (10.3-24.8) (6.0-40.0)	16.0 (10.0-21.0) (3.0-21.0)	11.5 (8.0-19.3) (2.0-58.0)		15.0 (10.0-24.0) (2.0-61.0)
Score, median (IQR)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)		3.0 (3.0-3.0)

*AB, antibiotic; EHR, electronic health record; GI, gastro-intestinal; ICPC, International Classification of Primary Care; IQR, interquartile range; n, number; PAD, primary antibody deficiency; RTI, respiratory tract infection; Sig, significance. Score per category of the algorithm is based on the presence of items (e.g. ICPC-code for pneumonia) multiplied by the weight of that item (see Table E1). *Statistically significant difference between groups of included patients. † Included patients who did not receive an advice for referral for suspected PAD, or who were not referred by their GP based on valid reasons (Table E4). Referral was advised based on reduced immunoglobulin results or if patients were within the top 10% of highest scoring patients on the algorithm.‡ Diagnosis was inconclusive if 1) referral was advised, but the patient was not referred based on invalid reasons (Table E4) or 2) the patient was referred, but no definite diagnosis could be made (see Table E5). ‡‡ See Table E1 for the diagnostic codes that belong to this category, includes e.g. meningitis and septic arthritis.*

The serum immunoglobulin results are shown in Table IV. PAD patients had significantly lower IgM, total IgG, and IgG1, IgG2, and IgG3 subclass levels than 'Unlikely PAD patients'. There were no statistical differences for IgA nor IgG4 subclass levels. Concerning calculated globulin, PAD patients had a significantly lower median value, but none of the patients had a value below the diagnostic cut-off of 18 grams/litre.[25]

To optimize the efficiency of the screening approach, we aimed to improve the predictive performance within the subset of high-risk patients identified by the original algorithm, i.e. the 88 confirmed PAD/unlikely PAD cases, see Table V. When considering the total dataset of 61 172 patients with 10 new PAD patients and all others assumed to be free of PAD, the original algorithm has an estimated AUC-ROC of 0.99 (95% CI 0.99–0.99). However, when the original algorithm (version 1) is applied to the subset of high-risk patients, the AUC-ROC is 0.58 (95% CI 0.39–0.78). In version 2 of the algorithm, category weights were adjusted as follows based on the Ridge regression coefficients: 'GI-complaints' was reduced to 0.5, 'Antibiotic prescriptions,' 'RTIs,' and 'Malignancy, lymphoproliferative- and other symptoms' remained 1, and 'Auto-immune symptoms,' 'GP-visits,' and 'Other infections' was increased to 2. This did not improve algorithm performance; the AUC-ROC remained 0.58. The weights per item group and per category in version 3 of the algorithm, based on the Ridge regression coefficients, are shown in Table E7. This version improved the AUC-ROC to 0.80. In version 4 only the variables 'algorithm score version 3' and 'EWS-score' were retained, but this did not improve the AUC compared to version 3 (AUC-ROC 0.80).

Based on these results, we suggest a two-step PAD-screening approach using algorithm versions 1 and 3, see Figure 2. First, version 1 can be applied to a primary care database to distinguish low-risk from high-risk patients using a cut-off of ≥ 17. This cut-off is

based on the lowest score of the identified PAD- and inconclusive patients of 18, maintaining a safety margin of 1 point. Within this subset of high-risk patients, algorithm version 3 can be applied to improve the distinction between PAD and non-PAD patients, using a cut-off of ≥ 15.5 based on Youden's index (sensitivity 90%, specificity 68%). A cut-off of ≥ 9.5 points based on a 100% sensitivity may also be considered, although this greatly reduced the specificity to 9%. For the remaining patients, the EHR should be screened to confirm the absence of exclusion criteria. Patients satisfying in-/exclusion criteria can then be invited for immunoglobulin analysis by their GP. Referral to an immunologist can be advised for patients with reduced immunoglobulins, with the exception of isolated reduced IgG4 as this has little clinical relevance[26]. Referral can also be advised for patients with a high algorithm version 3 score of ≥ 21.5 (based on specificity ~90%) to account for SpAD, as this can present without reduced immunoglobulin levels.

Table IV. Serum measurements of immunoglobulins (gram/litre)

	PAD-diagnosis (n=10)	**Unlikely PAD† (n=78)**	**Inconclusive ‡ (n=16)**	**Sig.**
IgM median (IQR)	0.52 (0.40-0.87)	0.98 (0.74-1.38)	0.77 (0.43-1.22)	*/**
IgM n (%) reduced	3 (30.0%)	5 (6.4%)	4 (25.0%)	*/**
IgA median (IQR)	2.07 (1.52-2.32)	2.09 (1.38-2.84)	2.20 (1.68-2.97)	
IgA n (%) reduced	0 (0.0%)	2 (2.6%)	0 (0.0%)	
IgG total median (IQR)	7.28 (6.12-8.52)	9.89 (8.24-11.18)	9.77 (8.73-11.38)	*/**
IgG total n (%) reduced	5 (50.0%)	3 (3.8%)	1 (6.3%)	*/**
IgG1 median (IQR)	5.00 (4.43-6.90)	6.80 (5.90-7.78)	6.75 (6.05-7.85)	*/**
IgG1 n (%) reduced	5 (50.0%)	4 (5.1%)	2 (12.5%)	*/**
IgG2 median (IQR)	1.61 (1.18-2.08)	2.60 (1.90-3.20)	2.40 (1.43-2.98)	*/**
IgG2 n (%) reduced	5 (50.0%)	4 (5.1%)	5(31.3%)	*/**
IgG3 median (IQR)	0.22 (0.16-0.28)	0.22 (0.16-0.28)	0.38 (0.32-0.49)	*/**
IgG3 n (%) reduced	4 (40.0%)	5 (6.5%)	4(25.0%)	*/**
IgG4 median (IQR)	0.24 (0.15-0.34)	0.33 (0.14-0.52)	0.50 (0.30-1.04)	
IgG4 n (%) reduced	1 (10.0%)	9 (11.5%)	2 (12.5%)	
Calculated globulin median (IQR)	27.0 (25.0-28.0)	30.0 (28.0-33.0)	32.0 (28.8-34.3)	*/**
Calculated globulin n (%) reduced	0 (0.0%)	0 (0.0%)	0 (0.0%)	

*Ig, immunoglobulin; IQR, interquartile range; PAD, primary antibody deficiency; Sig., significance; yrs, years. Values expressed as gram/litre. * overall statistically significant difference between the 3 groups of patients, ** statistically significant difference between 'PAD-diagnosis' and 'Unlikely PAD' groups. † Included patients who did not receive an advice for referral for suspected PAD, or who were not referred by their GP based on valid reasons (Table E4). Referral was advised based on reduced immunoglobulin results or if patients were within the top 10% of highest scoring patients on the algorithm. ‡Diagnosis was inconclusive if 1) referral was advised,*

but the patient was not referred based on invalid reasons (Table E4) or 2) the patient was referred, but no definite diagnosis could be made (see Table E5). Values were considered reduced when: IgM < 0.4(< 0.28 age 12–16 yrs), IgA < 0.7, IgG total < 7(< 5.2 age 12–16 yrs), IgG1 < 4.9 (< 3.7 age 12–16 yrs), IgG2 < 1.5 (< 1.06 age 12–18 yrs), IgG3 < 0.20 (< 0.18 age 12–18 yrs), IgG4< 0.08 (< 0.035 age 12–18 yrs), calculated globulin < 18.

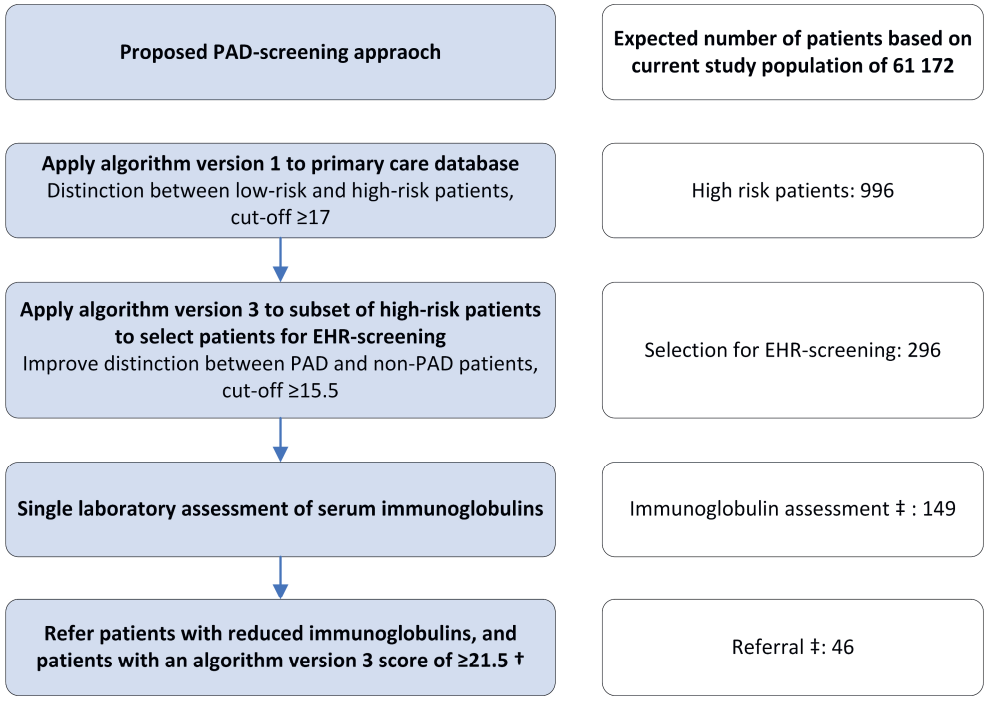
Table V. Predictive performance of different versions of the algorithm

	Version 1	**Version 2**	**Version 3 ‡**	**Version 4 ‡**
Estimated† AUC-ROC in total population of 61,172 patients	0.99 (95% CI 0.99-0.99)	-	-	-
Within subset of high-risk patients *				
AUC-ROC	0.58 (95% CI 0.39-0.78)	0.58 (95% CI 0.38-0.79)	0.80 (95% CI 0.63-0.97)	0.80 (95% CI 0.63-0.97)
Optimal cut-off based on Youden's index	23.0 Sensitivity: 70% Specificity: 50%	21.0 Sensitivity: 70% Specificity: 53%	15.5 Sensitivity: 90% Specificity: 68%	16.2 Sensitivity: 90% Specificity: 69%
Optimal cut-off based on 100% sensitivity	18.0 Specificity: 8%	16.0 Specificity: 12%	9.5 Specificity: 9%	9.6 Specificity: 9%

*AUC-ROC, area under the receiver-operator curve; CI, confidence interval. † When considering the total dataset of 61 172 patients with 10 new PAD patients identified and all others assumed to be free of disease. ‡ For versions 3 and 4 of the algorithm, a principal component analysis was performed, using all available individual patient data from the Julius General Practitioner Network (n=580). For version 4, a Lasso regression analysis was performed with λ = minimal λ, as no variables were retained in the model with λ = 1SE. * 88 confirmed PAD/unlikely PAD cases (i.e. 104 included minus 16 inconclusive patients).*

We estimated the costs for the proposed screening approach based on our population of 61 172 patients. See Figure 2 for the expected numbers of patients per step of the proposed screening approach. The total estimated cost for this screening approach is €52 586.68, assuming that all invited patients participate. These costs include manual EHR-screening for 296 patients, immunoglobulin assessment for 149 patients, and 2 visits to an academic hospital including additional laboratory assessments for 46 patients, see Table E8 for details. Based on the 10 patients we identified in this study, the estimated cost per detected patient is €5258.66. Initial costs for development of a certified digital screening tool were not taken into account. Potentially, the costs for EHR-screening could be reduced by automatizing the process using text-mining techniques.

Figure 2. Proposed PAD-screening approach



EHR, electronic health record; PAD, primary antibody deficiency; SpAD, specific antibody deficiency. †Reduced immunoglobulins, with the exception of isolated reduced IgG4 as this has little clinical relevance[26]. Refer patients with a high algorithm version 3 score of ≥ 21.5 (specificity ~90%) to account for SpAD. ‡ Based on extrapolation from our current study data using multiple imputation

Discussion

In the current study we aimed to detect PAD patients using a screening algorithm in primary care. From a population of 61 172 patients, we included 104 high-risk patients for laboratory analysis, of whom finally 10 PAD patients were identified. A priori, we expected 2-36 PAD patients to be present in our total study population of 61 172, based on the prevalence of PAD in the general population.[4-6] As we identified a total of 18 PAD-patients (i.e. 10 new diagnoses and 8 previous diagnoses), this is well within the range of expected PAD-patients. Our original screening algorithm (version 1) performed well when distinguishing low-risk from high-risk patients, as the PAD-prevalence in included patients was 1:10, compared to 1:1700–1:25 000 in the general population. Within the subset of high-risk patients, the original algorithm did not perform well (AUC-ROC 0.58), but predictive performance improved with an optimized version of the algorithm (AUC-ROC 0.80). We therefore propose a screening approach combining these two algorithms. To our knowledge, this is the first study to identify new PAD-

patients within a primary care adolescent/adult population using a screening algorithm.

The 10 PAD-patients identified in this study had either an isolated IgG-subclass deficiency or SpAD. Similar to Rider et al., we did not encounter selective IgA-deficiencies, which is a more prevalent but also milder type of PAD.[29] Possibly IgA-deficient patients were not classified as high-risk because the majority is asymptomatic.[30] On the other hand, we also did not encounter more severe diagnoses such as CVID, which may explain the absence of patients with reduced calculated globulin levels.[25,31] As CVID has a relatively low prevalence, cases may be encountered when applying the algorithm to a larger population. In addition, PAD cannot be excluded for high-risk patients who did not respond to the invitation from their GP nor for patients for whom referral advice was ignored. It is therefore possible that undetected (more severe) PAD-cases are present in our included study population.

In our screening approach we aimed to account for SpAD by advising referral for patients with normal serum immunoglobulins, but with very high algorithm scores. This is however based on assumptions, as vaccination responses were not performed for all patients with normal immunoglobulins. Further limitations include that we initially intended to screen the 400 highest screening EHRs, but due to a data-extraction error lower ranking EHRs were also screened. An unintended benefit of this occurrence was that it allowed to study patients with a wider range of algorithm scores. As three newly identified PAD-patients had a rank lower than 400, we can conclude that our initially estimated cut-off point was too strict. This would have remained unnoticed if we had only screened the top ranking 400 patients. Lastly, as our optimized algorithm (version 3) was developed in a relatively small dataset of high-risk patients, it should be validated in another population.

As for all screening methods, cost-effectiveness should be taken into account before considering implementation. Our initial rough estimate of the costs per detected PAD-patient with our proposed screening approach is €5258. This is comparable to the costs per detected patient for other screening programs in the Netherlands, such as for breast cancer (€9300), colon cancer (€5000), and cervical cancer (€8400).[32-34] Of note, these screening programs include more invasive interventions such as colonoscopy, compared to the laboratory assessment in our approach. The estimated annual cost savings for early PID diagnosis are $85 882 (≈€81 157) for patients without immunoglobulin replacement therapy, and between $6500–$55 882 (≈€6066–€52 158) for patients with this therapy.[35-37] When assuming that our approach would reduce the diagnostic delay of PAD with 3 years (median diagnostic delay 2–10 years[7-12]), this would imply a cost-saving of €18 198–€243 471 over 3 years per detected PAD-

patient. This seems well proportionate to the expected cost per detected PAD-patient of €5 258. It is however important to note that the estimations for cost-savings are based on PID patients in general rather than specifically PAD, and that these studies were performed in other countries. A full cost-effectiveness analyses specific for our PAD-screening approach would thus be of interest.

Other efforts to reduce the diagnostic delay of immunodeficiencies include the approaches developed by Rider et al. and Mayampurath et al., which could be complementary to the screening algorithm described in the current study.[29,38-40] For example, these approaches focus mainly on secondary/tertiary care and use ICD diagnostic codes, while our approach specifically focusses on primary care and incorporates ICPC codes. Considering that diagnostic coding practices differ per country (ICPC being used in 27 countries), it is important to develop screening tools for both systems.[41] In addition, the approach by Rider et al. is based on paediatric data, while we specifically focus on PAD in an adolescent/adult primary care population aged 12–70 years. We chose to target this population as different cut-offs are to be expected in a paediatric population due to the higher frequency of RTIs and because most PADs present between the 2nd–4th decade of life.[8,20-23] X-linked agammaglobulinemia (XLA) is an exception, as this presents during the first few years of life. However, the diagnostic delay of XLA is limited (e.g. reported median of 1 year compared to 7.5 years for PAD in general[8]), and the proposed addition of XLA to newborn screening is likely a more effective diagnostic strategy for this particular PAD. [42,43]

In conclusion, in the current study we were able to identify 10 new PAD-patients from a primary care population of 61 172 patients using a PAD-screening algorithm. We also present an optimized screening approach including a revised algorithm to improve predictive performance within high-risk patients. This approach may aid in the prevention of morbidity and mortality by reducing diagnostic delay of PAD, and appears to be cost-effective based on a limited analysis. Future studies should address further validation of the proposed screening approach in other populations and a full cost-effectiveness analysis.

References

1. Wood PM. Primary antibody deficiency syndromes. Curr Opin Hematol 2010;17:356-61.
2. Durandy A, Kracker S, Fischer A. Primary antibody deficiencies. Nat Rev Immunol 2013;13:519-33.
3. Fried AJ, Bonilla FA. Pathogenesis, diagnosis, and management of primary antibody deficiencies and infections. Clin Microbiol Rev 2009;22:396-414.
4. Boyle JM, Buckley RH. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. J Clin Immunol 2007;27:497-502.
5. Bousfiha AA, Jeddane L, Ailal F, Benhsaien I, Mahlaoui N, Casanova JL, et al. Primary immunodeficiency diseases worldwide: more common than generally thought. J Clin Immunol 2013;33:1-7.
6. Wood P, Stanworth S, Burton J, Jones A, Peckham DG, Green T, et al. Recognition, clinical diagnosis and management of patients with primary antibody deficiencies: a systematic review. 2007;149:410-23.
7. Seymour B, Miles J, Haeney M. Primary antibody deficiency and diagnostic delay. J Clin Pathol 2005;58:546-7.
8. Slade CA, Bosco JJ, Giang TB, Kruse E, Stirling RG, Cameron PU, et al. Delayed diagnosis and complications of predominantly antibody deficiencies in a cohort of Australian adults. Front Immunol 2018;14:9:694.
9. Aghamohammadi A, Bahrami A, Mamishi S, Mohammadi B, Abolhassani H, Parvaneh N, et al. Impact of delayed diagnosis in children with primary antibody deficiencies. J Microbiol Immunol Infect 2011;44:229-34.
10. Maarschalk-Ellerbroek LJ, Hoepelman AIM, Van Montfrans JM, Ellerbroek PM. The spectrum of disease manifestations in patients with common variable immunodeficiency disorders and partial antibody deficiency in a university hospital. J Clin Immunol 2012;32:901-21.
11. El-Helou SM, Biegner AK, Bode S, Ehl SR, Heeg M, Maccari ME, et al. The German National Registry of Primary Immunodeficiencies (2012-2017). Front Immunol 2019;10:1272.
12. Odnoletkova I, Kindle G, Quinti I, Grimbacher B, Knerr V, Gathmann B, et al. The burden of common variable immunodeficiency disorders: a retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. Orphanet J Rare Dis 2018;13:201.
13. Aghamohammadi A, Pouladi N, Parvaneh N, Yeganeh M, Movahedi M, Gharagolou M, et al. Mortality and morbidity in common variable immunodeficiency. J Trop Pediatr 2007;53:32-8.
14. Barlogis V, Mahlaoui N, Auquier P, Pellier I, Fouyssac F, Vercasson C, et al. Physical health conditions and quality of life in adults with primary immunodeficiency diagnosed during childhood: A French Reference Center for PIDs (CEREDIH) study. J Allergy Clin Immunol 2017;139:1275-81.e7.
15. Modell V, Gee B, Lewis DB, Orange JS, Roifman CM, Routes JM, et al. Global study of primary immunodeficiency diseases (PI)-diagnosis, treatment, and economic impact: an updated report from the Jeffrey Modell Foundation. Immunol Res 2011;51:61-70.
16. Messelink M.A. BRM, van Montfrans J.M., Ellerbroek, P.M., Gladiator A., Welsing P.M.J., Leavis H.L. Development of a primary care screening algorithm for the early detection of patients at risk of primary antibody deficiency [Manuscript submitted for publication] Department of Rheumatology & Clinical Immunology, University Medical Centre Utrecht 2022.
17. Smeets HM, Kortekaas MF, Rutten FH, Bots ML, van der Kraan W, Daggelders G, et al. Routine primary care data for scientific research, quality of care programs and educational purposes: the Julius General Practitioners' Network (JGPN). BMC Health Serv Res 2018;18:735.
18. Cunningham-Rundles C, Sidi P, Estrella L, Doucette J. Identifying undiagnosed primary immunodeficiency diseases in minority subjects by using computer sorting of diagnosis codes. J Allergy Clin Immunol 2004;113:747-55.
19. Brodszki N, Jönsson G, Skattum L, Truedsson L. Primary immunodeficiency in infection-prone children in southern Sweden: Occurrence, clinical characteristics and immunological findings. BMC Immunol 2014:15:31.
20. Cunningham-Rundles C. The many faces of common variable immunodeficiency. Hematology Am Soc Hematol Educ Program 2012;2012:301-5.

21. Wiesik-Szewczyk E, Jahnz-Rozyk K. From infections to autoimmunity: diagnostic challenges in common variable immunodeficiency. World J Clin Cases 2020;8:3942-55.
22. Song J, Lleo A, Yang GX, Zhang W, Bowlus CL, Gershwin ME, et al. Common Variable Immunodeficiency and Liver Involvement. Clin Rev Allergy Immunol 2018;55:340-51.
23. Bjelac JA, Yonkof JR, Fernandez J. Differing Performance of the Warning Signs for Immunodeficiency in the Diagnosis of Pediatric Versus Adult Patients in a Two-Center Tertiary Referral Population. J Clin Immunol 2019;39:90-8.
24. McCusker C, Upton J, Warrington R. Primary immunodeficiency. Allergy Asthma Clin Immunol 2018;14:61.
25. Jolles S, Borrell R, Zouwail S, Heaps A, Sharp H, Moody M, et al. Calculated globulin (CG) as a screening test for antibody deficiency. Clin Exp Immunol 2014;177:671-8.
26. Ochs HD SE, Winkelstein JA, et al. Antibody deficiencies. In: Immunologic disorders in infants and children. In: Ochs HD SE, Winkelstein JA. , editor. 5th edition ed. Philadelphia: Elsevier; 2004.
27. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol 2020;40:24-64.
28. Bousfiha A, Jeddane L, Picard C, Al-Herz W, Ailal F, Chatila T, et al. Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification. J Clin Immunol 2020;40:66-81.
29. Rider NL, Coffey M, Kurian A, Quinn J, Orange JS, Modell V, et al. A validated artificial intelligence-based pipeline for population-wide primary immunodeficiency screening. J Allergy Clin Immunol 2023;151:272-9.
30. Luca L, Beuvon C, Puyade M, Roblot P, Martin M. Selective IgA deficiency. Rev Med Interne 2021;42:764-71.
31. Cunningham-Rundles C. Common variable immunodeficiency. Curr Allergy Asthma Rep 2001;1:421-9.
32. Bevolkingsonderzoek Borstkanker Facsheet 2021 Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 2022 [updated 02-2022]. Available from: https://www.rivm.nl/documenten/factsheet-borstkanker.
33. Bevolkingsonderzoek Darmkanker Factsheet 2021 Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 2022 [updated 02-2022]. Available from: https://www.rivm.nl/sites/default/files/2022-02/22400658_013665_FS%20Bevolkingsonderzoek%20Darmkanker%202021_V3_TG.pdf.
34. Bevolkingsonderzoek Baarmoederhalskanker Factsheet 2021 Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 2022 [updated 02-2022]. Available from: https://www.rivm.nl/documenten/factsheet-baarmoederhalskanker.
35. Modell V, Orange JS, Quinn J, Modell F. Global report on primary immunodeficiencies: 2018 update from the Jeffrey Modell Centers Network on disease classification, regional trends, treatment modalities, and physician reported outcomes. Immunol Res 2018;66:367-80.
36. Sadeghi B, Abolhassani H, Naseri A, Rezaei N, Aghamohammadi A. Economic burden of common variable immunodeficiency: annual cost of disease. Expert Rev Clin Immunol 2015;11:681-8.
37. Elsink K, van Montfrans JM, van Gijn ME, Blom M, van Hagen PM, Kuijpers TW, et al. Cost and impact of early diagnosis in primary immunodeficiency disease: a literature review. Clin Immunol 2020;213:108359.
38. Rider NL, Cahill G, Motazedi T, Wei L, Kurian A, Noroski LM, et al. PI Prob: a risk prediction and clinical guidance system for evaluating patients with recurrent infections. PLoS One 2021;16:e0237285.
39. Rider NL, Miao D, Dodds M, Modell V, Modell F, Quinn J, et al. Calculation of a Primary Immunodeficiency "Risk Vital Sign" via Population-Wide Analysis of Claims Data to Aid in Clinical Decision Support. Front Pediatr 2019;7:70.
40. Mayampurath A, Ajith A, Anderson-Smits C, Chang SC, Brouwer E, Johnson J, et al. Early Diagnosis of Primary Immunodeficiency Disease Using Clinical Data and Machine Learning. J Allergy Clin Immunol Pract 2022;10:3002-7.e5.

41. Basílio N, Ramos C, Figueira S, Pinto D. Worldwide Usage of International Classification of Primary Care. Revista Brasileira de Medicina de Família e Comunidade 2016;11:1-9.
42. Lackey AE, Ahmad F. X-linked Agammaglobulinemia. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022.
43. Blom M, Bredius RGM, van der Burg M. Future Perspectives of Newborn Screening for Inborn Errors of Immunity. Int J Neonatal Screen 2021;7.
44. Orange JS, Ballow M, Stiehm ER, Ballas ZK, Chinen J, De La Morena M, et al. Use and interpretation of diagnostic vaccination in primary immunodeficiency: a working group report of the Basic and Clinical Immunology Interest Section of the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol 2012;130:S1-24.
45. L. Hakkaart-van Roijen NvdL, C. Bouwmans, T. Kanters, S. Swan Tan. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam; 2015.
46. Centraal Bureau voor de Statistiek (CBS) / Statistics Netherlands. Consumer Price Index. [updated 14-03-2023]. Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83131NED/table?ts=1679511519388.
47. Zorginstituut Nederland. Farmacotherapeutisch Kompas. [Available from: www.farmacotherapeutischkompas.nl.

Supplemental Material

Table E1. Screening algorithm for early detection of primary antibody deficiencies in a primary care setting (original version): See pages 74-78 of chapter 4 of this thesis.

Table E2. TRIPOD checklist for prediction model development and validation

Section/ Topic	Checklist Item	Section(s)
Title	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Title page
Abstract	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Abstract
Background and objectives	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Introduction, Discussion
	Specify the objectives, including whether the study describes the development or validation of the model or both.	Introduction
Source of data	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	Methods
	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Methods
Participants	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	Methods
	Describe eligibility criteria for participants.	Methods
	Give details of treatments received, if relevant.	NA
Outcome	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Methods
	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Methods, Tables E1 and E7
	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	Explain how the study size was arrived at.	Methods
Missing data	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	†
Statistical analysis methods	Describe how predictors were handled in the analyses.	Methods, Results, Tables E1 and E7
	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	Methods, Results, Tables E1 and E7
	For validation, describe how the predictions were calculated.	Methods
	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Methods, Results
	Describe any model updating (e.g., recalibration) arising from the validation, if done.	Methods, Results, Table E7
Risk groups	Provide details on how risk groups were created, if done.	NA

Development vs. validation	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Methods, ‡
Participants	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Tables II and III, †
	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	‡
Model development	Specify the number of participants and outcome events in each analysis.	Results, Figure 1
	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Tables E1 and E7
	Explain how to the use the prediction model.	Methods, Results, Tables E1 and E7
Model performance	Report performance measures (with CIs) for the prediction model.	Results, Table V
Model-updating	If done, report the results from any model updating (i.e., model specification, model performance).	Results, Table E7
Limitations	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	Discussion
Interpretation	For validation, discuss the results with reference to performance in the development data, and any other validation data.	‡
	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Discussion
Implications	Discuss the potential clinical use of the model and implications for future research.	Discussion
Supplementary information	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Availability of data
Funding	Give the source of funding and the role of the funders for the present study.	Title page

D, development; NA, not applicable; V, validation. † As our algorithm was based on primary care electronic health record data on registered diagnostic codes, medication prescriptions, visits to the general practitioner and laboratory values, there were no missing values other than incomplete registration, which could not be verified. For the cost-effectiveness analysis, multiple imputation was performed in order to extrapolate the results from our current study implementing algorithm version 1, to estimate the impact of a proposed screening method using algorithm version 1 and subsequently version 3. This is described in the results section. ‡ The original model (version 1) was developed based on aggregate subgroup primary care data and on expert opinion. Therefore no comparison on individual patient-level data could be made. Messelink M.A. BRM, van Montfrans J.M., Ellerbroek, P.M., Gladiator A., Welsing P.M.J., Leavis H.L. Development of a primary care screening algorithm for the early detection of patients at risk of primary antibody deficiency Currently Under Submission. 2022

Table E3. Included patients with a PAD-diagnosis

Rank	Serum immunoglobulin results (Presented as a range in g/L due to privacy considerations)	Clinical presentation & treatment	Vaccine responses	Diagnosis
9	**Total IgG 6-7**, **IgG1 4-5**, **IgG2 0.5-1.5**, IgG3 0.3-0.4, IgG4 0.2-0.3, IgM 0.4-0.5, IgA 1.0-2.0	Recurrent RTIs, treated with prophylactic AB and additional AB on demand for breakthrough infections. Medical history: asthma	Borderline normal pneumococcal response in 2017 (75%)	Isolated IgG subclass deficiency
42	Total IgG 8-9, IgG1 7-8, IgG2 2-3, **IgG3 0.1-0.2**, IgG4 0.3-0.4, IgM 0.9-1.0, IgA 2-3	Recurrent RTIs, treated with prophylactic AB and additional AB on demand for breakthrough infections. Medical history: asthma	Insufficient pneumococcal vaccine response (30%)	Isolated IgG Subclass deficiency
98	**Total IgG 6-7, IgG1 4-5,** IgG2 1-2, **IgG3 0.1-0.2**, IgG4 0.2-0.3, IgM 0.9-0.1, IgA 1-2	Recurrent RTIs, otitis externa, recurrent fungal infections. Medical history: asthma	Normal (100%)	Isolated IgG subclass deficiency
191	Total IgG 7-8, IgG1 5-6, **IgG2 1-2**, **IgG3 0.1-0.2**, IgG4 1-2, IgM 1-2, IgA 1-2. At referral: **total IgG 6-7, IgG1 3.5-4.5, IgG2 0.5-1.5, IgG3 0.1-0.2**	Recurrent RTIs. Advice AB use on demand in case of RTIs, 5-yearly pneumococcal vaccine and IgG check in 1 year Medical history: COPD	Borderline normal pneumococcal response (75%)	Isolated IgG subclass deficiency
228	Total IgG 9-10, IgG1 7-8, **IgG2 0.5-1.5**, IgG3 0.2-0.3, IgG4 0.2-0.3, IgM 0.7-0.8, IgA 2-3	Chronic otitis media. Advice local treatment and AB on demand for other RTIs.	Normal (90%)	Isolated IgG subclass deficiency
279	Total IgG 13-14, IgG1 10-11, IgG2 2-3, **IgG3 0.1-0.2**, IgG4 1-2, IgM 0.4-0.5, IgA 3-4	Recurrent RTIs. Advice AB on demand. Medical history: asthma	Insufficient pneumococcal vaccine response (65%)	Isolated IgG subclass deficiency
317	**Total IgG 6-7, IgG1 3-4,** IgG2 2-3, IgG3 0.2-0.3, IgG4 0.1-0.2, **IgM 0.3-0.4**, IgA 1-2	Recurrent RTIs. Treatment by discretion of GP. Medical history: colitis	Insufficient pneumococcal vaccine response (50%)	Isolated IgG subclass deficiency
528	**Total IgG 4-5, IgG1 3-4, IgG2 0.5-1.5**, IgG3 0.2-0.3, IgG4 0.1-0.2, IgM 0.5-0.6, IgA 2-3	Recurrent RTIs. Start maintenance AB, consider immunoglobulin replacement therapy in case of insufficient infection control. Medical history: COPD	Insufficient pneumococcal vaccine response (20%)	Isolated IgG subclass deficiency
657	Total IgG 7-8, IgG1 5-6, IgG2 1.5-2.5, IgG3 0.2-0.3, **IgG4 <0.02, IgM 0.1-0.2,** IgA 2-3	Recurrent RTIs. Start prophylactic ABs. Medical history: IBS	Insufficient pneumococcal vaccine response (35%)	Specific Antibody Deficiency
791	**Total IgG 6-7, IgG1 3.5-4.5, IgG2 0.5-1.5,** IgG3 0.4-0.5, IgG4 0.1-0.2, **IgM 0.3-0.4,** IgA 1-2	AB on demand, monitoring of immunoglobulins Medical history: COPD	Insufficient pneumococcal vaccine response (40%)	Isolated IgG subclass deficiency

AB, antibiotics; COPD, chronic obstructive pulmonary disease; IBS, irritable bowel syndrome; Ig, immunoglobulin; RTI, respiratory tract infection. Immunoglobulins are shown as a range in gram/liter due to privacy regulations. Reduced values are depicted in bold. Pneumococcal vaccine response is based on comparison

of pneumococcal antibody response against different serotypes (before and) after pneumovax23 (PPV23) vaccination, adequate rise in case of titres 1,0 ug/mL for ≥ 70% of serotypes tested.[44] Immunoglobulins were considered reduced when: IgM < 0.4(< 0.28 age 12–16 yrs), IgA < 0.7, IgG total < 7(< 5.2 age 12–16 yrs), IgG1 < 4.9 (< 3.7 age 12–16 yrs), IgG2 < 1.5 (< 1.06 age 12–18 yrs), IgG3 < 0.20 (< 0.18 age 12–18 yrs), IgG4< 0.08 (< 0.035 age 12–18 yrs), calculated globulin < 18.

Table E4. Non-referred patients who had an advice for referral

Rank	Reason referral advice (Immunoglobulins presented as a range due to privacy considerations)	Reason non-referral	(In)valid Non-referral
3	Top 10%	Referral not deemed relevant by GP, antibiotics are prescribed by dermatologist due to rosacea, GI-complaints due to IBS and recent cholecystectomy	Valid
6	Top 10%, IgG2 <0.2 g/l and IgG4 <0.01	Referral not deemed relevant by GP as patient has a COPD diagnosis	Invalid
22	Top 10%	Referral not deemed relevant by GP. Patient used maintenance AB for asthma, which has been terminated since half a year.	Invalid
35	Top 10%	Referral declined by patient, too much effort	Invalid †
88	IgG2 0.5-1.5 g/l	Referral not deemed relevant by GP as patient only has recurrent urinary tract infections, but no other infections	Valid
128	IgG3 0.1-0.2 g/l	Referral declined by patient, too much effort	Invalid †
134	IgM 0.3-0.4 g/l	Not deemed relevant by GP as IgM was previously reduced when measured by another specialist, recurrent RTIs present	Invalid
146	IgG2 0.5-1.5 g/l, IgG3 0.1-0.2 g/l	Referral not deemed relevant by GP as patient has around 6 infections a year, not deemed enough to label as recurrent infections	Invalid
226	Total IgG 5-6 g/l, IgM 0.3-0.4 g/l, IgG1 4-5 g/l, IgG2 0.5-1.5 g/l, IgG3 0.2-0.3 g/l	Referral declined by patient, too much effort	Invalid †
289	IgM 0.3-0.4 g/l	Referral declined by patient, too much effort	Invalid †
317	Total IgG 6-7 g/l, IgM 0.3-0.4 g/l, IgG1 3-4 g/l	Patient previously referred for recurrent infections, retrospectively PAD diagnosis can be made based on abnormal vaccine responses, hypogammaglobulinemia and recurrent infections.	Valid
360	IgG3 0.1-0.2 g/l	No referral made due to change in GPs during study resulting in miscommunication. Possible future referral would be after study closure.	Invalid
430	IgG3 <0.03 g/l	Referral declined by patient, already too many hospital visits	Invalid †
458	IgA 0.5-0.6 g/l	Referral not deemed relevant by GP as there are no ongoing recurrent infections and laboratory aberrances are mild	Valid
652	Total IgG 6-7 g/l, IgM 0.3-0.4 g/l, IgG2 0.5-1.5 g/l	Patient and GP together decided there was no indication for referral, as there were no ongoing recurrent infections	Valid
1050	IgG3 0.1-0.2 g/l, IgG4 <0.04 g/l	Referral not deemed relevant by GP as patient does not have ongoing recurrent infections, currently only psoriasis complaints	Valid

COPD, chronic obstructive pulmonary disease; GI, gastro-intestinal, g/l: gram/liter, GP, general practitioner; IBS, irritable bowel syndrome; Ig, immunoglobulin; PAD, primary antibody deficiency. † Labeled as invalid as PAD cannot be excluded for these patients. Immunoglobulins are shown as a range in gram/liter due to privacy regulations.

Table E5. Overview of referred patients where PAD could neither be confirmed nor excluded

Rank	**Reason referral advice** (Immunoglobulins presented as a range due to privacy considerations)	**Outcome referral**	**Reason inconclusive diagnosis**
7	Top 10%	Patient with recurrent infections and COPD diagnosis, for which prophylactic AB. Normal immunoglobulins. Vaccine responses were not assessed.	Specific Antibody Deficiency cannot be excluded
24	Top 10%	Patient with recurrent infections and COPD diagnosis, for which prophylactic AB. Despite pneumococcal vaccination PVR23 response not assessed. Normal immunoglobulins.	Specific Antibody Deficiency cannot be excluded
77	IgG1 4-5 g/l	Patient with recurrent infections, bronchiectasis and COPD diagnosis, for which maintenance antibiotic therapy. Igs upon repeated measurement: IgG1 3.6 g/l, IgM 0.39 g/l. Normal vaccine responses.	Isolated IgG subclass deficiency cannot be excluded due to use of immunosuppressants
107	IgG2 0.5-1.5 g/l	Patient with recurrent infections and COPD diagnosis, for which maintenance antibiotic therapy. Igs upon repeated measurement IgG total 6.36 g/l, IgG2 1.31 g/l. Vaccine responses were not performed.	Isolated IgG subclass deficiency cannot be excluded due to prednisone use
227	IgG2 0.5-1.5 g/l, IgG4 <0.03 g/l	Patient with recent pregnancy (discovered post-inclusion), therefore no vaccine responses were performed. Reduced Igs may be physiological. Follow-up after pregnancy.	PAD cannot be excluded, follow-up after study closure
803	IgM 0.3-0.4 g/l	Late referral, vaccine responses have yet to be performed after study closure	Specific antibody deficiency cannot be excluded, follow-up after study closure

AB, antibiotics; COPD, chronic obstructive pulmonary disease; g/l, gram/litre; Ig, immunoglobulin; PAD, primary antibody deficiency.

TABLE E6. Comparison of EHRs that were used for recruitment of patients, and those that were within the top 400 but were not used for screening and recruitment of patients

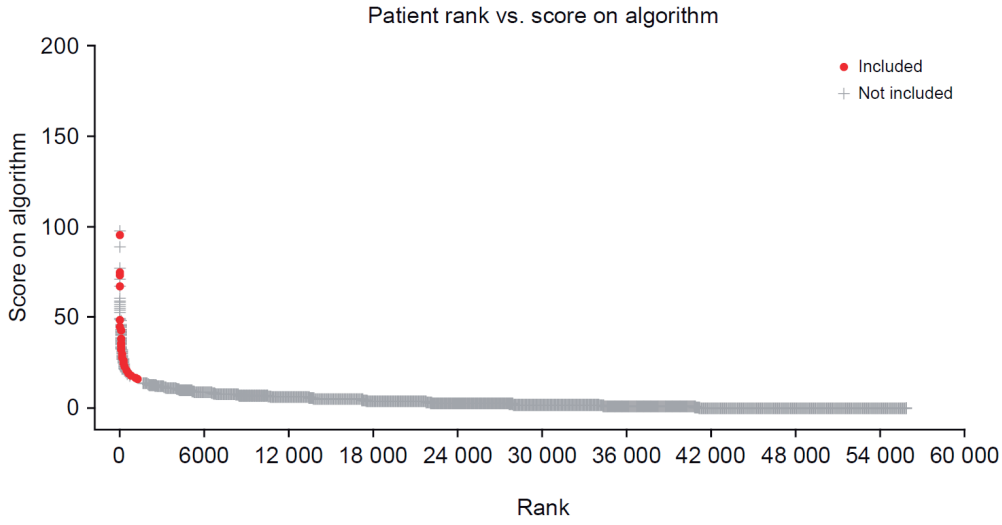
	EHRs within top 400, used for patient recruitment	**EHRs outside of top 400, used for patient recruitment**	**EHRs within top 400, *not* used for patient recruitment**	**P-value**
Number of EHRs	220	180	180	NA
Excluded records during remote screening, number (%)	38 (17.3%)	19 (10.6%)	18 (10.0%)	0.051
Nephrotic syndrome / stage 3-4 liver cirrhosis	*4 (1.8%)*	*2 (1.1%)*	*2 (1.1%)*	
Current chemotherapy	*2 (0.9%)*	*1 (0.6%)*	*2 (1.1%)*	
Current pregnancy	*2 (0.9%)*	*1 (0.6%)*	*1 (0.6%)*	
Short life expectancy/ deceased/moved away	*4 (1.8%)*	*2 (1.1%)*	*3 (1.7%)*	
Severe alcohol/drug abuse	*16 (7.3%)*	*10 (5.6%)*	*5 (2.8%)*	1.000
Already treated by immunologist	*2 (0.9%)*	*0 (0.0%)*	*2 (1.1%)*	
Anorexia/bulimia	*1 (0.5%)*	*1 (0.6%)*	*2 (1.1%)*	
Post-transplantation	*5 (2.3%)*	*1 (0.6%)*	*1 (0.6%)*	
Leukemia / multiple myeloma	*2 (0.9%)*	*1 (0.6%)*	*0 (0.0%)*	
Excluded records during screening at GP, number (%)	42 (19.1%)	40 (22.2%)	*NA*	0.440
Short life expectancy/ deceased/moved away	*7 (3.2%)*	*2 (1.1%)*	*NA*	
Nephrotic syndrome	*3 (1.4%)*	*2 (1.1%)*	*NA*	
Deemed unsuitable by GP	*25 (11.4%)*	*32 (17.8%)*	*NA*	
Current pregnancy	*1 (0.5%)*	*2 (1.1%)*	*NA*	0.962
Multiple myeloma	*1 (0.5%)*	*0 (0.0%)*	*NA*	
Already treated by immunologist	*2 (0.9%)*	*1 (0.6%)*	*NA*	
Cystic fibrosis	*1 (0.5%)*	*0 (0.0%)*	*NA*	
Stage 3-4 liver cirrhosis	*1 (0.5%)*	*1 (0.6%)*	*NA*	
Age, mean (SD)	54.0 (14.0)	50.1 (15.6)	40.4 (16.3)	<0.001
Female, n(%)	164 (74.5)	130 (72.2)	127 (70.6)	0.667
Score on algorithm 1, median (IQR)	28.1 (24.0-35.8)	18.3 (17.0-20.0)	24.9 (21.0-27.9)	<0.001
Score antibiotics, median (IQR)	12.5 (8.0-23.0)	3.0 (2.0-6.0)	16.1 (12.0-20.2)	<0.001
Score ICPC codes related to RTI, median (IQR)	9.0 (7.0-12.0)	9.0 (6.0-10.0)	5.0 (3.0-7.0)	<0.001
Score ICPC codes related to GI-complaints, median (IQR)	2.0 (0.0-2.0)	2.0 (0.0-2.0)	0.0 (0.0-2.0)	<0.001
Score ICPC codes related to other infections, median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.204

Score ICPC codes related to auto-immune symptoms, median (IQR)	0.0 (0.0-1.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.116
Score ICPC codes related to lymphoproliferative symptoms, median (IQR)	1.0 (0.0-1.0)	1.0 (0.0-1.5)	1.0 (0.0-1.0)	<0.001
Patients with a previously known reduced number of immunoglobulin levels in their EHR, number (%)	2.0 (0.9%)	0.0 (0.0%)	0.0 (0.0%)	0.335
Score GP-visits in the year before the censoring date, median (IQR)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	<0.001

AB, antibiotics; EHR, electronic health record; GI, gastro-intestinal; GP, general practitioner; ICPC, international classification of primary care; IQR, interquartile range; n,number; RTI, respiratory tract infection; SD, standard deviation.

For continuous data, an ANOVA (parametric) or Kruskall-Wallis (non-parametric) test was performed. For categorical outcomes Chi-square tests were performed, unless with small (< 5) expected cell frequencies, in which case Fisher's exact test was performed.

Figure E1. Overview of included patients vs. total population



Total population based on 55 888 patients, i.e. initial study population of 61 172 minus 5284 patients excluded based on registered ICPC codes (International Classification of Primary Care). See also Figure 1 in main manuscript.

Table E7. Screening algorithm for early detection of primary antibody deficiencies in a primary care setting, optimized based on principal component and subsequent ridge regression analyses per dimension and per category

ANTIBIOTICS Sum of prescriptions in the 4 years before the censoring date. If a patient was enrolled in the GP clinic less than 4 years, this was corrected for with the formula: (total number of prescriptions*4)/(number of days enrolled/365). Prescriptions before the age of 6 were not taken into account.		**Coefficient Ridge regression analysis per category** 3.90 e^{-05}	**Weight for category** 1
ATC-code	**Description**	**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
	Dimension 1 (sum of prescriptions in 4 years)	1.53 e^{-04}	2
J01CF05	Flucloxacillin		
J01CR02	Amoxicillin / clavulanic acid		
J01MA02	Ciprofloxacin		
	Dimension 2 (sum of prescriptions in 4 years)	8.24 e^{-05}	1
J01AA02	Doxycycline		
J01FA10	Azithromycin		
	Dimension 3 (sum of prescriptions in 4 years)	-6.27 e^{-06}	0.5
J01CA04	Amoxicillin		
J01EE01	Cotrimoxazole		
S02CA03	Hydrocortisone / colistin / bacitracin ear suspension		
	Dimension 4 (sum of prescriptions in 4 years)	-1.99 e^{-04}	0.5
J01FA09	Clarithromycin		
J01MA12	Levofloxacin		
	Dimension 5 (sum of prescriptions in 4 years)	2.55 e^{-04}	2
J01CE05	Pheneticillin		
J01DC02	Cefuroximaxetil		
J01FA01	Erythromycin		
	Dimension 6 (sum of prescriptions in 4 years)		
J01MA14	Moxifloxacin		
	Dimension 7 (sum of prescriptions in 4 years)	3.83 e^{-04}	2
P01AB01	Metronidazole		
S02AA16	Ofloxacin ear suspension		
	Dimension 8 (sum of prescriptions in 4 years)	-5.17 e^{-04}	0.5
J01CE02	Phenoxymethylpenicillin		
	Dimension 9 (sum of prescriptions in 4 years)	NA*	
J01DB01	Cefalexin		
J01DC04	Cefaclor		
J01DD14	Ceftibuten		
RESPIRATORY TRACT INFECTIONS Score was attributed if the ICPC code was registered in the 10 years before the censoring date.		**Coefficient Ridge regression analysis per category** 5.24 e^{-05}	**Weight for category** 1
ICPC code	**Description**	**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
	Dimension 1 (1 point if 1 ICPC, 2 points if ≥2 ICPCs)	4.56 e^{-05}	1
H01	Ear pain		
H71	Acute otitis media / myringitis		
H74	Chronic otitis media / other ear infections		
R74	Acute upper respiratory tract infection		
	Dimension 2 (1 point if 1 ICPC, 2 points if ≥2 ICPCs)	-2.73 e^{-04}	0.5
R75	Acute / chronic rhinosinusitis		
R75.01	Acute rhinosinusitis		
R75.02	Chronic rhinosinusitis		
R05	Coughing		
	Dimension 3 (1 point if present)	-7.66 e^{-04}	0.5
R76	Acute tonsillitis / peritonsillar abscess		
	Dimension 4 (1 point if ≥1 present)	1.47 e^{-03}	3
R72.01	Streptococcal pharyngitis		

R76.02	Peritonsillar abscess		
R72	Streptococcal pharyngitis / scarlet fever		
R72.02	Scarlet fever		
	Dimension 5 (1 point if ≥1 present)	5.03 e^{-04}	2
H04	Discharge from ear		
H72	Otitis media with effusion		
H74.01	Chronic otitis media		
	Dimension 6 (1 point if ≥1 present)	9.48 e^{-05}	1
R76.01	Acute tonsillitis		
R96	Asthma		
R90	Hypertrophy / chronic infection tonsils / adenoid		
	Dimension 7 (1 point if ≥1 present)	2.99 e^{-04}	2
R81	Pneumonia		
R78	Acute bronchitis / bronchiolitis		
R91	Chronic bronchitis / bronchiectasis		
R91.01	Chronic bronchitis		
	Dimension 8 (1 point if present)	-9.91 e^{-05}	0.5
R74.01	Common cold		
	Dimension 9 (1 point if ≥1 present)	8.87 e^{-04}	2
R09	Symptoms/complaints sinuses		
R73	Furuncle / abscess nose		
R77.01	Subglottic laryngitis / pseudo croup		
	Dimension 10 (1 point if present)	-7.48 e^{-04}	0.5
R91.02	Bronchiectasis		
	Dimension 11 (1 point if ≥1 present)	6.29 e^{-04}	2
R07	Sneezing / nasal congestion / running nose		
R77	Acute laryngitis / tracheitis		
	Dimension 12 (1 point if ≥1 present)	-7.70 e^{-04}	0.5
H74.02	Mastoiditis		
R74.02	Acute pharyngitis		
	Dimension 13 (1 point if present) Code did not occur in dataset of included PAD/unlikely PAD patients, categorized on clinical relevance	NA*	2
R77.02	Acute epiglottitis		
GASTRO-INTESTINAL COMPLAINTS Score was attributed if the ICPC code is registered in the 10 years before the censoring date.		**Coefficient Ridge regression analysis per category** 1.72 e^{-05}	**Weight for category** 1
ICPC code	**Description**	**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
	Dimension 1 (1 point if ≥1 present)	3.01 e^{-04}	2
D94	Ulcerative colitis / Chronic enteritis		
D94.01	Ulcerative colitis		
D11	Diarrhea		
	Dimension 2 (1 point if ≥1 present)	-7.19 e^{-04}	0.5
D86	Other peptic ulcer		
D70.01	Salmonella		
	Dimension 3 (1 point if ≥1 present)	-6.84 e^{-04}	0.5
D73	Presumed gastro-intestinal infection		
D70	Infectious diarrhea, dysentery		
	Dimension 4 (1 point if ≥1 present)	1.50 e^{-05}	1
D93	Inflammatory Bowel Syndrome		
D70.03	Giardia		
	Dimension 5 (1 point if present)	-7.14 e^{-04}	0.5
D70.02	Shigella-/Yersinia-/Campylobacter intestinal infection		
OTHER INFECTIONS Score was attributed if the ICPC code was registered in the EHR at any time-point before the censoring date.		**Coefficient Ridge regression analysis per category** NA*	**Weight for category** 3
ICPC code	**Description**	**Coefficient Ridge regression**	**Weight per dimension**

		analysis per dimension	
	Dimension 1 (1 point if ≥1 present)	NA*	1
L70.01	Osteomyelitis		
L70.02	Septic arthritis		
	Dimension 2 (1 point if ≥1 present)	NA*	1
N71	Meningitis / Encephalitis		
N71.01	Bacterial meningitis		
N71.02	Viral meningitis		
N71.03	Encephalitis		
N71.04	Myelitis		
AUTO-IMMUNE SYMPTOMS Score was attributed if the ICPC code was registered in the EHR at any time-point before the censoring date.		**Coefficient Ridge regression analysis per category** 5.77 e^{-04}	**Weight for category** 3
ICPC code	**Description**	**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
	Dimension 1 (1 point if present)	-8.82 e^{-04}	0.5
L88	Rheumatoid arthritis / related diseases		
	Dimension 2 (1 point if ≥1 present)	NA*	2
B83.02	Idiopathic thrombocytopenic purpura (ITP)		
T99.02	Thyroiditis		
	Dimension 3 (1 point if ≥1 present)	1.09 e^{-03}	1
B81	Pernicious / Folic acid anemia		
B82	Other / Non specified anemia		
T86	Hypothyroidism		
	Dimension 4 (1 point if ≥1 present)	1.04 e^{-03}	1
L88.01	Rheumatoid arthritis		
B83	Purpura / coagulation disorder / aberrant thrombocytes		
D99.06	Coeliac disease		
	Dimension 5 (1 point if ≥1 present)	1.04 e^{-03}	1
B04	Symptoms / complaints blood / blood forming organs		
S23.01	Alopecia areata		
	Dimension 6 (1 point if present)	-8.47 e^{-04}	0.5
D94.02	Crohn's Disease		
	Dimension 7 (1 point if present)	-8.59 e^{-04}	0.5
S99.04	Vitiligo		
N99	Myasthenia Gravis		
	Dimension 8 (1 point if present)	NA*	1
R83.02	Sarcoidosis		
	Dimension 9 (1 point if present)	NA*	0.5
T99.12	Adrenal insufficiency		
MALIGNANCIES, LYMPHOPROLIFERATIVE- AND OTHER SYMPTOMS Score was attributed if the ICPC code was registered in the EHR at any time-point before the censoring date.		**Coefficient Ridge regression analysis per category** 1.46 e^{-04}	**Weight for category** 2
ICPC code	**Description**	**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
	Dimension 1 (1 point if ≥1 present)	6.63 e^{-04}	2
B72	Hodgkin's disease		
B72.01	Hodgkin's disease		
B72.02	Non-Hodgkin lymphoma		
	Dimension 2 (1 point if ≥1 present)	NA*	1
T08	Weight loss		
B02	Lymphadenopathy		
A04	Fatigue / weakness		
	Dimension 3 (1 point if ≥1 present)	-2.70 e^{-05}	0.5
B84	Aberrant leukocytes		
N94	Other peripheral neuritis / neuropathy		

	Dimension 4 (1 point if ≥1 present)	NA*	2
B87	Splenomegaly		
D96	Hepatomegaly		
	Dimension 5 (1 point if present)	NA*	1
T10	Failure to thrive		
	Dimension 6 (1 point if present)		1
D74	Gastric cancer		
LABORATORY VALUES Score was attributed if a reduced lab value was registered in the EHR at any time-point before the censoring date.		**Coefficient Ridge regression analysis per category** NA*	**Weight for category** 2
Reduced laboratory measure		**Coefficient Ridge regression analysis per dimension**	**Weight per dimension**
Dimension 1 (1 point if ≥1 present)		NA*	1
IgG4 < 0.08 g/L			
IgM-total < 0.4 g/L			
IgA-total < 0.7 g/L			
Dimension 2 (1 point if ≥1 present)		NA*	2
IgG – total < 7 g/L			
IgG1 < 4.9 g/L			
IgG2 < 1.5 g/L			
IgG3 < 0.2 g/L			
Calculated globulin (total protein – albumin) < 18g/L			
VISITS TO GENERAL PRACTITIONER >=6 visits to the general practitioner clinic in the year before the censoring date. Both physical and electronic visits were taken into account.		**Coefficient Ridge regression analysis per category** 2.81 e^{-04}	**Weight for category** 2

AMBIGUOUS CODES

For these codes, EHRs were screened up to the moment of the ambiguous diagnosis

ICPC code	Description
B72	Hodgkin's lymphoma
B72.01	Hodgkin's lymphoma
B72.02	Non-Hodgkin lymphoma
A87.02	Post-transplantation
D74	Gastric cancer
D74	Colon or rectal cancer

EXCLUSION CRITERIA

ICPC code / other	Description
B73	Leukemia
B74.01	Multiple Myeloma
B90	HIV-infection
B90.01	HIV seropositive without symptoms
B90.02	AIDS / AIDS-related complex
P15.01	Alcoholism
P15.02	Delirium tremens
P15.03	Wernicke – Korsakoff
P19.03	Addiction to hard drugs
T06	Anorexia nervosa / bulimia
T06.01	Anorexia nervosa
T06.02	Bulimia
T99.01	Immunodeficiency
T99.10	Cystic fibrosis
Age	< 12 years or > 70 years

AIDS, acquired immune deficiency syndrome; ATC, anatomic therapeutic chemical; EHR, electronic health record; HIV, human immunodeficiency virus; ICPC, International Classification of Primary Care; Ig, immunoglobulin. Individual items were reduced to dimensions based on a principal component analysis using the data of 580 primary care patients. Subsequently, the weights per dimension (within a category) was determined based on the coefficients from a Ridge regression analysis for each category, using the data from 88 PAD/unlikely PAD patients that were included in the GP-PAD II study. Lastly, the weights per category were

determined by performing an overarching Ridge regression analysis, where the variables consisted of the scores per category of the algorithm (adjusted based on weights dimensions).
**If a prescription, ICPC code, or laboratory value did not occur in the dataset of included PAD/unlikely-PAD patients, Ridge regression analysis could not be performed. In this case, the dimension and weight per dimension were based on clinical relevance.*

Table E8. Overview of estimated costs of the PAD-screening algorithm when applying version 1 and subsequently version 3 of the algorithm to our study population of 61 172 patients

Description	**Estimated cost for 1 patient †**	**Estimated number of patients to whom this step would apply**	**Total estimated cost**
EHR-screening, performed by doctor's assistant, 10 minutes per EHR at a gross hourly rate of €14,07	€2.34	296	€694.12
Blood withdrawal and serum analysis of immunoglobulins (IgA, IgM, IgG total, IgG1, IgG2, IgG3, IgG4, total protein and albumin)	€191.55	149	€28 540.66
Two outpatient visits to an immunologist at an academic hospital ‡ [45,46]	€398.26	46	€18 319.96
Laboratory testing including blood count, leukocyte differentiation, ferritin, HIV-test and urine sediment	€56.09	46	€2580.14
Pneumococcal vaccination	€25.94	46	€1193.24
2x pneumococcal vaccination response[47]	€27.36	46	€1258.56
Total			**€52 586.68**

EHR, electronic health record; HIV, human immunodeficiency virus; Ig, immunoglobulin. † Based on expenses in current study or internal prices of the University Medical Centre Utrecht, unless references in the description specify otherwise. ‡ Costs based on estimation by the Institute for Medical Technology Assessment of the Erasmus Universiteit Rotterdam, corrected for inflation.

PART III
Treatment decision support

Chapter 6

What is the best target in a treat-to-target strategy in rheumatoid arthritis? Results from a systematic review and meta-regression analysis

Marianne A. Messelink
Alfons A. den Broeder
Florine E. Marinelli
Edwin Michgels
Patrick Verschueren
Daniel Aletaha
Paco M.J. Welsing

RMD Open, 2023;9(2):e003196

Abstract

Objectives A treat-to-target (T2T) strategy has been shown to be superior to usual care in rheumatoid arthritis (RA), but the optimal target remains unknown. Targets are based on a disease activity measure (e.g. DAS28, SDAI/CDAI), and a cut-off such as remission or low disease activity (LDA). Our aim was to compare the effect of different targets on clinical and radiographic outcomes.

Methods Cochrane, Embase and (pre)MEDLINE databases were searched (1-Jun-2022) for RCTs and cohort studies after 2003 that applied T2T in RA patients for ≥12 months. Data was extracted from individual T2T study-arms; risk of bias was assessed with the Cochrane Collaboration tool. Using meta-regression, we evaluated the effect of the target used on clinical and radiographic outcomes, correcting for heterogeneity between and within studies.

Results 115 treatment arms were used in the meta-regression analyses. Aiming for SDAI/CDAI-LDA was statistically superior to targeting DAS-LDA regarding DAS- and SDAI/CDAI/Boolean-remission outcomes over 1-3 years. Aiming for SDAI/CDAI-LDA was also significantly superior to DAS-remission regarding both SDAI/CDAI/Boolean-remission (over 1-3 years) and mean SDAI/CDAI (over 1 year). Targeting DAS-remission rather than DAS-LDA only improved the percentage of patients in DAS-remission, and only statistically significantly after 2-3 years of T2T. No differences were observed in HAQ and radiographic progression.

Conclusions Targeting SDAI/CDAI-LDA, and to a lesser extent DAS-remission, may be superior to targeting DAS-LDA regarding several clinical outcomes. However, due to the risk of residual confounding and the lack of data on (over)treatment and safety, future studies should aim to directly and comprehensively compare targets.

Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disorder characterized by inflammation of synovial joints, although other organ systems can also be involved.[1] Treatment is aimed at limiting and controlling disease activity, as prolonged high levels of activity increase the risk of progressive joint damage and mortality.[2-4] In the proactive treat-to-target (T2T) strategy, disease activity is frequently and systematically assessed using a validated measure, which is then compared to a pre-specified treatment target. If the target is not reached within a particular timeframe, treatment is intensified accordingly.[5]

Targets are commonly based on a composite disease activity score with a cut-off value for remission or low disease activity (LDA).[6] Examples of composite scores include the Disease Activity Score (DAS, counting 28 or 44 joints in the DAS28/DAS44 respectively), the Simplified- and the Clinical Disease Activity Indices (SDAI/CDAI).[7] The DAS variants are based on tender and swollen joint counts (TJC/SJC), the erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP), and the patient's global assessment of disease activity (PGA) on a visual analogue scale (VAS).[8] Although the DAS is well-established and validated, it has been criticized for allowing a high SJC whilst fulfilling the definition of remission, due to calculation effects.[9] The SDAI and CDAI weigh the individual components equally and have added the physicians assessment of disease activity. The CDAI does not include any laboratory marker, making it easier to apply in clinical practice, but also less objective.

Alternative to the composite scores, a target may also be defined using a Boolean definition, e.g. ACR/EULAR remission. Here, a set of core variables (TJC/SJC, PGA and CRP) must all have a value of ≤1. As the original Boolean criteria were criticized for being too stringent, the recently revised criteria have loosened the maximal PGA to 2 cm on a 10 cm VAS.[10]

While T2T has been shown to be superior to the (previous) standard of usual care, the direct comparison of different treatment targets is insufficiently studied.[5, 11-13] Therefore, it remains unknown what the optimal target in a T2T-strategy is, and thus different targets are used in clinical practice and recommended in international guidelines.[5, 13, 14] Specifying the optimal treatment target is important: too lenient a target may result in undertreatment and a higher disease burden. Conversely, too stringent a target may lead to overtreatment, side-effects, patient dissatisfaction and unnecessary costs. Both situations may negatively impact patients' quality of life. Accordingly, determining the optimal treatment target has been included in the research agenda of the 2022 EULAR recommendations.[13]

As the T2T-strategy has been generally accepted and recommended for some time, many recent clinical studies evaluating a specific drug or treatment strategy as primary objective apply a T2T-approach.[5, 15] In the current study, we exploited this available evidence by performing a systematic literature review and meta-regression analysis. Our aim was to compare the effect of different treatment targets on clinical and radiographic outcomes in patients with rheumatoid arthritis.

Methods

Protocol and registration

Prior to commencing this study, the protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO, number: CRD42021249015). This systematic review was performed and reported in accordance with the PRISMA guidelines.[16]

Literature search

The databases (pre)MEDLINE, Cochrane and Embase were searched, combining synonyms and MeSH terms for T2T and RA. Studies published in English (due to lack of fluency in other languages) after 2003 were included, as after this date the use of biological disease modifying anti-rheumatic drugs (bDMARDs) and the T2T-principle became common in clinical studies. In addition to original research papers, reviews were included for reference screening of other relevant articles. For a full overview of the search strategy as performed on June 1st 2022, see Supplemental Figures S1-S4.

Study selection

Identified articles were deduplicated and uploaded to a reference management program (Rayyan), where MM and FM independently performed title-, abstract-, and subsequent full-text screening. Randomized controlled trials (RCTs) and cohort studies were included if they applied a T2T-strategy in RA patients for ≥12 months. T2T was defined as the frequent (≤ 4 monthly) assessment of disease activity using a validated measure (i.e. DAS28/44-ESR/CRP, SDAI, CDAI, RADAI, RAPID3, Boolean remission or a SJC), which is compared to a pre-specified treatment target. If the target is not met, DMARD-treatment should be intensified. Studies with a sample size of ≥50 patients and ≥1 T2T-treatment arm could be included. Outcomes of interest were disease activity, radiographic progression, functional status, bDMARD use and safety after ≥12 months of T2T. These measures could be expressed continuously or as a response percentage. Disagreements were discussed among FM, MM, PW and AB until consensus was reached.

Data-extraction

Data-extraction was performed independently by MM and EM using a predefined form; discrepancies were double-checked with the source data. Data-extraction was performed per treatment arm that applied a T2T-strategy (e.g. if an RCT reported 2-year results for 3 T2T-treatment arms, these were extracted separately). Only time-points when T2T-treatment was continued were considered. When outcomes were present at multiple time-points (e.g. 1 and 2 years) these were extracted as separate arms. See Table 1 for an overview of the extracted data.

Table 1: Overview of extracted data from included articles

Study characteristics	**Baseline characteristics**	**Outcome measures****
Title	*Mean/median*	
Publication year	SDAI, CDAI or DAS and their components	(Change in) SDAI, CDAI or DAS and their components
Study type (RCT or cohort)	SF-survey	SF-survey
Treatment target	Fatigue and pain on a VAS	Fatigue and pain on a VAS
Drug treatment protocol	HAQ-score	HAQ score
Interval between assessments	SHS-score and components	SHS-score and components
Number of patients	Age and symptom duration	
	Proportion of patients	
	Female	Remission (drug-free, Boolean, DAS, SDAI or CDAI) and LDA
	Early* vs. Established RA	Relevant radiographic progression
	Erosions	Erosions
	RF/ACPA positive patients	AE/SAE
		Use of bDMARDs

** Early RA was based on the definition given in the articles, which was usually either <1 year or <2 years disease duration. ** Regarding outcome measures, the time-point, the outcome value and the measure of spread (i.e. SD, CI or IQR) were extracted. ACPA: anti-citrullinated protein antibody, AE: adverse events, bDMARD: biological disease modifying anti-rheumatic drug, CDAI: Clinical Disease Activity Index, CI: confidence interval, DAS: disease activity score, HAQ: Health Assessment Questionnaire, IQR: interquartile range, LDA: low disease activity, RA: rheumatoid arthritis, RCT: randomized controlled trial, RF: rheumatoid factor, SAE: serious adverse event, SD: standard deviation, SDAI: Simplified Disease Activity Index, SF: Short Form health survey, SHS: Sharp van der Heijde score, VAS: visual analogue scale.*

Risk of bias assessment

Risk of bias (RoB) assessment was performed independently by FM and MM using the Cochrane Collaboration tool.[17] All domains were assessed, except the randomization domain for cohort studies as this is not applicable to this design. Disagreements were discussed among FM, MM, PW and AB until consensus was reached.

Data pre-processing for meta-regression analysis

To prepare the data for meta-regression analysis, we selected arms that reported the same outcome measure (e.g. DAS-remission) at the same time-point (i.e. "1", "2-3", "4-6" or ">6" years) for each analysis. A minimum of 10 arms was deemed required for a relevant analysis. Treatment targets that were used in ≤2 studies were excluded from analysis, which was the case for Boolean- and SDAI-remission and a SJC of 0. The *targets* SDAI-LDA and CDAI-LDA were grouped together, as these were uncommon and were deemed sufficiently similar.[18] Similarly, for the targets of DAS-remission and DAS-LDA, the different DAS-variants were combined (i.e. DAS28/44 using ESR or CRP). Regarding *outcome measures,* SDAI-, CDAI- and Boolean remission were grouped together, for which we corrected in the meta-regression analysis (see below). Of note, the trials included in this study did not yet incorporate the revised 2022 ACR/EULAR Boolean criteria.[10] To standardize scales of different DAS variants, mean DAS44-ESR and DAS28-CRP measures were converted to DAS28-ESR according to published translation formulae, as DAS28-ESR was most commonly reported (27 of 52 mean DAS analyses).[19, 20] For the SHS outcome, the mean change per year was calculated in order to correct for limited differences in the duration of follow up, as the SHS is a cumulative measure and can typically only increase over time.

Changes from baseline in disease activity/HAQ/SHS were transformed to outcome values at endpoint when needed using the baseline value. The SD of the score at endpoint was calculated using a correlation coefficient, in accordance with the Cochrane handbook for systematic reviews (needed for 78 of 165 arms).[21] For the HAQ, a correlation coefficient of 0.6 was assumed to determine the SD of the outcome due to insufficient data for 17 of 58 arms. When only medians and IQRs were reported (28 of 165 mean outcome analyses), means and SDs were estimated based on the formulas provided by Wan et al (2014).[22]

Meta-regression analysis

Meta-regression analyses were performed with the R package 'metafor' (rma.mv function) for multilevel meta-analytic regression models. Both 'study' and 'arm' were added as random effects in all analyses, where 'study' refers to an overarching clinical trial (e.g. BeST or OPERA) and 'arm' refers to the individual treatment arms. This was done to account for the heterogeneity between and within studies, as often multiple arms from one study were included. In the metaregression analyses each arm is attributed a weight based on the standard error, thus correcting for small sample sizes.

We first estimated the effect of the treatment target on the outcome at a certain timepoint in a univariate model. Subsequently, a full (adjusted) model was composed,

adding the following covariates: early or established RA, the availability of bDMARDs, the baseline value of the outcome variable, and whether the treatment intensifications were formalized in a protocol. For remission outcomes we corrected for the baseline DAS and the specific type of remission (e.g., DAS44-ESR- or SDAI-based), for which we also performed subgroup analyses. For yearly SHS progression a non-linear association for baseline SHS was also explored using a squared term, given its importance and the known ceiling effect of the SHS.[23] The selection of covariates was based on clinical expertise, variables' availability and impact on the outcome in bivariate analyses from the extracted variables (see Data-extraction). The primary variables of interest for the full model (i.e. target, outcome, covariates: early/established RA, availability of bDMARDs, baseline value outcome variable and presence of a formal treatment protocol) were available for all arms included in this study.

If there was no need to correct for one or two of the selected covariates due to a lack of variation (e.g. if all treatment arms were early RA), the mean percentage RF and/or ACPA-positive patients and symptom duration were explored as covariates. This was the case for 3 analyses: SHS progression at year 1, mean DAS28-ESR at 2-3 years and DAS remission at 4-6 years.

From the full model, a parsimonious model was derived in which covariates with little effect were removed from the model. Covariates with a p-value >0.2 were removed, starting with the covariate with the highest p-value. If this removal resulted in a change of (any of) the regression coefficient(s) for the treatment target(s) of ≥15% with a minimum absolute effect of 0.05/0.005 for means/proportions respectively, the covariate was kept in the model. If these criteria were not met, the covariate was removed. Subsequently, the next covariate with the highest p-value >0.2 was removed and evaluated in the same way. This process was iterated until no more variables could be removed, resulting in the parsimonious model. The baseline value of the outcome variable was always retained.

Although normal distributions of the outcome variables are to be expected based on the central limit theorem, a natural log transformation was performed if the residuals of the parsimonious model had a skewness exceeding -2/+2 or kurtosis exceeding -7/+7, in addition to an untransformed sensitivity analysis.[24-26] Subgroup analyses were performed for early vs. established RA, remmision types and RCT vs. cohort studies, if ≥5 arms per subgroup were available. Additionally, we performed sensitivity analyses, using the parsimonious models, by excluding high risk of bias papers and by performing 'leave-one-out' analyses. In the latter, the same analysis is performed numerous times, excluding one arm in each analysis, in order to assess the influence of individual arms. The minimum and maximum effects of the 'leave-one-out' analyses are reported.

Results

Study selection and characteristics
Of the 3,879 articles identified through the literature search, 66 articles were selected after title-, abstract-, full-text- and reference screening (Figure 1).[27-92] Characteristics of selected articles including RoB assessments are shown in Supplemental Table S1. These 66 selected articles concerned results from 40 studies, and from these the data of 169 treatment arms were extracted. Of the 66 articles/169 arms extracted, the data of 52 articles/114 arms were used in the meta-regression analyses. Reasons not to use treatment arms in the analysis are stated in Supplemental Table S1 and include: reported outcome measure was present in an insufficient number of arms (<10, e.g. bDMARD use and adverse events), insufficient arms at a particular time-point (e.g. after >6 years of T2T), or duplicate results with another arm (e.g. 2 arms that both report mean DAS after 1 year in the BeSt-trial). The characteristics of the 114 treatment arms used in the meta-regression analyses are shown in Table 2.

Meta-regression analyses
For a full overview of the target regression coefficients and confidence intervals of the parsimonious-, full- and univariate models, see Supplemental Tables S2-S10.

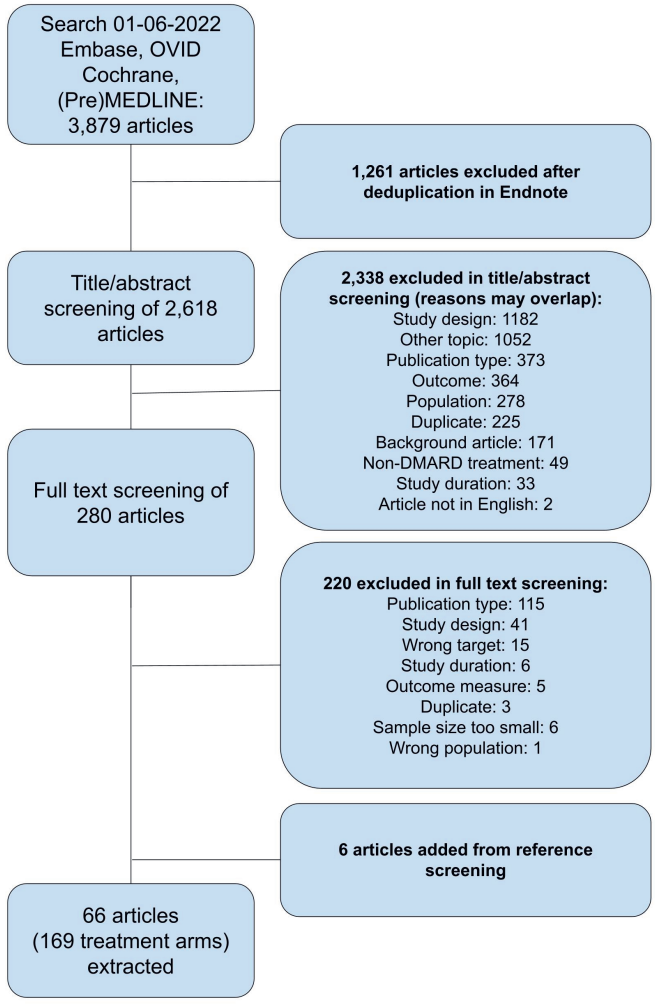
Outcome mean DAS28-ESR
Targeting either SDAI/CDAI-LDA or DAS-remission rather than DAS-LDA gave a non-statistically significant improvement of the mean DAS28-ESR after 1-3 years of T2T, see Figure 2A. There were no differences between the target of SDAI/CDAI-LDA and DAS-remission (Supplemental Tables S2-S3).

Outcome percentage DAS-remission
Aiming for SDAI/CDAI-LDA rather than DAS-LDA significantly improved the percentage of patients in DAS-remission after 1-3 years of T2T, see Figure 2B. There were no differences between the SDAI/CDAI-LDA and DAS-remission targets (Supplemental Tables S2-S3). Compared to DAS-LDA, the target of DAS-remission significantly improved the percentage of patients in DAS-remission at 2-3 years with 21% (percentage points, e.g. from 50% to 71%, p=0.03). At year 1 only a trend was observed (p=0.11), with only a statistically significant effect on the outcome DAS44-ESR-remission (38%, p<0.0001). After 4-6 years no significant improvement was observed when targeting DAS-remission vs. DAS-LDA (10%, p=0.53, Supplemental Table S4).

Outcome Mean SDAI/CDAI
After 1 year of T2T, targeting SDAI/CDAI-LDA compared to DAS-remission significantly improved the mean SDAI/CDAI with 5.08 units (p=0.03), see Supplemental Table S2.

Compared to to DAS-LDA, only a non-statisically significant mean improvement of 2.03 units (p=0.23) was observed (Figure 2C). Targeting DAS-remission compared to DAS-LDA, gave a non-statistically significant deterioration of mean SDAI/CDAI of -3.05 (p=0.06).

Figure 1. Flowchart of article selection



DMARD: disease-modifying anti-rheumatic drug

Table 2: Characteristics of treatment arms used in the meta-regression analyses

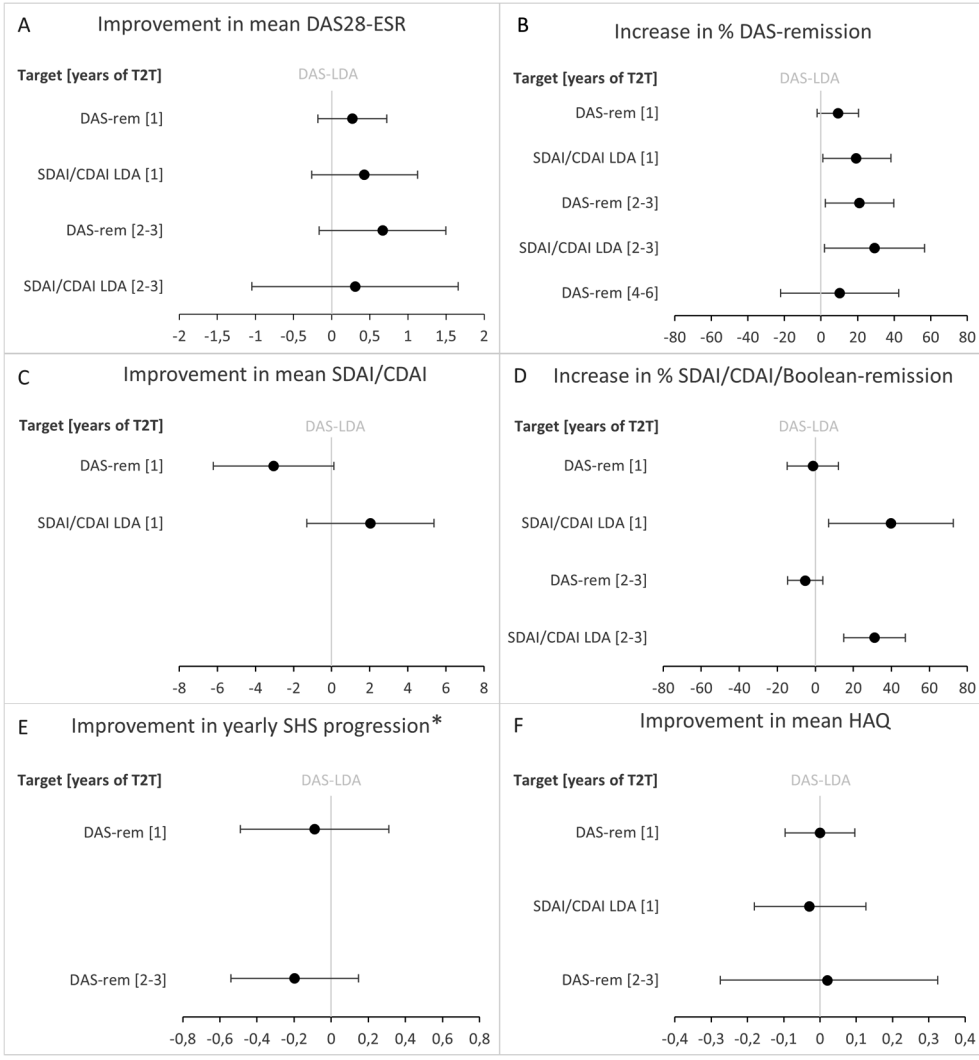
Arms used in analyses, n	114
Arms with T2T target DAS28/44-LDA, n (%)	71 (62.3%)
Arms with T2T target DAS28/44-remission, n (%)	36 (31.6%)
Arms with T2T target SDAI-CDAI LDA, n (%)	7 (6.1%)
Publication year	
2003-2009, n(%)	13 (11.4)
2010-2015, n(%)	35 (30.7)
2016-2022, n(%)	66 (57.9)
% Female, mean (SD) [115 arms]	70.0 (6.7)
Age, mean (SD) [103 arms]	54.3 (4.6)
%RF positive, mean (SD) [107 arms]	64.7 (15.9)
%ACPA positive, mean (SD) [94 arms]	66.5 (16.2)
Baseline DAS, mean (SD) [109 arms]	5.0 (0.8)
Baseline HAQ, mean (SD) [66 arms]	1.2 (0.3)
Baseline SHS, median (IQR) [57 arms]	3.0 (1.3-7.0)
Early RA*, n (%) [115 arms]	100 (87.7)
bDMARD available, n (%) [115 arms]	80 (70.2)
Formal treatment protocol, n (%) [115 arms]	89 (78.1)
Outcome DAS28/44 remission, mean % (SD) [88 arms]	55.9 (16.8)
Outcome SDAI/CDAI/Boolean remission, mean % (SD) [67 arms]	37.2 (14.5)
Outcome DAS28-ESR, mean (SD) [51 arms]	2.8 (0.6)
Outcome SDAI/CDAI, median (IQR) [16 arms]	4.5 (2.6-5.6)
Outcome HAQ, mean (SD) [66 arms]	0.6 (0.2)
Outcome SHS, median (IQR) [57 arms]	6.0 (1.9-12.4)

** Early RA was based on the definition given in the articles, which was usually either <1 year or <2 years disease duration. ACPA: Anti-citrullinated protein antibodies, bDMARD: biological disease modifying antirheumatic drug, CDAI: Clinical Disease Activity Index, DAS: Disease Activity Score, HAQ: Health Assessment Questionnaire, IQR: interquartile range, LDA: low disease activity, RA: rheumatoid arthritis, RF: rheumatoid factor, SD: standard deviation, SDAI: Simplified Disease Activity Index, SHS: Sharp van der Heijde Score.*

Outcome percentage SDAI/CDAI/Boolean remission

Targeting SDAI/CDAI-LDA significantly improved the percentage of patients in SDAI/CDAI/Boolean remission compared to both DAS-LDA and DAS-remission targets, after 1 year (borderline significant for DAS-LDA) and 2-3 years (year 1: 34% p=0.05 and 35% p=0.03 for comparison to DAS-LDA and DAS-remission respectively, year 2-3: 31% p=0.0002 and 36% p<0.0001). See Figure 2D and Supplemental Tables S2-S3. There were no clear differences between the targets of DAS-remission and DAS-LDA, nor did the subgroup analysis of the remission outcome types (SDAI/CDAI vs. Boolean remission) differ substantially from the overall analysis.

Figure 2: Effect of the DAS28/44-remission and SDAI/CDAI-LDA treatment targets compared to the target of DAS28/44-LDA, on different outcomes and at different time-points based on meta-regression analyses (parsimonious models).



*Results are presented in comparison to a treatment target of DAS-LDA. Exact coefficients and confidence intervals can be found in Supplemental Tables S2-S4. *The natural log of yearly SHS progression was used in the analysis. The models were corrected for the following covariates. Mean DAS28-ESR year 1: formal treatment, bDMARD, BL DAS28-ESR, year 2-3: mean RF/ACPA, early/established RA, BL DAS28-ESR. DAS remission year 1: remission type, bDMARD, early/established RA, BL DAS28-ESR, year 2-3: remission type, formal treatment, BL DAS28-ESR, year 4-6: remission type, symptom duration, BL DAS28-ESR. Mean SDAI/CDAI year 1: SDAI/CDAI, formal treatment, bDMARD, early/established RA, BL DAS28-ESR. SDAI/CDAI/Boolean remission year 1: remission type, formal treatment, early/established RA, BL DAS28-ESR, year 2-3: remission type, bDMARD, early/established, BL DAS28-ESR. SHS progression year 1: formal treatment, mean RF/ACPA, symptom duration, log(BL SHS), log(BL SHS)^2, year2-3: mean RF/ACPA, log(BL SHS), log(BL SHS)^2. Mean HAQ year 1: bDMARD, early/established RA, BL HAQ, year 2-3: BL HAQ. ACPA: anti-citrullinated protein antibody, bDMARD: biological disease modifying antirheumatic drug, BL: baseline, CDAI: Clinical Disease Activity Index, DAS(28/44): Disease*

activity score (28/44 indicating the joint count), ESR: erythrocyte sedimentation rate, HAQ: Health assessment questionnaire, LDA: low disease activity, RA: rheumatoid arthritis, rem: remission, RF: rheumatoid factor, SDAI: Simplified Disease Activity Index, SHS: Sharp van der Heijde Score, T2T: treat-to-target.

Outcome yearly progression of SHS

No differences between targeting DAS-remission and DAS-LDA were found for yearly SHS progression, although surprisingly, a numerical deterioration was found for targeting DAS-remission vs. DAS-LDA, after both 1 year of T2T (-0.09, p=0.66) and 2-3 years of T2T (-0.20, p=0.26). See Figure 2E. A similar trend was present in the untransformed sensitivity analyses at years 1 and 2 (-0.19, p=0.75 and -0.40, p=0.21 respectively). For the analysis of SHS progression at year 1, one arm that did not report the mean percentage RF nor ACPA-positive patients was removed from the primary analysis, in order to add this covariate to the model. As a sensitivity analysis, we added this arm to the model, and left out the RF/ACPA variable, which did not greatly affect results (-0.09 p=0.66 to -0.15 p=0.44).

Outcome mean HAQ

We observed no differences between targets at years 1-3 for the mean HAQ, see Figure 2F.

Outcome mean CRP

Given the subjective nature of many of the outcome measures, we also analyzed the more objective outcome of CRP. The target of SDAI/CDAI-LDA but not the target of DAS-remission non-significantly reduced mean CRP at year 1 (-2.01, p=0.54 and -0.01, p=0.99) compared to the DAS-LDA target.

Subgroup analysis early vs. established RA

After 1 year of T2T, targeting DAS-remission rather than DAS-LDA showed a numerical improvement in mean DAS for the subgroup with early RA (0.29, p=0.24), and this effect was more pronounced for established RA (1.21, p=0.25). The interaction between early vs. established RA and the target was not significant (p=0.53). The percentage of patients in DAS-remission non-significantly improved when targeting DAS-remission (8%, p=0.16) or SDAI-CDAI LDA (4%, p=0.79) rather than DAS-LDA in the early RA subgroup. In the established RA subgroup, this effect was again more pronounced: DAS-remission vs. DAS-LDA (20%, p=0.18), and significant for SDAI-CDAI LDA vs. DAS-LDA (48%, p=0.003). The interaction term for the DAS-remission model was significant (p=0.01).

Subgroup analysis RCT vs. cohort studies
The beneficial effects of SDAI/CDAI-LDA and DAS-remission compared to DAS-LDA were more pronounced in the cohort subgroups than in the RCT subgroups, when considering the outcomes mean DAS and HAQ at year 1, and DAS remission at years 1-3. See Supplemental Table S11. For the comparison of DAS-remission to DAS-LDA regarding SDAI/CDAI/Boolean remission at years 1-3, this effect was not present. None of the interaction terms were significant, except for mean HAQ year 1 (p=0.02).

Comparison of parsimonious and full models
The results of the parsimonious and full models were similar regarding the direction and statistical significance of the coefficients. The only exception was the improvement in the proportion of patients in DAS-remission when comparing the targets of DAS-remission (after 2-3 years of T2T) and SDAI/CDAI-LDA (after 1-3 years of T2T) to DAS-LDA: these did reach statistical significance in the parsimonious models, but were not statistically significant in the full models. See supplemental tables S2-S7.

High risk of bias studies
As a sensitivity analysis, we ran the parsimonious models excluding the high risk of bias studies (see Supplemental Table S1). The improvement in the proportion of patients in DAS-remission after 1 year of T2T when targeting DAS-remission vs. DAS-LDA increased from 0.09 (p=0.11) to 0.11 (p=0.06). When targeting SDAI/CDAI-remission vs. DAS-LDA, the improvement in the proportion of patients in DAS-remission remained 0.21 (p-value from 0.04 to 0.03). Regarding mean HAQ after 1 year of T2T, there were no differences between targets, and this did not change when the high risk of bias study was excluded (DAS-remission target vs. DAS-LDA from 0.00 (p=1.00) to 0.01 (p=0.82), SDAI/CDAI LDA target vs. DAS-LDA from -0.03 (p=0.73) to -0.07 (p=0.58)).

Leave-one-out analyses
For the majority of the analyses, the direction and significance of the effect remained the same after performing leave-one-out analyses. See Supplemental Table S12 for details.

Narrative results adverse events and medication use
Insufficient data was available to perform meta-regression analysis of (S)AEs and medication use. Narratively, the percentages of patients with ≥1 reported AE at year 1 varied from 35%-59% for the target DAS-LDA[71, 74, 90], and from 82%-96% for DAS-remission[28, 61]. For SAEs at year 1, this varied from 5%-19% for DAS-LDA[48, 74, 90], from 0%-19% for DAS-remission[28, 61, 69] , and from 2%-15% for SDAI/CDAI LDA.[39, 40] After 2 years of T2T, the SAEs for the target of DAS-LDA varied from 6%-23%[49, 58, 60], and from 0.6%-18.4% for DAS-remission[34, 45, 52]. The reported use

of bDMARDs during the first 1-2 years of T2T varied from 2%-32% for the target DAS-LDA[43, 74], from 7%-17% for DAS-remission[52, 73], and from 3%-8% for SDAI-LDA[76]. After 5 years of T2T, SAEs varied from 5%-21% for DAS-LDA[79], and 17% was reported for the target DAS-remission[56].

Narrative results target comparison studies

Two studies directly compared treatment targets. Tam et al. found no significant differences after 1 year when targeting SDAI vs. DAS28-ESR remission (mean change DAS28-ESR: -2.5(SD 1.3) vs. -2.3(SD 1.3), p=0.51, mean change SDAI: -22.2(SD 12.6) vs. -20.0(SD 11.8), p=0.35, median change in HAQ: -0.6(IQR -1.1 to -0.38) vs -0.5(IQR -1.1 to 0.0, p=0.25), DAS remission (51% vs. 55%, p=0.66), SDAI-remission: 37% vs. 40%, p=0.73).[68] Hodkinson et al. found no significant differences after 1 year when targeting SDAI-LDA vs. CDAI-LDA (mean DAS28: 3.0 (SD 1.2) vs. 3.3 (SD 1.2), p=0.29, DAS28-remission: 34% vs. 33%, p=1.00, HAQ: 1.0 (SD 0.7) vs. 1.0 (SD 0.7), p=0.94).[40]

Discussion

In this study we assessed the effect of different treatment targets on clinical and radiographic outcomes using meta-regression analyses. Our results indicate that aiming for SDAI/CDAI-LDA was superior to targeting DAS-LDA regarding the percentages of both DAS- and SDAI/CDAI/Boolean remission. Aiming for SDAI/CDAI-LDA was also superior to targeting DAS-remission regarding SDAI/CDAI based outcomes (SDAI/CDAI/Boolean-remission and mean SDAI/CDAI). When comparing the target of DAS-remission to DAS-LDA, the former only significantly improved the percentage of patients in DAS-remission after 2-3 years of T2T, and only a trend was present for the improvement in DAS-remission at other time-points and mean DAS28-ESR. Functioning and radiographic damage did not differ between targets.

A strength of the current study is that it is the first to perform a quantitative analysis of the optimal treatment target based on the available evidence in literature. The technique of meta-regression allows to partly correct for heterogeneity between and within studies for measured confounders, thus optimally utilizing the available evidence on T2T-strategies in trials and cohort studies in RA. This resulted in the largest number of T2T cohorts and trials included in such a review to date.

Limitations of our study firstly include that our analyses are based on an indirect comparison of treatment targets. We aimed to correct for at least the most important confounders on study-level, namely the availability of bDMARDs, the use of a formal treatment protocol, early vs. established RA and the baseline value of the outcome. Nevertheless, residual confounding by insufficiently or unmeasured factors and

treatment regimen specifics can certainly be present. Furthermore, insufficient data was available for the analysis of medication use, quality of life and adverse events. These are important factors that should be taken into account when selecting a treatment target for a patient. Based on the limited reported evidence, we have no indication to assume that SAEs or the use of bDMARDs were increased for DAS-remission or SDAI/CDAI LDA versus DAS-LDA as a target, although the number of AEs may be higher for DAS-remission. However, as these results are from a small number of studies, and could not be adjusted for confounding, they should be interpreted with caution.

Further limitations include that information bias due to systematic differences in the scoring of the disease activity and/or SHS between studies could have occurred. Also, a certain amount of circularity is inherent to all studies evaluating a disease activity score based treatment target and outcome. For example, targeting DAS-remission may inherently increase the chance of achieving DAS-remission. Indeed in our results targets often performed better on related outcomes, although the target of SDAI/CDAI-LDA also performed better on DAS-based outcomes. No differences were observed in the more independent HAQ and SHS outcomes, which may be partly due to a limited variation in these measures with modern-day intensive therapy. Lastly, it is important to bear in mind that in routine clinical practice individual patient-factors should always be considered. The optimal target is therefore a piece of the puzzle for clinical decision making, but it is not a replacement of clinical decision making for individual patients.

Interestingly, our results suggest a limited benefit of targeting DAS-remission compared to DAS-LDA. We unexpectedly even found a negative trend on the SHS-score when targeting DAS-remission compared to DAS-LDA, which could potentially be due to residual confounding. Based on previous studies, it could also be hypothesized that a stricter treatment target may cause premature drug cycling or reduced therapy adherence, both resulting in higher remaining disease activity.[60, 93, 94] Of note, the SHS is known to have a strong right-skewed distribution, making the mean more susceptible to outliers despite our efforts to compensate this with a natural log transformation. Unfortunately we could not analyze the effect of a SDAI/CDAI LDA target on the SHS outcome, as this was not reported.[39, 40, 76] Another finding of interest arose from our subgroup analysis of early vs. established RA patients, where it seemed that targeting DAS-remission rather than DAS-LDA might be more beneficial for established RA than early RA patients. This may suggest that a less stringent target may be sufficient early in the disease, possibly in part because it allows for the effects of treatment adjustments to be fully established. Once a steady state has been reached, a stricter target may be required to optimize results. This is compatible with the ACR-

recommendations on T2T, but in contrast with the advice of the international T2T task force.[5, 14]

From the articles included in this review, two studies directly compared treatment targets. Tam et al. found no difference between the SDAI and DAS28-ESR remission targets, a finding which we could not replicate in our study as this was the only article applying a SDAI-remission target.[95] Of note, the DMARD treatment protocol differed substantially between the two target arms, and this was not corrected for in the analysis. Hodkinson et al. found no difference between the targets of SDAI- and CDAI-LDA, which may not be surprising given these targets are quite similar.[40] We therefore combined these uncommon targets in the current review.

Previous reviews have also considered the optimal target in a T2T-strategy, although only narratively.[96, 97] Hock et al. report no preference for any particular target. Bergstra et al. conclude that based on the limited available and indirect evidence, aiming for remission rather than LDA (types not specified) seemed to result in more patients achieving remission, but not better physical functioning. This is in line with our results regarding DAS-remission as a target for treatment steering, but SDAI/CDAI-LDA targets appeared to be superior in several aspects as described above. Similar to Bergstra et al., we found no effects on physical functioning as measured with the HAQ. This could potentially be due to the known 'floor-effect' of the HAQ, as it may be insensitive to changes at the lower end of the spectrum.[98]

The international T2T task force recommends to target a state of clinical remission, described as the absence of signs and symptoms of significant inflammatory disease.[5] They suggest that ACR-EULAR remission (i.e. Boolean- or SDAI/CDAI-remission[10]) may be the best suitable definition for this criterion. Similarly, the EULAR guidelines recommend ACR-EULAR remission as the main therapeutic target, with LDA (type not specified) as an alternative, especially in established RA.[13] Surprisingly, no more than 1 of 66 T2T articles applied a SDAI- or Boolean remission target, and none applied a CDAI-remission target. This was insufficient for inclusion in our analyses, and therefore our results cannot support these recommendations. The ACR guidelines, in contrast, recommend to initially target LDA, and to subsequently consider targeting remission (types not specified).[14] Thus, there is no consensus in international guidelines regarding the initial treatment target, reflecting equipoise. Lastly, once a statisfactory stable level of disease activity has been reached after a period of T2T (perhaps irrespetive of whether the target was actually reached), tapering (b)DMARDs may be considered. As the ACR and EULAR guidelines differ regarding when to initiate tapering (being either remission or LDA), determining the optimal target may also be relevant in the context of tapering.[13, 14]

To fully determine the optimal treatment target in a T2T-strategy, an RCT comparing different targets head-to-head will be necessary. In line with our results and the ACR-guidelines, we would recommend to include an LDA target as a reference arm.[14] Based on our results SDAI- or CDAI-LDA would be the LDA-target of choice, although DAS(28)-LDA may be an alternative as it is the most commonly used. In addition, DAS28-ESR remission may be considered, as our results showed a (limited) benefit of DAS-remission over DAS-LDA. Based on the EULAR and T2T task force guidelines, a target of SDAI- or Boolean remission is of interest. Lastly, a predefined subgroup analysis regarding early versus established RA is recommended.

References

1. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. Bone Res. 2018;6:15.
2. Book C, Saxne T, Jacobsson LT. Prediction of mortality in rheumatoid arthritis based on disease activity markers. J. Rheumatol. 2005;32:430-4.
3. Mikuls TR, Fay BT, Michaud K, Sayles H, Thiele GM, Caplan L, et al. Associations of disease activity and treatments with mortality in men with rheumatoid arthritis: results from the VARA registry. Rheumatology (Oxford). 2011;50:101-9.
4. Listing J, Kekow J, Manger B, Burmester GR, Pattloch D, Zink A, et al. Mortality in rheumatoid arthritis: the impact of disease activity, treatment with glucocorticoids, TNFα inhibitors and rituximab. Ann. Rheum. Dis. 2015;74:415-21.
5. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann. Rheum. Dis. 2016;75:3-15.
6. Fleischmann RM, Van Der Heijde D, Gardiner PV, Szumski A, Marshall L, Bananis E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open. 2017;3:e000382.
7. Aletaha D, Smolen JS. The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract. Res. Clin. Rheumatol. 2007;21:663-75.
8. Fleischmann R, Van Der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann. Rheum. Dis. 2015;74:1132-7.
9. Jacobs JW, Ten Cate DF, van Laar JM. Monitoring of rheumatoid arthritis disease activity in individual patients: still a hurdle when implementing the treat-to-target principle in daily clinical practice. Rheumatology (Oxford). 2015;54:959-61.
10. Studenic P, Aletaha D, de Wit M, Stamm TA, Alasti F, Lacaille D, et al. American College of Rheumatology/EULAR remission criteria for rheumatoid arthritis: 2022 revision. Ann. Rheum. Dis. 2022:ard-2022-223413.
11. Wailoo A, Hock ES, Stevenson M, Martyn-St James M, Rawdin A, Simpson E, et al. The clinical effectiveness and cost-effectiveness of treat-to-target strategies in rheumatoid arthritis: a systematic review and cost-effectiveness analysis. Health Technol. Assess. 2017;21:1-258.
12. Stoffer MA, Schoels MM, Smolen JS, Aletaha D, Breedveld FC, Burmester G, et al. Evidence for treating rheumatoid arthritis to target: Results of a systematic literature search update. Ann. Rheum. Dis. 2016;75:16-22.
13. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann. Rheum. Dis. 2023;82:3-18.
14. Fraenkel L, Bathon JM, England BR, St Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Care Res. (Hoboken). 2021;73:924-39.
15. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann. Rheum. Dis. 2020;79:685-99.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
17. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

18. Fleischmann R, Strand V, Wilkinson B, Kwok K, Bananis E. Relationship between clinical and patient-reported outcomes in a phase 3 trial of tofacitinib or MTX in MTX-naïve patients with rheumatoid arthritis. RMD Open. 2016;2:e000232.
19. Carpenter L, Norton S, Nikiphorou E, Kiely P, Walsh DA, Dixey J, et al. Validation of methods for converting the original Disease Activity Score (DAS) to the DAS28. Rheumatol. Int. 2018;38:2297-305.
20. Leong KP, Tan JWL, Gao X, Koh ET. Conversion among the 28-joint count activity indices for rheumatoid arthritis. Eur J Rheumatol. 2020;7:105-11.
21. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions, version 6.3: Cochrane; 2022 [updated (updated February 2022). Available from: www.training.cochrane.org/handbook.
22. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med. Res. Methodol. 2014;14:135.
23. Kuper IH, van Leeuwen MA, van Riel PL, Sluiter WJ, Houtman NM, Cats HA, et al. Influence of a ceiling effect on the assessment of radiographic progression in rheumatoid arthritis during the first 6 years of disease. J. Rheumatol. 1999;26:268-76.
24. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. Korean J. Anesthesiol. 2017;70:144-56.
25. Hair J, Black, W. C., Babin, B. J. & Anderson, R. E. Multivariate data analysis. 7th ed: Upper Saddle River, New Jersey: Pearson Educational International; 2010.
26. Byrne BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. . 3rd Edition ed: New York: Routledge; 2016. 460 p.
27. De Cock D, Vanderschueren G, Meyfroidt S, Joly J, Westhovens R, Verschueren P. Two-year clinical and radiologic follow-up of early RA patients treated with initial step up monotherapy or initial step down therapy with glucocorticoids, followed by a tight control approach: Lessons from a cohort study in daily practice. Clin. Rheumatol. 2014;33:125-30.
28. Urata Y, Uesato R, Tanaka D, Nakamura Y, Motomura S. Treating to target matrix metalloproteinase 3 normalisation together with disease activity score below 2.6 yields better effects than each alone in rheumatoid arthritis patients: T-4 Study. Ann. Rheum. Dis. 2012;71:534-40.
29. Bosello S, Fedele AL, Peluso G, Gremese E, Tolusso B, Ferraccioli G. Very early rheumatoid arthritis is the major predictor of major outcomes: Clinical ACR remission and radiographic non-progression. Ann. Rheum. Dis. 2011;70:1292-5.
30. Brenol CV, da Chakr RMS, Andrade NPB, Toni M, Laurindo IMM, Brenol JCT, et al. Daily practice feasibility and effectiveness of treating long-standing rheumatoid arthritis to target with synthetic disease-modifying antirheumatic drugs: a prospective cohort study. Clin. Rheumatol. 2015;34:1781-5.
31. Castrejón I, Pincus T, Soubrier M, Lin YC, Rat A, Combe B, et al. GUEPARD treat-to-target strategy is significantly more efficacious than ESPOIR routine care in early rheumatoid arthritis according to patient-reported outcomes and physician global estimate. Rheumatology (United Kingdom). 2013;52:1890-7.
32. Soubrier M, Lukas C, Sibilia J, Fautrel B, Roux F, Gossec L, et al. Disease activity score-driven therapy versus routine care in patients with recent-onset active rheumatoid arthritis: Data from the GUEPARD trial and ESPOIR cohort. Ann. Rheum. Dis. 2011;70:611-5.
33. Brinkmann GH, Norvang V, Norli ES, Grøvle L, Haugen AJ, Lexberg ÅS, et al. Treat to target strategy in early rheumatoid arthritis versus routine care – A comparative clinical practice study. Semin. Arthritis Rheum. 2019;48:808-14.
34. Norvang V, Brinkmann GH, Yoshida K, Lillegraven S, Aga AB, Sexton J, et al. Achievement of Remission in Two Early Rheumatoid Arthritis Cohorts Implementing Different Treat-to-Target Strategies. Arthritis and Rheumatology. 2020;72:1072-81.
35. Kuijper TM, Luime JJ, de Jong PH, Gerards AH, van Zeben D, Tchetverikov I, et al. Tapering conventional synthetic DMARDs in patients with early arthritis in sustained remission: 2-year follow-up of the tREACH trial. Ann. Rheum. Dis. 2016;75:2119-23.

36. Wabe NT, Sorich MJ, Wechalekar MD, Cleland LG, McWilliams L, Lee AT, et al. Effect of Adherence to Protocolized Targeted Intensifications of Disease-modifying Antirheumatic Drugs on Treatment Outcomes in Rheumatoid Arthritis: Results from an Australian Early Arthritis Cohort. J. Rheumatol. 2016;43:1643-9.
37. Maksymowych WP, FitzGerald O, Østergaard M, Homik J, Van Der Heijde D, Lambert RG, et al. Outcomes and findings of the international rheumatoid arthritis (RA) BIODAM cohort for validation of soluble biomarkers in RA. J. Rheumatol. 2020;47:796-808.
38. Ramiro S, Landewé RBM, Van Der Heijde D, Sepriano A, Fitzgerald O, Ostergaard M, 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. 2020;79:453-9.
39. Harrold LR, Reed GW, John A, Barr CJ, Soe K, Magner R, et al. Cluster-Randomized Trial of a Behavioral Intervention to Incorporate a Treat-to-Target Approach to Care of US Patients With Rheumatoid Arthritis. Arthritis Care Res. 2018;70:379-87.
40. Hodkinson B, Musenge E, Tikly M. Tight control of rheumatoid arthritis in a resource-constrained setting: A randomized controlled study comparing the clinical disease activity index and simplified disease activity index. Rheumatology (United Kingdom). 2015;54:1033-8.
41. Akdemir G, Markusse IM, Bergstra SA, Goekoop RJ, Molenaar ET, Van Groenendael JHLM, et al. Comparison between low disease activity or das remission as treatment target in patients with early active rheumatoid arthritis. RMD Open. 2018;4:e000649.
42. Allaart CF, Geokoop-Ruiterman YPM, de Vries-Bouwstra JK, Breedveld FC, Dijkmans BAC. Aiming at low disease activity in rheumatoid arthritis with initial combination therapy or initial monotherapy strategies: The BeSt study. Clin. Exp. Rheumatol. 2006;24:S77-S82.
43. Goekoop-Ruiterman YPM, De Vries-Bouwstra JK, Kerstens PJSM, Nielen MMJ, Vos K, Van Schaardenburg D, et al. DAS-driven therapy versus routine care in patients with recent-onset active rheumatoid arthritis. Ann. Rheum. Dis. 2010;69:65-9.
44. van der Kooij SM, Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Guler-Yuksel M, Zwinderman AH, Kerstens PJ, et al. Drug-free remission, functioning and radiographic damage after 4 years of response-driven treatment in patients with recent-onset rheumatoid arthritis. Ann. Rheum. Dis. 2009;68:914-21.
45. Bijlsma JWJ, Welsing PMJ, Woodworth TG, Middelink LM, Pethö-Schramm A, Bernasconi C, et al. Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. The Lancet. 2016;388:343-55.
46. Nam JL, Villeneuve E, Hensor EMA, Conaghan PG, Keen HI, Buch MH, 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-naive, rheumatoid arthritis (the IDEA study). Ann. Rheum. Dis. 2014;73:75-85.
47. Brahe CH, Dehlendorff C, Østergaard M, Johansen JS, Ørnbjerg LM, Hørslev-Petersen K, et al. Circulating serum interleukin-6, serum chitinase-3-like protein-1, and plasma vascular endothelial growth factor are not predictive for remission and radiographic progression in patients with early rheumatoid arthritis: post-hoc explorative and validatio. Scand. J. Rheumatol. 2018;47:259-69.
48. Hørslev-Petersen K, Hetland ML, Junker P, Pødenphant J, Ellingsen T, Ahlquist P, 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, parallel-group, placebo-controlled Trial. Ann. Rheum. Dis. 2014;73:654-61.
49. Hørslev-Petersen K, Hetland ML, Ørnbjerg LM, Junker P, Pødenphant J, Ellingsen T, et al. Clinical and radiographic outcome of a treat-to-target strategy using methotrexate and intra-articular glucocorticoids with or without adalimumab induction: A 2-year investigator-initiated, double-blinded, randomised, controlled trial (OPERA). Ann. Rheum. Dis. 2016;75:1645-53.

50. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): A single-blind randomised controlled trial. Lancet. 2004;364:263-9.
51. Saunders SA, Capell HA, Stirling A, Vallance R, Kincaid W, McMahon AD, et al. Triple therapy in early active rheumatoid arthritis: A randomized, single-blind, controlled trial comparing step-up and parallel treatment strategies. Arthritis Rheum. 2008;58:1310-7.
52. Haavardsholm EA, Aga AB, Olsen IC, Lillegraven S, Hammer HB, Uhlig T, et al. Ultrasound in management of rheumatoid arthritis: ARCTIC randomised controlled strategy trial. BMJ. 2016;354:i4205.
53. Steunebrink LMM, Versteeg GA, Vonkeman HE, ten Klooster PM, Kuper HH, Zijlstra TR, 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 Research and Therapy. 2016;18:60.
54. Ten Klooster PM, Versteeg LGA, Oude Voshaar MAH, De La Torre I, De Leonardis F, Fakhouri W, et al. Radiographic progression can still occur in individual patients with low or moderate disease activity in the current treat-to-target paradigm: Real-world data from the Dutch Rheumatoid Arthritis Monitoring (DREAM) registry. Arthritis Research and Therapy. 2019;21:237.
55. Vermeer M, Kuper HH, Hoekstra M, Haagsma CJ, Posthumus MD, Brus HLM, et al. Implementation of a treat-to-target strategy in very early rheumatoid arthritis: Results of the Dutch Rheumatoid Arthritis Monitoring remission induction cohort study. Arthritis Rheum. 2011;63:2865-72.
56. Versteeg GA, Steunebrink LMM, Vonkeman HE, ten Klooster PM, van der Bijl AE, van de Laar MAFJ. Long-term disease and patient-reported outcomes of a continuous treat-to-target approach in patients with early rheumatoid arthritis in daily clinical practice. Clin. Rheumatol. 2018;37:1189-97.
57. Stouten V, Westhovens R, Pazmino S, De Cock D, Van der Elst K, Joly J, et al. Five-year treat-to-target outcomes after methotrexate induction therapy with or without other csDMARDs and temporary glucocorticoids for rheumatoid arthritis in the CareRA trial. Ann. Rheum. Dis. 2021;02:02.
58. Stouten V, Westhovens R, Pazmino S, De Cock D, Van der Elst K, Joly J, et al. Effectiveness of different combinations of DMARDs and glucocorticoid bridging in early rheumatoid arthritis: two-year results of CareRA. Rheumatology (Oxford). 2019;58:2284-94.
59. Verschueren P, De Cock D, Corluy L, Joos R, Langenaken C, Taelman V, 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-l. Ann. Rheum. Dis. 2017;76:511-20.
60. Møller-Bisgaard S, Hørslev-Petersen K, Ejbjerg B, Hetland ML, Ørnbjerg LM, Glinatsi D, et al. Effect of magnetic resonance imaging vs conventional treat-to-target strategies on disease activity remission and radiographic progression in rheumatoid arthritis the IMAGINE-RA randomized clinical trial. JAMA - Journal of the American Medical Association. 2019;321:461-72.
61. Ter Wee MM, Den Uyl D, Boers M, Kerstens P, Nurmohamed M, Van Schaardenburg D, 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. 2015;74:1233-40.
62. Brown Z, Metcalf R, Bednarz J, Stavrou C, Spargo L, James M, et al. Modifiable factors associated with response to treatment in early rheumatoid arthritis. Arthritis and Rheumatology. 2019;71:4198-200.
63. El Miedany Y, El Gaafary M, Ahmed I, Youssef S, Nasr A. US guided treat-to-target approach in early RA: Implications for uncoupling of disease activity and structural joint damage. Curr. Rheumatol. Rev. 2015;11:18-27.
64. Fedele AL, Petricca L, Tolusso B, Alivernini S, Canestri S, Di Mario C, et al. Interleukin-6 and IgA-rheumatoid factor are crucial for baseline erosiveness, and anti-citrullinated peptide antibodies for radiographic progression in early rheumatoid arthritis treated according to a treat-to-target strategy. Scand. J. Rheumatol. 2018;47:351-9.

65. Montecucco C, Todoerti M, Sakellariou G, Scirè CA, Caporali R. Low-dose oral prednisone improves clinical and ultrasonographic remission rates in early rheumatoid arthritis: Results of a 12-month open-label randomised study. Arthritis Research and Therapy. 2012;14:R112.
66. Pope JE, Haraoui B, Rampakakis E, Psaradellis E, Thorne C, Sampalis JS. Treating to a target in established active rheumatoid arthritis patients receiving a tumor necrosis factor inhibitor: Results from a real-world cluster-randomized adalimumab trial. Arthritis Care Res. 2013;65:1401-9.
67. Song JJ, Song YW, Bae SC, Cha HS, Choe JY, Choi SJ, et al. Treat-to-Target Strategy for Asian Patients with Early Rheumatoid Arthritis: Result of a Multicenter Trial in Korea. J. Korean Med. Sci. 2018;33:e346.
68. Tam LH, Shang Q, Li EK, Wong PC, Kwok KY, Kun EW, et al. Effect of Treat-to-target Strategies Aiming at Remission of Arterial Stiffness in Early Rheumatoid Arthritis: A Randomized Controlled Study. J. Rheumatol. 2018;45:1229-39.
69. Teh CL, Wong JS. Impact of tight control strategy on rheumatoid arthritis in Sarawak. Clin. Rheumatol. 2011;30:615-21.
70. Verschueren P, Esselens G, Westhovens R. Daily practice effectiveness of a step-down treatment in comparison with a tight step-up for early rheumatoid arthritis. Rheumatology. 2008;47:59-64.
71. Zhao J, Wang Y, Geng Y, Zhang X, Deng X, Ji L, et al. Intensive therapy alleviates subclinical synovitis on ultrasound and disease activity and reduces flare in rheumatoid arthritis patients who have achieved clinical target - a randomized controlled trial. Semin. Arthritis Rheum. 2020;50:673-9.
72. Bugatti S, Manzo A, Benaglio F, Klersy C, Vitolo B, Todoerti M, et al. Serum levels of CXCL13 are associated with ultrasonographic synovitis and predict power Doppler persistence in early rheumatoid arthritis treated with non-biological disease-modifying anti-rheumatic drugs. Arthritis Res. Ther. 2012;14:R34-R.
73. Horton SC, Tan AL, Freeston JE, Wakefield RJ, Buch MH, Emery P. Discordance between the predictors of clinical and imaging remission in patients with early rheumatoid arthritis in clinical practice: implications for the use of ultrasound within a treatment-to-target strategy. Rheumatology (Oxford). 2016;55:1177-87.
74. Sugihara T, Ishizaki T, Hosoya T, Iga S, Yokoyama W, Hirano F, et al. Structural and functional outcomes of a therapeutic strategy targeting low disease activity in patients with elderly-onset rheumatoid arthritis: A prospective cohort study (CRANE). Rheumatology (United Kingdom). 2014;54:798-807.
75. Sugihara T, Ishizaki T, Onoguchi W, Baba H, Matsumoto T, Iga S, et al. Effectiveness and safety of treat-to-target strategy in elderly-onset rheumatoid arthritis: a 3-year prospective observational study. Rheumatology (Oxford). 2021;60:4252-61.
76. Huang H, Xie W, Geng Y, Fan Y, Wang Y, Zhao J, et al. Towards a Better Implementation of Treat-to-Target Strategy in Rheumatoid Arthritis: A Comparison of Two Real-World Cohorts. Rheumatology and Therapy. 2022;9:907-17.
77. de Andrade NPB, da Silva Chakr RM, Xavier RM, Viecceli D, Correa RHB, de Oliveira Filho CM, et al. Long-term outcomes of treat-to-target strategy in established rheumatoid arthritis: a daily practice prospective cohort study. Rheumatol. Int. 2017;37:993-7.
78. Allaart CF, Breedveld FC, Dijkmans BAC. Treatment of recent-onset rheumatoid arthritis: Lessons from the BeSt study. J. Rheumatol. 2007;34:25-33.
79. Klarenbeek NB, Guler-Yuksel M, van der Kooij SM, Han KH, Ronday HK, Kerstens PJ, et al. The impact of four dynamic, goal-steered treatment strategies on the 5-year outcomes of rheumatoid arthritis patients in the BeSt study. Ann. Rheum. Dis. 2011;70:1039-46.
80. Markusse IM, Akdemir G, Dirven L, Goekoop-Ruiterman YPM, Van Groenendael JHLM, Han KH, 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. 2016;164:523-31.
81. Van Den Broek M, Lems WF, Allaart CF. BeSt practice: The success of early-targeted treatment in rheumatoid arthritis. Clin. Exp. Rheumatol. 2012;30:S35-S8.
82. Van Der Kooij SM, De Vries-Bouwstra JK, Goekoop-Ruiterman YPM, Ewals JAPM, Han KH, Hazes JMW, et al. Patient-reported outcomes in a randomized trial comparing four different treatment strategies in recent-onset rheumatoid arthritis. Arthritis Care Res. 2009;61:4-12.

83. van der Woude D, Visser K, Klarenbeek NB, Karel Ronday H, Peeters AJ, Kerstens PJSM, et al. Sustained drug-free remission in rheumatoid arthritis after DAS-driven or non-DAS-driven therapy: A comparison of two cohort studies. Rheumatology (United Kingdom). 2012;51:1120-8.
84. Teitsma XM, Jacobs JWG, Welsing PMJ, Pethö-Schramm A, Borm MEA, van Laar JM, et al. Radiographic joint damage in early rheumatoid arthritis patients: comparing tocilizumab- and methotrexate-based treat-to-target strategies. Rheumatology (Oxford). 2018;57:309-17.
85. Verhoeven MMA, Tekstra J, Welsing PMJ, Pethö-Schramm A, Borm MEA, Bruyn GAW, 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 (United Kingdom). 2020;59:2325-33.
86. Schipper LG, Vermeer M, Kuper HH, Hoekstra MO, Haagsma CJ, Den Broeder AA, et al. A tight control treatment strategy aiming for remission in early rheumatoid arthritis is more effective than usual care treatment in daily clinical practice: A study of two cohorts in the Dutch Rheumatoid Arthritis Monitoring registry. Ann. Rheum. Dis. 2012;71:845-50.
87. Versteeg GA, Steunebrink LMM, Vonkeman HE, Ten Klooster PM, Van Der Bijl AE, Van De Laar MAFJ. Early radiological progression remains associated with long-term joint damage in real-world rheumatoid arthritis patients treated to the target of remission. Scand. J. Rheumatol. 2022;51:87-96.
88. Gullick NJ, Ibrahim F, Scott IC, Vincent A, Cope AP, Garrood T, et al. Real world long-term impact of intensive treatment on disease activity, disability and health-related quality of life in rheumatoid arthritis. BMC Rheumatology. 2019;3:6.
89. Ebrahimian S, Salami A, Malek Mahdavi A, Esalatmanesh K, Khabbazi A, Hajialilo M. Can treating rheumatoid arthritis with disease-modifying anti-rheumatic drugs at the window of opportunity with tight control strategy lead to long-term remission and medications free remission in real-world clinical practice? A cohort study. Clin. Rheumatol. 2021;40:4485-91.
90. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, 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. 2008;58:S126-35.
91. Vermeer M, Kuper HH, Moens HJ, Drossaers-Bakker KW, van der Bijl AE, van Riel PL, et al. Sustained beneficial effects of a protocolized treat-to-target strategy in very early rheumatoid arthritis: three-year results of the Dutch Rheumatoid Arthritis Monitoring remission induction cohort. Arthritis Care Res. (Hoboken). 2013;65:1219-26.
92. Allaart CF, Markusse IM, Lems WF. What have we learned from BeSt? Clin. Immunol.;186:74-8.
93. Dale J, Stirling A, Zhang R, Purves D, Foley J, Sambrook M, et al. Targeting ultrasound remission in early rheumatoid arthritis: The results of the TaSER study, a randomised clinical trial. Ann. Rheum. Dis. 2016;75:1043-50.
94. Akdemir G, Markusse IM, Goekoop-Ruiterman YP, Steup-Beekman GM, Grillet BA, Kerstens PJ, et al. Rheumatologists' adherence to a disease activity score steered treatment protocol in early arthritis patients is less if the target is remission. Clin. Rheumatol. 2017;36:317-26.
95. Tam LH, Shang Q, Li EK, Wong PC, Kwok KY, Kun EW, et al. Effect of treat-to-target strategies aiming at remission of arterial stiffness in early rheumatoid arthritis: A randomized controlled study. J. Rheumatol. 2018;45:1229-39.
96. Hock ES, Martyn-St James, M., Wailoo, A., Scott D.L., Stevensen M., Rawdin A., Simpson E., Dracup N., Young A. . Treat-to-Target Strategies in Rheumatoid Arthritis: a Systematic Review and Cost-Effectiveness Analysis. SN Comprehensive Clinical Medicine. 2021;3:838-54.
97. Bergstra SA, Allaart CF. What is the optimal target for treat-to-target strategies in rheumatoid arthritis? Curr. Opin. Rheumatol. 2018;30:282-7.
98. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: A review of its history, issues, progress, and documentation. J. Rheumatol. 2003;30:167-78.

Supplemental material

Supplemental figure S1: search performed in OVID (MEDLINE database) on 01-06-2022

# ▲	Searches	Results	Type	Actions	Annotations
1	(T2T or "treatment target$1" or (treat$ adj3 target) or "targeted treatment" or ("treating rheumatoid arthritis" adj3 target) or "target oriented approach" or (("disease activity" or DAS or DAS28 or remission or DFR or CDAI or SDAI or RADAI or RAPID3) adj2 (steered or driven or guided)) or "objectively steered" or "intensive management" or "intensive outpatient management" or "intensive disease management" or "step up" or (tight adj3 control$) or "tight$2 controlled" or (aggressive adj (care or treatment or therapy)) or "goal directed therapy" or "goal steered treatment" or ((systematic or standardized or standardised) adj monitoring) or (target$ adj3 (remission or "low disease activity" or LDA)) or (("aiming at" or "aiming for" or "aimed at" or "aim of treatment") adj2 (remission or "disease activity" or DAS))).ab,ti.	80015	Advanced	Display Results | More	
2	rheumatoid arthritis.ab,kf,ti.	115858	Advanced	Display Results | More	
3	*Arthritis, Rheumatoid/	87058	Advanced	Display Results | More	
4	("randomized controlled trial" or "randomised controlled trial" or RCT or "cohort study" or "clinical trial" or "systematic review" or "meta-analysis").ab,ti.	847747	Advanced	Display Results | More	
5	exp Clinical Trials as Topic/	374453	Advanced	Display Results | More	
6	exp Random Allocation/	106851	Advanced	Display Results | More	
7	exp Therapeutics/	4995109	Advanced	Display Results | More	
8	exp Drug Therapy/	1467895	Advanced	Display Results | More	
9	exp Cohort Studies/	2354530	Advanced	Display Results | More	
10	("clinical trial" or review or "systematic review" or "meta-analysis" or "randomized controlled trial" or "randomised controlled trial" or "controlled clinical trial" or "clinical study").pt.	4044413	Advanced	Display Results | More	
11	4 or 5 or 6 or 7 or 8 or 9 or 10	9638122	Advanced	Display Results | More	
12	2 or 3	133885	Advanced	Display Results | More	
13	exp Animals/	25530668	Advanced	Display Results | More	
14	exp Humans/	20517682	Advanced	Display Results | More	
15	13 not 14	5012986	Advanced	Display Results | More	
16	1 and 11 and 12	1392	Advanced	Display Results | More	
17	16 not 15	1381	Advanced	Display Results | More	
18	16 not 15	1381	Advanced	Display Results | More	
19	limit 18 to (english language and yr="2003 -Current")	1176	Advanced	Display Results | More	

Supplemental figure S2: search performed in Embase on 01-06-2022

☐	**#11**	#1 AND #6 AND #9 NOT #7 AND [english]/lim AND [2003-2022]/py	1,740
☐	**#10**	#1 AND #6 AND #9	1,911
☐	**#9**	#2 OR #8	202,191
☐	**#8**	**'rheumatoid arthritis'**/mj	133,487
☐	**#7**	**'animal'**/exp NOT **'human'**/exp	5,792,476
☐	**#6**	#3 OR #4 OR #5	5,711,787
☐	**#5**	**review**:it OR **'clinical trial'**:it OR **'systematic review'**:it OR **'meta-analysis'**:it OR **'randomized controlled trial'**:it OR **'randomised controlled trial'**:it OR **'controlled clinical trial'**:it OR **'clinical study'**:it	2,935,445
☐	**#4**	**'randomized controlled trial'**/exp OR **'randomization'**/exp OR **'clinical trial'**/exp OR **'cohort analysis'**/exp OR **'systematic review'**/exp OR **'meta analysis'**/exp	2,899,190
☐	**#3**	**'randomized controlled trial'**:ab,ti OR **'randomised controlled trial'**:ab,ti OR **rct**:ab,ti OR **'cohort study'**:ab,ti OR **'clinical trial'**:ab,ti OR **'systematic review'**:ab,ti OR **'meta analysis'**:ab,ti	1,161,022
☐	**#2**	**'rheumatoid arthritis'**:ti,ab,kw	178,895
☐	**#1**	**t2t**:ab,ti OR **'treatment target*'**:ab,ti OR ((**treat*** NEAR/3 **target**):ab,ti) OR **'targeted treatment'**:ab,ti OR ((**'treating rheumatoid arthritis'** NEAR/3 **target**):ab,ti) OR **'target oriented approach'**:ab,ti OR (((**'disease activity'** OR **das28** OR **das** OR **remission** OR **dfr** OR **cdai** OR **sdai** OR **radai** OR **rapid3**) NEAR/2 (**steered** OR **driven** OR **guided**)):ab,ti) OR **'objectively steered'**:ab,ti OR **'intensive management'**:ab,ti OR **'intensive outpatient management'**:ab,ti OR **'intensive disease management'**:ab,ti OR **'step up'**:ab,ti OR ((**tight** NEAR/3 **control***):ab,ti) OR **'tight* controlled'**:ab,ti OR ((**aggressive** NEXT/1 (**care** OR **treatment** OR **therapy**)):ab,ti) OR **'goal directed therapy'**:ab,ti OR **'goal steered treatment'**:ab,ti OR (((**systematic** OR **standardized** OR **standardised**) NEXT/1 **monitoring**):ab,ti) OR ((**target*** NEAR/3 (**remission** OR **'low disease activity'** OR **lda**)):ab,ti) OR (((**'aiming at'** OR **'aiming for'** OR **'aimed at'** OR **'aim of treatment'**) NEAR/2 (**remission** OR **'disease activity'** OR **das**)):ab,ti)	120,745

Supplemental figure S3: search performed in Cochrane database on 01-06-2022

ID	Search	Hits
#1	((T2T OR "treatment target*" OR (treat* NEAR/2 target) OR "targeted treatment" OR ("treating rheumatoid arthritis" NEAR/3 target) OR "target oriented approach" OR (("disease activity" OR DAS OR DAS28 OR remission OR DFR OF CDAI OR SDAI OR RADAI OR RAPID3) NEAR/2 (steered OR driven OR guided)) OR "objectively steered" OR "intensive management" OR "intensive outpatient management" OR "intensive disease management" OR "step up" OR (tight NEAR/3 control*) OR (tight* NEXT controlled) OR (aggressive NEXT (care OR treatment OR therapy)) OR "goal directed therapy" OR "goal steered treatment" OR ((systematic OR standardized OR standardised) NEXT monitoring) OR (target* NEAR/3 (remission OR "low disease activity" OR LDA)) OR (("aiming at" OR "aiming for" OR "aimed at" OR "aim of treatment") NEAR/2 (remission OR "disease activity" OR DAS))):ti,ab,kw (Word variations have been searched)	17800
#2	(("rheumatoid arthritis")):ti,ab,kw (Word variations have been searched)	16699
#3	MeSH descriptor: [Arthritis, Rheumatoid] this term only	6150
#4	MeSH descriptor: [Clinical Trials as Topic] explode all trees	48676
#5	MeSH descriptor: [Random Allocation] explode all trees	20677
#6	MeSH descriptor: [Therapeutics] explode all trees	327999
#7	MeSH descriptor: [Drug Therapy] explode all trees	147897
#8	MeSH descriptor: [Cohort Studies] explode all trees	159921
#9	(("clinical trial" OR "systematic review" OR "meta-analysis" OR "randomized controlled trial" OR "randomised controlled trial" OR "controlled clinical trial" OR "clinical study")):pt (Word variations have been searched)	640747
#10	MeSH descriptor: [Animals] explode all trees	647954
#11	(("randomized controlled trial" OR "randomised controlled trial" OR RCT OR "cohort study" OR "clinical trial" OR "systematic review" OR "meta analysis")):ti,ab,kw (Word variations have been searched)	775820
#12	MeSH descriptor: [Humans] explode all trees	647921
#13	#2 OR #3	17332
#14	#4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #11	1143210
#15	#10 NOT #12	33
#16	#1 AND #13 AND #14	664
#17	#16 NOT #15	664

Supplemental figure S4: search performed in preMEDLINE database on 01-06-2022

#13	···	›	Search: **#11 AND #12**	113	07:24:38
#12	···	›	Search: **"pubstatusaheadofprint"[All Fields] OR "publisher"[Filter] OR "pubmednotmedline"[Filter]**	4,707,474	07:23:47
#11	···	›	Search: **#10 NOT #7**	1,759	07:23:16
#10	···	›	Search: **#1 AND #8 AND #9**	1,767	07:23:00
#9	···	›	Search: **#4 OR #5 OR #6**	9,669,394	07:22:43
#8	···	›	Search: **#2 OR #3**	159,576	07:22:20
#7	···	›	Search: **"animals"[MeSH Terms] NOT "humans"[MeSH Terms]**	5,010,889	07:16:38
#6	···	›	Search: **"clinical trial"[Publication Type] OR "review"[Publication Type] OR "systematic review"[Publication Type] OR "meta analysis"[Publication Type] OR "randomized controlled trial"[Publication Type] OR "controlled clinical trial"[Publication Type] OR "clinical study"[Publication Type]**	4,205,976	07:16:27
#5	···	›	Search: **"clinical trials as topic"[MeSH Terms] OR "random allocation"[MeSH Terms] OR "therapeutics"[MeSH Terms] OR "drug therapy"[MeSH Terms] OR "cohort studies"[MeSH Terms]**	6,789,890	07:16:14
#4	···	›	Search: **"randomized controlled trial"[Title/Abstract] OR "randomised controlled trial"[Title/Abstract] OR RCT[Title/Abstract] OR "cohort study"[Title/Abstract] OR "clinical trial"[Title/Abstract] OR "systematic review"[Title/Abstract] OR "meta analysis"[Title/Abstract]**	866,551	07:16:04
#3	···	›	Search: **rheumatoid arthritis[MeSH Terms]**	121,464	07:15:52
#2	···	›	Search: **"rheumatoid arthritis"[Title/Abstract]**	115,887	07:15:43
#1	···	›	Search: **T2T[Title/Abstract] OR "treatment target*"[Title/Abstract] OR "treat to target"[Title/Abstract] OR "treating to target"[Title/Abstract] OR "targeted treatment"[Title/Abstract] OR "target oriented approach"[Title/Abstract] OR (("disease activity"[Title/Abstract] OR DAS[Title/Abstract] OR DAS28[Title/Abstract] OR remission[Title/Abstract] OR DFR[Title/Abstract] OR CDAI[Title/Abstract] OR SDAI[Title/Abstract] OR RADAI[Title/Abstract] OR RAPID3[Title/Abstract]) AND (steered[Title/Abstract] OR driven[Title/Abstract] OR guided[Title/Abstract])) OR "intensive management"[Title/Abstract] OR "intensive outpatient management"[Title/Abstract] OR "intensive disease management"[Title/Abstract] OR "step up"[Title/Abstract] OR "tight control*"[Title/Abstract] OR "tightly controlled"[Title/Abstract] OR "aggressive care"[Title/Abstract] OR "aggressive therapy"[Title/Abstract] OR "aggressive treatment"[Title/Abstract] OR "goal directed therapy"[Title/Abstract] OR "systematic monitoring"[Title/Abstract] OR "standardized monitoring"[Title/Abstract] OR "standardised monitoring"[Title/Abstract] OR (target*[Title/Abstract] AND (remission[Title/Abstract] OR "low disease activity"[Title/Abstract] OR LDA[Title/Abstract])) OR ((aiming[Title/Abstract] OR "aimed at"[Title/Abstract] OR "aim of treatment"[Title/Abstract]) AND (remission[Title/Abstract] OR "disease activity"[Title/Abstract] OR DAS[Title/Abstract]))**	56,934	07:15:21

Supplemental Table S1. Characteristics of selected studies

Author/trial	Year	Cohort /RCT	Treatment arms extracted	Target	Early RA, Formal treatment protocol, bDMARD available	Risk of Bias	Used in analyses
de Cock et al.[27]	2014	Cohort	1. Initial monotherapy (n=42) 2. Initial combination therapy (n=32)	DAS28-CRP<3.2	Yes, No, Yes	Low	DAS-remission, mean DAS, HAQ and SHS year 1 and 2-3
T-4 study [28]	2012	RCT	DAS-driven arm (n=56)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	DAS- and SCB-remission, mean DAS, HAQ and SHS year 1
Bosello et al.[29]	2011	Cohort	DAS-driven (n=121)	DAS44ESR ≤2.4	Yes, Yes, Yes	Low	DAS- and SCB-remission, mean DAS and HAQ year 1
Brenol et al. [30]	2015	Cohort	All patients (n=241)	DAS28ESR < 3.2	No, No, No	Some concerns	DAS- and SCB-remission, mean HAQ, SDAI/CDAI and DAS year 1
GUEPARD [31]	2013	RCT	GUEPARD arm (n=65)	DAS28-ESR<3.2	Yes, Yes, Yes	Some concerns	Mean DAS, HAQ and SDAI/CDAI year 1
GUEPARD [32]	2011	RCT	GUEPARD arm (n=65)	DAS28-ESR<3.2	Yes, Yes, Yes	Some concerns	Mean SHS year 1
NOR-VEAC [33]	2019	Cohort	T2T cohort (n=293)	DAS28-ESR<2.6	Yes, No, Yes	Some concerns	DAS-remission, mean SDAI/CDAI and DAS at year 1, SCB remission and mean DAS at year 2-3
NOR-VEAC [34]	2020	Cohort	NOR-VEAC (n=330)	DAS28-ESR<2.6	Yes, No, Yes	Some concerns	DAS-remission year 2-3, SCB-remission year 1 and 2-3
tREACH[35]	2016	Cohort	1. Initial triple therapy (n=281) 2. Initial monotherapy (n=281)	DAS44-ESR≤2.4	Yes, Yes, Yes	Low	DAS-remission year 2-3
Wabe et al.[36]	2016	Cohort	All patients (n=198 at year 1 and n=149 at year 3)	DAS28-ESR<2.6	Yes, Yes, No	Some concerns	DAS- and SCB-remission, mean HAQ and mean SHS year 1 and year 2
BIODAM[37]	2019	Cohort	All patients (n=571)	DAS44-ESR<1.6	No, No, Yes	Some concerns	DAS-remission at year 2-3 Mean DAS, HAQ at years 1 and 2-3
BIODAM[38]	2020	Cohort	All patients (n=571)	DAS44-ESR<1.6	No, No, Yes	Some concerns	DAS remission year 1, SCB remission year 1 and 2-3
Harrold et al. [39]	2018	RCT	Treat-to-target (n=246)	CDAI≤10	No, No, Yes	High	Mean HAQ year 1
Hodkinson et al. [40]	2015	RCT	1. SDAI-target (n=42) 2. CDAI-target (n=60)	1. SDAI≤11 2. CDAI≤10	Yes, Yes, No	Some concerns	DAS-remission, mean DAS and HAQ year 1
BeSt [41]	2018	RCT	Initial combination therapy with prednisone (arm 3 of BeSt) (n=133)	DAS44-ESR ≤2.4	Yes, Yes, Yes	Low	DAS- and SCB-remission and mean DAS year 1
BeSt [42]	2006	RCT	1. Sequential monotherapy (n=126) 2. Step up combination therapy (n=121) 3. Initial combination therapy with prednisone (n=133) 4. Initial combination therapy with infliximab (n=128)	DAS44-ESR ≤2.4	Yes, Yes, Yes	Low	Mean HAQ and SHS year 2-3
BeSt [90]	2005	RCT	1. Sequential monotherapy (n=126) 2. Step up combination therapy (n=121)	DAS44-ESR ≤2.4	Yes, Yes, Yes	Low	Mean HAQ and SHS year 1

			3. Initial combination therapy with tapered high-dose prednisone (n=133) 4. Initial combination therapy with infliximab (n=128)				
BeSt [43]	2009	RCT	1. Sequential monotherapy + Step up combination therapy (group 1+2 BeSt), (n=234)	DAS44-ESR ≤2.4	Yes, Yes, Yes	Low	DAS-remission and mean DAS year 1
BeSt [44]	2015	RCT	1. Sequential monotherapy (n=115) 2. Step up combination therapy (n=101) 3. Initial combination therapy with prednisone (n=119) 4. Initial combination therapy with infliximab (n=119)	DAS44ESR ≤2.4	Yes, Yes, Yes	Low	DAS-remission 4-6 years
U-Act-Early [45]	2016	RCT	1. Tocilizumab + methotrexate (n=106) 2. Tocilizumab + placebo methotrexate (n=103) 3. Methotrexate + placebo tocilizumab (n=108)	DAS28-ESR<2.6	Yes, Yes, Yes	Low	Mean HAQ and SHS year 1 and 2-3
IDEA [46]	2013	RCT	1. Methotrexate + methylprednisolone (n=57) 2. Methotrexate + infliximab (n=55)	DAS44-ESR ≤2.4	Yes, Yes, Yes	Low	DAS- and SCB remission, mean HAQ and SHS year 1
CIMESTRA and OPERA [47]	2018	RCT	1. CIMESTRA (n=150) 2. OPERA (all patients) (n=180)	1. (SJC 0) 2. DAS28CRP ≤3.2	1. Yes, Yes, No 2. Yes, Yes, Yes	1. Some concerns 2. Low	SCB remission year 2-3
OPERA [48]	2013	RCT	1. Adalimumab + methotrexate (n=89) 2. Methotrexate + placebo adalimumab (n=91)	DAS28CRP ≤3.2	Yes, Yes, Yes	Low	DAS- and SCB-remission year 1 and mean DAS, HAQ, SDAI/CDAI year 1
OPERA [49]	2015	RCT	1. Adalimumab + methotrexate (n=89) 2. Methotrexate + placebo adalimumab (n=91)	DAS28-CRP ≤3.2	Yes, Yes, Yes	Low	DAS-remission and mean DAS, HAQ and SHS year 2-3
TICORA [50]	2004	RCT	Intensive group (n=55)	DAS28-ESR<3.2	Yes, Yes, No	Low	SCB remission, mean DAS, HAQ and SHS year 1
Saunders et al. [51]	2008	RCT	1. Step-up therapy (n=47) 2. Triple therapy (n=49)	DAS28-ESR<3.2	Yes, Yes, Yes	Some concerns	SCB remission, mean DAS, HAQ and SHS year 1
ARCTIC [52]	2016	RCT	Conventional tight control (n=112)	DAS44-ESR<1.6	Yes, Yes, Yes	Low	DAS- and SCB-remission, mean DAS and SHS years 1 and 2-3 Mean SDAI/CDAI year 1
DREAM [53]	2016	Cohort	1. Step-up therapy (n=128) 2. Initial combination therapy (n=128)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	DAS remission year 1
DREAM [54]	2019	Cohort	All patients (n=229)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	Mean DAS year 2-3
DREAM [55]	2011	Cohort	All patients (n=534)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	Mean SHS year 1
DREAM [91]	2013	Cohort	All patients (n=342)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	SCB-remission year 1 and 2-3

							DAS-remission, mean HAQ and SHS year 2-3
DREAM [56]	2017	Cohort	All patients (n=229)	DAS28-ESR<2.6	Yes, Yes, Yes	Some concerns	DAS-remission year 4-6, mean HAQ and DAS year 1
CareRA [57]	2021	RCT	1. COBRA Classic (n=69) 2. COBRA Slim, high risk (n=75) 3. COBRA Avant-Garde (n=59) 4. COBRA Slim, low risk (n=23) 5. Methotrexate tight step-up (n=26)	DAS28-CRP≤3.2	Yes, Yes, No	Some concerns	DAS remission years 1, 2-3 and 4-6, SCB remission year 1
CareRA [58]	2019	RCT	1. COBRA Classic (n=98) 2. COBRA Slim, high risk (n=98) 3. COBRA Avant-Garde (n=93) 4. COBRA Slim, low risk (n=43) 5. Methotrexate tight step-up (n=47)	DAS28-CRP≤3.2	Yes, Yes, No	Some concerns	SCB remission, mean SHS and HAQ year 2-3
CareRA [59]	2016	RCT	1. COBRA Classic (n=98) 2. COBRA Slim, high risk (n=98) 3. COBRA Avant-Garde (n=93) 4. COBRA Slim, low risk (n=43) 5. Methotrexate tight step-up (n=47)	DAS28-CRP≤3.2	Yes, Yes, No	Some concerns	Mean DAS, HAQ and SHS year 1
IMAGINE-RA [60]	2019	RCT	Conventional treat-to-target (n=100)	DAS28-CRP≤3.2	No, Yes, Yes	Low	DAS- and SCB-remission, mean DAS, HAQ and SHS year 2-3
COBRA-light [61]	2014	RCT	1. COBRA (n=81) 2. COBRA-light (n=81)	DAS44-ESR<1.6	Yes, Yes, Yes	Low	DAS- and SCB-remission, mean DAS, HAQ and SHS year 1
Brown et al. [62]	2019	Cohort	All patients (n=300)	DAS28-ESR<2.6	Yes, Yes, Yes	Low	DAS-remission year 1
El Miedany et al. [63]	2015	Cohort	1. Biological DMARD + conventional DMARD (n=192) 2. Conventional DMARD (n=288)	DAS28-ESR≤3.2	Yes, No, Yes(arm1) No(arm2)	Low	Mean DAS and HAQ year 1
Fedele et al. [64]	2018	Cohort	ERA-patients (n=408)	DAS44-ESR≤2.4	Yes, Yes, Yes	Some concerns	DAS- and SCB-remission year 1
Montecucco et al. [65]	2012	RCT	1. Methotrexate (n=110) 2. Methotrexate + low dose prednisone (n=110)	DAS28-ESR≤3.2	Yes, Yes, Yes	Some concerns	DAS- and SCB-remission and mean DAS year 1
Pope et al. [66]	2013	RCT	1. DAS-group (n=100) 2. SJC-group (n=99)	1. DAS28-ESR<2.6 2. (SJC 0)	No, No, Yes	Low	DAS-remission, mean DAS and HAQ year 1
Song et al. [67]	2018	Cohort	1. Remission-steered (n=163) 2. (LDA or remission – steered, n=163))	1. DAS28-ESR<2.6 (2. DAS28-ESR < 2.6 or <3.2)	Yes, No, Yes	High	DAS-remission year 1
Tam et al. [68]	2018	RCT	1. SDAI-steered (n=57) 2. DAS-steered (n=60)	1. SDAI≤3.3 2. DAS28-CRP<2.6	Yes, Yes, Yes	Some concerns	DAS- and SCB remission year 1, mean DAS, HAQ, and SDAI/CDAI year 1
Teh et al. [69]	2011	Cohort	All patients (n=142)	DAS28-ESR<2.6	No, No, Yes	Some concerns	DAS-remission and mean DAS year 1
Verschueren et al. [70]	2008	Cohort	Step-up therapy (n=52)	DAS28-CRP<3.2	Yes, No, No	Low	DAS-remission and mean DAS year 1

Zhao et al. [71]	2020	RCT	Tight control arm (n=54)	DAS28-CRP≤3.2	No, Yes, Yes	Some concerns	DAS-remission, mean DAS and SDAI/CDAI year 1
Bugatti et al. [72]	2012	Cohort	All patients (n=161)	DAS44-ESR≤2.4	Yes, Yes, No	Low	DAS- and SCB-remisison year 1
Horton et al. [73]	2016	Cohort	All patients (n=105)	DAS28-CRP < 2.6	Yes, No, Yes	Some concerns	DAS- and SCB-remission year 1
CRANE [74]	2014	Cohort	All patients (n=145)	DAS28-ESR < 3.2	Yes, Yes, Yes	Some concerns	Mean HAQ year 1
CRANE [75]	2021	Cohort	All patients (n=197)	DAS28-ESR < 3.2	Yes, Yes, Yes	Some concerns	DAS- and SCB-remission year 1 and 2-3
Huang et al. [76]	2022	Cohort	1. TARRA cohort (n=389 at year 1, n=283 at year 2) 2. CENTRA cohort (n=111 at year 1, n=67 at year 2)	SDAI <= 11	No, No, Yes	1. Some concerns 2. Some concerns	DAS- and SCB-remission and mean DAS year 1 and 2-3. SDAI-CDAI year 1.

(b)DMARD: (biological) disease modifying antirheumatic drug, CDAI: Clinical Disease Activity Index, DAS(28/44): Disease activity score (based on a 28/44-joint count), CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, HAQ: Health assessment questionnaire, ITS: intention to study analysis, ITT: intention to treat analysis, LDA: low disease activity, NM:not mentioned, PP: per protocol analysis, RA: rheumatoid arthritis, SCB-remission: SDAI/CDAI/Boolean remission, SDAI: Simplified Disease Activity Index, RCT: Randomized Controlled Trial, SHS: Sharp van der Heijde Score, T2T: treat-to-target.

Articles that met the inclusion criteria but were not used in the meta-regression analyses, and the reason not to use them

Author/trial	Year	Cohort/ RCT	Treatment arms extracted	Target	Risk of Bias	Outcome measures reported	Reason not used in analysis
de Andrade et al.[77]	2017	Cohort	All patients (n=229)	DAS28-ESR<2.6	Some concerns	Mean DAS, CDAI and HAQ at year 7	Not enough arms at this time-point (7years) for analysis
BeSt [78]	2007	RCT	Sequential monotherapy group and Step up combination therapy group (arms 1 and 2 of BeSt) (n=247)	DAS44-ESR ≤2.4	Low	Mean change in DAS, HAQ, and SHS at year 1	Duplicate results with other extracted BeST papers
BeSt [92]	2017	RCT	1. Sequential monotherapy (n=109) 2. Step up combination therapy (n=104) 3. Initial combination therapy with prednisone (n=109) 4. Initial combination therapy with infliximab (n=107)	DAS44-ESR ≤2.4	Low	Mean change in SHS at year 10	Not enough arms at this time-point for analysis
BeSt [79]	2014	RCT	1. Sequential monotherapy (n=111) 2. Step up combination therapy (n=94) 3. Initial combination therapy with prednisone (n=113) 4. Initial combination therapy with infliximab (n=116)	DAS44-ESR ≤2.4	Low	Mean change in SHS at year 5	Not enough arms at this time-point for analysis
BeSt [80]	2016	RCT	1. Sequential monotherapy (n=126) 2. Step up combination therapy (n=121) 3. Initial combination therapy with prednisone (n=133) 4. Initial combination therapy with infliximab (n=128)	DAS44-ESR ≤2.4	Low	DAS44ESR-remisison, mean HAQ and SHS at year 10	Not enough arms at this time-point for analysis
BeSt [81]	2012	RCT	1. Sequential monotherapy (n=126) 2. Step up combination therapy (n=121) 3. Initial combination therapy with prednisone (n=133) 4. Initial combination therapy with infliximab (n=128)	DAS44-ESR ≤2.4	Low	Drug free remission at year 7	Not enough arms for analysis for this outcome measure and this time-point
BeSt [82]	2009	RCT	1. Sequential monotherapy (n=126) 2. Step up combination therapy (n=121) 3. Initial combination therapy with prednisone (n=133) 4. Initial combination therapy with infliximab (n=128)	DAS44-ESR ≤2.4	Low	Patient Reported Outcomes	Not enough arms for analysis of patient reported outcomes
BeSt [83]	2012	RCT	All arms of BeSt, DAS-driven therapy (n=508)	DAS44ESR ≤2.4	Low	Drug free remission at year 5	Not enough arms for analysis of drug free remission

U-Act-Early [84]	2017	RCT	1. Tocilizumab + methotrexate (n=106) 2. Tocilizumab + placebo methotrexate (n=103) 3. Methotrexate + placebo tocilizumab (n=108)	DAS28-ESR<2.6	Low	Mean change SHS years 1 and 2	Duplicate results with other extracted publication
U-Act-Early [85]	2020	RCT	1. Tocilizumab + methotrexate (n=106) 2. Tocilizumab + placebo methotrexate (n=103) 3. Methotrexate + placebo tocilizumab (n=108)	DAS28-ESR<2.6	Low	Median change in SHS at year 5	Duplicate results with other extracted publication
DREAM [86]	2015	Cohort	Tight control (n=126)	DAS28-ESR<2.6	Some concerns	DAS-remission, mean change in DAS and HAQ at year 1	Duplicate results with other extracted publication
DREAM [87]	2021	Cohort	All patients (n=219)	DAS28-ESR<2.6	Some concerns	Median SHS at years 1, 3 and 6	Duplicate results with other extracted publication
Gullick et al. [88]	2019	Cohort	All patients (n=1693)	DAS28-ESR<2.6	Some concerns	DAS-remission, mean DAS at year 10	Not enough arms for analysis at this time-point
Ebrahimian et al. [89]	2021	Cohort	1. Very early treatment (n=196) 2. Early treatment (n=275) 3. Late treatment (n=138)	ACR-EULAR Boolean remission	Some concerns	Mean DAS, Boolean remission at years 5-7	Not enough arms for analysis at this time-point

DAS(28/44): Disease activity score (based on a 28/44-joint count), CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, HAQ: Health assessment questionnaire, LDA: low disease activity, RA: rheumatoid arthritis, RCT: Randomized Controlled Trial, SHS: Sharp van der Heijde Score.

Supplemental table S2: effect of treatment target on clinical and radiographic outcome measures after 1 year of T2T with 95% CI, based on parsimonious models

Target / Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.27 (-0.18 - 0.72)	0.43 (-0.26 - 1.13)	0.16 (-0.58 - 0.91)
Proportion DAS-remission	0.09 (-0.02 - 0.21)	0.21 (0.01 - 0.40)	0.11 (-0.08 - 0.31)
*DAS28-ESR subgroup**	*0.08 (-0.03 - 0.20)*	*0.22 (0.05 - 0.39)*	*0.14 (-0.03 - 0.30)*
*DAS44-ESR subgroup**	*0.38 (0.21 - 0.55)*	*NA*	*NA*
*DAS28-CRP subgroup**	*-0.18 (-0.38 - 0.02)*	*NA*	*NA*
Proportion SDAI/CDAI/Boolean remission	-0.01 (-0.15 - 0.12)	0.40 (0.07 - 0.73)	0.41 (0.08 - 0.74)
*SDAI/CDAI subgroup**	*-0.01 (-0.19 - 0.16)*	*0.34 (0.00 - 0.68)*	*0.35 (0.03 - 0.68)*
*Boolean remission subgroup**	*0.03 (-0.13 - 0.19)*	*NA*	*NA*
Improvement mean SDAI/CDAI	-3.05 (-6.22 - 0.12)	2.03 (-1.31 - 5.37)	5.08 (0.61 - 9.55)
Improvement mean HAQ	0.00 (-0.10 - 0.10)	-0.03 (-0.18 - 0.13)	-0.03 (-0.19 - 0.14)
Improvement in progression mean SHS per year (log transformed)	-0.09 (-0.49 - 0.31)	NA	NA

**The subgroups refer to the type of reported outcome measure. ACR: American College of Rheumatology, CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score. Dark green=statistically significant improvement (p<0.05), light green= statistically non-significant improvement (p≥0.05), orange= statistically non-significant deterioration (p≥0.05).*

Supplemental table S3: effect of treatment target on clinical and radiographic outcome measures after 2-3 years of T2T with 95% CI, based on parsimonious models.

Target / Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.67 (-0.17 - 1.50)	0.31 (-1.05 - 1.66)	-0.36 (-1.72 - 1.01)
Proportion DAS-remission	0.21 (0.02 - 0.40)	0.29 (0.02 - 0.57)	0.08 (p=0.48)
*DAS-ESR subgroup (28 or 44)**	*0.14 (-0.87 - 1.14)*	*0.01 (-1.55 - 1.58)*	*-0.12 (-1.31 - 1.06)*
Proportion SDAI/CDAI/Boolean remission	-0.05 (-0.15 - 0.04)	0.31 (0.15 - 0.47)	0.36 (0.21 - 0.52)
*CDAI/SDAI subgroup**	*-0.05 (-0.16 - 0.06)*	*0.30 (0.12 - 0.47)*	*0.34 (0.21 - 0.48)*
Improvement mean HAQ	0.02 (-0.28 - 0.32)	NA	NA
Improvement in progression mean SHS per year (log transformed)	-0.20 (-0.54 - 0.15)	NA	NA

**The subgroups refer to the type of reported outcome measure. CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score. Dark*

green=statistically significant improvement (p<0.05), light green= statistically non-significant improvement (p≥0.05), orange= statistically non-significant deterioration (p≥0.05).

Supplemental table S4: effect of treatment target on clinical and radiographic outcome measures after 4-6 years of T2T with 95% CI, based on parsimonious models.

Target / Outcome	DAS-remission vs. DAS-LDA	CDAI/SDAI LDA vs. DAS-LDA	CDAI-SDAI LDA vs. DAS-remission
Proportion DAS-remission	0.10 (-0.22 - 0.43)	NA	NA

**According to any of the DAS-variants (e.g. 28/44 joints, ESR/CRP). The parsimonious model has been corrected for the DAS-variant.. CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, LDA: low disease activity, SDAI: simplified disease activity index. Light green= statistically non-significant improvement (p≥0.05).*

Supplemental table S5: effect of treatment target on clinical and radiographic outcome measures after 1 year of T2T with 95% CI, based on full models.

Target / Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.28 (-0.19 - 0.75)	0.41 (-0.32 - 1.15)	0.13 (-0.67 - 0.94)
Proportion DAS-remission	0.09 (-0.03 - 0.21)	0.21 (0.00 - 0.41)	0.12 (-0.08 - 0.31)
*DAS28-ESR subgroup**	*0.10 (-0.03 - 0.23)*	*0.23 (0.05 - 0.40)*	*0.13 (-0.05 - 0.30)*
*DAS44-ESR subgroup**	*0.37 (0.19 - 0.55)*	*NA*	*NA*
*DAS28-CRP subgroup**	*-0.20 (-0.55 - 0.15)*	*NA*	*NA*
Proportion SDAI/CDAI/Boolean remission	-0.01 (-0.16 - 0.13)	0.40 (0.04 - 0.75)	0.41 (0.06 - 0.76)
*SDAI/CDAI subgroup**	*-0.01 (-0.19 - 0.16)*	*0.34 (0.00 - 0.68)*	*0.35 (0.03 - 0.68)*
*Boolean remission subgroup**	*0.03 (-0.13 - 0.19)*	*NA*	*NA*
Improvement mean SDAI/CDAI	-3.05 (-6.22 - 0.12)	2.03 (-1.31 - 5.37)	5.08 (0.61 - 9.55)
Improvement mean HAQ	-0.02 (-0.11 - 0.08)	-0.04 (-0.19 - 0.11)	-0.02 (-0.19 - 0.14)
Improvement in progression mean SHS per year (log transformed)	-0.09 (-0.49 - 0.31)	NA	NA

**The subgroups refer to the type of reported outcome measure. ACR: American College of Rheumatology, CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score. Dark green=statistically significant improvement (p<0.05), light green= statistically non-significant improvement (p≥0.05), orange= statistically non-significant deterioration (p≥0.05).*

Supplemental table S6: effect of treatment target on clinical and radiographic outcome measures after 2-3 years of T2T with 95% CI, based on full models.

Target / Outcome	DAS-remission vs. DAS-LDA	SDAI/SDAI LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.66 (-0.32 - 1.64)	0.31 (-1.37 - 1.98)	-0.36 (-1.99 - 1.28)
Proportion DAS-remission	0.21 (-0.08 - 0.51)	0.27 (-0.23 - 0.76)	0.05 (-0.35 - 0.45)
*DAS-ESR subgroup (28 or 44)**	*0.14 (-0.87 - 1.14)*	*0.01 (-1.55 - 1.58)*	*-0.12 (-1.31 - 1.06)*
Proportion SDAI/CDAI/Boolean remission	-0.05 (-0.15 - 0.05)	0.32 (0.14 - 0.50)	0.37 (0.20 - 0.54)
*CDAI/SDAI subgroup**	*-0.05 (-0.17 - 0.06)*	*0.29 (0.10 - 0.48)*	*0.34 (0.20 - 0.49)*
Improvement mean HAQ	0.03 (-0.32 - 0.39)	NA	NA
Improvement in progression mean SHS per year (log transformed)	-0.18 (-0.54 - 0.17)	NA	NA

**The subgroups refer to the type of reported outcome measure. CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score. Dark green=statistically significant improvement (p<0.05), light green= statistically non-significant improvement (p≥0.05), orange= statistically non-significant deterioration (p≥0.05).*

Supplemental table S7: effect of treatment target on clinical and radiographic outcome measures after 4-6 years of T2T with 95% CI, based on full models.

Target / Outcome	DAS-remission vs. DAS-LDA	CDAI/SDAI LDA vs. DAS-LDA	CDAI-SDAI LDA vs. DAS-remission
Proportion DAS-remission*	0.10 (-0.23 - 0.44)	NA	NA

**According to any of the DAS-variants (e.g. 28/44 joints, ESR/CRP). The full model has been corrected for the DAS-variant.. CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, LDA: low disease activity, SDAI: simplified disease activity index. Light green= statistically non-significant improvement (p≥0.05).*

Supplemental table S8: effect of treatment target on clinical and radiographic outcome measures after 1 year of T2T with 95% CI, based on univariate models.

Target Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.42 (-0.12 - 0.96)	0.29 (-0.60 - 1.18)	-0.13 (-1.06 - 0.79)
Proportion DAS-remission	0.02 (-0.10 - 0.13)	0.06 (-0.15 - 0.27)	0.04 (-0.17 - 0.25)
*DAS28-ESR subgroup**	*0.10 (-0.06 - 0.26)*	*0.16 (-0.07 - 0.39)*	*0.06 (-0.16 - 0.28)*
*DAS44-ESR subgroup**	*0.08 (-0.13 - 0.28)*	*NA*	*NA*
*DAS28-CRP subgroup**	*-0.08 (-0.30 - 0.14)*	*NA*	*NA*
Proportion SDAI/CDAI/Boolean remission	-0.03 (-0.15 - 0.09)	0.28 (0.02 - 0.54)	0.31 (0.04 - 0.58)
*SDAI/CDAI subgroup**	*0.01 (-0.15 - 0.16)*	*0.27 (0.03 - 0.52)*	*0.27 (0.02 - 0.52)*
*Boolean remission subgroup**	*-0.08 (-0.31 - 0.14)*	*NA*	*NA*
Improvement mean SDAI/CDAI	1.42 (-4.29 - 7.14)	3.97 (-4.04 - 11.99)	2.55 (-5.84 - 10.94)
Improvement mean HAQ	0.18 (-0.05 - 0.42)	-0.18 (-0.56 - 0.21)	-0.36 (-0.76 - 0.04)
Improvement in progression in mean SHS per year (log transformed)	-0.34 (-0.80 - 0.11)	NA	NA

**The subgroups refer to the type of reported outcome measure. ACR: American College of Rheumatology, CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score.*

Supplemental table S9: effect of treatment target on clinical and radiographic outcome measures after 2-3 years of T2T with 95% CI, based on univariate models.

Target Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Improvement mean DAS28-ESR	0.40 (-0.45 - 1.24)	0.71 (-0.55 - 1.97)	0.32 (-0.91 - 1.54)
Proportion DAS-remission	-0.03 (-0.22 - 0.16)	0.11 (-0.22 - 0.43)	0.14 (-0.20 - 0.47)
*DAS-ESR subgroup (28 or 44)**	*0.12 (-0.08 - 0.32)*	*0.27 (0.00 - 0.53)*	*0.15 (-0.09 - 0.40)*
Proportion SDAI/CDAI/Boolean remission	0.02 (p=0.66)	0.31 (p<0.0001)	0.29 (0.15 - 0.43)
*CDAI/SDAI subgroup **	*0.03 (-0.07 - 0.13)*	*0.31 (0.16 - 0.45)*	*0.28 (0.15 - 0.41)*
Improvement mean HAQ	0.03 (-0.31 - 0.37)	NA	NA
Improvement in progression in mean SHS per year (log transformed)	-0.33 (-0.62 - 0.06)	NA	NA

**The subgroups refer to the type of reported outcome measure. CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index, SHS: Sharp/ van der Heijde score.*

Supplemental table S10: effect of treatment target on clinical and radiographic outcome measures after 4-6 years of T2T with 95% CI, based on univariate models.

Target Outcome	DAS-remission vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-LDA	SDAI/CDAI-LDA vs. DAS-remission
Proportion DAS-remission*	0.06 (-0.43 - 0.55)	NA	NA

**According to any of the DAS-variants (e.g. 28/44 joints, ESR/CRP). CDAI: clinical disease activity index, CI: confidence interval, CRP: C-reactive protein, DAS: disease activity score, based on either a 28- or 44 joint count and either on CRP or ESR, ESR: erythrocyte sedimentation rate, LDA: low disease activity, SDAI: simplified disease activity index.*

Supplemental table S11: subgroup analyses of cohort studies vs. RCT showing the effect of treatment target on clinical outcome measures with 95% CI, based on parsimonious models.

Target Outcome	Target (compared to DAS-LDA)	Cohort studies	RCTs
Improvement mean DAS28-ESR, year 1	DAS-remission	0.40 (-0.17 - 0.97)	0.27 (-0.37 - 0.91)
	SDAI/CDAI LDA	0.58 (-0.42 - 1.57)	-0.34 (-1.48 - 0.81)
Proportion DAS-remission, year 1	DAS-remission	0.15 (-0.02 – 0.31)	0.10 (-0.10 – 0.30)
	SDAI/CDAI LDA	0.41 (0.10 – 0.71)	-0.06 (-0.41 – 0.30)
Proportion DAS-remission, year 2-3	DAS-remission	0.12 (-0.18 – 0.42)	-0.11 (-0.89 – 0.68)
	SDAI/CDAI LDA	0.20 (-0.15 – 0.54)	NA
Improvement mean HAQ, year 1	DAS-remission	0.14 (-0.15 – 0.43)	-0.06 (-0.15 – 0.04)
	SDAI/CDAI LDA	NA	-0.11 (-0.27 – 0.04)
Improvement mean HAQ, year 2-3	DAS-remission	0.13 (-0.08 – 0.34)	-0.11 (-0.78 – 0.57)
	SDAI/CDAI LDA	NA	NA
Proportion SDAI/CDAI/Boolean remission, year 1	DAS-remission	-0.02 (0.89)	0.04 (-0.21 – 0.28)
	SDAI/CDAI LDA	0.38 (0.07 – 0.68)	NA
Proportion SDAI/CDAI/Boolean remission, year 2-3	DAS-remission	-0.28 (-0.42 – -0.14)	0.10 (-0.01 – 0.20)
	SDAI/CDAI LDA	0.05 (-0.10 – 0.20)	-0.05 (-0.15 – 0.05)

CDAI: Clinical Disease Activity Index, CI: confidence interval, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, HAQ: Health Assessment Questionnaire, LDA: low disease activity, RCT: randomized controlled trial, SDAI: Simplified Disease Activity Index

Supplemental table S12: sensitivity 'leave-one-out' analyses using parsimonious models.

Outcome [target vs. DAS-LDA]	Original coefficient parsimonious model	Minimum coefficient leave-one-out analysis	Maximum coefficient leave-one-out analysis
Improvement in mean DAS28-ESR year 1 [DAS remission]	0.27 (p=0.24)	0.14 (p=0.58)	0.40 (p=0.09)
Improvement in mean DAS28-ESR year 1 [SDAI/CDAI LDA]	0.43 (p=0.22)	0.34 (p=0.31)	0.57* (p=0.03)
Improvement in mean DAS28-ESR year 2-3 [DAS remission]	0.67 (p=0.12)	0.50 (p=0.41)	0.81 (p=0.06)
Improvement in mean DAS28-ESR year 2-3 [SDAI/CDAI LDA]	0.31 (p=0.66)	-0.03 (p=0.98)	0.55 (p=0.55)
Increase in % DAS-remission year 1 [DAS remission]	0.09 (p=0.11)	0.08 (p=0.19)	0.12 (p=0.05)
Increase in % DAS-remission year 1 [SDAI/CDAI LDA]	0.21* (p=0.04)	0.16 (p=0.11)	0.23* (p=0.03)
Increase in % DAS-remission year 2-3 [DAS remission]	0.21* (p=0.03)	0.17* (p=0.003)	0.37* (p=0.03)
Increase in % DAS-remission year 2-3 [SDAI/CDAI LDA]	0.29* (p=0.04)	0.21* (p=0.009)	0.38* (p=0.02)
Increase in % DAS-remission year 4-6 [DAS-remission]	0.10 (p=0.53)	0.00 (p=0.99)	0.21 (p=0.20)
Improvement in mean SDAI/CDAI year 1 [DAS remission]	-3.05 (p=0.06)	-4.72* (p=0.04)	1.12 (p=0.87)
Improvement in mean SDAI/CDAI year 1 [SDAI/CDAI LDA]	2.03 (p=0.23)	-2.80 (p=0.72)	5.56* (p<0.0001)
Increase in % SDAI/CDAI/Boolean remission year 1 [DAS remission]	-0.01 (p=0.85)	-0.04 (p=0.60)	0.02 (p=0.78)
Increase in % SDAI/CDAI/Boolean remission year 1 [SDAI/CDAI LDA]	0.40* (p=0.02)	0.29 (p=0.10)	0.52* (p=0.005)
SDAI/CDAI Boolean rem year 2-3 [DAS-remission]	-0.05 (p=0.26)	-0.08 (p=0.19)	-0.01 (p=0.79)
SDAI/CDAI Boolean rem year 2-3 [SDAI/CDAI LDA]	0.31* (p=0.0002)	0.26* (p=0.02)	0.40* (p=0.006)
Mean HAQ year 1 [DAS-remission]	0.00 (p=0.99)	-0.02 (p=0.66)	0.03 (p=0.60)
Mean HAQ year 1 [SDAI/CDAI LDA]	-0.03 (p=0.73)	-0.09 (p=0.24)	-0.01 (p=0.90)
Mean HAQ year 2-3 [DAS-remission]	0.02 (p=0.87)	-0.03 (p=0.85)	0.07 (p=0.71)

In the leave-one-out analyses, each analysis is performed numerous times, excluding one arm at each analysis to assess the influence of individual arms. Leave-one-out analyses for the outcome improvement in yearly SHS progression could not be performed as the model could not achieve convergence in this analysis. CDAI: clinical disease activity index, DAS: disease activity score based on 28/44 joints, ESR: erythrocyte sedimentation rate, HAQ: health assessment questionnaire, LDA: low disease activity, SDAI: simplified disease activity index.

Chapter 7

Using real-world data to dynamically predict flares during tapering of biological DMARDs in rheumatoid arthritis: development, validation, and potential impact of prediction-aided decisions

Marianne A. Messelink*
Matthijs S. van der Leeuw*
Janneke Tekstra
Ojay Medina
Jaap M. van Laar
Saskia Haitjema
Floris Lafeber
Josien J. Veris-van Dieren
Marlies C. van der Goes
Alfons A. den Broeder
Paco M.J. Welsing

**these authors contributed equally*

Arthritis Research & Therapy, 2022;24(1):74

Abstract

Background Biological disease modifying antirheumatic drugs (bDMARDs) are effective in the treatment of rheumatoid arthritis. However, as bDMARDs may also lead to adverse events and are expensive, tapering them is of great clinical interest. Tapering according to disease activity guided dose optimization (DGDO) does not seem to affect long term remission rates, but flares are frequent during this process. Our objective was to develop a model for the prediction of flares during bDMARD tapering using data from routine care and to evaluate its potential clinical impact.

Methods We used a joint latent class model to repeatedly predict the probability of a flare occurring within the next 3 months. The model was developed using longitudinal data on disease activity (DAS28) and other routine care data from two clinics. Predictive accuracy was assessed in cross-validation and external validation was performed with data from the DRESS (Dose Reduction Strategy of Subcutaneous tumor necrosis factor inhibitors) trial. Additionally, we simulated the reduction in number of flares and bDMARD dose when implementing the model as a decision aid during bDMARD tapering in the DRESS trial.

Results Data from 279 bDMARD courses were used for model development. The final model included two latent DAS28-trajectories, bDMARD type and dose, disease duration and seropositivity. The Area Under the Curve of the final model was 0.76 (0.69-0.83) in cross-validation and 0.68 (0.62–0.73) in external validation. In simulation of prediction-aided decisions, the mean number of flares over 18 months decreased from 1.21 (0.99-1.43) to 0.75 (0.54-0.96). The reduction in he bDMARD dose was mostly maintained, increasing from 54% to 64% of full dose.

Conclusions We developed a dynamic flare prediction model, exclusively based on data typically available in routine care. Our results show that using this model to aid decisions during bDMARD tapering may significantly reduce the number of flares whilst maintaining most of the bDMARD dose reduction.

Introduction

Many rheumatoid arthritis (RA) patients who are treated with biological disease modifying anti-rheumatic drugs (bDMARDs) achieve long periods of low disease activity or remission.(1) However, bDMARDs may also lead to adverse events, call for self-injections or hospital visits and are expensive.(2–4) Thus, tapering bDMARDs to the lowest effective dose is of great clinical interest and may support the sustainability of the healthcare system as a whole.

The guidelines of the European League against Rheumatism (EULAR) on the management of RA advise to consider tapering in patients that are in persistent remission.(5) In addition, numerous clinical trials and reviews provide supportive evidence to also consider tapering in patients with stable low disease activity (LDA).(6,7) This is in line with routine clinical practice, as maintaining a satisfactory low level of disease activity with a reduced medication dose is also of value.

The most successful and cost-effective strategy for tapering appears to be 'disease activity guided dose optimization' (DGDO).(8–10) This means the dose is gradually tapered (usually by increasing the administration interval), until either disease activity flares or the bDMARD is discontinued. Two randomized trials have demonstrated that, using this strategy, 63-80% of patients can taper or even stop their bDMARD.(8,9) No important difference was observed in the proportion of patients with LDA or remission after 18 months between DGDO and usual care.

However, since DGDO is a 'trial and error' approach, flares occur frequently during the tapering process. In the case of a flare, the previously effective dose needs to be reinstated or additional therapy is necessary. Although these short-lived flares do not seem to relevantly affect radiographic progression or long term disease activity, there is conflicting evidence regarding functional outcome and impact on quality of life.(9,11) Therefore, it would be beneficial to predict whether, and to which extent, a bDMARD can be tapered in a particular patient without a flare occurring.

Several predictors for successful dose reduction or discontinuation of bDMARDs have been explored.(12,13) However, these studies only included 'baseline predictors' from before the start of the tapering process, and the strength of the evidence for these predictors is limited. Furthermore, 'successful tapering' is often defined as reaching a lower bDMARD dose at some time point after the start of tapering, regardless of whether a flare occurred during the tapering process.

Therefore, this study aims to predict the likelihood of a flare occurring during bDMARD tapering at each consecutive dose reduction step. Such a dynamic prediction may be used to optimize the DGDO strategy for bDMARDs for an individual patient, as the decision for a further tapering step can be based on the predicted risk of a flare. This could minimize the number of flares during tapering, whilst retaining most of the bDMARD dose reduction. To facilitate future implementation of this approach in routine practice, we decided to exclusively use information easily obtainable in regular care.

Methods

Data extraction and preparation

EHR data for model development

For the development of the prediction model, electronic health record (EHR) data of two rheumatology clinics in the Netherlands were extracted for the period 2012-2019 and 2013-2019 respectively: the University Medical Center Utrecht (UMCU; an academic hospital) and Reumazorg Zuid West Nederland (RZWN; a non-academic treatment center for rheumatic diseases). In both centers bDMARD tapering is regularly performed, but not yet standard practice. Data were extracted for all RA patients (based on ICD-10 codes) starting a bDMARD and reaching a Disease Activity Score assessing 28 joints (DAS28) <3.2, i.e. LDA, after at least 24 weeks of treatment. The following bDMARDs were included: infliximab, adalimumab, etanercept, golimumab, certolizumab, tocilizumab, sarilumab and abatacept. We selected patients with at least the following information available: bDMARD type and dose, seropositivity, disease duration and ≥2 DAS28 measurements per year available. In addition, we aimed to extract the following data: age, gender, body mass index, concurrent and previous DMARD and glucocorticoid use, smoking status and erosive disease.

To handle missing individual DAS28 components, we used all validated DAS28 formulae by calculating the mean of the 3- and 4-variable DAS28 formulae using ESR (erythrocyte sedimentation rate) as well as CRP (C-reactive protein).(14) We allowed a 4-week time window between components. Flares were defined using a validated criterion: an increase in DAS28 > 1.2 compared to the previous visit, or an increase of 0.6 with a resulting DAS28 > 3.2. (15) In addition, an 'increase in bDMARD dose' was also considered a flare, to also capture flares if insufficient information was present to calculate the DAS28.

All data was extracted according to current ethical and privacy regulations in the specific hospitals. The Medical Research Ethics Committee Utrecht waived the need for

informed consent, as the development data was already collected in routine care and was pseudonymized before analysis.

DRESS data for external validation

For external validation, we extracted data from the DRESS trial.(9) In this trial, RA patients with stable LDA or remission using adalimumab or etanercept were randomized to either DGDO (n=121) or routine care (n=59) and followed for 18 months. The study was performed between 2011 and 2014 in two Dutch clinics (Sint Maartenskliniek Nijmegen and Woerden). The DGDO group tapered the bDMARD in three steps by increasing the administration interval every three months, followed by discontinuation after 6 months as long as the patient did not flare. In case of flare, the last effective dose was reinstated, and no further dose reduction attempts were undertaken. If nevertheless flares persisted, the bDMARD dose was increased to the full dose and thereafter treatment was at the rheumatologists' discretion. In DRESS, flares were defined by the DAS28-CRP increase from baseline values.

The DRESS study (Dose REduction Strategy of Subcutaneous TNF inhibitors) was approved by the local ethics committee (Committee on Research Involving Human Subjects region Arnhem-Nijmegen), and informed consent was signed by all included patients. (9)

Model development

We developed a dynamic model to repeatedly predict the risk of a flare occurring in the next 3 months. This corresponds to a routine outpatient visit interval.(13) The model was developed using joint latent class mixed modelling, which combines a linear mixed effects- and a time-to-event model (R-package lcmm). Details of joint latent class models have been described elsewhere. (16,17)

First, in the linear mixed effects part of the model, the course ('trajectories') of the DAS28 values over time are modelled for each patient. This is done by categorizing these trajectories into a number of subgroups: latent classes. The general form of these trajectories is defined using polynomials for the time variable. We explored models with a random intercept using 1-3 latent classes and 1st to 3rd order polynomials for the time variable, using a random slope for time. The best fitting model was selecting based on the lowest Bayesian Information Criterion (BIC).(19). Based on the final model each individual patient has its own predicted DAS28 trajectory.

Next, these DAS28-trajectories are used as variables in the time-to-event part of the model. The time-to-event part of the model also incorporates other variables. We explored all variables as mentioned above in "*EHR data for model development*" and

selected those that had sufficient data to be extracted from the EHR. The time-to-event model was developed stepwise starting with a full model, excluding variables one by one to arrive at a final model. The decision to exclude a variable was based on clinical rationale, data availability and improvement in model fit in cross-validation, defined by the decrease in the BIC. In short, to make individual predictions, an estimation is made about the individuals trajectory of the DAS28 over time. This trajectory is then combined with additional variables to calculate the probability of a flare occurring in the next 3 months.

We adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.(18)

Model validation

We assessed the accuracy of 3-monthly flare predictions in the development data with 5-fold cross-validation, using all visits at which a DAS28 was available. The Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC) was calculated over all time-points. Other performance indicators were assessed based on an optimal cut-off as defined by Youden's Index in the development data.(15) This index is a summary measure for sensitivity and specificity.

External validation was performed by assessing the accuracy of flare predictions in data from the DRESS trial.(9) The AUC-ROC and other performance indicators were calculated using the optimal cut-off points as determined in the development data and in DRESS data, both defined by Youden's Index.(19)

Simulation of prediction-aided treatment

To evaluate the clinical utility of the flare predictions, we assessed the model's potential impact on the number of flares and on the bDMARD dose used over 18 months. We simulated a new tapering strategy where the model's predictions were used as a decision aid in the DGDO arm of the DRESS trial. At every 3-monthly visit, the predicted risk of a flare was taken into account when deciding to continue or to stop tapering. The predicted risk of flare was categorized into a high predicted risk (above or equal to the optimal cut-off point), or a low predicted risk (below the cut-off point). The simulation was based on the following assumptions:

1. If a flare occurred in the DRESS trial before the model predicted a high risk of flare, this flare also occurs in the simulation. The bDMARD dose is the same as in the trial. Thus, there is no impact of the predictions is observed in this case.
2. If the model predicted a high risk of flare in simulation and no flare had occurred in the DRESS trial thus far, the bDMARD is not tapered further (kept

at a constant dose). No flares occur in simulation during the remaining follow-up, except for the scenario described in 4.

3. If a patient had completely discontinued the bDMARD in the DRESS trial when the model predicted a high predicted risk of flare, the bDMARD dose in simulation is increased to and kept at 50% of the full registered dose. This corresponds to the last tapering step. No flares occur during the remaining follow-up, except for the scenario described in 4.
4. If in the DRESS trial a flare occurred after the model predicted a high risk of flare and the bDMARD dose in DRESS was equal to or higher than the bDMARD dose in simulation, that flare also occurs in the simulation. The bDMARD dose is equal to the DRESS trial during the remaining follow-up.

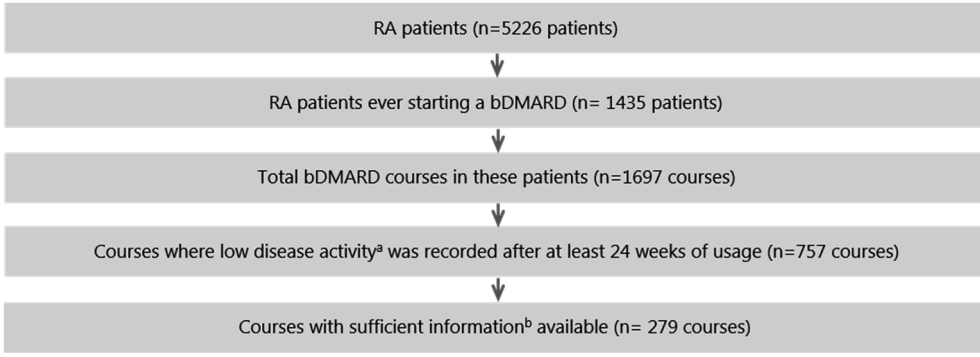
The number of flares occurring, the proportion of patients experiencing at least one flare, and the proportion of the full registered dose were calculated. These were then compared between the simulation and the DGDO arm of the DRESS trial over 18 months. Confidence intervals (CI) were calculated using 1000-fold bootstrapping. As there is no obvious optimum in the trade-off between the reduction in the number of flares and the increase in bDMARD dose, we evaluated the clinical impact of prediction-aided treatment for several cut-offs around the optimal cut-offs as defined by Youden's Index.(19)

Results

Patient characteristics

Of the total number of 5226 RA patients in the EHR data, there were 757 bDMARD courses in which LDA was recorded after at least 24 weeks of usage (Figure 1). In 279 bDMARD courses of 255 patients, sufficient data was available for model development (see Methods). Data for smoking, erosive disease, concurrent and previous DMARDs and glucocorticoids were of insufficient quality (>50% missing data) and/or could not be (easily) extracted from the EHR. The median follow-up time of the included bDMARD courses was 21 months, and the mean bDMARD dose was 76.7% of the full dose. Table 1 displays general patient characteristics of the development data and the data from the DRESS trial used for external validation. Significant differences between the populations were observed for age, DAS28 at baseline, the number of DAS28-measurements, flare rate and bDMARD dose, among others.

Figure 1. Selection of bDMARD courses from EHR data for model development



a: low disease activity was defined as a DAS28 (ESR or CRP) ≤ 3.2. b: based on the availability of at least two DAS28 measurements per year, bDMARD type and dose, disease duration and seropositivity bDMARD: biological disease modifying antirheumatic drug, EHR: electronic health record, RA: rheumatoid arthritis

Table 1. Patient characteristics in data for model development and external validation

	Development data (n=279)	**DRESS data for external validation (n=164)**	**P-value of difference**
General characteristics			
Age, mean in years (SD)	50.2 (17.1)	58.7 (9.9)	<0.01
Female, N (%)	203 (72.8%)	105 (64%)	0.05
BMI, median in kg/m^2 (IQR)	25.4 (22.6-30.0)	26.8 (23.3-29.5)	0.22
Height, mean in cm (SD)	170.4 (12.5)	172 (9.2)	0.15
Weight, mean in kg (SD)	77.0 (17.2)	78.8 (15.5)	0.27
Follow-up time, median in months (SD)	21 (2.0)	18.7 (1.6)	<0.01
RA characteristics			
Disease duration at start of bDMARD, median in years (IQR)	9.0 (5.0-16.5)	6.0 (2.7-12.7)	0.38
Positivity for RF and/or ACPA, N (%)	236 (84.6%)	140 (85.4%)	0.83
Biological			
bDMARD type, N (%)			
Etanercept	77 (27.6%)	107 (65.2 %)	<0.01
Infliximab	5 (1.8%)	-	-
Adalimumab	113 (40.5%)	57(34.8%)	0.23
Certolizumab	10 (3.6%)	-	-
Golimumab	18 (6.5%)	-	-
Tocilizumab	8 (2.9%)	-	-
Sarilumab	2 (0.7%)	-	-
Abatacept	37 (13.3%)	-	-

Time from start bDMARD baseline, mean in weeks (SD)[b]	10 (7.7)	42.1 (29.1)	<0.01
Prescribed dose during follow-up (mean, expressed as % of full dose)	76.7%	61.6%	<0.01
Disease activity			
DAS28 at baseline, mean (SD)	2.79 (1.34)	2.15 (0.70)	<0.01
VAS GH at baseline, median (IQR)	30 (11-40)	20 (10-34)	0.76
TJC at baseline, median (IQR)	0 (0-1)	0 (0-1)	-
SJC at baseline, median (IQR)	0 (0-1)	0 (0-1)	-
ESR at baseline, median in mm/hour (IQR)	7 (3-12)	13 (7-22)	<0.01
CRP at baseline, median mg/ml (IQR)	2.7 (1.3-5.0)	3 (3-3)	0.22
Increase in TJC (yes/no)[c], N (%)	49 (17.6%)	NA	-
Increase in SJC (yes/no)[c], N (%)	28 (10.0%)	NA	-
No. of DAS28 measurements, mean (SD)	6.1 (3.8)	7.3 (1.2)	<0.01
Time between DAS28, mean in weeks (SD)	22.3 (12.3)	12.0 (5.4)	<0.01
Flare rate (# flares per patient year)	0.47	0.62	<0.01
DAS28 measurement rate (#DAS28 measurements per patient year)	2.18	4.72	0.04

a: P-values based on T-test for normally distributed continuous data, Mann-Whitney U test for continuous non-normally distributed data and Chi²-test for nominal data. b: In the development data, baseline is defined as the first DAS28 ≤3.2 c: An increase in TJC and/or SJC (yes/no) at baseline relative to the previous TJC/SJC measurement. ACPA: anti-citrullinated protein antibodies, bDMARD: biological disease modifying antirheumatic drug, CRP: C-reactive protein, DAS28: disease activity score based on 28 joint count, EHR: electronic health record. ESR: erythrocyte sedimentation rate, IQR: interquartile range, RF: rheumatoid factor, SD: standard deviation, TJ(C)/SJ(C): tender/swollen joint count, VAS GH: an assessment of general health on a visual analogue scale (0-100mm).

Model development

The variables that were retained in the final prediction model and the corresponding hazard ratios are displayed in Table 2. The final model identified two latent DAS28-trajectories, defined by a linear and a quadratic time coefficient. Figure 2 shows the mean of these two trajectories (left), together with their respective time to flare (right). The course of disease activity in the class 2 DAS28-trajectory shows an increase in disease activity over time and a shorter time to flare, compared to the class 1 DAS28-trajectory. Variables that significantly increased the likelihood of a flare were seropositivity, bDMARD dose <50% and an increase in tender joint count at baseline (compared to previous visit).

As the DAS28-trajectories represent the course of disease activity over time, this is a time-dependent variable. By default, only one continuous time-dependent variable can

be included in a joint latent class model. In order to also add bDMARD dose as a second time-dependent variable, it was dichotomized in < or ≥ 50% of the full registered dose.

Table 2. Variables of the final flare prediction model

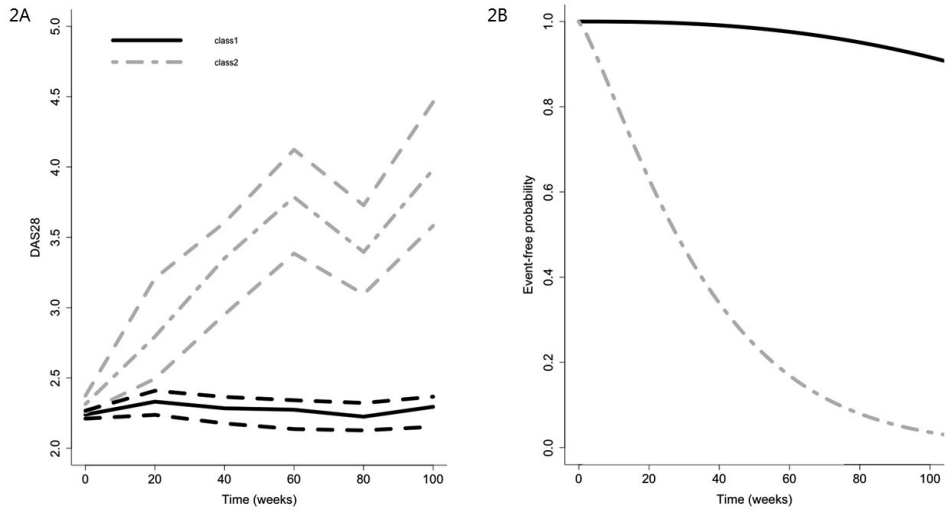
Parameter	Hazard Ratio (95% CI)
Linear time coefficient DAS28 trajectory latent class 1	1.04 (1.02 – 1.06)
Quadratic time coefficient DAS28 trajectory latent class 1	1.66 (0.42 – 6.55)
Linear time coefficient DAS28 trajectory latent class 2	1.14 (1.08 – 1.20)
Quadratic time coefficient DAS28 trajectory latent class 2	4.52 (3.83 – 5.33)
Time to reach stable low disease activity (weeks)[a]	0.97 (0.96 – 0.98)
DAS28 at baseline	1.18 (0.90 – 1.54)
Prescribed dose (% of standard dose) at baseline	1.21 (0.88 – 1.67)
SJ increase at baseline (yes/no)[b]	1.72 (0.94 – 3.17)
TJ increase at baseline (yes/no) [b]	2.07 (1.13 – 3.81)
Disease duration (years) at start of bDMARD	1.02 (0.99 – 1.05)
Seropositivity (RF and/or ACPA)	2.51 (1.39 – 4.53)
bDMARD TNFi type (yes/no)	0.90 (0.54 – 1.49)
bDMARD dose ≤ 50% of full registered dose (time-varying variable)	2.21 (1.73 – 2.82)

In development data, baseline is defined as the first DAS28 ≤3.2 (low disease activity). a. in development data: time from start biological until DAS28 <3.2 for the first time. In DRESS data: time from start biological until baseline visit. b. An increase in TJC or SJC (yes/no) at baseline, compared to the previous DAS28 measurement. ACPA: anti-citrullinated protein antibody, bDMARD: biological disease modifying antirheumatic drug, DAS28: disease activity score based on 28-joint count, RF: rheumatoid factor, TJ(C)/SJ(C): tender/swollen joint count, TNFi: tumor necrosis factor inhibitor.

Model validation

Predictive performance in cross-validation and external validation is summarized in Table 3. In cross-validation, the model achieved an AUC-ROC of 0.76 (CI 0.69-0.83). The optimal cut-off in the development data was at a predicted probability of flare of 14.3% within the next three months. For external validation, sixteen patients were excluded from DRESS because of missing predictor information. Furthermore, as the variable 'increase in SJC/TJC' was not available in the DRESS data, these were set to 0 (i.e. no increase) for the validation, with the rationale that patients that met the DRESS inclusion criteria had a stable low level of disease activity at baseline. Supplementary Figure S1 shows the AUC-ROC of the model in external validation (0.68 (CI 0.62–0.73), see Supplementary Figure S2 for the calibration plot). The optimal cut-off point in DRESS data was found to be at a predicted chance of flare of 31.5% within the next 3 months.

Figure 2. Mean DAS28-trajectories of identified latent classes and their relation to the occurrence of a flare



A: the mean course of the disease activity score (DAS28) over time in patients assigned to one of the two 'latent trajectory classes'. In class 1 (n= 182), a stable low disease activity is observed, whereas patients in class 2 (n= 97) display an increasing disease activity over time. B: the probability of remaining free from flares over time for patients assigned to one the 'latent trajectory classes' for disease activity, as displayed on the left. Patients in class 2 display a shorter time to flare as compared to patients in class 1.

Because the model cannot truly function as a 'joint' model at baseline, since no longitudinal information is yet available, we also explored the performance when removing baseline predictions. This indeed improved the AUC in external validation to 0.71 (CI 0.64-0.77, Supplementary Figure S3 and Supplementary Table S1).

Table 3. Predictive performance in cross validation and external validation

	Cross validation (Cut-off 14.3%)	**External validation (Cut-off 14.3%)**	**External validation (Cut-off 31.5%)**
AUC	0.76 (0.69-0.83)	0.68 (0.62–0.73)	0.68 (0.62–0.73)
Sensitivity (%)	86.1 (81.9-90.1)	73.2 (64.4–82.0)	58.8 (49.0-68.6)
Specificity (%)	66.5 (60.1-72.5)	52.0 (0.48–56.0)	68.7 (64.9-72.4)
Positive Predictive Value (%)	33.0 (29.3-38.5)	20.1 (15.9–24.3)	23.7 (18.3-29.0)
Negative Predictive Value (%)	96.2 (95.4-98.4)	92.1 (89.2–95.0)	91.0 (88.3-93.6)
Accuracy (%)	70.6 (65.6-75.6)	55.0 (51.2–58.7)	67.3 (63.6-70.8)

Results from the 5-fold cross-validation in development data are presented for an optimal cut-off point of 14.3% as determined with Youden's index. The results from external validation in the DRESS trial(9) *are presented for 2 different cut-off points: the optimal cut-off point from the development data (14.3%) and the optimal cut-off point in the DRESS data as determined by Youden's index (31.5%). 95% confidence intervals are presented between brackets. AUC: Area under the curve.*

Table 4. Flares and bDMARD dose in simulation of prediction-aided treatment

	DRESS Routine Care	Simulation (cut-off: 35%)	DRESS DGDO
Mean no. of flares (95% CI)	0.48 (0.24-0.72)	0.75 (0.54-0.96)	1.21 (0.99-1.43)
Decrease in flares compared to DRESS DGDO (95% CI)	0.73 (0.40-1.0)	0.46 (0.16-0.74)	-
Mean bDMARD dose (95% CI)	0.91 (0.86-0.96)	0.64 (0.61-0.68)	0.54 (0.50–0.58)
Increase in bDMARD dose compared to DRESS DGDO (95% CI)	0.37 (0.31-0.44)	0.10 (0.05–0.16)	-
Percentage of patients flaring (95% CI)	27% (15-40)	45% (36-54)	71% (63–79)
Increase in bDMARD dose per flare prevented vs. DRESS DGDO[a] (95% CI)	0.51 (0.44–0.59)	0.22 (0.15-0.32)	-
Number of extra flares per full bDMARD dose saved vs. routine care [b] (95% CI)	-	1.0 (0.3–1.8)	2.0 (1.4-2.6)

a. The difference in mean bDMARD dose divided by the difference in mean flares compared with DRESS(9) *DGDO. This represents the increase in bDMARD dose that was needed to prevent a flare over 18 months for this tapering strategy. b. The mean difference in the number of flares, divided by the mean difference in bDMARD dose, compared to routine care. This represents the number of extra flares that occurred for each full dose of bDMARD that is tapered compared to routine care over 18 monhts using this tapering strategy. bDMARD: biological disease modifying antirheumatic drug, DGDO: disease activity-guided dose optimisation.*

Simulation of prediction-aided treatment

We assessed the potential clinical impact of the model on the number of flares and the amount of bDMARD dose reduction, when used as a decision aid within a DGDO strategy. The clinical impact of prediction-aided treatment in simulation was evaluated for cut-offs from 15%-45% in steps of 10% (Supplementary Table S2) and results were discussed to determine the optimal cut-off for clinical practice. A risk cut-off of 35% was deemed optimal, as this significantly reduced the number of flares per patient over 18 months from 1.21 (0.99-1.43) to 0.75 (0.54-0.96), whilst retaining most of the bDMARD dose reduction (64% vs 54% of full registered dose used). See Table 4. When using this optimal cut-off of 35%, only 1.0 flare occurred for each full dose that was tapered in the simulation of prediction-aided treatment, versus 2.0 flares in the DRESS DGDO arm. Furthermore, in the DRESS routine care arm each prevented flare (compared with DRESS DGDO) came at a cost of 51% of a full bDMARD dose over 18 months, while this was only 22% in the simulated prediction-aided group. As the AUC-ROC improved when the predictions at baseline were not taken into account, we explored the simulation of prediction-aided treatment when removing the baseline predictions. However, the simulation results were hardly influenced by this (Supplementary Table S3).

Discussion

The goal of this study was to develop and validate a flare prediction model to reduce the number of flares during bDMARD tapering, exclusively using data that can easily be obtained in routine care. Our simulation results show that the addition of our flare prediction model to a DGDO tapering strategy is both superior to routine care and to DGDO alone, when considering the ratio between the number of flares and amount of bDMARD dose reduction. To our knowledge, this is the first study not only developing a dynamic flare prediction model, but also performing an external validation and subsequent simulation of clinical impact in the context of bDMARD tapering.

As tapering bDMARDs is of great clinical interest, other studies have also investigated predictors in the context of tapering. Several studies and systematic reviews have investigated the predictive value of biomarkers, serum drug levels or PET-scans during bDMARD tapering. (12,20–22) However, none of these studies showed a clear predictive value of these markers. In addition, the study by Verhoef et al. showed that for a biomarker to be cost-effective during bDMARD tapering, it must be inexpensive and have high sensitivity and specificity.(23) If future studies do show a predictive value of (bio)markers during tapering, these can be included in the prediction model. The added predictive value of such markers and their cost-effectiveness should then be assessed. An important advantage of the current model is that it only includes variables that are routinely collected in RA clinical practice, thereby enhancing feasibility and cost-effectiveness.

A recent review (13) focused on predictors for successful discontinuation, rather than tapering, of bDMARDs. Similar to the current study, they found seropositivity, LDA, disease duration and CRP/ESR to be possible predictors of value. In addition, they mention physical functioning and ultrasound measures as possible predictors. However, the studies included in this review were often small and too heterogeneous to compare in meta-analysis. Furthermore, only fixed baseline variables were included, rather than performing dynamic predictions using information over time.

Two studies have incorporated such dynamic variables to predict RA disease activity over time.(24,25) The study by Norgeot et al.(24) found the Clinical Disease Activity Index(CDAI), CRP/ESR, glucocorticoid use and other DMARD use to be important predictors. However, this study is not performed in the specific context of tapering bDMARDs. The model developed by Vodenčarević et al.(25) does focus specifically on bDMARD tapering. However, this model is developed and validated on the clinical trial data of 41 patients only, and may therefore be difficult to extrapolate to routine care.

Both of these dynamic prediction models were developed using machine learning techniques. We have previously also explored the potential of a machine learning model similar to Vodenčarević et al.(26) However, we chose to pursue the joint latent class model as the performance was similar, and the joint latent class model is more transparent regarding the DAS28-trajectories used and the effects of covariates in the model (i.e. providing hazard ratios).

A major unique strength of this study is that the model's performance is assessed in external validation. There were several significant differences between the patient populations from routine care used for developing the model and the DRESS pragmatic trial data for external validation regarding baseline characteristics, disease activity and bDMARD treatment. However, despite these differences the model retained an adequate performance in the external validation, indicating that these differences do not invalidate the model. Another strength is that the clinical impact is evaluated in simulation. In this simulation, successful tapering was not only defined by reaching a lower bDMARD dose, but also by the number of flares during tapering. Furthermore, our model was developed using easily obtainable parameters from routine care EHR data, rather than e.g. clinical trial data or specific biomarkers.(27)

The AUC in cross-validation and external validation (0.76 and 0.68, respectively) may be interpreted as only a moderate performance. However, the AUC may not be the most suitable measure to assess the model's clinical utility. The added value in clinical practice is determined by the effects of prediction-aided treatment on the rate of flares and the amount of bDMARD dose reduction, when compared to the available alternatives. The currently existing alternatives are either continuing the bDMARD at full dose or tapering until a flare occurs in a trial-and-error approach. Our simulation results show that prediction-aided treatment is superior to both these alternatives regarding the ratio between the number of flares and the amount of bDMARD dose reduction. Therefore, prediction-aided treatment may present the best available bDMARD tapering strategy. This is currently being investigated in the PATIO randomized controlled clinical trial (Dutch Trial Register number NL9798).

Interestingly, the AUC of the prediction model improved in external validation from 0.68 to 0.71 when baseline predictions were removed. This is likely because the model can only function as a 'joint' model when longitudinal information is available. This effect on AUC was also observed in the development data, but due to the relative overrepresentation of baseline visits in the DRESS data compared to the development data, this was less pronounced. As the removal of baseline predictions had almost no effect on the simulation of clinical impact, we chose to retain these predictions. Including disease activity measures prior to the start of tapering could potentially

improve the performance of our model, as this would ensure that longitudinal information is available at baseline.

A challenge in this study was the limited data quality regarding the frequency of DAS28 measurements in the development data. This might also have contributed to the different flare rates and resulting discrepancy between the optimal cut-off points in the development data and external validation data from the DRESS trial. When implementing a prediction-aided bDMARD tapering strategy in clinical practice or clinical studies, a treat-to-target (T2T) strategy with regular (e.g. 3 monthly) DAS28 measurements should be used, in line with EULAR recommendations.(5) As the DAS28 measurement frequency in the DRESS trial best reflects these recommendations, the optimal cut-off point found in simulation (i.e. 35%) is likely the most suitable for implementation of the model in clinical practice.

Besides the DAS28 measurements, several other parameters were also difficult to extract as structured data from the EHR, such as smoking, concurrent csDMARDs and erosiveness of disease. We explored imputation to increase the amount of these data points, but this did not improve the model's performance in cross-validation. Improved registration of these parameters and the optimization of free text mining techniques could allow for future inclusion of these parameters in model development and possibly a better performance. Importantly, the results from external validation are not biased by missing data, since the DRESS data had a standard measurement frequency and very few data missing on disease activity. Therefore, we think our simulation should be an accurate representation of the potential clinical impact of using the models predictions as an decision aid added to a DGDO strategy.

Since prediction-aided treatment could reduce the number of flares during bDMARD tapering, patients and physicians may be more willing to start tapering with such a prediction model than without.(28) Furthermore, our prediction model can be used as an add-on to DGDO, retains most of the bDMARD reduction as attained by DGDO and is a low cost intervention. Therefore, the model might prove to be an even more cost-effective strategy than DGDO alone.(10) The clinical implementation may be relatively straightforward, as it uses only predictors usually available in the EHR.

Conclusions

In conclusion, we developed and validated a dynamic prediction model to predict the risk of a flare occurring within 3 months during a bDMARD tapering strategy. In simulation, we showed that a prediction-aided treatment strategy has the potential to significantly reduce the number of flares, whilst maintaining most of the bDMARD dose

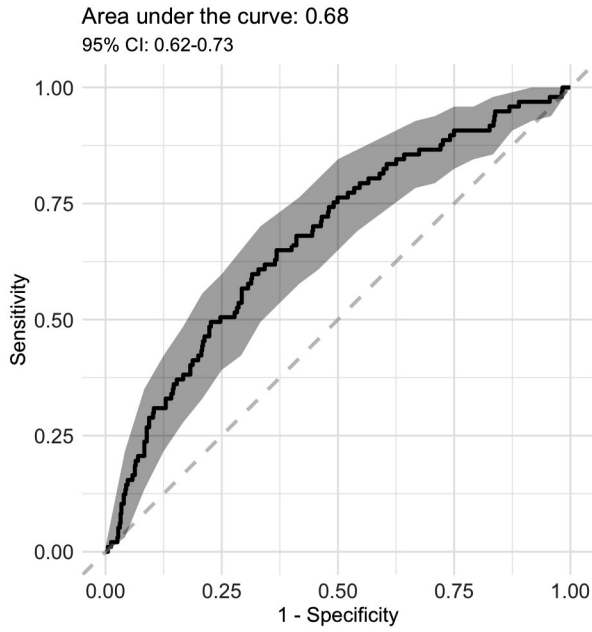
reduction. As this simulation is inevitably based on certain assumptions, we are currently investigating the clinical impact of prediction aided treatment in the PATIO randomized controlled trial. The current study and the PATIO-trial provide the next step towards the successful implementation of personalized medicine using clinical decision support systems.

References

1. Aga AB, Lie E, Uhlig T, Olsen IC, Wierød A, Kalstad S, et al. Time trends in disease activity, response and remission rates in rheumatoid arthritis during the past decade: Results from the NOR-DMARD study 2000-2010. Ann Rheum Dis. 2015;74(2):381–8.
2. Joensuu JT, Huoponen S, Aaltonen KJ, Konttinen YT, Nordström D, Blom M. The cost-effectiveness of biologics for the treatment of rheumatoid arthritis: a systematic review. PLoS One [Internet]. 2015;10(3):e0119683. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25781999
3. Bittner B, Richter W, Schmidt J. Subcutaneous Administration of Biotherapeutics: An Overview of Current Challenges and Opportunities. BioDrugs [Internet]. 2018;32(5):425–40. Available from: https://doi.org/10.1007/s40259-018-0295-0
4. Ramiro S, Sepriano A, Chatzidionysiou K, Nam JL, Smolen JS, Van Der Heijde D, et al. Safety of synthetic and biological DMARDs: A systematic literature review informing the 2016 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis. 2017;76(6):1093–101.
5. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020;1–15.
6. Verhoef LM, Van Den Bemt BJF, Van Der Maas A, Vriezekolk JE, Hulscher ME, Van Den Hoogen FHJ, et al. Down-titration and discontinuation strategies of tumour necrosis factor-blocking agents for rheumatoid arthritis in patients with low disease activity. Cochrane Database Syst Rev. 2019;5(5):CD010455.
7. Ruscitti P, Sinigaglia L, Cazzato M, Grembiale RD, Triolo G, Lubrano E, et al. Dose adjustments and discontinuation in TNF inhibitors treated patients: when and how. A systematic review of literature. Rheumatology (Oxford). 2018;57:vii23–31.
8. Fautrel B, Pham T, Alfaiate T, Gandjbakhch F, Foltz V, Morel J, et al. Step-down strategy of spacing TNF-blocker injections for established rheumatoid arthritis in remission: Results of the multicentre non-inferiority randomised open-label controlled trial (STRASS: Spacing of TNF-blocker injections in Rheumatoid ArthritiS St. Ann Rheum Dis. 2016;75(1):59–67.
9. Van Herwaarden N, Van Maas A Der, Minten MJM, Van Den Hoogen FHJ, Kievit W, Van Vollenhoven RF, et al. Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis:Open label, randomised controlled, non-inferiority trial. BMJ. 2015;350:1–8.
10. Den Broeder N, Bouman CAM, Kievit W, Van Herwaarden N, Van Den Hoogen FHJ, Van Vollenhoven RF, et al. Three-year cost-effectiveness analysis of the DRESS study: Protocolised tapering is key. Ann Rheum Dis. 2019;78:141–2.
11. van Mulligen E, Weel AEAM, Kuijper TM, Hazes JMW, van der Helm- van Mil AHM, de Jong PHP. The impact of a disease flare during tapering of DMARDs on the lives of rheumatoid arthritis patients. Semin Arthritis Rheum. 2020;50:423–31.
12. Tweehuysen L, van den Ende CH, Beeren FMM, Been EMJ, van den Hoogen FHJ, den Broeder AA. Little Evidence for Usefulness of Biomarkers for Predicting Successful Dose Reduction or Discontinuation of a Biologic Agent in Rheumatoid Arthritis: A Systematic Review. Arthritis Rheumatol. 2017;69(2):301–8.
13. Schlager L, Loiskandl M, Aletaha D, Radner H. Predictors of successful discontinuation of biologic and targeted synthetic DMARDs in patients with rheumatoid arthritis in remission or low disease activity: A systematic literature review. Rheumatol (United Kingdom). 2020;59:324–34.
14. Salaffi F, Ciapetti A. Clinical disease activity assessments in rheumatoid arthritis. Int J Clin Rheumtol. 2013;8:347–60.
15. Van Der Maas A, Lie E, Christensen R, Choy E, De Man YA, Van Riel P, et al. Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria: An OMERACT cohort validation study. Ann Rheum Dis. 2013;72(11):1800–5.
16. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. J Stat Softw. 2017;78.

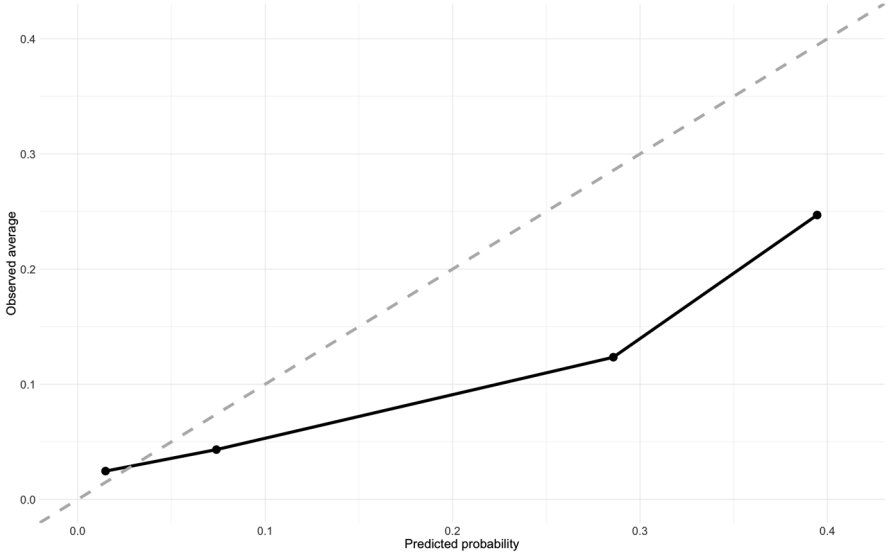
17. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues. BMC Med Res Methodol [Internet]. 2016;16(1):1–15. Available from: http://dx.doi.org/10.1186/s12874-016-0212-5
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ. 2015;350(January):1–9.
19. Youden WJ. Index for rating diagnostic tests. Cancer. 1950 Jan;3(1):32–5.
20. Bouman CAM, van Herwaarden N, Blanken AB, Van der Laken CJ, Gotthardt M, Oyen WJG, et al. 18F-FDG PET-CT scanning in rheumatoid arthritis patients tapering TNFi: reliability, validity and predictive value. Rheumatology (Oxford) [Internet]. 2021 Nov 13; Available from: http://www.ncbi.nlm.nih.gov/pubmed/34791068
21. Van Herwaarden N, Van Den Bemt BJF, Wientjes MHM, Kramers C, Den Broeder AA. Clinical utility of therapeutic drug monitoring in biological disease modifying anti-rheumatic drug treatment of rheumatic disorders: a systematic narrative review. Expert Opin Drug Metab Toxicol [Internet]. 2017 Aug;13(8):843–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28686523
22. Tweehuysen L, den Broeder N, van Herwaarden N, Joosten LAB, van Lent PL, Vogl T, et al. Predictive value of serum calprotectin (S100A8/A9) for clinical response after starting or tapering anti-TNF treatment in patients with rheumatoid arthritis. RMD open [Internet]. 2018;4(1):e000654. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29657832
23. Verhoef LM, Bos D, van den Ende C, van den Hoogen F, Fautrel B, Hulscher ME, et al. Cost-effectiveness of five different anti-tumour necrosis factor tapering strategies in rheumatoid arthritis: a modelling study. Scand J Rheumatol [Internet]. 2019 Nov;48(6):439–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31220991
24. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw open. 2019;2(3):e190606.
25. Vodencarevic A, Tascilar K, Hartmann F, Reiser M, Hueber AJ, Haschka J, et al. Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. Arthritis Res Ther. 2021;23:67.
26. Vodenčarević A, Goes MC van der, Medina OAG, Groot MCH de, Haitjema S, Solinge WW van, et al. Predicting Flare Probability in Rheumatoid Arthritis using Machine Learning Methods. 2018;(Data):187–92.
27. Bouman CAM, van der Maas A, van Herwaarden N, Sasso EH, van den Hoogen FHJ, den Broeder AA. A multi-biomarker score measuring disease activity in rheumatoid arthritis patients tapering adalimumab or etanercept: predictive value for clinical and radiographic outcomes. Rheumatology (Oxford). 2017 Jun;56(6):973–80.
28. Verhoef LM, Selten EMH, Vriezekolk JE, de Jong AJL, van den Hoogen FHJ, den Broeder AA, et al. The patient perspective on biologic DMARD dose reduction in rheumatoid arthritis: A mixed methods study. Rheumatol (United Kingdom). 2018;57(11):1947–55.

Supplemental Figure S1. Receiver operating characteristic (ROC) curve in external validation



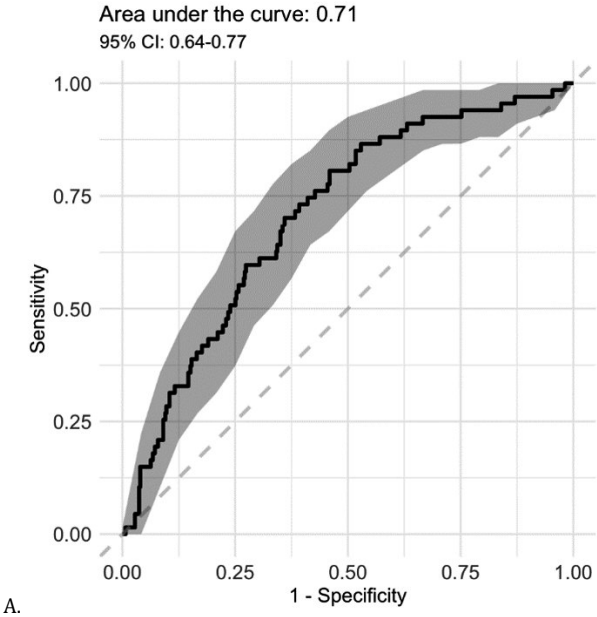
ROC-curve of the model in external validation in data of the Dose Reduction Strategy of Subcutaneous TNF inhibitors (DRESS) trial.(9)

Supplementary Figure S2. Calibration plot of flare prediction model including baseline predictions

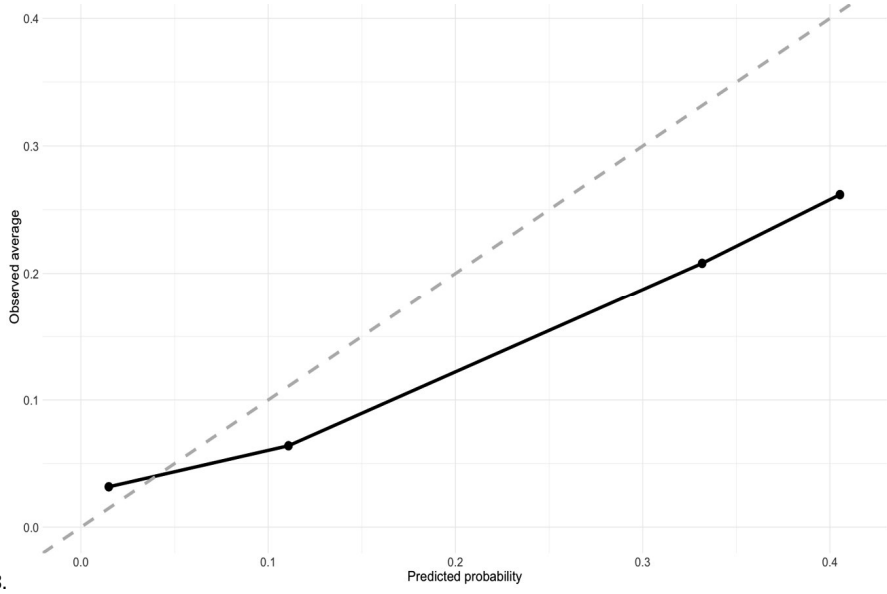


Calibration plot in external DRESS-data(9)*. Patients were grouped based on their predicted probability from lowest to highest predicted 3-monthly risk of flare (x-axis) using the median, 25th and 75th percentile. On the y-axis these groups are compared with the observed frequency of flare within 3 months. Perfectly calibrated predictions would be expected to be at the diagonal.*

Supplementary Figure S3. AUC and calibration plot without baseline predictions



A.



B.

A. *Receiver operator characteristic (ROC)-curve of external validation of the flare prediction model in DRESS data*(9)*, where baseline predictions are removed. The rationale is that the prediction model cannot truly function as a 'joint' model at baseline, as no longitudinal data is available. B: Calibration plot in DRESS-data, excluding baseline predictions. Patients were grouped based on their predicted probability from lowest to highest predicted 3-monthly risk of flare (x-axis) using the median, 25th and 75th percentile. On the y-axis these groups are compared with the observed frequency of flare within 3 months. Perfectly calibrated predictions would be expected to be at the diagonal. AUC: Area Under the Curve*

Supplementary Table S1. Predictive performance without baseline predictions in DRESS data

	External validation cut-off 14.3%	**External validation cut-off 31.5%**
AUC	0.71 (0.64–0.77)	0.71 (0.64–0.77)
Sensitivity (%)	86.6 (78.4–94.7)	71.6 (60.8–82.4)
Specificity (%)	46.9 (42.2–51.6)	61.3 (56.7–65.9)
Positive Predictive Value (%)	20.2 (15.6–24.6)	22.3 (16.8–27.9)
Negative Predictive Value (%)	95.7 (93.0-98.5)	93.3 (90.4–96.2)
Accuracy (%)	52.2 (47.7–56.7)	62.7 (58.2–66.9)

95% confidence intervals are presented between brackets. The results from external validation in the DRESS trial(9) *without baseline predictions. The rationale for leaving out baseline predictions is that the prediction model cannot truly function as a 'joint' model at baseline, as no longitudinal data is available. The results for 2 different cut-off points are presented: the optimal cut-off point from the development data (14.3%) and the optimal cut-off point in the DRESS data as determined by Youden's index (31.5%). AUC: Area under the curve*

Supplementary Table S2. Simulation results for different cut-off points (baseline predictions included)

	15%	**25%**	**35%**	**45%**
Mean no. of flares	0.45 (0.26-0.63)	0.51 (0.32–0.70)	0.75 (0.55-0.95)	1.16 (0.94–1.39)
Decrease in flares compared to DRESS DGDO	0.76 (47.0-1.05)	0.70 (0.42–0.99)	0.45 (0.16-0.74)	0.05 (-0.25–0.36)
Mean bDMARD dose	0.80 (0.77-0.84)	0.75 (0.72–0.78)	0.64 (0.61-0.68)	0.55 (0.50-0.59)
Increase in bDMARD dose compared to DRESS DGDO	0.27 (0.21-0.32)	0.21 (0.16–0.26)	0.10 (0.05–0.16)	0.01 (-0.04–0.07)
Percentage of patients flaring	23% (25%-32%)	28% (19%-37%)	45% (36%-54%)	68% (59%-77%)
Increase in bDMARD dose per flare prevented vs. DRESS DGDO[a]	0.34 (0.30-0.39)	0.30 (0.26-0.34)	0.23 (0.15-0.32)	0.21 (-0.68–1.10)
Number of extra flares per full bDMARD dose saved vs. routine care[b]	-0.3 (-2.1–1.4)	0.2 (-1.0–1.3)	1.0 (0.3–1.8)	1.9 (1.3–2.5)

95% confidence intervals are presented between brackets. a. The mean difference in bDMARD dose divided by the mean number of flares compared with the DRESS(9) *DGDO arm. The number therefore represents the increase in bDMARD dose that was needed to prevent a flare for this specific tapering strategy. b. The mean difference in the number of flares, divided by the mean difference in bDMARD dose, compared to routine care. The ratio thus represents the number of extra flares that occurred for each extra full dose of bDMARD that is tapered compred to routine care over 18 monhts using this specific tapering strategy. bDMARD: biological disease modifying antirheumatic drug, DGDO: disease activity guided dose optimisation.*

Supplementary Table S3. Simulation results with and without baseline predictions

	With baseline predictions cut-off 35%	**Without baseline predictions cut-off 35%**
Mean no. of flares	0.75 (0.55-0.95)	0.76 (0.56–0.97)
Decrease in flares compared to DRESS DGDO	0.45 (0.16-0.74)	0.45 (0.15–0.74)
Mean bDMARD dose	0.64 (0.61-0.68)	0.64 (0.60–0.68)
Increase in bDMARD dose compared to DRESS DGDO	0.10 (0.05–0.16)	0.10 (0.05-0.16)
Percentage of patients flaring	45% (36%-54%)	46% (37%-56%)
Increase in bDMARD dose per flare prevented vs. DRESS DGDO[a]	0.23 (0.15-0.32)	0.23 (0.15–0.31)
Number of extra flares per full bDMARD dose saved vs. routine care[b]	1.0 (0.3–1.8)	1.0 (0.3–1.8)

95% confidence intervals are presented between brackets. The results from external validation in the DRESS trial(9) *without baseline predictions, for the optimal cut-offpoint of 35% as determined in simulation (see Supplementary Table S2). The rationale for leaving out baseline predictions is that the prediction model cannot truly function as a 'joint' model at baseline, as no longitudinal data is available. a. The mean difference in bDMARD dose divided by the mean number of flares compared with the DRESS DGDO arm. The number therefore represents the increase in bDMARD dose that was needed to prevent a flare for this specific tapering strategy. b. The mean difference in the number of flares, divided by the mean difference in bDMARD dose, compared to routine care. The ratio thus represents the number of extra flares that occurred for each extra full dose of bDMARD that is tapered compred to routine care over 18 monhts using this specific tapering strategy. bDMARD: biological disease modifying antirheumatic drug, DGDO: disease activity guided dose optimisation.*

Chapter 8

Prediction Aided Tapering In rheumatoid arthritis patients treated with biOlogicals (PATIO): Protocol for a randomized controlled trial

Marianne A. Messelink
Matthijs S. van der Leeuw
Alfons A. den Broeder
Janneke Tekstra
Marlies C. van der Goes
Marloes W. Heijstek
Floris P.J.G. Lafeber
Paco M.J. Welsing

Trials, 2022, 23(1):494

Abstract

Background Biological Disease Modifying Anti Rheumatic Drugs (bDMARDs) are effective in the treatment of rheumatoid arthritis (RA), but are expensive and increase the risk of infection. Therefore, in patients with a stable low level of disease activity or remission, tapering bDMARDs should be considered. Although tapering does not seem to affect long term disease control, (short-lived) flares are frequent during the tapering process. We have previously developed and externally validated a dynamic flare prediction model for use as a decision aid during stepwise tapering of bDMARDs to reduce the risk of a flare during this process.

Methods In this investigator-initiated, multicenter, open-label, randomized (1:1) controlled trial, we will assess the effect of incorporating flare risk predictions into a bDMARD tapering strategy. 160 RA patients treated with a bDMARD with stable low disease activity will be recruited. In the control group, the bDMARD will be tapered according to 'disease activity guided dose optimization' (DGDO). In the intervention group, the bDMARD will be tapered according to a strategy that combines DGDO with the dynamic flare prediction model, where the next bDMARD tapering step is not taken in case of a high risk of flare. Patients will be randomized 1:1 to the control or intervention group. The primary outcome is the number of flares per patient (DAS28-CRP increase > 1.2, or DAS28-CRP increase >0.6 with a current DAS28-CRP ≥ 2.9) during the 18-month follow-up period. Secondary outcomes include the number of patients with a major flare (flare duration ≥12 weeks), bDMARD dose reduction, adverse events, disease activity (DAS28-CRP) and patient reported outcomes such as quality of life and functional disability. Health Care Utilization and Work Productivity will also be assessed.

Discussion This will be the first clinical trial to evaluate the benefit of applying a dynamic flare prediction model as a decision aid during bDMARD tapering. Reducing the risk of flaring during tapering may enhance the safety and (cost)effectiveness of bDMARD treatment. Furthermore, this study pioneers the field of implementing predictive algorithms in clinical practice.

Introduction

Background and rationale

Biological disease modifying anti-rheumatic drugs (bDMARDs) are effective in the treatment of rheumatoid arthritis (RA), improving clinical, functional and radiographic outcomes.(1) Examples of bDMARDs include infliximab, adalimumab, etanercept, abatacept and tocilizumab. Many RA patients who are treated with bDMARDs achieve long periods of stable low disease activity or remission.(2) However, bDMARDs may also lead to adverse events, call for self-injections or hospital visits and are expensive.(3–5) Thus, tapering bDMARDs to the lowest effective dose is of great clinical interest and may support the sustainability of the healthcare system as a whole.

The most successful and cost-effective strategy for tapering appears to be 'disease activity guided dose optimization' (DGDO).(6–8) This means the dose is gradually tapered, until either disease activity flares or the bDMARD is discontinued. Two randomized trials have demonstrated that, using this strategy, 63%-80% of patients can taper or even stop their bDMARD.(6,7) No important differences were observed in the proportion of patients with low disease activity or remission after 18 months between DGDO and usual care(7).

However, since DGDO is a 'trial and error' approach, flares occur frequently during the tapering process, for which the previously effective dose needs to be reinstated or additional therapy is necessary. Although these short-lived flares do not seem to relevantly affect radiographic progression or long term disease activity, there is conflicting evidence regarding functional outcome and impact on quality of life.(7,9) Therefore, it would be beneficial to predict whether, and to which extent, a bDMARD can be tapered in a particular patient without a flare occurring.

For this purpose, we recently developed and externally validated a dynamic flare prediction model. (10,11) The model combines both fixed and longitudinal patient and disease characteristics to predict the risk of a flare at every visit to the outpatient clinic. In case of a high predicted risk, tapering can be halted in time to prevent a flare. We simulated the incorporation of the prediction model in a DGDO tapering strategy in data of the DRESS trial, which significantly reduced flares whilst maintaining bDMARD dose reduction.(7,11) In the presented randomized controlled trial (RCT) we will evaluate if these promising results can be confirmed in clinical practice. In this paper we will discuss the design of this RCT and its rationale.

Objectives

Our aim is to assess whether incorporating dynamic flare risk predictions into a

bDMARD tapering strategy can reduce the number of flares whilst maintaining bDMARD dose reduction in RA patients with a stable low level of disease activity.

Trial design
Pragmatic, open, randomized, superiority, multi-centre strategy trial with 18 months follow-up. Patients will be randomized in a 1:1 ratio.

Methods: Participants, interventions and outcomes

Study setting
Patients will be recruited through outpatient rheumatology clinics of participating centers in the Netherlands. Currently we aim at participation of seven centers in the Netherlands in this trial.

Eligibility criteria
Inclusion criteria

- A clinical diagnosis of rheumatoid arthritis as assessed by the treating rheumatologist.
  Compliance with (subcomponents of) the ACR 1987 or EULAR/ACR 2010 criteria will be reported.
- Treatment of their RA with one of the following bDMARDs in ≥66% (i.e. maximally one dose reduction step previously taken) of the defined daily dose based on the Dutch pharmacological guidelines(16): adalimumab, certolizumab, golimumab, infliximab, etanercept, sarilumab, tocilizumab or abatacept, either as monotherapy or in combination with a conventional synthetic (cs)DMARD.
- Patient is eligible to taper bDMARD according to the treating physician (e.g. no other indication for bDMARD such as psoriasis, no recent relevant radiographic progression).
- Stable low disease activity (LDA) on current bDMARD for ≥ 6 months according to treating physician
- Current stable low disease activity according to treating physician and a maximum DAS28-CRP of 3.5 (i.e. the cut-off point for low disease activity of 2.9 + measurement error in DAS28 (~0.6)(12–14))
- Patient is willing to taper (and if possible, stop) his/her bDMARD as well as to continue his/her current bDMARD dose.
- At least 18 years of age and consenting

Exclusion criteria

- Recent earlier (<6 months) tapering attempt(s) with the same bDMARD that failed according to treating physician.
- Inability to comply with protocol, e.g. no possibility to measure outcome over 18 months, insufficient knowledge of the Dutch language.

The in- and exclusion criteria aim to resemble the conditions in which bDMARD tapering would be considered in clinical practice. When both the patient and physician are satisfied with the level of disease activity and do not wish to intensify, maintaining this level of disease activity with less medication is of added clinical value. Therefore, if a patient has a DAS28-CRP slightly above the DAS28-CRP low disease activity (LDA) threshold of 2.9, but the treating physician and the patient judge that there is a stable low level of disease activity, the patient is still eligible for inclusion. This is also in line with the DRESS study(7), the fact that the DAS28-score has a measurement error of around 0.6 (12), and the possibility of overestimation of the DAS28-score in patients with comorbidities such as fibromyalgia(15).

If patients do not meet the ACR or EULAR/ACR diagnostic criteria, the individual components of these criteria will be registered to enable exploring this subgroup in more detail in later analysis. Furthermore, we have chosen not to include patients treated with rituximab, due to the long (often ≥ 6 months) dosing intervals of this bDMARD.

Who will take informed consent?

Patients will be approached for participation by their treating physician at the rheumatology department. In case a patient is interested to participate, a research physician or nurse will inform the patient and if appropriate obtain informed consent.

Additional consent provisions for collection and use of participant data and biological specimens

We will also ask patients for permission to approach them for future related research.

Interventions

Explanation for the choice of comparators

The aim of the current RCT is to assess whether the incorporation of flare risk predictions in a bDMARD tapering strategy can reduce the number of flares during tapering, whilst preserving the tapering potential as optimal as possible. For this purpose it is necessary to compare the current optimal tapering strategy with and without the incorporation of the results of a flare prediction model. The most successful

and cost-effective strategy for tapering appears to be DGDO.(6–8) This is a treat-to-target strategy, in which the bDMARD dose is gradually tapered, until either disease activity flares or the bDMARD is discontinued. During this process disease activity is monitored closely to enable a swift increase in bDMARD dose when a flare occurs, in order to quickly resolve the flare. In the current trial we will compare the DGDO strategy (control group) with a tapering strategy that combines DGDO with flare risk predictions (intervention group).

In the control arm of our trial (DGDO alone), the bDMARD is tapered stepwise every 3 months until the patient has a flare or until the bDMARD is discontinued. The tapering steps are defined as a rounded percentage of defined daily dose in the following order: 100% - 70% - 50% -33% - 0% (Figure 1). The defined daily dose is based on the Dutch national guidelines of standard dosages.(16) We chose this 4-step tapering approach rather than the 3-step tapering approach as used in DRESS(6) as this may reduce the risk of (severe) flaring in both groups(17), increases the number of decision steps where the prediction model can be applied, and is in line with other tapering studies.(5) Patients that have already made a tapering step and enter the study at ~70% (minimum 66%) of defined daily dose will follow the same schedule, with the exception of the first step (70% - 50% - 33% - 0%).

Tapering is done for most bDMARDs by increasing the administration interval, rather than the administration dose. This is because 1) for some bDMARDs lower dosages are not available (e.g. prefilled syringes of some bDMARDs only come in one dosage); 2) lower dosages can be flat priced, and thus more expensive per mg (e.g. sarilumab 150mg), and 3) this approach reduces the burden for the patient by reducing the amount of injections. The only exception to this rule is intravenously administered infliximab, as this already has a long standard dosing interval of 8 weeks. Further increasing this interval could reduce pharmacological efficacy(18), and would cause a dosing interval greater than the 3-monthly study visit interval. Dosing intervals are rounded to 0.5 weeks for feasibility reasons.

In case of a flare, the bDMARD dose is increased to the last effective dose, for both the control group and the intervention group. In addition, short-term glucocorticoids are allowed within the protocol. Flares are defined by an increase in DAS28-CRP of >1.2, or an increase of >0.6 where the resulting DAS28-CRP is >2.9. At each study visit, the DAS28-CRP will be compared with baseline, rather than the last visit, to prevent undertreatment of patients with a gradual increase of disease activity. When a flare occurs, no further tapering attempts will be taken during the follow-up period. If a flare persists during 3 months, the bDMARD is increased to full dose. If a flare persists at the full dose of the bDMARD, further treatment is at the discretion of the treating

rheumatologist.

If a patient experiences symptoms of a flare the patient will be encouraged to plan an extra (i.e. unscheduled) study visit (USV). During an USV no further tapering steps will be taken.

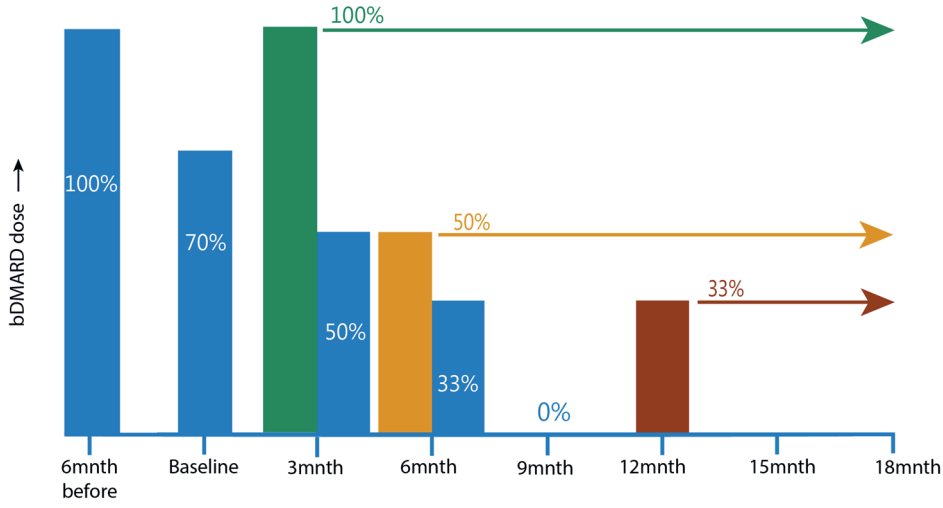
Intervention description
At each 3-monthly study visit, it will be assessed whether the tapering schedule as described above can be continued. In the intervention group, tapering is continued until a flare occurs (similar to the control group), *or* until there is a high predicted risk of a flare occurring in the next 3 months when taking the next tapering step.

The cut-off for a high predicted risk of flare is >= 35%, determined as optimal in the simulation of clinical impact of the prediction model. (11) In case of a high predicted risk of a flare, the bDMARD is kept at the same dose, and no further tapering attempts are taken (Figure 1). If the bDMARD is already discontinued and there is a high predicted risk of flare, the bDMARD will be restarted at 33% of the defined daily dose. The reasons for this exception are twofold. Firstly, the model predicts the risk of a flare when taking the next tapering step. For example, if there is a high predicted risk at the step from 50% to 33%, it will be advised to remain at 50% of the defined daily dose. But when the bDMARD is discontinued, the current dose and the next tapering step are however equal (both 0%), thus justifying a dose increase to 33%. Secondly, from a clinical perspective it was deemed undesirable for both rheumatologists and patients to keep a previously stopped bDMARD discontinued when there is a high predicted risk of a flare, especially in the case of bDMARD monotherapy.

Criteria for discontinuing or modifying allocated interventions
If the treating physician (together with the patient) determines the treatment advice will not be followed this will be registered together with the reason for deviating from the advice.

Strategies to improve adherence to interventions
For the purpose of this study, a web-based dashboard has been developed, that will be used in both the control group and the intervention group to facilitate adherence to the treatment protocol. This dashboard displays the disease activity (including the current presence of a flare) and DMARD use over time throughout the study, as well as the treatment advice specific for each patient. Rheumatologists are encouraged and trained to adhere to the treatment protocol, unless medical reasons require a deviation. These deviations must be registered in the study database. Furthermore, at each visit we will register whether the patient adhered to the treatment advice of the previous visit.

Figure 1: Tapering schedule.



The x-axis shows the time in months. The y-axis shows the bDMARD dose expressed as percentage of the defined daily dose.(16) At every study visit it will be assessed if tapering should be continued. Below are several examples, where the colors of the text correspond to the colors of the bars in the figure above:
Standard tapering scheme (both groups): If no flares occur and there is no high predicted risk of flare, then the bDMARD is tapered from 100% and discontinued at 9 months until the end of the study.
Example 1 (both groups): The patient tapered until 70%. At the 3-month visit a flare occurs. The patient increases the dose to 100% and remains at this dose until the end of the study.
Example 2 (intervention group): The patient tapered until 50%. At the 6-month visit there is a high predicted risk of a flare. The patient does not taper further and remains at 50% until the end of the study.
Example 3 (intervention group): The patient tapered until 0%. At the 12-month visit there is a high predicted risk of a flare. The patient increases the dose to 33% and remains at this dose until the end of this study.

Relevant concomitant care permitted or prohibited during the trial

If a patient uses conventional synthetic (cs)DMARDs and/or glucocorticoids at inclusion are aimed to be continued in a stable dose during the study. An exception is made for short-term (max. 2 weeks) oral prednisone use to treat a flare to a maximum of 10 mg. csDMARD, glucocorticoid and NSAID use will be registered throughout the duration of the study.

Provisions for post-trial care

As both the control group and the intervention group will be treated with bDMARDs in dosages that are currently common in clinical practice, we do not expect (post-trial) harm from participation in the study. There are no specific provisions for post-trial care. Post-trial treatment is at the discretion of the treating rheumatologist.

Outcomes

Primary outcome
As we aim to determine whether incorporating flare risk predictions in a bDMARD tapering strategy can reduce flares during tapering to the lowest effective dose, our primary outcome is the number of flares per patient during the follow-up period of 18 months.

A flare is defined in line with previously validated definitions as follows(13,14,19):

Compared to DAS28-CRP at baseline

- an increase in DAS28-CRP > 1.2 or
- an increase in DAS28-CRP > 0.6, where the resulting DAS28-CRP > 2.9

Both compared to DAS28-CRP at baseline (i.e. before start of tapering).

Secondary outcomes
To determine the clinical benefit of the incorporation of flare risk predictions in a bDMARD tapering strategy, the number of flares per patient must be interpreted together with the bDMARD dose reduction. If the addition of our flare prediction model is able to reduce flares, but does not reduce the bDMARD dose, it is of no use in a tapering strategy. Furthermore, the consequences of possible undertreatment or overtreatment with bDMARDs should be taken into account. These include the patient's quality of life and adverse events related to either bDMARDs or disease activity.

- Clinical outcome measures
    - Presence of any (one or more) flare during the study
    - Presence of any major flare during the study (flare duration >12 weeks)
    - DAS28-CRP over time
    - Mean bDMARD dose reduction, expressed as percentage of defined daily dose over 18 months (monthly full dose equivalents)
    - Use (dose and duration) of anti-rheumatic drugs other than bDMARDs during the study period
- Patient-reported outcomes
  Measured every 3 months:
    - Functional disability using the Health Assessment Questionnaire (HAQ DIv2 Dutch version)(20)
    - Quality of Life using the EQ5D5L(21)
    - Provider assessed general disease activity (GDA) on a visual analogue scale (VAS)
    - Patient assessment of pain on a VAS 0-100mm

  - Patient Acceptable Symptom State (PASS)(22)
  - 7 scale Likert transition question(23), assessing the change in symptoms compared to the last study visit.

  Measured every month and in case of symptoms of a flare:
  - Flare severity score: OMERACT RA Flare questionnaire(24)

  Measured at the last study visit:
  - Patient satisfaction with treatment (SAPS)(25)
  - Physician satisfaction with care on a VAS 0-100mm
- Safety outcomes
  - Infections for which antibiotic, anti-viral treatment or antimycotic therapy is prescribed
  - (Serious) Adverse events (probably or definitively) related to the bDMARD as assessed by the treating physician other than infections using the Rheumatology Common Toxicity Criteria(26)
  - (Serious) Adverse events related to increased disease activity due to tapering other than joint complaints and DAS28-CRP-based flares as assessed by the treating physician.
- Other
  - Tapering attempted (yes/no) up to flare/high-risk of flare or discontinuation of the bDMARD
  - Proportion of clinical visits where treatment advice is not followed
  - Reasons for not following the advised treatment steps in both arms

Cost-effectiveness

Observed anti-rheumatic drug use and visits to the rheumatology outpatient clinic will be recorded. Direct medical (e.g., general practitioner visits) and nonmedical (e.g. travel expenses) costs as well as indirect costs (e.g., productivity loss) will be obtained using a Health Care Utilization and Work Productivity Questionnaire.

General characteristics

We will collect the following general patient characteristics at baseline:

- Demographic data: sex, age height and weight, level of education
- Smoking status (current, ever, never), packyears and alcohol use (current units/day)
- Medical history: year of RA diagnosis, rheumatoid factor and anti-CCP positivity, the Charlson Comorbidity index(27)
- Anti-rheumatic treatment: current use of anti-rheumatic treatment, including dose

  and frequency of b/csDMARDs, glucocorticoids and NSAIDs. Previous DMARD use will also be recorded.

Participant timeline

Table 1: overview of participant timeline.

Visit number	**V0 Screening**	**V1 Baseline**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**USV**
Month	**-3 to 0**	**0**	**3**	**6**	**9**	**12**	**15**	**18**	**....**
Informed consent	X								
Eligibility criteria	X								
Demographic data		X							
Medical history	X								
Anti-rheumatic treatment	X	X	X	X	X	X	X	X	X
Smoking and alcohol use		X							
Height, Weight		X							
Randomization		X							
Clinical disease parameters									
28TJC and 28SJC	X	X	X	X	X	X	X	X	X
Provider VAS GDA		X	X	X	X	X	X	X	X
Patient VAS GDA	X	X	X	X	X	X	X	X	X
Questionnaires									
HAQ-DI		X	X	X	X	X	X	X	
EQ-5D-5L		X	X	X	X	X	X	X	
RA flare questionnaire*		X	X	X	X	X	X	X	X*
PASS		X	X	X	X	X	X	X	X*
Likert transition question		X	X	X	X	X	X	X	X*
Health care utilization and work participation		X	X	X	X	X	X	X	
Patients satisfaction with care (Dutch SAPS)								X	
Physician satisfaction with care (VAS)								X	
Laboratory assessments									
CRP		X	X	X	X	X	X	X	X
Safety/ adverse events									
(Serious) Adverse Events		X	X	X	X	X	X	X	X

USV: unscheduled visit. Patients with complaints of a flare will be encouraged to plan an unscheduled visit.

* : Patients will also be asked to fill out this questionnaire in between visits at 4-weekly intervals and if patients experience any disease related complaints

CRP: C-reactive protein, EQ-5D-5L: questionnaire assessing quality of life, GDA: global disease activity, HAQ: health assessment questionnaire, PASS: patient acceptable symptom state, RA: rheumatoid arthritis, SJC: swollen joint count, TJC: tender joint count, VAS: visual analogue scale 0-100mm.

Sample size
Based on our simulation results, over the follow-up of the trial (18 months) a mean of about 1.2 flares per patient can be expected in the DGDO control group. The number of flares is estimated to reduce from 1.2 to 0.75 flares over 18 months when incorporating the flare predictions. (11) We thus expect a reduction in flares of about 38% (relative risk of 0.63)

Assuming a more conservative relative risk of 0.65 and a base flare rate of 1.2, using a sample size calculation for a Poisson regression analysis and the program G*Power version 3.1.9.2, 152 patients (76 per group) are needed to detect this difference with a power of 80% and a two-sided alpha of 0.05. Therefore, we will include 160 patients (80 per arm) in our study, taking possible loss to follow-up into account.

Recruitment
Several outpatient rheumatology clinics in the Netherlands (aim 7) will participate in this trial. We have planned an enrolment period of 18 months. As the inclusion criteria apply to a large proportion of RA patients, the recruitment is deemed feasible.

Assignment of interventions: allocation
Sequence generation
Patients will be randomly assigned to either control or intervention group in a 1:1 ratio, stratified by centre using random block sizes.

Concealment mechanism
Randomization will be performed by the validated randomization algorithm in the Castor electronic data capture (EDC) system. The exact randomization algorithm is unknown to any of the investigators, thereby ensuring allocation concealment.

Implementation
Patients will be enrolled by a research physician or nurse of the rheumatology departments of participating centres. All patients who give consent for participation and who fulfil the inclusion criteria will be randomized. The allocation sequence will be generated by Castor EDC.

Assignment of interventions: Blinding
Who will be blinded
This study will not be blinded. In the control group, the tapering process is halted when a flare occurs. In the intervention group, the tapering process is halted when a flare occurs *or* when there is a high predicted risk of flare. Therefore, if it is advised to halt the tapering process in the absence of a flare, it will be evident that this is due to a high

predicted risk of flare. Blinding is thus not feasible in this study. The outcome measures will be assessed by an (unblinded) research physician or nurse, and by the patient. These include (partly) objective measures, such as the bDMARD dose and the DAS28-CRP. Knowledge of the assigned group might influence the subjective outcome measures in the benefit of the intervention group. For this we will investigate the subjective vs. objective components of the disease activity scores within both groups.

Procedure for unblinding if needed
Not applicable as this study is not blinded.

Data collection and management
Plans for assessment and collection of outcomes
The clinical and safety outcome measures will be determined by a physician or research nurse of the rheumatology department and will be registered in Castor EDC. Participating physicians and nurses will be trained prior to the start of the study.

Flares are defined based on the validated measure as described by van der Maas et al. (19) We have chosen for the validated measure of DAS28-CRP rather than DAS28-BSE as CRP levels are more sensitive to short-term changes in disease activity, and ESR can be more influenced by a number of unrelated factors. (28,29) Patient reported outcomes and questionnaires are collected either electronically via e-mail (preferred) or on paper.(19–25)

Plans to promote participant retention and complete follow-up
Participating centers will regularly receive updates of trial progress and promotion/practical material to enhance inclusion and follow-up of participants. If patients do not follow the tapering steps in line with the treatment protocol patients will remain in follow-up according to the protocol and reasons for deviations from the treatment protocol will be recorded.

Data management
We created a data management plan in line with the General Data Protection Regulation, which can be viewed upon request. Data will be managed within the Castor EDC system. This is a secure cloud-based platform that contains automatic range checks, and study IDs are used to pseudonymize all data.

Confidentiality
The type of data that is collected is in line with the General Data Protection Regulation. In Castor study IDs are used to pseudonymize all data. The key to the study ID is safely kept by the (local) coordinating investigator in each participating center. Data will be

stored for 15 years.

Plans for collection, laboratory evaluation and storage of biological specimens for genetic or molecular analysis in this trial/future use
Not applicable. No biological specimens will be stored.

Statistical methods
Statistical methods for primary and secondary outcomes

- Primary endpoint
  The number of flares over 18 months per patient will be compared (superiority testing) between strategies using Poisson regression as appropriate for a 'count' outcome variable. Center (as stratification factor used in randomisation), bDMARD line (1st, 2nd or >2nd bDMARD) and baseline disease activity will be used as (prognostic) covariates in this analysis. The appropriateness of the assumption regarding the Poisson distribution will be checked before performing the analysis. In case overdispersion is present, alternative analysis methods like the use of a negative binomial model will be considered according to the (at that time) state of the art. This will be defined in a formal statistical analysis plan to be finalized before database lock.
  The primary analysis will be performed on the intention to treat (ITT) population consisting of all patients who were randomized to one of the strategies. All tests of significance will be performed two-sided with α=0.05.
- Secondary endpoints
  Binary secondary outcomes will be compared between strategies using logistic regression analysis. For secondary continuous outcomes over time a mixed effects model will be used to account for clustering of measurements within patients over time. The secondary analyses will be corrected for the same covariates as stated in the description of the primary endpoint analysis and according to the ITT principle, with a secondary analysis in the per protocol population.

Interim analyses
No interim analyses will be performed. As the treatments in both study arms are within the range of usual care, we do not anticipate differences between the arms that warrant early cessation of the study due to detrimental effects to the participant.

Methods for additional analyses (e.g. subgroup analyses)
We will perform an exploratory subgroup analysis to compare the effect of the use of the flare risk predictions between patients using a TNF-inhibitor (TNFi) and patients using a different biological. This will also be tested using a modelling approach with an

interaction term between the type of bDMARD (TNFi or non-TNFi) and the intervention-arm.

A trial based economic evaluation as well as a budget impact analysis will be performed. A Dutch Healthcare perspective as well as a societal perspective will be used in these analyses. The analysis will be performed according to Dutch guidelines for economic evaluations. (30) Extensive sensitivity analyses regarding e.g. the price of biological DMARDs will be performed.
All analyses will be further specified in the Statistical Analysis Plan, which can be viewed upon request.

Methods in analysis to handle protocol non-adherence and any statistical methods to handle missing data
In the case that more than 10% of patients have missing outcome values, data will be imputed before analysis, using multiple imputation by chained equations with baseline characteristics and disease activity characteristics of previous study visits as predictor.

Plans to give access to the full protocol, participant level-data and statistical code
The trial protocol is available from the corresponding author on reasonable request. Data will be handled according to the FAIR principles. After completion of the study metadata will be available upon request. For access to participant level-data and statistical code an application can be submitted to the corresponding author which will be reviewed by the trial steering committee.

Oversight and monitoring
Composition of the coordinating centre and trial steering committee
The trial steering committee consists of PW, AB, JT, AM and MM, as stated on the title page. Their responsibilities include agreement on the final protocol, reviewing progress of study and data collection, and if necessary agreeing changes to the protocol or budget to facilitate the smooth execution of the study. The trial steering committee will discuss the progress of the trial at least twice a year, and more often if necessary. In addition they will actively search for new published data that may be relevant for this trial.
(Local) research physicians/nurses, (local) study coordinators and (local) principal investigators are, together with the trial steering committee, responsible for running the trial day-to-day and providing organisational support.

Composition of the data monitoring committee, its role and reporting structure
A data management plan has been created in collaboration with the data manager of the coordinating centre and is available upon request. As both arms of this study are within the spectrum of regular care, no data safety monitoring board is installed.

Adverse event reporting and harms
Serious adverse events of special interest for bDMARD treatment (AESI) will be collected via Castor EDC and will be reported to the CCMO (Central Committee on Research Involving Human Subjects). All (serious) adverse events ((S)AEs) reported will be classified according to the Rheumatology Common Toxicity Criteria v.2.1. (26)

Frequency and plans for auditing trial conduct
The study will be monitored by a central monitor of the University Medical Centre Utrecht, according to the monitoring plan in line with the guidelines of the Dutch Federation of University Medical Centres (NFU). (31)

Plans for communicating important protocol amendments to relevant parties (e.g. trial participants, ethical committees)
Any modifications to the protocol which may impact on the conduct of the study, potential benefit of the patient or may affect patient safety, including changes of study objectives, study design, patient population, sample sizes, study procedures, or significant administrative aspects will require a formal amendment to the protocol. Such amendments will be agreed upon by the trial steering committee, need to be approved by the Ethics Committee prior to implementation and notified to the health authorities in accordance with local regulations. When applicable, amendments will be made to the registration in the Netherlands Trial Register. (32)

Dissemination plans
Overall trial results will be communicated to participants and will be published in a scientific journal. There are no publication restrictions. Thereafter the research team will proactively disseminate the results through (inter)national congresses and in appropriate recommendations and guidelines, amongst others by A.A. den Broeder, who is a member of the EULAR (European League Against Rheumatism) RA recommendations group.

Discussion

The PATIO trial will be the first study to assess the effect of incorporating the results of a dynamic flare prediction model into a bDMARD tapering strategy. With this study we will address two major challenges in modern day health care: improving (cost-)effectiveness of treatment and implementation of predictive algorithms in clinical practice. We will discuss these topics separately.

Over the past decades the steep increase in health care costs has led to a growing

interest in health care cost-effectiveness research. Tapering bDMARDs has the potential to increase cost-effectiveness by reducing medication costs, side effects and patient burden whilst maintaining the same patient outcome as treatment with a full bDMARD dose. The currently available bDMARD tapering strategies, however, still give an increased risk of short lived flares.(6,7) The conflicting evidence regarding the impact of these flares on functional outcome and impact on quality of life (7,9), may make physicians and patients hesitant to start the tapering process.(33) The implementation of the flare prediction model into a bDMARD tapering strategy has the potential to reduce the risk of a flare, and may thus also increase the willingness to start bDMARD tapering.

The second challenge is the implementation of predictive algorithms in medicine. From the abundance of prediction models developed for health care purposes, very few are actually implemented in clinical practice.(34) This may be due several reasons, such as uncertainties on how to specifically use resulting predictions in care, the willingness of patients and physicians to trust these models and the scarcity of studies providing evidence for the effectiveness of using such prediction models in clinical practice. In addition, technical requirements may be challenging, such as complying with the medical device regulations and the need for a user-friendly interface. The PATIO trial addresses all these challenges, and may thus facilitate the safe and effective implementation of predictive algorithms in clinical practice.

We have previously demonstrated the potential of the flare prediction model to reduce the number of flares during bDMARD tapering a simulation study.(11) In the current randomized controlled trial we will assess whether these results are maintained when actually using the flare prediction model in a bDMARD tapering strategy in clinical practice. In a future study we will also address the views of both RA patients and rheumatologists on the implementation of predictive algorithms in clinical practice.

References

1. Nam JL, Takase-Minegishi K, Ramiro S, Chatzidionysiou K, Smolen JS, van der Heijde D, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2016 update of the EULAR recommendations for the management of rheumatoid arthritis. Annals of the rheumatic diseases. 2017 Jun;76(6):1113–36.
2. Aga A-B, Lie E, Uhlig T, Olsen IC, Wierød A, Kalstad S, et al. Time trends in disease activity, response and remission rates in rheumatoid arthritis during the past decade: results from the NOR-DMARD study 2000-2010. Annals of the rheumatic diseases. 2015 Feb;74(2):381–8.
3. Joensuu JT, Huoponen S, Aaltonen KJ, Konttinen YT, Nordström D, Blom M. The cost-effectiveness of biologics for the treatment of rheumatoid arthritis: a systematic review. PloS one. 2015;10(3):e0119683.
4. Bittner B, Richter W, Schmidt J. Subcutaneous Administration of Biotherapeutics: An Overview of Current Challenges and Opportunities. BioDrugs. 2018;32(5):425–40.
5. Ramiro S, Sepriano A, Chatzidionysiou K, Nam JL, Smolen JS, Van Der Heijde D, et al. Safety of synthetic and biological DMARDs: A systematic literature review informing the 2016 update of the EULAR recommendations for management of rheumatoid arthritis. Annals of the Rheumatic Diseases. 2017;76(6):1093–101.
6. Fautrel B, Pham T, Alfaiate T, Gandjbakhch F, Foltz V, Morel J, et al. Step-down strategy of spacing TNF-blocker injections for established rheumatoid arthritis in remission: Results of the multicentre non-inferiority randomised open-label controlled trial (STRASS: Spacing of TNF-blocker injections in Rheumatoid ArthritiS Study). Annals of the Rheumatic Diseases. 2016;75(1):59-67.
7. van Herwaarden N, van Maas A der, Minten MJM, van den Hoogen FHJ, Kievit W, van Vollenhoven RF, et al. Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis:Open label, randomised controlled, non-inferiority trial. BMJ. 2015; doi: 10.1136/bmj.h1389.
8. Den Broeder N, Bouman CAM, Kievit W, Van Herwaarden N, Van Den Hoogen FHJ, Van Vollenhoven RF, et al. Three-year cost-effectiveness analysis of the DRESS study: Protocolised tapering is key. Annals of the Rheumatic Diseases. 2019;78:141–2.
9. van Mulligen E, Weel AEAM, Kuijper TM, Hazes JMW, van der Helm- van Mil AHM, de Jong PHP. The impact of a disease flare during tapering of DMARDs on the lives of rheumatoid arthritis patients. Seminars in Arthritis and Rheumatism. 2020;50:423–31.
10. Welsing P, Broeder A den, Tekstra J, Goes M van der, Lafeber F, Jacobs JWG, et al. SAT0116 DYNAMIC PREDICTION OF FLARES IN RHEUMATOID ARTHRITIS USING JOINT MODELLING AND MACHINE LEARNING: SIMULATION OF CLINICAL IMPACT WHEN USED AS DECISION AID IN A DISEASE ACTIVITY GUIDED DOSE REDUCTION STRATEGY. Annals of the Rheumatic Diseases. 2019;78:1125–6.
11. Leeuw, MS, Tekstra J, Medina OJ, Laar JM, Messelink M., Lafeber F, den Broeder, A.A. WPMJ. Currently under submission: Using real-world data to dynamically predict flares during tapering of biological DMARDs in rheumatoid arthritis: development, validation and potential impact of prediction aided decisions. Arthritis Research and Therapy. 2022;24(1):74.
12. Fransen J, Stucki G, van Riel PLCM. Rheumatoid arthritis measures: Disease Activity Score (DAS), Disease Activity Score-28 (DAS28), Rapid Assessment of Disease Activity in Rheumatology (RADAR), and Rheumatoid Arthritis Disease Activity Index (RADAI). Arthritis Care & Research. 2003;49(S5):S214–24.
13. Fleischmann RM, Van Der Heijde D, Gardiner P V., Szumski A, Marshall L, Bananis E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open. 2017;3(1):e000382.
14. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Annals of the Rheumatic Diseases. 2015;74(6):1132–7.
15. Coury F, Rossat A, Tebib A, Letroublon MC, Gagnard A, Fantino B, et al. Rheumatoid arthritis and fibromyalgia: A frequent unrelated association complicating disease management. Journal of Rheumatology. 2009;36(1):58–62.
16. Farmacotherapeutisch Kompas. www.farmacotherapeutischkompas.nl. 2021. Accessed 15 July 2021.
17. Verhoef LM, Bos DPG, van den Ende CHM, van den Hoogen FHJ, Fautrel B, Hulscher ME, et al. Cost-effectiveness of five different anti-tumour necrosis factor tapering strategies in rheumatoid arthritis: a modelling study. Scandinavian Journal of Rheumatology. 2019;48(6):439–47.
18. van der Maas A, Kievit W, van den Bemt BJF, van den Hoogen FHJ, van Riel PL, den Broeder AA. Down-titration and discontinuation of infliximab in rheumatoid arthritis patients with stable low disease activity and stable treatment: An observational cohort study. Annals of the Rheumatic Diseases. 2012;71(11):1849–54.
19. Van Der Maas A, Lie E, Christensen R, Choy E, De Man YA, Van Riel P, et al. Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria: An OMERACT cohort validation

study. Annals of the Rheumatic Diseases. 2013;72:1800–5.
20. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: A review of its history, issues, progress, and documentation. Journal of Rheumatology. 2003;30(1):167–78.
21. Hernández-Alava M, Pudney S. Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. Journal of health economics. 2017 Sep;55:139–52.
22. Salaffi F, Carotti M, Gutierrez M, di Carlo M, de Angelis R. Patient Acceptable Symptom State in Self-Report Questionnaires and Composite Clinical Disease Index for Assessing Rheumatoid Arthritis Activity: Identification of Cut-Off Points for Routine Care. BioMed Research International. 2015;930756.
23. Rensis Likert. A technique for the measurement of attitudes. Archives of Psychology. 1932;22(140):1–55.
24. Bartlett SJ, Barbic SP, Bykerk VP, Choy EH, Alten R, Christensen R, et al. Content and construct validity, reliability, and responsiveness of the rheumatoid arthritis flare questionnaire: OMERACT 2016 workshop report. Journal of Rheumatology. 2017;44(10):1536–43.
25. Hawthorne G, Sansoni J, Hayes L, Marosszeky N, Sansoni E. Measuring patient satisfaction with health care treatment using the Short Assessment of Patient Satisfaction measure delivered superior and robust satisfaction estimates. Journal of Clinical Epidemiology. 2014;67(5):527–37.
26. Stach CM, Sloan VS, Woodworth TG, Kilgallen B, Furst DE. Rheumatology Common Toxicity Criteria (RCTC): An Update Reflecting Real-World Use. Drug Safety. 2019;42(12):1499–506.
27. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987;40(5):373–83.
28. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythr. Annals of the Rheumatic Diseases. 2009;68(6):954–60.
29. Kushner I. C-reactive protein in rheumatology. Vol. 34, Arthritis & Rheumatism. 1991. p. 1065–8.
30. Zorginstituutnederland. https://www.zorginstituutnederland.nl/. Accessed 15 Jul 2021.
31. NFU Richtlijn Kwaliteitsborging Mensgebonden Onderzoek, update december 2020. https://www.nfu.nl/sites/default/files/2021-01/21.00023_Richtlijn_Kwaliteitsborging_Mensgebonden_Onderzoek_2020.pdf. Accessed 1 Aug 2021.
32. Nederlands Trial Register. https://www.trialregister.nl/. Accessed 10 Oct 2021.
33. Verhoef LM, Selten EMH, Vriezekolk JE, de Jong AJL, van den Hoogen FHJ, den Broeder AA, et al. The patient perspective on biologic DMARD dose reduction in rheumatoid arthritis: A mixed methods study. Rheumatology (United Kingdom). 2018;57(11):1947–55.
34. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC medicine. 2019;17(1):195.

Appendix A

Drug	Standard dose definition
Infliximab	IV: 3mg/kg / 8 weeks
	s.c.: 120mg / 2 weeks
Etanercept	s.c.: 50mg / week
Golimumab	s.c.: 50mg / 4 weeks
Certolizumab	s.c.: 200mg / 2 weeks
Adalimumab	s.c.: 40mg / 2 weeks
Sarilumab	s.c.: 200mg / 2 weeks
Tocilizumab	IV: 8mg / kg per 4 weeks, with a maximum of 800mg
	s.c.: 162 mg / week
Abatacept	IV: < 60 kg: 500 mg / 4 weeks 60–100 kg: 750 mg / 4 weeks > 100 kg: 1000 mg / 4 weeks
	s.c.: 125 mg / week

IV = intravenous
s.c. = subcutaneous

Chapter 9

Summary & general discussion

Summary

The focus of this thesis was the development and implementation of clinical decision support systems (CDSSs) in rheumatoid arthritis (RA) and primary antibody deficiencies (PAD). The main findings are summarized below.

Part I: Patient perspective

Chapter 2 evaluates the rheumatoid arthritis (RA) patients' perspective regarding the use of prediction models in clinical decision-making using semi-structured interviews. Six themes were identified related to the need for information, trust in model-supported treatment, the role of the model in medical decision-making, hopes to personally benefit from model-supported treatment, the willingness to invest time and effort to provide model input, and the relationship with the caregiver. The results of this study can be applied during the development and implementation of model-based CDSSs in order to enhance patient acceptability. Examples include providing patient information about possible treatment consequences rather than technical details, considering patient-preferred model variables such as stress and diet, and addressing concerns about commercial party involvement.

Part II: Diagnostic decision support

One of the potential applications of CDSSs is early disease identification. In **chapter 3** we explored different methods for the identification and prediction of difficult-to-treat (D2T) RA patients in structured and unstructured routine care data (e.g. medication prescriptions and clinical notes). D2T RA refers to a subgroup of patients that remain symptomatic despite treatment with several biological- or targeted synthetic disease modifying antirheumatic drugs (b/tsDMARDs) according to EULAR guidelines.[1] Using a rule-based approach with the D2T RA criteria, we were able to identify 123 potential D2T RA patients. Different machine-learning techniques were applied to develop a D2T RA identification model (AUC-ROC 0.88) and to predict the future development of D2T RA already early in the RA disease course (i.e. before starting the first b/tsDMARD, AUC-ROC 0.73). In a research setting, correct identification and prediction can enable retrospective analysis of the development of RA into D2T RA and the progression of the D2T RA state over time. In a clinical setting, it can serve as a (early) warning system that a patient is, or may become, D2T. Rather than cycling through yet another b/tsDMARD, the physician can then take a step back and consider e.g. misdiagnosis of disease (activity), addressing possible non-adherence and optimizing (non)pharmacological treatment.

Chapters 4 and **5** describe the development and clinical validation of a CDSS for the early detection of primary antibody deficiencies (PAD). A screening algorithm was developed based on literature, clinical expertise, and primary care metadata of PAD patients and control groups. The weights of individual components were based on analysis of aggregated data from routine care of PAD patients and controls. For example, the mean number of antibiotic prescriptions per year was compared between PAD-patients and patients with chronic obstructive pulmonary disease (COPD) or asthma. The final algorithm included antibiotic prescriptions, diagnostic codes for respiratory tract and other infections, gastro-intestinal complaints, auto-immune symptoms, malignancies, lymphoproliferative symptoms, laboratory values and the number of visits to the general practitioner. In the clinical validation study, the algorithm was applied to a dataset of 61,172 electronic health records (EHRs) from 13 primary care practices, of which 104 high-risk patients were included for a laboratory assessment of immunoglobulins. Based on the laboratory results and algorithm score, a referral advice was provided to the general practitioner. Sixteen patients were referred for a suspected PAD, of whom 10 had a PAD-diagnosis. With the screening algorithm, we were therefore able to select a high-risk population with a PAD-prevalence of 1:10, versus 1:1,700-1:25,000 in the general population. Using principal component analysis (PCA) and ridge regression, we developed an additional model to improve PAD-identification within the high-risk patients with an AUC-ROC of 0.80. By adding this model as a subsequent step, the population selected for laboratory evaluation could be further reduced. This proposed screening approach may aid in the prevention of morbidity and mortality by reducing the diagnostic delay for PAD patients.

Part 3: Treatment decision support

The third part of this thesis focused on decision support during treatment of RA with DMARDs. In this context, it is important to know the optimal target within a treat-to-target (T2T) strategy, as this can guide the decision to intensify or taper DMARDs. In **chapter 6** we performed a systematic review and meta-regression analysis to compare the effect of different targets on clinical outcomes. Targets were based on the simplified and clinical disease activity indices (SDAI/CDAI) or the disease activity score (DAS) with either remission or low disease activity (LDA) as cut-off points. Targeting SDAI/CDAI-LDA, and to a lesser extent DAS-remission, was superior to targeting DAS-LDA regarding several clinical outcomes, although no difference in structural/functional outcomes could be identified. A clinical trial comparing these targets directly will be necessary, due to the risk of residual confounding and lack of data on (over)treatment and safety.

Chapters 7 and **8** cover the development and validation of a model-based CDSS to dynamically predict flares during the stepwise tapering of bDMARDs. The aim is to reduce the number of flares during tapering, whilst still reducing the dose as much as possible. Previous studies have shown that baseline biomarkers alone are not very predictive of successful dose reduction.[2] We therefore combined baseline variables with follow-up data in a joint latent class model to make dynamic predictions over time. This model was developed using routine care data of RA patients to predict the risk of a flare occurring in the next 3 months, with an AUC-ROC of 0.76 in cross-validation and 0.68 in external validation. A simulation of clinical impact was performed using the external data by comparing a stepwise tapering strategy supported by the flare prediction model to the current best-available tapering strategy (disease activity guided dose optimization, DGDO). In this simulation, the addition of the flare prediction model was superior to DGDO alone when considering the ratio between the number of flares and the amount of bDMARD dose reduction. **Chapter 8** describes the protocol for a pragmatic test-treatment randomized controlled trial (RCT) to assess the effect of implementing the flare prediction model as a CDSS during stepwise bDMARD tapering. In the control group, bDMARDs are tapered according to DGDO. In the intervention group, a CDSS provides a treatment advice based on the flare prediction model in addition to DGDO. Stepwise bDMARD tapering is recommended until there is a high predicted risk of a flare, until an actual flare occurs, or until the bDMARD is stopped. If the bDMARD has been stopped, but a high predicted risk of a flare occurs, it is recommended to restart the bDMARD. The primary outcome is the number of flares per patient over the 18-month follow-up, secondary outcomes include bDMARD dose reduction, adverse events and quality of life. To the best of our knowledge, this is the first RCT to study a prediction model-based CDSS in rheumatology.

General discussion

In the past decades, there has been a steady increase in the number of developed and implemented CDSSs.[3] Considering the recent advances in predictive analytics, I expect this trend to continue. Personally, this was an important reason for me to start this PhD. Healthcare systems worldwide face pressing challenges, such as unaffordable costs, high workloads and an information overload. CDSSs may provide a piece of the puzzle for these challenges by integrating information, providing data-driven personalized treatment advice and increasing efficiency.

Currently the majority of implemented CDSSs are rule-based, such as medication interaction warnings.[5] But the hype around artificial intelligence (AI) and machine learning has also fueled the hopes for prediction model based CDSSs to improve healthcare. This enthusiasm for big data and derived (AI) prediction model applications

is shared by the European Alliance of Associations for Rheumatology (EULAR), formulated in an overarching principle: '*Big data provides unprecedented opportunities to deliver transformative discoveries in RMD* [rheumatic and musculoskeletal disorders] *research and practice'*.[6] The implementation of AI-based CDSSs has already begun: in the USA, there are currently 171 FDA-approved medical devices based on AI methods[7]. These mostly concern image-recognition software for radiology. Within the field of rheumatology and clinical immunology, there is one FDA approved AI-based CDSS, although the evidence for its effectiveness is sparse.[8] This tool suggests the most likely diagnosis based on antinuclear antibody (ANA) screening. There are many other published examples of rheumatology and clinical immunology prediction models aiming to improve diagnosis and treatment, although these are still in a developmental stage.[9-11]

Despite the promising potential of CDSSs, they also pose several risks and challenges. These will be discussed below, together with possible solutions (Table 1). A distinction is made between CDSS development and implementation. Concerning development, we will focus mainly on prediction model based CDSSs, as rule-based CDSSs are commonly developed using pre-existing literature and clinical guidelines. Interestingly, the solutions for problems during prediction model development can often come from yet other prediction models. The challenges of implementation apply to both rule-based on model-based CDSSs.

Challenges in development

Training- and validation datasets

Regarding training data, the mantra 'Garbage in, garbage out' applies. If a model is based on a small, unrepresentative dataset with many missing values, it will most likely not be useful in clinical practice. The lack of large high-quality datasets also hampers external validation, an essential step of model development. I will discuss possible solutions for optimizing the availability and quality of datasets for training and validation.

Unstructured data

Currently, an estimated 80% of clinical data is in an unstructured format such as clinical notes, making it difficult to add this information as a model variable.[21] As an example, in chapter 7 we were unable to add certain important factors to the prediction model, like smoking status and radiographic outcomes, as this information was only captured in unstructured data. In chapter 3, we were unable to derive the third criterium of the D2T RA definition "the management is perceived as problematic by the rheumatologist

and/or the patient" from the available data. Unstructured data also impedes the integration of different datasets into one large dataset.

Table 1. Challenges for the development and implementation of clinical decision support systems and proposed solutions

Development	
Challenges	*Possible solutions*
Unstructured data	→Short term: only using structured data →Promote registration structured data, e.g. HPO terms[12], EU Rare Disease Platform data elements[13], EULAR RA core dataset[14] →Text mining and Natural Language Processing →Integration of new digital health technologies in EHR that register features of interest, e.g. ePROs or wearables
Missing data	→Optimizing use of (un)structured data, see above →Imputation →Using missing data as an asset
Small datasets	→Registries: All of Us[15] research program, RISE[16] registry →Improving interoperability: HL-7[17], EHDS[18] →Federated learning →Transfer learning →Data augmentation
Unrepresentative training data	→Registering variables such as race/ethnicity, language, social determinants to allow for subgroup analyses to assess bias →PROBAST[19] tool to assess bias →External validation
Incorrect labeling	→Identification models
Implementation	
Challenges	*Possible Solutions*
Unfeasible model variables	→Use of routine care data only →For new markers: careful consideration of additional predictive value, cost-effectiveness and impact on clinical workflow
Real-world benefit	→Assess clinical impact (e.g. using simulation), rather than just performance metrics →Randomized controlled trials →Cost-effectiveness analysis
Acceptability by patients and clinicians	→Involvement of patients in model and CDSS development →User-friendly application with a clear treatment advice →Integration of CDSS in clinical workflow →Explainability for machine learning methods using e.g. SHAP[20]

References and abbreviations: All of us research program[15], EHR: electronic health record, ePRO: electronic patient reported outcomes, EU (European Union) Rare Disease Platform data elements[13], EULAR RA (rheumatoid arthritis) core dataset[14], HL-7 (Health Level 7)[17], HPO (human phenotype ontology) terms[12], PROBAST (prediction model study risk of bias assessment tool)[19], RISE (rheumatology informatics system for effectiveness) registry[16], SHAP (Shapley Additive exPlanations)[20]

A short-term simple solution that we have applied in chapters 4 and 7 is to develop models only based on structured data. This does however not use the available data to its fullest potential. Another relatively straightforward solution is to promote the

registration of data in a structured manner. For example, the European Platform on Rare Disease Registration has released a set of 16 data elements and their exact coding format (e.g. yes/no) that should be registered by all rare disease databases across Europe.[13] Similarly, the EULAR provides recommendations for an RA core dataset.[14] Another example is the Human Phenotype Ontology (HPO), a standardized vocabulary of phenotypic abnormalities associated with 7000+ diseases.[12] However, a risk of this type of solution is that it increases the administrative pressure on clinicians, which is already high.

As an alternative, techniques such as text mining and natural language processing (NLP) are in development to transform unstructured data into structured data. In chapter 3, we used these techniques to search for signs of active disease in clinical notes. Other studies have used NLP techniques to estimate CDAI scores and the SLE disease activity index.[31,32] Looking ahead, it is also possible that the amount of structured data will increase by integrating the results of wearables (e.g. fitness trackers) and electronic patient reported outcomes (ePROs, e.g. pain on a visual analog scales (VAS)) into the EHR. As we have shown in chapter 2 patients are willing to register and share such healthcare related data for the purpose of prediction models, which is in line with other studies.[24,25] This is already applied in the REMORA project in the UK, where RA patients track daily symptoms with a smartphone app, which is automatically integrated into their EHR.[26] There is however conflicting evidence regarding the long-term patient adherence to such apps.[27-29]

Missing data

Another common problem with routine care data is the amount of missing data. To cope with this limitation, various imputation techniques exist to estimate the missing values, based on either machine learning or more conventional statistical methods.[30] For example, in chapter 3 we developed a model to estimate missing DAS28-values based on other variables. In addition, missing structured values can sometimes be extracted from the unstructured data, as discussed above. As the structured registration of medication prescriptions in chapter 3 was incomplete, we complemented this data with information from clinical notes (i.e. unstructured data). Interestingly, missing EHR data can also be used as an asset, as the missingness is often not random and can reveal patterns of interest. For example, a sudden absence of further blood tests in a critically ill patient can reflect a change in the goal of treatment.[10] The missingness can however be very dependent on the context (e.g. local registration habits), and therefore such a predictor should be interpreted with caution.

Small datasets

The lack of large datasets impedes the development and external validation of prediction models. Especially in rare rheumatic and immunological disorders, such as the primary antibody deficiencies (PADs) discussed in chapters 4 and 5 of this thesis, it can be difficult to obtain a decent sample size. The availability of multiple large datasets would allow to not only perform external validation, but also to finetune the model based on this second dataset, as we have done in chapter 5. Such an approach requires a third dataset for external validation of the revised model, which is currently very rare.[33] In order to acquire larger datasets, there are several registries that pool data from different institutes. For example, the All of Us research program aims to combine data from e.g. the EHR, wearables, and health questionnaires of >1 million US citizens, including ethnic minorities.[15] Similarly, the Rheumatology Informatics System for Effectiveness (RISE) registry combines the EHR-data of 3.1 million US rheumatology patients.[16] The European Union (EU) also has an initiative to improve interoperability between health data systems: the European Health Data Space.[18] This is currently however still in development. A worldwide initiative to improve interoperability between health data systems is the Health Level 7 (HL-7) project, which is supported by members from >50 countries.[17] Increasing the amount of structured data, as described above, would also facilitate interoperability. In chapters 4 and 5, we collaborated with the Julius Practitioner Network in order to pool the EHR-data of several general practitioners in the Netherlands, which use different registration systems. We used this data to develop an algorithm specifically for primary care, but there are also PAD algorithms in development for secondary and tertiary care.[34] Integrating the EHR-data from these different health systems could potentially aid the development of one overarching model with improved performance.

It is however often not possible to pool routine care data from different sources into large datasets due to technical and privacy considerations. Alternatively, a technique called federated learning can be utilized, as done in the Personal Health Train project.[35,36] Instead of bringing all the data together to develop a model in a central place, the model ('train') is brought to the different data sites ('stations'). These individual sites, such as hospitals, general practitioners, or even individual patients, provide FAIR (Findable, Accessible, Interoperable, Reusable) data, Thus, multiple sites can contribute to model development by sharing (non-identifiable) analysis results rather than raw data. In another alternative called transfer learning, a previously trained model is re-purposed for a second related task. A study by Gao and Cui[37] describe the training of a model on ethnic majority group data, after which they transferred the knowledge learnt to assist the development of models for different ethnic minorities in small datasets. If data pooling, federated- or transfer learning are not possible, there are some techniques to cope with a small dataset. In data

augmentation the training set is artificially increased by creating modified copies of existing data.[38] As an example, generative adversarial networks use one network to generate synthetic data, and a second network to discriminate real from synthetic data. As it is trained, this discrimination becomes more and more difficult.[39] However, such synthetic data remains an estimation of real-world data, and its generalizability highly depends on the source data.

Unrepresentative training data

CDSSs may have the potential to improve equality by giving a 'neutral' advice and by sharing the integrated knowledge across digitalized health centers. However, there is also a substantial risk of introducing bias and discrimination if the training data of a model-based CDSS does not represent the target population. This is not unique to CDSSs, as it is already an established issue with the data from RCTs.[40-42] In this thesis we have focused on EHR data, which represents a greater proportion of the population compared to RCTs with strict in- and exclusion criteria. However, models trained on EHR data can certainly still introduce bias. For example, patients of lower socioeconomic status may already receive suboptimal care, which could be amplified by training a model on their EHR-data.[43] Bias can be detected in several ways, for example by subgroup analyses. This does however require data on e.g. ethnicity, race, language and social determinants, which is often not registered or only in unstructured notes.[44] The PROBAST tool (Prediction model Risk Of Bias Assessment Tool) offers a systematic way to assess both the risk of bias and concerns regarding applicability of prediction models.[19,45] Ultimately, external validation is necessary to determine model performance in a population that differs from the training data.

Incorrect labeling

When developing a model, class labels for cases and controls (e.g. 'PAD-diagnosis vs. no presence of PAD' or 'flare vs. no flare') should be clearly defined. The control group should be selected based on the intended use of the model, e.g. in primary care, the emergency department or outpatient rheumatology clinic. But also the selection of cases can be challenging. A problem that we encountered in chapter 4 was a lack of a golden standard for the selection of the PAD-cases. There is no specific diagnostic code for PAD, and even if this had been available, diagnostic- and billing codes can be unreliable.[46] The diagnosis as mentioned in the clinical notes of the physician can be an alternative, but as this is unstructured data it can be difficult to extract. Eventually we combined a strategy using both diagnostic codes and manually extracted clinical diagnoses, but this is not ideal. Identification models with the aim to select cases could be helpful to improve class labeling. These are different from identification models that aim for the early detection of a disease. When developing a model to support early diagnosis, only data from before the moment of (considering a) diagnosis are useful

predictors. In contrast, an identification model that aims to detect cases (i.e. patients that have already been diagnosed) can also use predictors from after the moment of diagnoses, such as billing codes or specific laboratory- or genetic tests. In chapter 3, we applied the D2T RA criteria to the available data in order to identify D2T RA cases. Other published examples include models for the identification of SLE and axial spondylarthritis.[47,48] Such identification models may help to clearly define cases, after which models for e.g. early diagnosis, disease progression or treatment response can be developed.

Challenges in implementation

Unfeasible model variables

It is important to consider if the variables included in a prediction model are feasible to collect in clinical practice. To illustrate, a review on prediction models for knee osteoarthritis (OA) found that half of the models included MRI-based measures. MRI is however not generally used for knee OA diagnosis, and therefore both external validation and implementation of these models will be difficult.[49] When considering a predictor variable that is not commonly available in routine care (e.g. MRI), first its additional predictive value in comparison to routinely available predictors should be carefully assessed. If its predictive value has been established, the impact on the clinical workflow should be assessed. If an additional scan or laboratory test hampers the workflow excessively, it will not be performed, and thus the predictor will be of no use. Lastly, the cost-effectiveness of an additional test should be considered before implementation. In order to limit the number of patients undergoing tests that are not usually performed in routine care, a 2-step approach can be considered. In the first step, a high-risk population is selected based on predictors that are readily available from routine care. In a second step, the additional test can be performed only for the subset of high-risk patients.

Real-world benefit

The strong focus on the performance metrics of prediction models risks overshadowing the actual impact for the patient, clinician and healthcare system. A review of diagnostic laboratory and imaging tests (non-CDSS) showed that many RCTs failed to show a significant improvement in patient outcomes.[50] Diagnostic accuracy alone does not guarantee a patient benefit. The effect of a CDSS on healthcare outcomes can initially be estimated based on simulation, but should thereafter be studied in a test-treatment RCT (if possible). In addition, a cost-effectiveness analysis should be performed. The overall assessment of CDSS performance, cost-effectiveness, risks, and the effects of a CDSS on the clinical workflow, is often referred to as a health technology assessment (HTA).

The diagnostic CDSS described in chapters 4 and 5 was in part inspired by work of the Jeffrey Modell Foundation[52], who have shown that early diagnosis of immunodeficiencies improves outcomes such as morbidity, mortality, and healthcare costs (taking the cost of treatment in the USA into account). The limited cost-analysis in chapter 5 shows that the costs of the proposed screening approach per patient is comparable to that of other (population-wide) screenings in the Netherlands (e.g. breast- and cervical cancer). Although the cost-effectiveness of these existing screening programs for asymptomatic individuals can (and should) be debated, many societies are prepared to bear these costs.[53] Importantly, the screening approach described in chapter 5 focuses on symptomatic patients with a wide range of complaints, rather than asymptomatic individuals. In addition, the screening intervention consists of a single blood withdrawal, rather than more invasive procedures such as colonoscopy. Of the patients included for the intervention, 1 in 10 received a PAD diagnosis. Although these results are promising, an ongoing evaluation of patient benefit and a full cost-effectiveness analysis should be performed.

During the development of the flare risk prediction model described in chapter 7, clinical impact was assessed by simulating the effect on the number of flares and amount of dose reduction, rather than only reporting metrics such as the AUC. The next step in assessing the clinical impact of a CDSS is an RCT, such as the one we describe in chapter 8. To the best of our knowledge, this is the first RCT to assess the effect of a prediction model-based CDSS in rheumatology. Setting up this RCT was not an easy task, partly due to the necessary administration required by the recently updated Medical Device Regulation (MDR).[54] The current sparsity of CDSS RCTs is likely also influenced by difficulties to obtain sufficient funding. In addition, it has been hypothesized that researchers are more interested in the purely technical innovation, rather than the administrative obligations that are necessary for clinical applications ('hit and run' research).[55] This tide may however turn in response to recent successes in AI implementation. Collaboration between universities and industry may be necessary to obtain sufficient funding and to utilize the available experience with software validation and implementation.

Acceptability by patients and clinicians

For successful implementation, CDSSs should not only have good performance metrics and improve healthcare outcomes, but both patients and clinicians must also be willing to use them. Acceptability may be influenced by the envisioned benefit and by the user-friendliness of the application. Healthcare is catching up with industry in this regard, as nowadays the early involvement of end users is often a requirement to receive (EU) funding for CDSS development. For the CDSS described in chapter 8, we collaborated with a tech company to develop a user-friendly dashboard, although the challenging

integration with the EHR has yet to follow. Clinicians may be apprehensive about the use of CDSSs due to previous negative experiences with digitalization (e.g. EHR implementation), thus a smooth integration with the current workflow is essential.[56] Furthermore, the CDSS treatment advice must be clearly formulated, as the CDSS is unlikely to have a clinical impact if the physician is not sure how to interpret the advice.

Regarding patients, we identified several factors in chapter 2 that would enhance RA patient acceptability of prediction models in clinical decision-making. For example, patients expressed their preference for models that included variables such as stress, diet, and the weather, which are currently not commonly explored as predictors. A publication with the illustrative title "Nothing about us without us" also describes a method for involving rheumatology patients in machine learning applications.[57] For both patients and physicians, explainability of 'black-box' machine learning models could improve acceptance, for which machine learning solutions such as 'Shapley additive explanations' (SHAP, as used in chapter 3) are in development.[20] Such perturbation methods adjust the input data and assess how this effects model output. It can however also be debated which level of explainability is necessary for patients and clinicians, as we currently accept many technologies in our professional (e.g. MRI) and day-to-day lives (e.g. Google) which most users do not fully understand.

Future prospects

Fortunately, all of the projects discussed in this thesis will be followed-up. In a consecutive study of chapter 2, a quantitative analysis of the qualitative results will be performed using a best-worst scaling. This will further elucidate which factors are most important for patients when considering model-based CDSSs. Regarding identification of D2T RA (chapter 3), the STRATA-FIT consortium will further develop computational models and integrate them in a CDSS in a EU-setting using federated learning.[58] An early HTA will be performed for the screening tool for the early detection of PAD (chapters 4 and 5), and opportunities to further validate this tool on a larger scale in primary care are being explored. Furthermore, an additional model is in development for the early detection of PAD in a hospital setting, which also integrates unstructured data. As for the treat-to-target (T2T) meta-regression analysis in chapter 6, the TETRA RCT from the Sint Maartenskliniek will provide further insight in the optimal target by comparing targets head-to-head.[59] Lastly, the final results from the PATIO RCT (chapters 7 and 8) and the accompanying cost-effectiveness analysis are expected within the next 1-2 years.

In addition to these ongoing projects, there are other interesting opportunities for further research. In chapter 2, RA patients stated that they would have more confidence

in prediction models that include stress, diet, and the weather as variables. I would therefore be interested to explore their additional predictive value in a RA flare prediction model. Predictions of RA disease activity could also potentially be useful in a T2T-strategy to select a personalized treatment target, i.e. a 'treat-to-predicted-target' (T2PT) approach. From the PAD screening study, I would be interested in a qualitative analysis of patients' experience with the invitation for additional testing based on their EHR data. This should include true positives, false positives, and patients who declined participation. Lastly, if the results from the PATIO-trial are favorable, it would be interesting to perform an external validation of this model in databases from countries with comparable healthcare systems to assess generalizability.

In the upcoming years, I expect a gradual increase in CDSS implementation. This will likely initially concern models where wrong predictions have very limited negative consequences compared to usual care. This is the case for the PATIO flare prediction model, where an unpredicted flare results in tapering according to DGDO (an accepted tapering strategy), and an incorrectly predicted flare results in maintaining the bDMARD at the same dose (also common in usual care). The impact of an incorrect prediction will also affect the necessary extent of safety testing before and during implementation. If indeed more and more CDSSs are implemented in clinical practice, physicians will be faced with the challenge of integrating these support systems into their clinical decision-making process. Such a change warrants an open, though critical, attitude of treating physicians towards these technological advances.

Conclusions

This thesis focused on the development and implementation of clinical decision support for both diagnosis and treatment in rheumatoid arthritis and primary antibody deficiencies. CDSSs may have the potential to support early diagnosis and optimize treatment decisions on a large scale. There are however numerous challenges in the development and implementation, for which we have proposed possible solutions. Ultimately, the focus should not be on predictive accuracy, but rather on a real-world benefit for both the patient and the healthcare system as a whole.

References

1. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82(1):3-18.
2. Tweehuysen L, van den Ende CH, Beeren FM, Been EM, van den Hoogen FH, den Broeder AA. Little Evidence for Usefulness of Biomarkers for Predicting Successful Dose Reduction or Discontinuation of a Biologic Agent in Rheumatoid Arthritis: A Systematic Review. Arthritis Rheumatol 2017;69(2):301-8.
3. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform 2016;Suppl 1(Suppl 1):S103-16.
4. Cerrato P, Halamka J. Reinventing clinical decision support: data analytics, artificial intelligence, and diagnostic reasoning: Taylor & Francis; 2020.
5. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17.
6. Gossec L, Kedra J, Servy H, Pandit A, Stones S, Berenbaum F, et al. EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann Rheum Dis 2020;79(1):69-76.
7. Administration FaD. Artifical Intelligence and Machine Learning in Software as a Medical Device;Accessed on 4-12-2023. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#news.
8. Administration FaD. Diagnostic Software, K-Nearest Neighbor Algorithm, Autoimmune Disease;Accessed on 4-12-2023. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K043341.
9. Madrid-García A, Merino-Barbancho B, Rodríguez-González A, Fernández-Gutiérrez B, Rodríguez-Rodríguez L, Menasalvas-Ruiz E. Understanding the role and adoption of artificial intelligence techniques in rheumatology research: An in-depth review of the literature. Semin Arthritis Rheum 2023;61:152213.
10. Nelson AE, Arbeeva L. Narrative Review of Machine Learning in Rheumatic and Musculoskeletal Diseases for Clinicians and Researchers: Biases, Goals, and Future Directions. J Rheumatol 2022;49(11):1191-200.
11. Rider NL, Srinivasan R, Khoury P. Artificial intelligence and the hunt for immunological disorders. Curr Opin Allergy Clin Immunol 2020;20(6):565-73.
12. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine JP, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019;47(D1):D1018-d27.
13. Set of Common Data Elements EU Rare Diseases Platform: European Comission;Accessed on 04-12-2023. Available from: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en.
14. Radner H, Chatzidionysiou K, Nikiphorou E, Gossec L, Hyrich KL, Zabalan C, et al. 2017 EULAR recommendations for a core data set to support observational research and clinical care in rheumatoid arthritis. Ann Rheum Dis 2018;77(4):476-9.
15. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, et al. The All of Us Research Program: Data quality, utility, and diversity. Patterns (N Y) 2022;3(8):100570.
16. Yazdany J, Bansback N, Clowse M, Collier D, Law K, Liao KP, et al. Rheumatology Informatics System for Effectiveness: A National Informatics-Enabled Registry for Quality Improvement. Arthritis Care Res (Hoboken) 2016;68(12):1866-73.
17. Health Level 7;Accessed on 2023Dec 4. Available from: https://www.hl7.org/about/index.cfm?ref=common.
18. European Health Data Space: European Union;Accessed on 2023Dec 4. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en.
19. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51-8.
20. Singh A, Jothi Balaji J, Rasheed MA, Jayakumar V, Raman R, Lakshminarayanan V. Evaluation of Explainable Deep Learning Methods for Ophthalmic Diagnosis. Clin Ophthalmol 2021;15:2573-81.

21. Bergier H, Duron L, Sordet C, Kawka L, Schlencker A, Chasset F, Arnaud L. Digital health, big data and smart technologies for the care of patients with systemic autoimmune diseases: Where do we stand? Autoimmun Rev 2021;20(8):102864.
22. Brunekreef TE, Otten HG, van den Bosch SC, Hoefer IE, van Laar JM, Limper M, Haitjema S. Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients With Systemic Lupus Erythematosus. ACR Open Rheumatol 2021;3(2):65-71.
23. Atsumi T, Ando Y, Matsuda S, Tomizawa S, Tanaka R, Takagi N, Nakasone A. Prodromal signs and symptoms of serious infections with tocilizumab treatment for rheumatoid arthritis: Text mining of the Japanese postmarketing adverse event-reporting database. Mod Rheumatol 2018;28(3):435-43.
24. Navarro-Millán I, Zinski A, Shurbaji S, Johnson B, Fraenkel L, Willig J, et al. Perspectives of Rheumatoid Arthritis Patients on Electronic Communication and Patient-Reported Outcome Data Collection: A Qualitative Study. Arthritis Care Res (Hoboken) 2019;71(1):80-7.
25. Rudin RS, Fanta CH, Qureshi N, Duffy E, Edelen MO, Dalal AK, Bates DW. A Clinically Integrated mHealth App and Practice Model for Collecting Patient-Reported Outcomes between Visits for Asthma Patients: Implementation and Feasibility. Appl Clin Inform 2019;10(5):783-93.
26. Austin L, Sharp CA, van der Veer SN, Machin M, Humphreys J, Mellor P, et al. Providing 'the bigger picture': benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record: findings from the Remote Monitoring of Rheumatoid Arthritis (REMORA) study. Rheumatology 2020;59(2):367-78.
27. Wiegel J, Seppen BF, Nurmohamed MT, Bos WH, Ter Wee MM. Who stop telemonitoring disease activity and who adhere: a prospective cohort study of patients with inflammatory arthritis. BMC Rheumatol 2022;6(1):73.
28. Colls J, Lee YC, Xu C, Corrigan C, Lu F, Marquez-Grap G, et al. Patient adherence with a smartphone app for patient-reported outcomes in rheumatoid arthritis. Rheumatology (Oxford) 2021;60(1):108-12.
29. Seppen B, Wiegel J, Ter Wee MM, van Schaardenburg D, Roorda LD, Nurmohamed MT, et al. Smartphone-Assisted Patient-Initiated Care Versus Usual Care in Patients With Rheumatoid Arthritis and Low Disease Activity: A Randomized Controlled Trial. Arthritis Rheumatol 2022;74(11):1737-45.
30. Raja P, Thangavel K. Missing value imputation using unsupervised machine learning techniques. Soft Computing 2020;24(6):4361-92.
31. Spencer AK, Bandaria J, Leavy MB, Gliklich B, Su Z, Curhan G, Boussios C. Validation of a machine learning approach to estimate Clinical Disease Activity Index Scores for rheumatoid arthritis. RMD Open 2021;7(3).
32. Alves P, Bandaria J, Leavy MB, Gliklich B, Boussios C, Su Z, Curhan G. Validation of a machine learning approach to estimate Systemic Lupus Erythematosus Disease Activity Index score categories and application in a real-world dataset. RMD Open 2021;7(2).
33. Kingsmore KM, Puglisi CE, Grammer AC, Lipsky PE. An introduction to machine learning and analysis of its use in rheumatic diseases. Nat Rev Rheumatol 2021;17(12):710-30.
34. Rider NL, Cahill G, Motazedi T, Wei L, Kurian A, Noroski LM, et al. PI Prob: a risk prediction and clinical guidance system for evaluating patients with recurrent infections. PLoS One 2021;16(2):e0237285.
35. Deist TM, Dankers F, Ojha P, Scott Marshall M, Janssen T, Faivre-Finn C, et al. Distributed learning on 20 000+ lung cancer patients - The Personal Health Train. Radiother Oncol 2020;144:189-200.
36. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST) 2019;10(2):1-19.
37. Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nature Communications 2020;11(1):5131.
38. Wong SC, Gatt A, Stamatescu V, McDonnell MD, editors. Understanding data augmentation for classification: when to warp? 2016 international conference on digital image computing: techniques and applications (DICTA); 2016: IEEE.
39. Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks. arXiv preprint arXiv:171104340 2017.

40. Yip K, Navarro-Millán I. Racial, ethnic, and healthcare disparities in rheumatoid arthritis. Curr Opin Rheumatol 2021;33(2):117-21.
41. Faison WE, Harrell PG, Semel D. Disparities across Diverse Populations in the Health and Treatment of Patients with Osteoarthritis. Healthcare (Basel) 2021;9(11).
42. Singh JA. Racial and gender disparities among patients with gout. Curr Rheumatol Rep 2013;15(2):307.
43. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine 2018;178(11):1544-7.
44. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866-72.
45. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al., editors. Model cards for model reporting. Proceedings of the conference on fairness, accountability, and transparency; 2019.
46. Moores KG, Sathe NA. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data. Vaccine 2013;31 Suppl 10:K62-73.
47. Murray SG, Avati A, Schmajuk G, Yazdany J. Automated and flexible identification of complex disease: building a model for systemic lupus erythematosus using noisy labeling. J Am Med Inform Assoc 2019;26(1):61-5.
48. Zhao SS, Hong C, Cai T, Xu C, Huang J, Ermann J, et al. Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. Rheumatology (Oxford) 2020;59(5):1059-65.
49. Ramazanian T, Fu S, Sohn S, Taunton MJ, Kremers HM. Prediction Models for Knee Osteoarthritis: Review of Current Models and Future Directions. Archives of Bone and Joint Surgery 2023;11(1):1.
50. El Dib R, Tikkinen KAO, Akl EA, Gomaa HA, Mustafa RA, Agarwal A, et al. Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes. J Clin Epidemiol 2017;84:61-9.
51. Siontis KC, Siontis GC, Contopoulos-Ioannidis DG, Ioannidis JP. Diagnostic tests often fail to lead to changes in patient outcomes. J Clin Epidemiol 2014;67(6):612-21.
52. Modell V, Quinn J, Ginsberg G, Gladue R, Orange J, Modell F. Modeling strategy to identify patients with primary immunodeficiency utilizing risk management and outcome measurement. Immunol Res 2017;65(3):713-20.
53. Saquib N, Saquib J, Ioannidis JP. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. Int J Epidemiol 2015;44(1):264-77.
54. Niemiec E. Will the EU Medical Device Regulation help to improve the safety and performance of medical AI devices? Digit Health 2022;8:20552076221089079.
55. Stoel B. Use of artificial intelligence in imaging in rheumatology - current status and future perspectives. RMD Open 2020;6(1).
56. Friedberg MW, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. Rand Health Q 2014;3(4):1.
57. Shoop-Worrall SJ, Cresswell K, Bolger I, Dillon B, Hyrich KL, Geifman N. Nothing about us without us: involving patient collaborators for machine learning applications in rheumatology. Ann Rheum Dis 2021;80(12):1505-10.
58. STRATA-FIT: European Union;Accessed on 2024Jan 5th. Available from: https://strata-fit.eu/strata-fit/project/.
59. TrEatment Targets in Rheumatoid Arthritis: a Randomized Multi-centre, Treat to Target Strategy Trial: TETRA Study. Accessed on 2024 Mar 20th. Available from: https://classic.clinicaltrials.gov/ct2/show/NCT06046950

Appendices

Nederlandse samenvatting

Introductie

Sinds de jaren '90 is er een gestage ontwikkeling in klinische beslisondersteunings-systemen (clinical decision support systems, CDSS). Een CDSS genereert een medisch advies op basis van de kenmerken van een individuele patiënt, welke ter overweging wordt gepresenteerd aan de zorgverlener. Adviezen kunnen gaan over bv. de diagnose, aanvullend onderzoek of de medicamenteuze behandeling van de patiënt. In eerste instantie werden CDSSen vooral gebaseerd op regels die zijn afgeleid van bestaande richtlijnen, andere gepubliceerde literatuur of medische expertise. Een voorbeeld hiervan is een waarschuwing over medicatie-interacties tijdens het voorschrijven van een medicijn. Tegenwoordig zijn er echter steeds meer CDSSen gebaseerd op predictiemodellen, die zijn ontwikkeld met behulp van conventionele statistische methodes of met machine learning. Een voorbeeld hiervan is een advies over de insuline dosering op basis van voorspelde glucose-waardes bij diabetes-patiënten.[1]

Om een CDSS te ontwikkelen en toe te passen zijn patiëntgegevens nodig. In dit proefschrift richten we ons voornamelijk op CDSSen die routine zorgdata gebruiken. Dit zijn gegevens die al worden verzameld in de reguliere zorg, in tegenstelling tot bijvoorbeeld complexe bloedtesten die alleen nog in onderzoeksverband worden gedaan. Routine zorgdata heeft verschillende voordelen. Zo is het beschikbaar voor vele patiënten, waaronder patiënten die vaak niet mee (kunnen) doen aan klinisch onderzoek zoals zwangeren of patiënten met meerdere ziektes. Daarnaast kan een CDSS die gebaseerd is op routine zorgdata eenvoudiger worden geïmplementeerd, omdat er geen aanvullende testen nodig zijn om een advies te genereren.

Voordat een CDSS kan worden ingevoerd in de praktijk, zijn er veel verschillende ontwikkelingsstappen nodig. Als er een predictiemodel wordt gebruikt, moet eerst de voorspellende waarde van dit model worden aangetoond. Dit dient te gebeuren binnen de dataset waarin het model is ontwikkeld (interne validatie), maar vooral ook in een andere dataset van bv. een ander ziekenhuis (externe validatie). Een goede voorspellende waarde is echter geen garantie dat een CDSS daadwerkelijk patiënten-uitkomsten verbetert, zoals bv. minder klachten of ziekenhuisopnames. Het effect van een CDSS op patiënten-uitkomsten kan in eerste instantie worden gesimuleerd met besliskundige modellen. Idealiter wordt het effect vervolgens aangetoond in een gerandomiseerd onderzoek waarbij behandeling met en zonder de CDSS wordt vergeleken en de patiënten-uitkomsten worden gemeten. Daarnaast dient een CDSS

gebruiksvriendelijk en kosteneffectief te zijn, en moet het voldoen aan Europese regelgeving over veiligheid en privacy.

In dit proefschrift richten we ons op de ontwikkeling, validatie en evaluatie van CDSSen voor reumatoïde artritis en primaire antistofdeficiënties. Reumatoïde artritis (RA) is een chronische auto-immuunziekte die zich voornamelijk kenmerkt door gewrichtsontstekingen. Er kan echter ook betrokkenheid optreden van andere delen van het lichaam zoals het hart, de longen, of de huid. De ziekte kan op elke leeftijd beginnen en komt vaker voor bij vrouwen. Roken en erfelijke factoren spelen een rol in het ontstaan van RA, al is de oorzaak niet volledig bekend. Primaire antistofdeficiënties (PADs) betreffen een heterogene groep van afweerstoornissen. De symptomen kunnen variëren van luchtweginfecties, tot maag-darmklachten, auto-immuniteit en een verhoogd risico op bepaalde vormen van kanker. Omdat PADs zeldzaam zijn en zoveel verschillende uitingsvormen hebben wordt de diagnose vaak gemist. Het verkorten van de tijd tot diagnose zou bij kunnen dragen aan het voorkomen van complicaties en vroegtijdig overlijden door het starten van de medische behandeling.

Deel 1: Het patiënten-perspectief

In hoofdstuk 2 hebben we RA patiënten geïnterviewd over hun visie op het gebruik van predictiemodellen als ondersteuning in hun medische behandeling. Er werden zes thema's geïdentificeerd over (1) de behoefte aan informatie, (2) het vertrouwen in ondersteuning door predictiemodellen, (3) de voorziene rol van predictiemodellen in medische besluitvorming, (4) de hoop om persoonlijk te profiteren van de beslisondersteuning, (5) de bereidheid om tijd en moeite te investeren om bij te dragen aan de variabelen van het model en (6) de effecten op de relatie met de zorgverlener. De resultaten van dit onderzoek kunnen worden gebruikt tijdens de ontwikkeling en implementatie van CDSSen die zijn gebaseerd op predictiemodellen, met als doelstelling om de acceptatie door patiënten te verbeteren. Voorbeelden zijn het overwegen van predictoren die het patiënten-vertrouwen vergroten, zoals stress, voeding en het weer; en het voorzien in de behoefte van patiënten om de input van het model te verifiëren.

Deel 2: Diagnostische beslisondersteuning

Een van de mogelijke toepassingen van CDSSen is de vroege herkenning van ziektes. In hoofdstuk 3 hebben we verschillende methodes onderzocht voor de identificatie en predictie van moeilijk behandelbare RA (difficult-to-treat RA, D2T RA). D2T RA verwijst naar patiënten die klachten blijven houden ondanks meerdere lijnen van behandeling.

We hebben verschillende machine learning methodes toegepast om D2T RA patiënten te identificeren in routinezorgdata, en om in een vroeg stadium te voorspellen wie er D2T RA zal ontwikkelen. Het (vroeg) identificeren van deze patiënten is van belang voor onderzoek naar de oorzaken van D2T RA. In de kliniek zou de tijdige voorspelling van een risico op D2T RA bij kunnen dragen aan het optimaliseren van de behandeling op basis van recente richtlijnen. De technieken uit hoofdstuk 3 zullen verder worden ontwikkeld tot een CDSS binnen het Europees samenwerkingsverband STRATA-FIT[2].

Hoofdstukken 4 en 5 richten zich op de ontwikkeling van een CDSS voor de vroege herkenning van PAD binnen de huisartsenpraktijk. In hoofdstuk 4 is er een screeningsalgoritme ontwikkeld op basis van literatuur, klinische expertise en metadata van PAD patiënten en verschillende controlegroepen. Het algoritme omvat antibiotica voorschriften, diagnostische codes voor PAD symptomen en het aantal bezoeken aan de huisarts. Tijdens de klinische validatie in hoofdstuk 5 is dit algoritme toegepast op een dataset van 61.172 elektronische patiëntendossiers (EPDs) van 13 huisartsenpraktijken. Van de 104 hoog-risico patiënten die werden geïncludeerd voor een bloedonderzoek op immunoglobulines, is bij 10 patiënten een PAD diagnose bevestigd. Met onze methode konden we dus een hoog-risico populatie selecteren met een PAD-prevalentie van 1:10, versus een prevalentie van 1:1.700-1:25.000 in de algemene bevolking. Om de selectie van hoog-risico patiënten nog verder te verscherpen hebben we een aanvullend model ontwikkeld. De voorgestelde screeningsmethode kan bijdragen aan de vroegtijdige herkenning van PAD-patiënten binnen de huisartsenpraktijk. Vervolgonderzoek zal zich richten op validatie op grotere schaal en op het evalueren van de kosteneffectiviteit van deze methode.

Deel 3: Beslisondersteuning bij medicamenteuze behandelingen

Naast diagnostische beslisondersteuning kunnen CDSSen ook worden toegepast als ondersteuning bij het kiezen van de medicamenteuze behandeling. De geadviseerde behandelstrategie bij RA is volgens het 'treat-to-target' concept. Hierbij wordt de ziekte-activiteit op vaste tijdsintervallen gemeten en vergeleken met een vooraf gekozen behandeldoel (target). Als het behandeldoel niet tijdig is gehaald, wordt de behandeling geïntensiveerd. Een behandeldoel kan gebaseerd zijn op verschillende ziekte-maten, zoals de 'disease activity score'(DAS), de 'clinical disease activity index' (CDAI) of de 'simplified disease activity index' (SDAI). Binnen deze ziekte-maten kan er ook weer gekozen worden voor verschillende afkappunten, zoals remissie of lage ziekte-activiteit (LDA). Er zijn dus vele verschillende behandeldoelen mogelijk, en er is geen consensus over welke optimaal is. In hoofdstuk 6 hebben we een systematische review en meta-regressie analyse uitgevoerd om verschillende behandeldoelen te vergelijken met betrekking tot hun effect op klinische uitkomstmaten. Het bleek dat het

nastreven van SDAI/CDAI-LDA, en in mindere mate DAS-remissie, op verschillende vlakken betere uitkomsten gaven dan het DAS-LDA behandeldoel. Er was echter geen verschil aantoonbaar wat betreft gewrichtsschade zoals gescoord op röntgenfoto's. Daarnaast was er onvoldoende data beschikbaar om de effecten van de verschillende behandeldoelen op de medicatiedosering en bijwerkingen te onderzoeken. Vanwege het ontbreken van deze data en het risico op residuele confounding bij dergelijke meta-regressie analysemethoden, zou een gerandomiseerd onderzoek het optimale behandeldoel met meer zekerheid vast kunnen stellen. Dit onderzoek (de TETRA studie[3]) is recentelijk gestart in de Sint Maartenskliniek in Nijmegen. Het selecteren van het juiste behandeldoel kan de beslissing om reumamedicatie te intensiveren, stoppen of af te bouwen ondersteunen.

Voor de behandeling van RA zijn verschillende medicijnen beschikbaar, waaronder 'biological disease modifying antirheumatic drugs' (bDMARDs). bDMARDs worden geproduceerd door levende cellen, en kunnen via verschillende wegen pro-inflammatoire processen remmen. Alhoewel bDMARDs effectief zijn in de behandeling van RA geven ze ook een risico op bijwerkingen, zijn er injecties nodig en gaan ze gepaard met aanzienlijke kosten. Voor zowel de patiënt als het gehele zorgsysteem is het daarom relevant om bDMARDs tot een zo laag mogelijke dosering af te bouwen. Het afbouwen van bDMARDs geeft echter een risico op opvlammingen van ziekte-activiteit. Daarom hebben we in hoofdstuk 7 een predictiemodel ontwikkeld om het risico op een opvlamming tijdens het stapsgewijs afbouwen van bDMARDs te voorspellen. Het model is gebaseerd op routine zorgdata van RA patiënten. Naast interne en externa validatie hebben we een simulatie uitgevoerd van de effecten op het aantal opvlammingen en de bDMARD dosering. Hieruit bleek dat afbouwen met behulp van het predictiemodel beter was dan afbouwen zonder het model. In hoofdstuk 8 beschrijven we het protocol voor een gerandomiseerd onderzoek naar het gebruik van dit predictiemodel in een CDSS, dat momenteel wordt uitgevoerd in verschillende ziekenhuizen in Nederland. Bij een hoog voorspeld risico op een opvlamming wordt er geadviseerd om niet verder af te bouwen. Bij een laag voorspeld risico op een opvlamming wordt geadviseerd om door te gaan met afbouwen. Het doel is om het aantal opvlammingen te verminderen, terwijl de bDMARD wel tot een zo laag mogelijke dosering kan worden afgebouwd. De resultaten van dit klinische onderzoek worden in de komende 2 jaar verwacht.

Conclusies

Dit proefschrift beschrijft de ontwikkeling, validatie en evaluatie van verschillende klinische beslisondersteunings-systemen voor RA en PAD. Dergelijke systemen hebben de potentie om diagnoses te vervroegen en behandelingen te optimaliseren. Er zijn echter ook meerdere uitdagingen en valkuilen in de ontwikkeling en implementatie van

CDSSen. Uiteindelijk dient de nadruk niet te liggen op de voorspellende waarde, maar op het werkelijke voordeel voor de patiënt en het zorgsystem.

Referenties

1. Huang S, Liang Y, Li J, Li X. Applications of Clinical Decision Support Systems in Diabetes Care: Scoping Review. J Med Internet Res 2023;25:e51024.
2. STRATA-FIT: European Union; Accessed on 2024 Jan 5th. Available from: https://strata-fit.eu/strata-fit/project/.
3. TrEatment Targets in Rheumatoid Arthritis: a Randomized Multi-centre, Treat to Target Strategy Trial: TETRA Study. Accessed on 2024 Mar 20th. Available from: https://classic.clinicaltrials.gov/ct2/show/NCT06046950

Dankwoord

Vele mensen hebben direct of indirect bijgedragen aan dit proefschrift, waarvan ik enkelen hierbij graag in het bijzonder zou willen bedanken.

Allereerst veel dank aan de patiënten die deel hebben genomen aan de PATIO-, IMPACT- en GP-PAD studies. Klinisch onderzoek is alleen mogelijk met het vertrouwen en de inzet van patiënten. Dank jullie wel!

Dr. Welsing, beste Paco. Sinds het begin van mijn promotietraject hebben we elkaar vrijwel wekelijks gesproken, soms zelfs meerdere keren per week. Je was altijd uitstekend op de hoogte van mijn projecten en gaf waardevolle en constructieve feedback. Als clinicus heb ik veel kunnen leren van jouw statistische kennis. Je hebt dan ook een belangrijke bijdrage geleverd aan alle hoofdstukken van deze thesis. Zo'n gedegen begeleiding is uitzonderlijk tijdens promotieonderzoek, veel dank daarvoor!

Dr. Leavis, beste Helen. Ik heb er bewondering voor dat je naast je drukke klinische verantwoordelijkheden altijd laagdrempelig beschikbaar was voor de begeleiding van de GP-PAD projecten. Je kennis van zeldzame immuundeficiënties en je uitgebreide netwerk waren hierbij van groot belang. Even leek het erop dat mijn verhuizing naar het buitenland het einde zou betekenen van mijn promotietraject. Mede dankzij jouw vertrouwen is het mogelijk geweest om mijn onderzoeken op afstand af te ronden. Dankjewel voor het vertrouwen en de goede begeleiding de afgelopen jaren.

Dr. den Broeder, beste Alfons. Vanuit Nijmegen was je altijd bereid om mee te denken over de verschillende projecten. Je kritische wetenschappelijke blik gaf vaak de aanleiding tot interessante discussies die de manuscripten tot een hoger niveau hebben getild. Dank voor deze inbreng.

Prof. dr. Lafeber, beste Floris. Als promotor wist je direct bij te springen als een van de projecten vast dreigde te lopen. Ook stond je open voor alternatieve oplossingen, zoals het afronden van mijn promotietraject vanuit het buitenland. Dank voor deze pragmatische instelling en begeleiding.

Dan natuurlijk m'n mede PhD-studenten! Matthijs, mede dankzij jouw goede opzet heeft de PATIO-studie succesvol van start kunnen gaan. Ik heb veel plezier gehad in de samenwerking voor ons gezamenlijke manuscript, en onze ijzersterke elevator pitch. Nadia, ook na de D2T RA hackathon was je altijd bereikbaar voor advies, dankjewel daarvoor! Je succesvolle promotietraject is een voorbeeld geweest. Juliëtte, ik heb

bewondering voor je enthousiasme en werklust, je bent inderdaad echt een dendritische cel. Frank en Tammo, we waren maar kort kamergenoten, maar het was wel heel gezellig. Bas, Amir en Sharisa, ik kijk met veel plezier terug op ons ESID-congres in Gotenburg! Sina, je hebt het PATIO-kindje met veel succes overgenomen, en ik ben trots op ons gezamenlijke IMPACT-project. Dankjewel voor de gezellige samenwerking!

Dr. Tekstra, beste Janneke, ik heb veel geleerd van het meedraaien met je polikliniek tussen de onderzoeksactiviteiten door. Je enthousiaste energie werkt aanstekelijk! Ik hoop dat jullie gezamenlijke-woonboerderij-droom inmiddels geslaagd is.

Anne-Karien, dank voor het helpen navigeren door het doolhof van de wet- en regelgeving rondom klinische trials. Ria, Karin en Anneloes, dank voor het bijspringen bij het zien van PATIO-patiënten wanneer dat nodig was, en vooral ook voor jullie gezelligheid.

Lieve pap en mam, dank voor jullie wijze raad op de momenten dat het nodig was. Jullie veilige vangnet heeft een stabiele basis gecreëerd van waaruit ik de wereld kon ontdekken. Dat is van onschatbare waarde! Lieve joor, grote broer, we hebben allebei heel wat afgereisd, maar nu komen onze levens toch weer heel dicht bij elkaar. Ik kijk er naar uit om met je samen te wonen op de Zandweg! Pel, je bent denk ik de eerste onderzoeker die ik ooit heb ontmoet, en je kent de academische wereld als geen ander. Ik heb er bewondering voor hoe je al je projecten combineert, en kijk er naar uit om in Hemelum op bezoek te komen.

Matthijs, m'n grote liefde, m'n beste vriend, je bent mijn maatje voor het leven. Nico, wat een feest dat je in ons leven bent gekomen! Het is magisch om te zien hoe je de wereld ontdekt, als een onderzoeker pur sang.

List of publications

- Messelink MA, Fadaei S, Verhoef LM, Welsing PMJ, Nijhof NC, Westland H. Rheumatoid arthritis patients' perspective on the use of prediction models in clinical decision-making. Accepted for publication in Rheumatology.
- Messelink MA, Welsing PMJ, Devercelli G, Marsden JWN, Leavis HL. Clinical Validation of a Primary Antibody Deficiency Screening Algorithm for Primary Care. J Clin Immunol. 2023 Sep 16. doi: 10.1007/s10875-023-01575-8. Epub ahead of print. PMID: 37715890.
- Messelink MA, Berbers RM, van Montfrans JM, Ellerbroek PM, Gladiator A, Welsing PMJ, Leavis H. Development of a primary care screening algorithm for the early detection of patients at risk of primary antibody deficiency. Allergy Asthma Clin Immunol. 2023 May 27;19(1):44. doi: 10.1186/s13223-023-00790-7. PMID: 37245042; PMCID: PMC10224324.
- Messelink MA, den Broeder AA, Marinelli FE, Michgels E, Verschueren P, Aletaha D, Tekstra J, Welsing PMJ. What is the best target in a treat-to-target strategy in rheumatoid arthritis? Results from a systematic review and meta-regression analysis. RMD Open. 2023 Apr;9(2):e003196. doi: 10.1136/rmdopen-2023-003196. PMID: 37116986; PMCID: PMC10152050.
- Messelink MA, van der Leeuw MS, den Broeder AA, Tekstra J, van der Goes MC, Heijstek MW, Lafeber F, Welsing PMJ. Prediction Aided Tapering In rheumatoid arthritis patients treated with biOlogicals (PATIO): protocol for a randomized controlled trial. Trials. 2022 Jun 16;23(1):494. doi: 10.1186/s13063-022-06471-x. PMID: 35710576; PMCID: PMC9202120.
- van der Leeuw MS, Messelink MA, Tekstra J, Medina O, van Laar JM, Haitjema S, Lafeber F, Veris-van Dieren JJ, van der Goes MC, den Broeder AA, Welsing PMJ. Using real-world data to dynamically predict flares during tapering of biological DMARDs in rheumatoid arthritis: development, validation, and potential impact of prediction-aided decisions. Arthritis Res Ther. 2022 Mar 23;24(1):74. doi: 10.1186/s13075-022-02751-8. PMID: 35321739; PMCID: PMC8941811.
- Messelink MA, Roodenrijs NMT, van Es B, Hulsbergen-Veelken CAR, Jong S, Overmars LM, Reteig LC, Tan SC, Tauber T, van Laar JM, Welsing PMJ, Haitjema S. Identification and prediction of difficult-to-treat rheumatoid arthritis patients in structured and unstructured routine care data: results from a hackathon. Arthritis Res Ther. 2021 Jul 8;23(1):184. doi: 10.1186/s13075-021-02560-5. PMID: 34238346; PMCID: PMC8265126.
- Messelink MA, Roodenrijs NMT, van Es B, Hulsbergen-Veelken CAR, Jong S, Overmars LM, Reteig LC, Tan SC, Tauber T, van Laar JM, Welsing PMJ, Haitjema S. Hackathon over de identificatie en exploratie van 'difficult-to-treat' reumatoïde artritis in reguliere zorgdata, Nederlands Tijdschrift voor de Reumatologie, 2021 Jan, 33-37

Curriculum Vitae

Marianne Alice Messelink was born December 2nd 1992 in Margraten, The Netherlands. She grew up in various cities in the Netherlands, St. Martin and the United States. In 2010 she graduated from secondary school at the Oostvaarders College Almere, after which she obtained her bachelor's degrees in medicine at the Vrije Universiteit van Amsterdam (2015) and Neuroscience at the Universiteit van Amsterdam (2016). During her studies she was the bachelor representative of the student council, the assistant coordinator of the Immunology & Oncology educational program and participated in various committees. After finishing medical school in 2018, she started her clinical career as a medical doctor (ANIOS) at the department of Internal Medicine at the Flevoziekenhuis hospital in Almere. In 2020 she continued as a PhD candidate under the supervision of dr. P.M.J. Welsing and dr. H.L. Leavis, with dr. A. A. den Broeder and prof. dr. F.P.J.G. Lafeber as promotors, at the Rheumatology & Clinical Immunology department of the University Medical Center Utrecht.