Diagnostic performance and clinical utility of referral rules to identify primary care patients at risk of an inflammatory rheumatic disease

Delft, E.T.A.M. van; Barreto, D.L.; Helm-van Mil, A.H.M. van der; Alves, C.; Hazes, J.M.W.; Kuijper, T.M.; Weel-Koenders, A.E.A.M.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 12, December 2022, pp 2100–2107
DOI 10.1002/acr.24789

Diagnostic Performance and Clinical Utility of Referral Rules to Identify Primary Care Patients at Risk of an Inflammatory Rheumatic Disease

Elke Theodora Antonia Maria van Delft,[1] Deirisa Lopes Barreto,[1] Annette Helena Maria van der Helm-van Mil,[2] Celina Alves,[3] Johanna Maria Wilhelmina Hazes,[4] Tjallingius Martijn Kuijper,[1] and Angelique Elisabeth Adriana Maria Weel-Koenders[5]

Objective. To determine the diagnostic performance and clinical utility of the Rotterdam Early Arthritis Cohort (REACH) and the Clinical Arthritis Rule (CARE) referral rules in an independent population of unselected patients from primary care.

Methods. This study consisted of adults who were suspected of the need for referral to a rheumatologist by their general practitioner. Diagnostic accuracy measures and a net benefit approach were used to compare both rules to usual care for recognizing inflammatory arthritis and inflammatory rheumatic diseases (IRDs). Using the least absolute shrinkage and selection operator method and cross-validation we created an optimal prediction rule for IRD.

Results. This study consisted of 250 patients, of whom 42 (17%) were diagnosed with inflammatory arthritis and 55 (22%) with an IRD 3 months after referral. Considering inflammatory arthritis, the area under the receiver operating characteristic curve (AUC) was 0.72 (95% confidence interval [95% CI] 0.64–0.80) for REACH and 0.82 (95% CI 0.75–0.88) for CARE. Considering IRD, the AUC was 0.66 (95% CI 0.58–0.74) for REACH and 0.76 (95% CI 0.69–0.83) for CARE. CARE was of highest clinical value when compared to usual care. The composite referral rule for IRD of 10 parameters included sex, age, joint features, acute onset of symptoms, physical limitations, and duration of symptoms (AUC 0.82 [95% CI 0.75–0.88]).

Conclusion. Both validated rules have a net benefit in recognizing inflammatory arthritis as well as IRD compared to usual care, but CARE shows superiority over REACH. Although the composite referral rule indicates a greater diagnostic performance, external validation is needed.

INTRODUCTION

At present there is a challenge for primary care physicians to refer patients who may have inflammatory arthritis as quickly as possible. Unfortunately, experience on who must be referred or for whom additional investigations are appropriate is lacking (1). Although musculoskeletal symptoms are one of the most common reasons for consulting primary care, suspected inflammatory arthritis is relatively unusual. General practitioners (GPs) register on average only 1 new patient with rheumatoid arthritis (RA) each year (2). Nevertheless, early diagnosis is vital for the response of inflammatory arthritis treatment to achieve a state of remission sooner, which may consequently prevent long-term joint damage and increase quality of life (3) and functional outcomes (4).

This challenge results in approximately 60–75% of patients referred to the rheumatology outpatient clinic being diagnosed with a noninflammatory musculoskeletal disease (5–7). These noninflammatory diagnoses might be considered inappropriate referrals since they unnecessarily consume time and money from the individual patient, but they also cause a great burden on secondary health care and costs for society (8). Western countries

ClinicalTrials.gov identifier: NCT03454438.

[1]Elke Theodora Antonia Maria van Delft, MSc, Deirisa Lopes Barreto, PhD, Tjallingius Martijn Kuijper, PhD, MD: Maasstad Hospital, Rotterdam, The Netherlands; [2]Annette Helena Maria van der Helm-van Mil, PhD, MD: Leiden University Medical Center, Leiden, and Erasmus Medical Center, Rotterdam, The Netherlands; [3]Celina Alves, PhD, MD: Ziekenhuisgroep Twente, Almelo, The Netherlands; [4]Johanna Maria Wilhelmina Hazes, PhD, MD: Erasmus Medical Center, Rotterdam, The Netherlands; [5]Angelique Elisabeth Adriana Maria Weel-Koenders, PhD, MD: Maasstad Hospital and Erasmus University, Rotterdam, The Netherlands.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24789&file=acr24789-sup-0001-Disclosureform.pdf.

Address correspondence to Elke Theodora Antonia Maria van Delft, MSc, Maasstad Hospital, Department of Rheumatology, Maasstadweg 21, 3079 DZ Rotterdam, The Netherlands. Email: elkevandelft@msn.com.

Submitted for publication June 25, 2021; accepted in revised form September 15, 2021.

SIGNIFICANCE & INNOVATIONS

- A validated referral rule can show that the proportion of referred patients with a definite suspicion of inflammatory rheumatic diseases will increase.
- Although not developed to pick up the entire spectrum of inflammatory rheumatic diseases, the Clinical Arthritis Rule shows good diagnostic accuracy in doing so.
- The composite Delft rule seems to have the largest potential considering diagnostic performance and clinical utility, besides being easy to use.

already experience an increasing demand for care. The number of referrals to the rheumatologist have been predicted to increase even further in the near future (9), while the health care budget will keep decreasing (10). Therefore innovative tools that support integrated care are necessary.

Recently, 2 separate and distinct referral rules for arthritis have been developed to select patients for referral to the rheumatologist. The aim of both rules was to assist in the decision-making process in patients with musculoskeletal symptoms with suspected inflammatory arthritis, to promote early identification of inflammatory arthritis. Both referral rules could promote early identification of inflammatory arthritis with the aim of increasing appropriate health care utilization.

Since both rules have been developed within a setting between primary and secondary care, adequate performance within other, purely primary care settings is not evident (1). In addition, the true population of patients in which GPs may consider referral to a rheumatologist is much more complex than solely suspected inflammatory arthritis. Patients experiencing any disease within the broad spectrum of inflammatory rheumatic diseases (IRDs) would likely benefit from treatment by an outpatient rheumatologist. Hence, establishing how well both rules perform in a group of primary care patients representative of the real life situation is warranted, alongside the possible clinical impact of the rules, in order to receive the right care at the right place for inflammatory arthritis and IRD patients.

Therefore, the aim of this study was to assess the diagnostic performance and clinical utility of the Rotterdam Early Arthritis Cohort (REACH) and the Clinical Arthritis Rule (CARE) referral rules in adult primary care patients suspected of the need for referral to an outpatient rheumatology clinic. Second, we aimed to develop an optimal rule for detecting IRD, consisting of parameters from both REACH and CARE.

PATIENTS AND METHODS

Study design and population. For this prospective observational diagnostic study we used data from the control group of the JOINT referral study (11). This randomized controlled trial (RCT) aimed to aid GPs in their decision to refer patients with musculoskeletal symptoms. The inclusion took place between April 2017 and November 2019. The control group was consecutive new patients who had been referred to the rheumatology outpatient clinic from the Maasstad Hospital. The referral process was as usual, without application of any referral strategy, based on the national guidelines. The Maasstad Hospital is a nonacademic trainee hospital that serves a population of approximately 600,000 in the greater region around Rotterdam, The Netherlands.

The new patients suspected of having inflammatory rheumatic disease were invited to participate in this validation study. Patients needed to fulfill the inclusion criteria, consisting of age ≥18 years at the time of their first consultation, suspicion by their GP of an IRD, and being able to understand and communicate in Dutch. Written informed consent was obtained from all study participants before any assessment was performed. A certification that this study was not subject to the full extent of the Medical Research Involving Human Subjects Act was obtained from the local medical ethics committee of the Maasstad Hospital in Rotterdam, The Netherlands.

Referral rules. *REACH.* REACH was developed within a population of patients who were age >16 years, who had joint symptoms for <1 year, with either synovitis in at least 1 joint or in the absence of synovitis, and who had pain in at least 2 joints, in addition to 2 or more of the following criteria: morning stiffness for >1 hour, inability to clench a fist in the morning, pain when shaking someone's hand, pins and needles in the fingers, difficulties wearing rings or shoes, family history of RA, or unexplained fatigue for <1 year. The rule consists of 9 items that can add up to a total score of 11 points (Figure 1A). The cutoff value for referral was set at a score of 6 points or higher (12).

CARE. Patients included in the derivation and validation study of CARE were patients in whom GPs were unsure about the presence of inflammatory arthritis (1). In case of a clear synovitis or very high suspicion of inflammatory arthritis, patients were excluded. Hence, deriving and validating CARE was performed in the difficult group of patients in whom GPs were uncertain of the presence of suspected inflammatory arthritis. This simplified rule consists of 7 items and a total score ranging from 0 to 7 and a half, with corresponding predicted risks (Figure 1B). A high sensitivity was preferred, and therefore the cutoff score was set at 4 points or higher (1).

Data collection. Patients suspected by their GP for the need of referral to a secondary care rheumatologist were seen by a research assistant prior to any consultation with a rheumatologist. During this intake, the questions of both REACH and CARE were asked. Additionally, data on demographic parameters were collected. After the intake, patients received their regular visit at the rheumatologist, who was unaware of the referral rule

A

Variables	Score
Male gender	1
Positive squeeze test of MCPs and/or MTPs	1
Any loss of motion*	1.5
1 to 10 painful joints*	1
11 or more painful joints*	1.5
10 or more swollen joints*	2
Joints red and/or warm	1
Acute onset	0.5
No fatigue	0.5
Delay in presentation < 3 months, or > 6months	1

*Self-reported

B

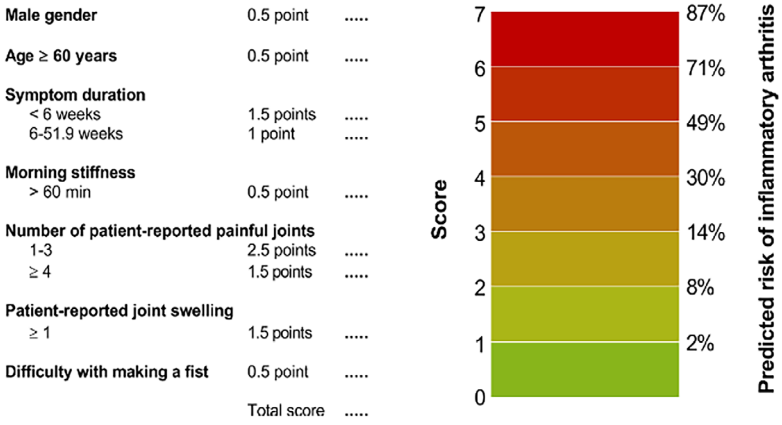

Figure 1. Referral rules and corresponding scores for **A**, Rotterdam Early Arthritis Cohort and **B**, Clinical Arthritis Rule patients. MCP = metacarpophalangeal joint; MTP = metatarsophalangeal joint. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24789/abstract.

outcome. The regular diagnostic workup, treatment, and follow-up were performed according to the clinical guidelines as usual.

Outcome. The primary outcome was the absence or presence of an inflammatory arthritis diagnosis (rheumatoid arthritis, psoriatic arthritis, arthritis urica, monoarthritis, oligoarthritis, and axial and peripheral spondyloarthritis) within the first 3 months after referral. The diagnosis made by the rheumatologist was used as the gold standard. The definition of IRD diagnoses (inflammatory arthritis plus systemic lupus erythematosus, systemic sclerosis, Sjögren's syndrome, Behçet's disease, mixed connective tissue disease, and polymyalgia rheumatica) as the secondary outcome in this study covered the broad spectrum of arthritis, and axial and peripheral spondyloarthritis plus systemic rheumatic diseases.

Statistical analysis. Demographic and clinical characteristics were presented using descriptive statistics. Three parameters were taken into account to assess the performance of the referral rules: discriminative capacity, calibration, and clinical utility. To determine the discriminative ability of REACH and CARE for inflammatory arthritis and IRD in an unselected population of patients from primary care suspected for IRD, the area under the receiver operating characteristic (ROC) curves (AUC) was assessed and diagnostic accuracy measures were evaluated. The ROC curve shows the discriminative ability of the test (13), and AUC represents exact accuracy.

To determine the most optimal cutoff point, the Youden index (J) was used, which represents the optimal balance between sensitivity and specificity (14,15). A high sensitivity is important for this prediction rule in daily practice, since it reflects the true positive rate. Correctly identifying those with an IRD is of utmost importance to be able to start treatment in an early stage of the disease. However, high specificity is of equal importance for the utility of this referral rule, since specificity can overcome inappropriate referrals to an outpatient rheumatology clinic. Hence, since both sensitivity and specificity are important in the current setting, the optimal cutoff point was considered to be the point at which there was an optimal balance between them.

Calibration of the referral rules was performed by 2 distinct methodologies due to differences in the design and applicability of the rules. For CARE, calibration was tested using a calibration plot including a chi-square test (13). Since REACH was designed as a dichotomous test rather than a risk model estimating a probability on a semicontinuous scale, a calibration plot could not be obtained. Instead cross tables were made to compare sensitivity and specificity in the population from which the referral rule was derived versus the current study population. Additionally a chi-square test was performed to statistically evaluate the calibration.

The possible clinical utility of the referral rules was obtained by performing a net benefit analysis, including a decision curve (16). The x-axis in a decision curve represents the percentage of referred patients being diagnosed with an inflammatory disease, e.g., the threshold probability. The y-axis shows the benefit of a model correctly identifying which patients do and do not have an inflammatory disease. This decision analytic approach can determine which of the 2 rules would lead to better clinical outcomes, and whether either would be better than a default strategy of treating all patients or not (16).

Finally, to see whether a combination of variables from both rules would lead to a more optimal prediction rule for recognizing IRD, logistic regression using the least absolute shrinkage and selection operator (LASSO) method was used (17), including leave-one-out cross-validation. An ROC curve with AUC and a density distribution plot were constructed to evaluate discriminative performance and to establish the optimal cutoff point using

the Youden index. Due to the relatively small number of participants, no subset of our data was available to validate this composite rule.

All analyses except for the LASSO were performed using STATA software, version 14.2. To fit the LASSO model and to generate the accompanying ROC curve, R software, version 3.6.1 (18), the glmnet package version 4.0-2 (19) and the proc package version 1.16.2 (20) were used. A 2-sided *P* value less than 0.05 was used to indicate statistical significance.

RESULTS

Study population. Of the 700 invited patients who were suspected by their GP of having an IRD, 560 (80%) responded to the invitation for this study (Figure 2). Of these responders, 310 expressed no interest in participating or did not fulfill the inclusion criteria. Informed consent was obtained from 250 participants.

Overall mean ± SD age of included patients was 50.8 ± 13.9 years (range 18–87 years), and 22.8% were male. Of all patients referred with the suspicion of an IRD, 42 (17%) were diagnosed with inflammatory arthritis within the first 3 months after referral. An IRD was diagnosed in 55 (22%) of all suspected patients. The most common inflammatory diagnosis was RA, and the most common noninflammatory diagnoses were osteoarthritis and fibromyalgia (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24789).

When using IRD as an outcome, in addition to the inflammatory arthritis diagnoses, we took into account the diagnosed systemic rheumatic diseases. A number of specific questions from both rules seemed to account for the recognition of those systemic disorders. All patients scored points for painful joints, and in addition, many scored points for swollen joints. Specifically within CARE, patients scored positively on the age variable, whereas within REACH the question on red or warm joints appeared to be discriminative.

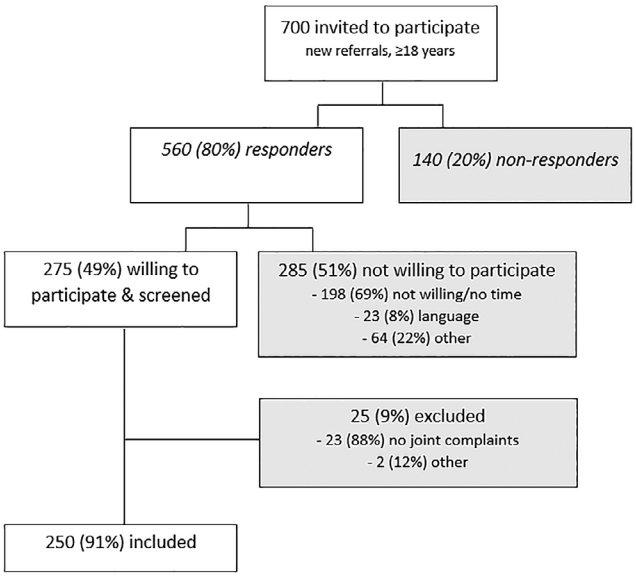

Figure 2. Recruitment flow chart.

Table 1. Diagnostic accuracy measures at optimal cutoff points for inflammatory arthritis and inflammatory rheumatic disease*

Outcome and diagnostic accuracy measure	REACH	CARE
Inflammatory arthritis		
Sensitivity	0.83	0.86
Specificity	0.53	0.70
AUC	0.72	0.82
Inflammatory rheumatic disease		
Sensitivity	0.75	0.75
Specificity	0.53	0.70
AUC	0.66	0.76

* AUC = area under the receiver operating characteristic curve; CARE = Clinical Arthritis Rule; REACH = Rotterdam Early Arthritis Cohort.

Discriminative capacity. For REACH, a cutoff of 6 points was advised by Alves et al (12). In this study the estimated optimal cutoff point of 6.5 based in the Youden index deviated 0.5 from the advised value (Table 1). For CARE, an optimal cutoff point of 4 was found. In the derivation study of the CARE referral rule, a similar cutoff value was advised (1).

Figure 3A shows the ROC curve for both referral rules when using inflammatory arthritis as an outcome. REACH, indicated by the blue line, had a fair accuracy in classifying those with and without inflammatory arthritis, with an AUC of 0.72 (95% confidence interval [95% CI] 0.64–0.80). The red line for CARE shows an AUC of 0.82 (95% CI 0.75–0.88). This result corresponds with a good accuracy of CARE to correctly classify those with and without inflammatory arthritis. Figure 3B shows the ROC curve for both referral rules when using IRD as an outcome. REACH had a poor accuracy in classifying IRD, with an AUC of 0.66 (95% CI 0.58–0.74). CARE showed an AUC of 0.76 (95% CI 0.69–0.83).

Calibration. For REACH, calibration was assessed comparing sensitivity and specificity between the population from which this rule was derived and our own population of patients suspected of IRD (see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24789). This calibration was done by using the cutoff value of 6 as recommended in the article from Alves et al (12). For the primary outcome (inflammatory arthritis), specificity showed a significant difference between the 2 populations ($P = 0.001$), with a higher specificity in REACH compared to this study. Also for the secondary outcome (IRD), specificity was significantly higher in REACH compared to the present study ($P = 0.001$).

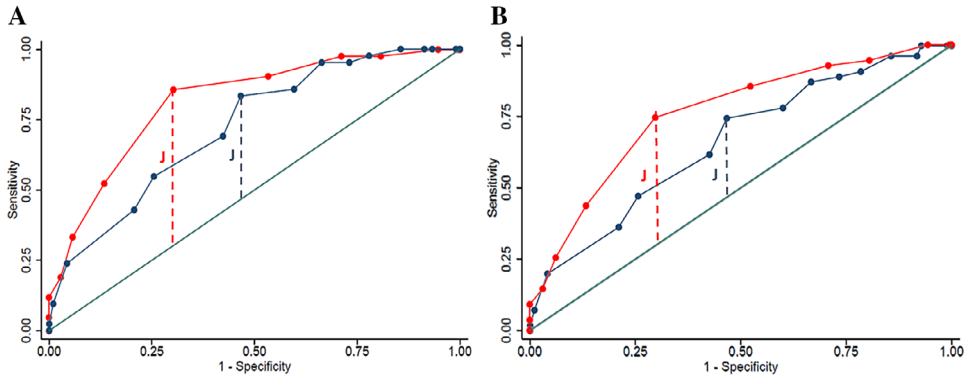

Figure 3. Receiver operating characteristic curve including Youden index (letters J) for both referral rules when using inflammatory arthritis (**A**) and inflammatory rheumatic disease (**B**) as outcome. Blue line indicates the Rotterdam Early Arthritis Cohort rule; red line indicates the Clinical Arthritis Rule.

Calibration plots for CARE can be found in Supplementary Figure 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24789. Results of the chi-square test showed that there was no significant miscalibration ($P = 0.42$) for the primary outcome. Using cross tables, a significant difference in specificity between the populations was found, with an increased specificity when using this referral rule within our population of referred patients compared to CARE ($P < 0.001$). Inspection of the calibration plot for CARE considering the secondary outcome also showed that there was no significant miscalibration ($P = 0.92$). Cross tables, however, indicated a significant difference in both sensitivity ($P = 0.001$) and specificity ($P < 0.001$).

Clinical utility. The clinical utility of both rules for the primary outcome compared to usual care is depicted in decision curves (Figure 4). Usual care shows zero net benefit, at a threshold probability of 16.8%, the percentage of definite inflammatory arthritis cases in the present study population. This finding corresponds to the percentage of referred patients diagnosed with inflammatory arthritis without application of any referral rule. At that threshold probability, there is a net benefit of 0.04 for REACH and 0.09 for CARE. Except for a small range of low threshold probabilities, we see that CARE always achieves the highest net benefit when compared to REACH and to usual care where all patients are being referred.

For the secondary outcome, the decision curve for usual care cuts the x-axis at 22.0%, which equals the percentage of referred patients diagnosed with an IRD without application of a referral rule. At that threshold probability, REACH shows a net benefit of 0.04, and CARE of 0.10. Except for a threshold probability lower than 10%, use of CARE leads to a higher benefit compared to the other 2 strategies.

Optimal rule (Delft rule). We included 13 parameters combined from the 2 referral rules for inclusion in the LASSO regression analysis. After the LASSO regression analysis, a set of 10 parameters remained as relevant predictors for recognizing IRD. From CARE, all parameters were part of the composite model, except for morning stiffness. Parameters from REACH that were covered in the composite rule include red or warm joints, acute onset of symptoms, restricted range of motion, and

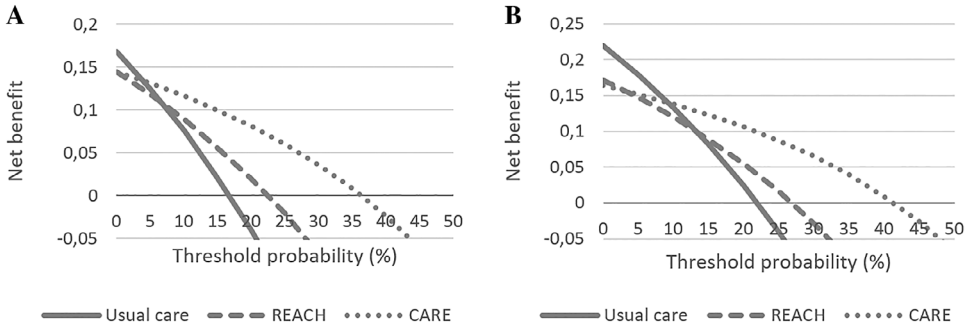

Figure 4. Decision curves showing net benefit for using a referral rule in suspected patients when using **A**, inflammatory arthritis and **B**, inflammatory rheumatic disease as outcome. The usual-care line crossing the x-axis represents the percentage of referred patients being diagnosed with an inflammatory disease in the normal situation without application of any referral rule. For this point we see that net benefit is zero. CARE = Clinical Arthritis Rule; REACH = Rotterdam Early Arthritis Cohort.

Table 2. Regression coefficients for selected parameters in the final optimal predictive rule

	Coefficient
Intercept	0.521
Sex (male vs. female)	–0.177
Age (per year)	0.003
Swollen joints (yes vs. no)	0.085
Painful joint count	
1–3 joints	0.028
≥4 joints	–0.001
Red or warm joints (yes vs. no)	0.082
Acute onset of symptoms (yes vs. no)	0.053
Inability to make a fist (yes vs. no)	0.012
Restricted range of motion (yes vs. no)	0.017
Absence of fatigue (yes vs. no)	–0.138
Duration of symptoms >1 year (yes vs. no)	–0.204

absence of fatigue. Hence, the most optimal predictive rule was composed of the following variables: sex, age, swollen joints, painful joint count, red or warm joints, acute onset of symtoms, inability to make a fist, restricted range of motion, absence of fatigue, and symptom duration of over a year (Table 2).

The ROC curve showed a good performance of this Delft rule for detecting IRD, with an AUC of 0.82 (95% CI 0.75–0.88) (see Supplementary Figures 2 and 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24789). The Youden index showed an optimal cutoff point at a predicted probability of 0.27. At this point the sensitivity was 0.69 and specificity was 0.84.

DISCUSSION

In the present study, good diagnostic accuracy was found for both REACH and CARE when considering our primary outcome; detecting patients with inflammatory arthritis in an unselected primary care setting. Despite not being developed to pick up the entire spectrum of IRD, we assessed the performance of these referral rules by using IRD as a secondary outcome. CARE yielded better results in terms of discriminative capacity, with an AUC indicative of good diagnostic accuracy for detecting IRD, even though the rule was not initially developed for this purpose. The proposed Delft rule that has been developed in the current study, specifically for detecting IRD in primary care, seems to indicate an even better accuracy in doing so when compared to the currently available rules.

Overall CARE achieved higher specificity when compared to REACH. Mainly contributing to this specificity are questions on joint involvement, like the presence of painful or swollen joints, which account for a larger part of the total score in CARE compared to REACH. High specificity is important, since for people not at risk of inflammatory arthritis or IRD the referral rule should indicate a negative outcome. A referral rule with high specificity can overcome inappropriate referrals to an outpatient rheumatology clinic. This specificity results in a lower individual and societal burden of musculoskeletal symptoms and a reduction of waiting times for rheumatology outpatient clinics (21). When looking for which referral rule is most effective in keeping inappropriate referrals from expensive outpatient care, CARE seems most accurate of the 2 existing rules. The composite Delft rule also yields a high specificity for detecting IRD and suggests great potential to diminish inappropriate referrals to outpatient rheumatology care.

Within a primary care population, both sensitivity and specificity are highly important to overcome substantial burden for both the individual and society (22). The sensitivity is reflected by the number of people with inflammatory arthritis or IRD, in whom the referral rule was indicative of a positive referral. In the present study, true positives occurred in an equal number of cases for both CARE and REACH. Since CARE showed superior specificity, and also showed superior overall performance when compared to REACH, we advise using CARE over REACH for detecting both inflammatory arthritis and IRD.

The composite Delft referral rule has been developed specifically to detect IRD. It consists of a combination of variables from both existing referral rules. All variables from CARE are maintained, except for the morning stiffness. Added to that were the presence of red or warm joints, acute onset of symptoms, loss of motion, and the absence of fatigue as variables from REACH, as patient-reported outcomes. The Delft rule does not require any physical examination or diagnostic tests; hence it still remains easy to use, besides indicating better diagnostic performance and clinical utility. Since this rule was fitted to our own data, overfitting is a possibility. Therefore, an external validation cohort within an independent larger sample is needed to evaluate its relevance and location in clinical practice before implementation.

When detecting IRD, we looked at inflammatory arthritis plus systemic rheumatic diseases. Although systemic rheumatic diseases can affect the entire body, joint involvement is frequently reported by patients with several systemic disorders (23–27). The fact that painful and swollen joints account for a large part of the total score within both CARE and REACH explains why those patients with IRD are also recognized by the referral rules. Painful, swollen, and red or warm joints also play a big role in the Delft rule. Therefore, a prerequisite to successfully use either of these referral rules in daily clinical practice for detecting IRD is the presence of some type of joint involvement.

A remarkable finding is that the incidence in the present study was rather different from the incidences in the derivation studies of both rules. The incidence of inflammatory arthritis was 41% in the CARE study and 44% in the REACH study. However, the presence of inflammatory arthritis among suspected inflammatory arthritis patients is expected to be lower in primary care compared to the derivation studies (1). The lower presence is reflected in the incidence of 17% inflammatory arthritis, which we found in the present study and which corresponds well to

incidences found in previous studies (5,28). This resemblance indicates that the study population in the present study likely closely resembles the true target population.

The finding above reflects the main strength of this study. The existing referral rules have now been tested within an independent population of unselected patients from primary care suspected to have inflammatory arthritis or IRD. Both referral rules for inflammatory arthritis have been developed in a prespecified population, reflected by the high incidence of inflammatory arthritis found in those studies. In contrast to these studies, the present study population consisted of unselected patients suspected for IRD. The strength of our study is that both referral rules have now been tested in the primary care population wherein the rules ought to be used. There was no selection bias for including patients and GPs. For GPs, no particular inclusion criteria were used. For patients, only the suspicion of an IRD and age ≥18 years, and no specific inflammatory arthritis or IRD inclusion criteria were required. Hence, this population is comparable to the population in which the referral rule will be used.

A potential weakness of this validation study is the number of cases, namely 42 inflammatory arthritis cases and 55 IRD cases. Several studies have suggested a minimum of 100 events for external validation of prediction models (29,30). As a result, the power to find significant deviations in this study might not be optimal. Since we used data from an ongoing RCT, we performed no separate sample size calculation for this validation. Nevertheless, when using decision curve analyses, the sample size is disregarded, as the most optimal strategy is chosen based on expected utility (16).

To conclude, especially in The Netherlands, where primary care physicians act as gatekeepers of the health care system, a referral rule like the rules assessed here is considered valuable. Performance of both REACH and CARE is sufficient for recognizing inflammatory arthritis in daily practice. CARE even shows sufficient performance in recognizing the entire spectrum of IRDs. CARE consists of only 7 questions that are all easy to use and easy to interpret; hence educating primary care physicians about the referral rule before implementation in daily practice will not be necessary. By using this referral rule, the proportion of referred patients with a definite suspicion of IRD can be increased. The high specificity implies that over half of all patients suspected of an IRD by their GP can be withheld from expensive outpatient rheumatology care. Therefore we advise evaluating CARE on its impact on cost-effectiveness in primary care before implementation. Although in this study the composite Delft referral rule seems more promising, external validation on a larger sample is needed to establish its real potential.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. van Delft had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Van Delft, Lopes Barreto, Weel-Koenders.

Acquisition of data. Van Delft, Kuijper, Weel-Koenders.

Analysis and interpretation of data. Van Delft, Lopes Barreto, van der Helm-van Mil, Alves, Hazes, Kuijper, Weel-Koenders.

REFERENCES

1. Ten Brinck RM, van Dijk BT, van Steenbergen HW, le Cessie S, Numans ME. Development and validation of a clinical rule for recognition of early inflammatory arthritis. BMJ Open 2018;8.
2. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford) 2002;41:793–800.
3. Pisetsky DS, Ward MW. Advances in the treatment of inflammatory arthritis. Best Pract Res Clin Rheumatol 2012;26:251–61.
4. Van Nies JA, Krabben A, Schoones JW, Huizinga TW, Kloppenburg M, van der Helm-van Mil AH. What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review. Ann Rheum Dis 2014;73:861–70.
5. Kallikorm R, Tender M. Rheumatic patients in primary and secondary care: differences in structure of diagnoses and working load within a five-year period. Croat Med J 2004;45:588–91.
6. Vanhoof J, Declerck K, Geusens P. Prevalence of rheumatic diseases in a rheumatological outpatient practice. Ann Rheum Dis 2002;61:453–5.
7. Guilar J, Wamuo I, Choy EH, Hossain M, Scott DL. Who is referred for specialist rheumatology care? Rheumatology (Oxford) 2000;39:1292–4.
8. Roberts S, Adebajo AO, Long S. Improving the quality of care of musculoskeletal conditions in primary care. Rheumatology (Oxford) 2002;41:503–8.
9. Cooper RA. There's a shortage of specialists: is anyone listening? Acad Med 2002;77:761–6.
10. Miller HD. From volume to value: better ways to pay for health care: the work begins. Health Affairs 2009;28:1418–28.
11. Van Delft ET, Lopes Barreto D, Roeterink JA, Han KH, Tchetverikov I, van der Helm-van Mil AH, et al. A study protocol on the evaluation of referral strategies for inflammatory arthritis in primary care patients at the level of healthcare organization, patient relevant outcomes and costs. Health 2020;12:240–52.
12. Alves C. Improving early referral of inflammatory arthritis. In: Early detection of patients at risk for rheumatoid arthritis: a challenge for primary and secondary care. Ridderkerk (The Netherlands): Ridderprint BV; 2015. p. 27–38.
13. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ. Discrimination and calibration of clinical prediction models: users' guide to the medical literature. JAMA 2017;318:1377–84.
14. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J 2005;47:458–72.
15. Martínez-Camblor P, Pardo-Fernández JC. The Youden Index in the generalized receiver operating characteristic curve context. Int J Biostat 2019;15(1).
16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
17. James G, Witten D, Hastie T, Tibshirani R. Linear model selection and regularization. In: An introduction to statistical learning with applications in R. New York: Springer; 2013. p. 219–227.

18. R Core Team. R: A language and environment for statistical computing. R foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.
19. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33: 1–22.
20. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
21. Caffrey A, Smart KM, FitzGerald O. Physiotherapist-led triage at a rheumatology-based musculoskeletal assessment clinic: an 18-month service evaluation of activity and outcomes. ACR Open Rheumatol 2019;1:213–8.
22. 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.
23. Amezcua-Guerra LM, Hofmann F, Vargas A, Rodriguez-Henriquez P, Solano C, Hernandez-Diaz C, et al. Joint involvement in primary Sjogren's syndrome: an ultrasound target area approach to arthritis. Biomed Res Int 2013;2013:640265.
24. Ingegnoli F, Ughi N, Mihai C. Update on the epidemiology, risk factors, and disease outcomes of systemic sclerosis. Best Pract Res Clin Rheumatol 2018;32:223–40.
25. Avouac J, Clements PJ, Khanna D, Furst DE, Allanore Y. Articular involvement in systemic sclerosis. Rheumatology (Oxford) 2012;51: 1347–56.
26. Fortuna G, Brennan MT. Systemic lupus erythematosus: epidemiology, pathophysiology, manifestations, and management. Dent Clin N Am 2013;57:631–55.
27. Tamirou F, Arnaud L, Talarico R, Scirè CA, Alexander T, Amoura Z, et al. Systemic lupus erythematosus: state of the art on clinical practice guidelines. RMD Open 2018;4:e000793.
28. Gamez-Nava JI, Gonzalez-Lopez L, Davis P, Suarez-Almazor ME. Referral and diagnosis of common rheumatic diseases by primary care physicians. Br J Rheumatol 1998;37:1215–9.
29. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58: 475–83.
30. Collins GS, Ogondimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016; 35:214–6.