RHEUMATOLOGY

Rheumatology 2019;0:1–11
doi:10.1093/rheumatology/kez542

Original article

An updated matrix to predict rapid radiographic progression of early rheumatoid arthritis patients: pooled analyses from several databases

Antoine Vanier [1,2], Josef S. Smolen[3], Cornelia F. Allaart[4], Ronald Van Vollenhoven[5], Patrick Verschueren[6], Nathan Vastesaeger[7], Saedis Saevarsdottir[5], Karen Visser[4], Daniel Aletaha[3], Bernard Combe[8] and Bruno Fautrel[9]

Abstract

Objective. In early RA, some patients exhibit rapid radiographic progression (RRP) after one year, associated with poor functional prognosis. Matrices predicting this risk have been proposed, lacking precision or inadequately calibrated. We developed a matrix to predict RRP with high precision and adequate calibration.

Methods. Post-hoc analysis by pooling individual data from cohorts (ESPOIR and Leuven cohorts) and clinical trials (ASPIRE, BeSt and SWEFOT trials). Adult DMARD-naïve patients with active early RA for which the first therapeutic strategy after inclusion was to prescribe methotrexate or leflunomide were included. A logistic regression model to predict RRP was built. The best model was selected by 10-fold stratified cross-validation by maximizing the Area Under the Curve. Calibration and discriminatory power of the model were checked. The probabilities of RRP for each combination of levels of baseline characteristics were estimated.

Results. 1306 patients were pooled. 20.6% exhibited RRP. Four predictors were retained: rheumatoid factor positivity, presence of at least one RA erosion on X-rays, CRP > 30mg/l, number of swollen joints. The matrix estimates RRP probability for 36 combinations of level of baseline characteristics with a greatly enhanced precision compared with previously published matrices (95% CI: from ± 0.02 minimum to ± 0.08 maximum) and model calibration is excellent ($P = 0.79$).

Conclusion. A matrix proposing RRP probability with high precision and excellent calibration in early RA was built. Although the matrix has moderate sensitivity and specificity, it is easily usable and may help physicians and patients to make treatment decisions in daily clinical practice.

Key words: early rheumatoid arthritis, rapid radiographic progression, risk prediction, prognosis factors

Rheumatology key messages

- An updated risk matrix predicting rapid radiographic progression in early rheumatoid arthritis was estimated.
- The matrix estimates risk of rapid radiographic progression of early RA with a high level of precision.
- The matrix displays the risk according to the combination of four common baseline characteristics only.

[1]Department of Biostatistics Public Health and Medical Informatics, Sorbonne University, APHP, University Hospitals Pitié-Salpêtrière Charles-Foix, Paris, [2]University Bretagne-Loire, University of Nantes, University of Tours, Inserm UMR U1246 SPHERE 'Methods in patient-centered outcomes and health research', Nantes, [3]Division of Rheumatology and Department of Medicine 3, University of Vienna, Vienna, Austria, [4]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands, [5]Rheumatology Unit, Department of Medicine, Karolinska Institute and Karolinska University Hospital, Solna, Stockholm, Sweden, [6]KU Leuven, Skeletal Biology and Engineering Research Center, Leuven, [7]MSD, Brussels, Belgium, [8]Department of Rheumatology, Montpellier 1 University, Montpellier University Hospital and [9]Department of Rheumatology, Sorbonne University, GRC-08 (EEMOIS), APHP, University Hospitals Pitié-Salpêtrière Charles-Foix, Paris, France

Submitted 19 June 2019; accepted 19 September 2019

Correspondence to: Antoine Vanier, Inserm UMR U1246 SPHERE, Université de Nantes – Institut de Recherche en Santé 2 (IRS2), 22, Boulevard Bénoni-Goullin, 44200 Nantes, France.
E-mail: antoine.vanier@univ-nantes.fr

CLINICAL SCIENCE

Introduction

Multiple conventional synthetic or biologic DMARDs have demonstrated efficacy to control disease activity and structural damage progression in early RA patients and several have thus been approved as potential first-line agents in early RA [1]. International recommendations propose methotrexate, in association with steroids if needed, as a first-line agent in early RA [1–3]. As 'the anchor drug', methotrexate can also be combined with other conventional synthetic , or biologic, DMARDs as a bridging strategy [4–6].

However, not all patients respond well to treatment with MTX and if not, apart from suffering a delay in symptom relief, these patients are at risk of developing irreversible joint damage. Several negative prognostic factors for radiologic damage progression have been identified. In the most recent version of the EULAR, clinical practice recommendations for either early arthritis or early rheumatoid arthritis [1, 3], the swollen joint count, acute phase reactants, RF, ACPA and X-ray bone erosions are mentioned because these variables have been associated with a higher risk of persistent or erosive arthritis [7–9] and poorer structural and functional outcomes [10, 11]. The presence of negative prognostic factors is used in the EULAR recommendations only in the second line for patients who have failed or inadequately responded to initial MTX therapy [1]. Ideally, the use of the negative prognostic factors could enable the identification of patients for whom more aggressive initial treatment strategy, e.g. conventional synthetic DMARD combination or biologic DMARD, could be more relevant.

During recent years, a proxy has been proposed to identify early RA patients with the most severe evolution: substantial structural damage progression at one year of follow-up during MTX therapy that had often started rapidly after disease onset. This refers to the concept of rapid radiographic progression (RRP), close to the notion of minimal clinically important change [12], which is defined as an increase of at least five points of the van der Heijde-modified Sharp score (vSHS) over one year of treatment [13–15]. RRP has been shown to be associated with a poor functional prognosis in the BeSt trial, in which RRP+ patients had persistently higher functional limitations and structural damage progression over 8 years of follow-up despite the application of a 'treat-to-target' strategy [16]. A similar observation was made in the ESPOIR cohort, whereby RRP+ patients at 1 year showed increased structural damage progression during the second and third years of follow-up [17].

The risk of RRP for a given patient is potentially quantifiable in daily clinical practice, thanks to the development of risk prediction matrices [13, 18–22], based on the same methodology as used for cardiovascular disease or osteoporosis [23, 24]. Some matrices have been developed in randomized clinical trials (RCTs) conducted in early RA patients where information about auto-antibodies and joint erosions was available. RCTs are representative of some of the patient population with early RA encountered in clinical practice but due to selection criteria are usually made of more homogeneous patients than everyday clinical practice [13, 19, 22]. Two other matrices were developed using data from observational cohorts that included early RA patients as they present in practice (the SONORA and the ESPOIR cohort) [18, 21], thus reflecting usual care without stringent therapeutic protocols or algorithms. Nonetheless, all these matrices were developed using data from a unique location, which can hamper their generalizability, because they do not consider the full heterogeneity of patients with early RA.

Comparison of the different matrices [25, 26] showed that discriminating power was interesting for three of them (Area Under the Curve (AUC) >0.7), but none was considered superior to the other. But AUC is a statistical indicator of discrimination (i.e. the ability of a model to dichotomize patients based on a specified threshold into two binary subgroups with the least classification errors possible) [27], which may not be the only characteristic of interest when estimating a risk matrix. Indeed, calibration, as referring to the ability of a model to estimate risks in different subgroups of patients along the risk continuum without under or overestimation [27], may be of importance because the purpose of a risk matrix is not to classify patients into two subgroups but rather to estimate risks for several combinations of baseline characteristics. Apart from the ESPOIR matrix [21], the calibration of the other matrices was not mentioned in the original reports [13, 18, 19, 22]. Nonetheless, the other matrices were tested on the ESPOIR data in a recent publication [25]. The hypothesis of acceptable calibration was rejected for all [25]. In addition, estimates of the probabilities of developing RRP suffered from lack of precision (as illustrated by the large range of the CI at a 95% level (95% CI)) due to the limited sample size. Inadequate calibration (risks are over or underestimated along the range) and lack of precision (the true value of the risk in the population can be far from the estimated value in the sample; e.g. the matrix indicates a risk of 0.60 but the true value can be somewhere between 0.30 and 0.90) are both obstacles to a reliable use of these matrices at a patient level.

Thus, the objective of the current study was to construct a new matrix for early RA, estimating the risk of developing RRP at one year, by pooling the data of cohorts (ESPOIR and Leuven cohorts) and several RCTs (BeSt, SWEFOT and ASPIRE trials) to get RRP probabilities for different combination of levels of baseline characteristics. We expect this new matrix to have a higher level of precision in estimation and a better generalizability (by including more heterogeneity) that will be illustrated by exhibiting appropriate calibration.

Methods

Patients

This current study is based on the pooling of individual data from two cohorts (ESPOIR and Leuven cohort) and

data from three RCTs (BeSt, SWEFOT and ASPIRE trials) [28–32]. In short, the three RCTs were all designed for testing the efficacy of a combination of MTX + infliximab *vs* one or several strategy(ies) starting with MTX monotherapy in patients with early and active RA [30–32]. The Leuven cohort was designed to compare patients receiving the combination therapy (COBRA therapy) in early and active RA with patients receiving a more conventional step-up strategy starting with MTX monotherapy [29]. The ESPOIR cohort was designed to follow the prognosis of patients with early and active RA from community practice [28]. The detailed summary of the key characteristics of the five pooled datasets can be found in the papers published previously [28–32] and in Supplementary Table S1, available at *Rheumatology* online

In the current study, the included patients were adult DMARD-naïve patients with recent diagnosis of active RA. Only patients for whom the first therapeutic strategy after inclusion in their respective study was to prescribe MTX or LEF as a monotherapy at least for 3 months were included. Thus, it corresponded with those patients in the ESPOIR and Leuven cohorts, with two arms of the BeSt trial (patients with sequential monotherapy or step-up combination), with one arm (control arm) in the ASPIRE trial and with the entire population of the SWEFOT trial. Of note, changes in therapeutic strategies between 3 months and 1 year after inclusion were not protocol-based in the ESPOIR and Leuven cohorts while they were (according to study design) in the other studies. Including different databases with heterogeneity regarding the treatment course after the first 3 months of initial MTX or LEF was made in order to get a matrix of baseline predictors that will be the most robust ones to various treatment courses.

Main outcomes and potential predictors

The follow-up time point in the current study was 1 year after inclusion. The main outcome was the presence of RRP after 1 year of follow-up. This was defined as increase in vSHS (van der Heijde-modified Sharp Score) of at least five points between baseline and 1 year [13, 26].

The baseline predictors of RRP, selected on the basis of the most recent literature, were the following: age, sex, disease duration, RF positivity, ACPA positivity, CRP level, ESR, number of swollen joints (SJC28), number of tender joints (TJC28), presence of at least one erosion on X-rays and vSHS total score. Smoking status, while found to be associated with RRP in the SWEFOT matrix study, was not considered because it was not available in the other databases.

Statistical analyses

Quantitative variables were described using mean (S.D.), median and inter-quartile range. Qualitative variables were described using counts (%). As the metric of most of the potential baseline predictors of RRP was initially quantitative, these predictors had to be categorized in order to further compute the risk matrix. Thus, potential predictors were categorized using thresholds determined by the sample distribution of the variables (using tertiles); the thresholds could then be slightly adjusted based on expert opinion.

The predictors of RRP were selected in two main steps. First, two-by-two comparisons were performed by the presence or not of RRP using χ^2 tests. The variables for which the *P*-values of the comparison were below 0.10 were retained as potential predictors for the second step of the selection procedure. Second, multivariate logistic regression with RRP as the dependent variable was applied. In this study, we wanted to get an optimal balance between estimating RRP probabilities with the most precision possible (i.e. with the narrowest 95% CI possible, therefore using all the available data), and getting a risk matrix with a sufficient level of proof of validation. Thus, to select the final model, a procedure of best subset selection using 10-fold stratified cross-validation was performed. The Area Under the Receiver Operating Characteristic Curve (AUC) of the model was used as the criterion to select the multivariate logistic model with the best combination of independent variables. Those AUCs were estimated for each possible combinations of independent variables (best subset selection) using a repeated (50 times) 10-fold cross validation, with blocks similar in distribution to the full sample for the outcome (i.e. providing a comparable rate of RRP in each block). The best combination of independent predictors that was retained led to a model with the highest AUC. The advantages of this method of model selection against the more traditionally used stepwise method are detailed in Supplementary Material, section Details about the selection procedure of the multivariate logistic model retained in this study, available at *Rheumatology* online [27]. The selected logistic model was fitted on the whole sample to check validity. Calibration and discrimination of the model were assessed using Hosmer and Lemeshow test as well as the Receiving Operating Curve. Finally, parameter estimates of the model (computed using 5000 bootstrapped samples) were used to estimate the probabilities (along with 95% CI) of RRP for each combination of predictor values to get the risk matrix of RRP. If, in a simulation study,10-fold stratified cross-validation was shown to lead to AUC estimates close to those obtained via adequate external validation [27, 33], it can be argued it does not provide a true assessment of model qualities on an independent external dataset. Thus, we also performed analyses via a data-splitting strategy where the model was trained and estimated on a randomly selected training dataset (75% of the original dataset) and validated on the remaining 25%. All the procedures (from selection of the variables to retain in the model, to checking calibration and discrimination, to estimating the final matrix) were therefore also performed under this data-splitting strategy as sensitivity analyses. All analyses were performed using R 3.2.3 [34].

The original studies from which the pooled datasets were extracted were all approved by their relevant ethics review board. Written informed consent was obtained from all patients before their participation in each original study.

Results

Patient characteristics at baseline

The data of 1306 patients were pooled. The proportion of patients coming from each pooled study along with baseline characteristics are detailed in Table 1. Disease duration (the delay from symptom onset to diagnosis) was 16.1 (21.6) weeks in average (median at 9 weeks). The distribution of disease duration was right skewed with most of patients having short disease duration. Disease duration longer or equal to 6 months corresponds to the 89th percentile of the sample.

Evolution over 1 year and main outcome

A decrease in disease activity and functional disability was observed over 1 year of follow-up. There was an average decrease of 2.4 (1.4) in DAS28 score (DAS28-ESR) [from 5.7 (1.1)–3.3 (1.4)] with 564 patients (49.5%) showing good response, 424 (37.2%) moderate response and 152 (13.3%) no response (European League Against Rheumatology criterion), leading to 391 patients (33.8%) in remission at year 1. There was also an average decrease of 0.6 (0.7) in HAQ score [from 1.2 (0.6)–0.6 (0.6)]. Within the whole sample, an average increase in vSHS total score of 3.2 (8.5) was observed over the year of follow-up [from 6.7 (10.2)–9.9 (14.3)] (Table 2). RRP at 1 year occurred in 236 patients (20.6% 95% CI [18.2, 22.9]) (Table 2). RRP occurred in 3/42 (7%) patients in the Leuven cohort, 43/370 (11.6%) in the ESPOIR cohort, 77/311 (24.7%) patients in the SWEFOT trial, 75/224 (33.5%) patients in the BeSt trial and 38/197 (19.3%) patients in the ASPIRE trial.

RRP determinants

The baseline variables associated with RRP on univariate analyses were gender, RF and ACPA status, alone or combined, CRP level, ESR level, SJC28, vSHS total score and the presence of at least one erosion on X-rays (Table 3). As vSHS total score is not a characteristic usually available in community daily practice and ACPA status was missing for a substantial proportion of the patients of one of the pooled populations, they were not retained for multivariate analyses. Moreover, as the inflammatory markers CRP and ESR are highly correlated variables, they were included separately in the process of model selection to avoid any multicollinearity issues. The final multivariate model was developed using the data of 1120 patients. It included RF status, erosive disease on X-rays, CRP level and SJC28 as predictors of RRP (Table 4). Figure 1A displays the number of predicted RRP patients according to the model and the number observed in the sample for 10 categories of increasing risk. If we can note a slight tendency to overestimate the number of RRP for low risk (12.5 expected *vs* eight observed for risk from 0.09–0.11), overall the calibration of the model was excellent with an observed/expected ratio of RRP events of 1 (Hosmer and Lemeshow test: $P = 0.79$). Receiving operating curve analysis showed moderate discriminative power with an AUC of 0.68 (Figure 1B).

Matrix elaboration

Using logistic model parameters, we developed a risk matrix (Fig. 2) estimating RRP probability for 36 combinations of level of baseline characteristics. The risk spreads from 5% in the lowest risk group (bottom-left of the matrix: probability of 0.05 95% CI [0.03, 0.08], patient with CRP <10mg/l, without RF, without erosive disease on X-ray, with <6 swollen joints) to 47% in the highest (top-right of the matrix: probability of 0.47 95% CI [0.39, 0.55], patient with CRP >30 mg/l, with RF, with erosive disease on X-ray, with >10 swollen joints).

Sensitivity analyses of model section and validation: results using a data-splitting strategy

When performing all the analyses via a data-splitting strategy (as aforementioned), the same four predictors were selected. Adequate calibration was shown on the validation dataset and an AUC of 0.63 (95% CI: [0.55, 0.71]) was estimated. Estimated RRP probabilities were very close to the main analysis, with a slight decrease in precision (i.e. larger 95% CI). Full results of these analyses are available in Supplementary Tables and Figures S3 to S8, available at *Rheumatology* online.

Discussion

Using four baseline characteristics, a matrix to predict the risk of RRP in early RA has been built. Although this new matrix is not strictly identical to any of the previously published matrices, all the characteristics have been identified at least once in a previous study as a predictor of RRP [13, 18, 19, 21, 22]. This pooled analysis seems to confirm characteristics associated with the level of inflammation (number of swollen joints or CRP), disruption of the immunologic system (RF positivity) or presence of structural damage at time of diagnosis as predictors of RRP while it does not advocate for an association with age at diagnosis, gender or disease duration. Moreover, the use of CRP instead of ESR was associated with a slight increase in discriminatory power (AUC at 0.68 *vs* 0.66 (data not shown)). In addition, in line with ESPOIR and ASPIRE matrices, this study confirms SJC 28 as a predictor of RRP rather than TJC 28 [13, 21].

In this study, the average probability of RRP was estimated to be 0.21 (95% CI [0.18, 0.23]). The matrix discriminates patients with a 4.1-fold lower risk of RRP compared with average risk, to patients with a 2.3-fold

Table 1 Patient characteristics at baseline

Characteristic	Whole sample (n = 1306 (100))		Leuven Cohort (n=42 (3.2))	ESPOIR Cohort (n=370 (28.3))	SWEFOT trial (n=405 (31.0))	BeSt trial (n=247 (18.9))	ASPIRE trial (n=242 (18.5))
	Estimate[a]	NA	Estimate[b]	Estimate[b]	Estimate[b]	Estimate[b]	Estimate[b]
Age, years	52.0 (13.3); 42-53-62	0 (0.0)	53.3 (16.8)	49.4 (11.4)	54.4 (14.0)	54.4 (13.3)	49.2 (12.9)
Female sex	936 (71.7)	0 (0.0)	26 (61.9)	271 (73.2)	288 (71.1)	172 (69.6)	179 (74.0)
Disease duration, weeks	16.1 (21.6); 5-9-19	1 (0.1)	9.0 (8.4)	15.2 (8.2)	6.3 (3.3)	10.3 (17.8)	41.2 (34.9)
IgM RF positivity	845 (65.2)	10 (0.8)	29 (69.0)	204 (55.1)	276 (68.8)	161 (65.2)	175 (74.2)
ACPA positivity	600 (58.6)	282 (21.6)	28 (68.3)	185 (50.0)	239 (62.9)	148 (63.5)	—
Erosive disease on X-rays[c]	730 (58.0)	48 (3.7)	8 (19.1)	233 (63.0)	146 (40.1)	171 (71.0)	172 (71.4)
vSHS erosion score	3.9 (7.4); 0-1-4	48 (3.7)	0.6 (2.3)	3.4 (5.5)	1.9 (4.2)	3.8 (5.4)	8.3 (12.4)
vSHS narrowing score	2.8 (4.8); 0-1-4	48 (3.7)	2.0 (4.3)	3.1 (4.5)	2.6 (5.5)	3.0 (4.2)	2.8 (4.8)
vSHS total score	6.7 (10.4); 0-3-8	48 (3.7)	2.6 (5.3)	6.5 (8.6)	4.5 (8.2)	6.8 (8.3)	11.1 (15.9)
ESR, mm/1st hr	39.0 (26.8); 18-33-54	14 (1.1)	35.8 (25.1)	32.7 (25.0)	39.7 (26.1)	42.7 (28.4)	44.6 (27.6)
CRP, mg/L	30.5 (38.1); 6-16-39	22 (1.7)	28.7 (34.8)	24.8 (37.7)	33.7 (37.2)	39.2 (45.9)	25.6 (39.4)
Swollen joint count (28 joints)	9.5 (5.2); 5-9-13	10 (0.8)	7.1 (5.1)	7.9 (5.4)	10.6 (5.2)	11.1 (4.9)	8.9 (4.2)
Tender joint count (28 joints)	10.7 (6.8); 5-9-15	8 (0.8)	7.0 (6.1)	8.7 (6.9)	9.4 (6.0)	8.7 (6.9)	14.2 (6.2)
HAQ score	1.2 (0.6); 0.7-1.1-1.6	263 (20.1)	1.1 (0.8)	1.0 (0.6)	1.2 (0.6)	—	1.5 (0.6)
DAS28 score	5.7 (1.1); 4.9-5.7-6.5	29 (2.2)	5.0 (1.2)	5.3 (1.2)	5.7 (1.0)	6.0 (1.0)	6.2 (1.0)
DAS28 remission[d]	6 (0.5)	29 (2.2)	1 (2.6)	5 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)

[a]Quantitative variables: mean (S.D.); 25th percentile-median-75th percentile; qualitative variables: count (%). [b]Quantitative variables: mean (S.D.). [c]At least one erosion on X-rays. [d]European League Against Rheumatology criterion. NA: not available; vSHS: van der Heijde-modified Sharp score.

Table 2 One-year main outcome of the study

Characteristic	Estimate[a]	NA
Whole sample (*n*=1306)		
Erosive disease on X-rays baseline[b]	730 (58.0%)	48 (3.7)
Erosive disease on X-rays year 1[b]	793 (67.5%)	131 (10.0)
vSHS erosion baseline	*3.8 (7.1); 0-1-4*	
vSHS erosion year 1	*5.7 (9.9); 0-2-7*	*162 (12.4)*
ΔvSHS erosion	*1.9 (5.7); 0-0-2*	*162 (12.4)*
vSHS narrowing baseline	*2.9 (4.8); 0-1-4*	
vSHS narrowing year 1	*4.1 (6.6); 0-2-5*	*162 (12.4)*
ΔvSHS narrowing	*1.2 (4.0), 0-0-1*	*162 (12.4)*
vSHS total baseline	6.7 (10.2); 0-3-8	
vSHS total year 1	9.9 (14.3); 1-5-13	*162 (12.4)*
ΔvSHS total	3.2 (8.5); 0-0-3.6	*162 (12.4)*
Main outcome		
RRP	236 (20.6) 95% CI: [18.2, 22.9]	162 (12.4)
In patients with RRP (*n*=236)		
vSHS total baseline	10.0 (12.3); 2-6-14	
vSHS total year 1	24.0 (20.1); 11-18-29	
ΔvSHS total	14.0 (13.4); 7-9-16	
In patients without RRP (*n*=908)		
vSHS total baseline	5.9 (9.5); 0-2-7	
vSHS total year 1	6.3 (9.3); 0-3-7	
ΔvSHS total	0.4 (2.5); 0-0-1	

[a]Quantitative variables: mean (S.D.); 25th percentile-median-75th percentile; qualitative variables: count (%). [b]At least one erosion on X-rays. vSHS: van der Heijde-modified Sharp score; Δ: difference between year 1 and baseline; RRP: rapid radiographic progression; 95% CI: CI at a 95% level; NA: not available.

higher risk than average. Of course, discriminating extreme cases with very high or very low risk of RRP is not the most useful as it does not add info to clinical expertise (at least it can be relevant to note it estimates a low absolute risk of RRP (0.05) in the absence of predictors of poor prognosis). Nonetheless, the matrix displays the risk of RRP for a wide range of combinations of baseline characteristics, which can add information to clinical expertise. According to the results, the presence of only one predictor is not sufficient to increase the risk of RRP above average (0.21). Moreover, the matrix shows that the main contribution to the increase in RRP risk can be attributed to the presence of structural damage at the time of diagnosis, then to RF positivity, then to CRP level and finally to the number of swollen joints. Indeed, an elevated number of swollen joints alone always corresponds to a risk of RRP $<$0.5-fold average while this risk increases to between 0.5 and 1.0-fold average when associated with the presence of just one another risk factor. Or, high level of CRP (above 30 mg/l) without the presence of structural damage is never associated with an increase in risk $>$1.5-fold average while structural damage can lead to such an increase despite mildly elevated CRP level (between 10 and 30 mg/l). Or, presence of structural damage plus RF positivity plus highly elevated CRP level leads to a much higher risk of RRP (0.37) than high number of swollen joints (above 10) plus RF positivity plus highly elevated CRP level (0.28). Also, the data illustrates the presence of structural damage at diagnosis combined with RF positivity leads to a RRP risk above average.

As the model built on these data has been used to express RRP in terms of an estimated risk for each combination of level of predictors and not to classify patients into two distinctive groups (absence or presence of RRP), it is paramount the estimated probabilities are reliable. Thus, two properties of the model are of great importance: calibration and precision of estimates [27]. Indeed, a model well calibrated is a model where there is a low discrepancy between what is predicted and what was observed. In our study, the calibration of the model was excellent. Overall, the model predicts in average the same number of RRP patients as was observed and along most of the continuum of risk level, the discrepancy between observed cases and predicted cases is low (the observed/expected ratio is close to 1). Apart from the ESPOIR matrix [21], the calibration of the other matrices was not mentioned in the original reports [13, 18, 19, 22]. Nonetheless, the other matrices were tested on the ESPOIR data in a recent publication [25]. The hypothesis of acceptable calibration was rejected for all [25]. Thus, these results seem to indicate this new matrix proposes reliable estimates of RRP risk. A second strength of this study is the improvement in the level of precision of the estimates (in terms of narrower 95% CI) compared with previously published matrices. Indeed, the precision of RRP probabilities estimates in

Table 3 Comparison of baseline characteristics by the presence or not of RRP, univariate analyses

Variable	Level	No RRP $n=908$ Count (%)	RRP $n=236$ Count (%)	*P*
Data available in all pooled databases				
Age	18–44	273 (30.1)	63 (26.7)	0.37
	45–60	367 (40.4)	107 (45.3)	
	60+	268 (29.5)	66 (28.0)	
Female sex	Yes	664 (73.1)	158 (66.9)	0.06
Disease duration, weeks	<6	241 (26.6)	76 (32.2)	0.21
	6–14	361 (39.8)	84 (35.6)	
	15 +	305 (33.6)	76 (32.2)	
RF Positivity	Yes	550 (60.9)	178 (75.7)	<0.001
CRP, mg/L	<10	433 (48.5)	81 (34.6)	<0.001
	10–30	239 (26.8)	58 (24.8)	
	30+	221 (24.7)	95 (40.6)	
ESR, mm/1st hr	<20	268 (29.7)	44 (18.9)	<0.001
	20–45	352 (39.0)	77 (33.0)	
	45+	283 (31.3)	112 (48.1)	
Swollen joint count 28	<6	247 (27.4)	52 (22.0)	0.02
	6–10	281 (31.1)	63 (26.7)	
	10+	375 (41.5)	121 (51.3)	
Tender joint count 28	<6	244 (27.0)	53 (22.5)	0.11
	6–10	231 (25.6)	53 (22.5)	
	10+	429 (47.5)	73 (55.1)	
Erosive disease on X-ray	Yes	500 (55.1)	172 (72.9)	<0.001
Characteristic not available in community daily practice				
vSHS total score	<1	274 (30.2)	41 (17.4)	<0.001
	1–4	301 (33.1)	54 (22.9)	
	5 +	333 (36.7)	141 (59.7)	
Data not available in all pooled databases				
ACPA positivity	Yes	386 (53.1)	144 (73.8)	<0.001
RF or ACPA positivity	Yes	601 (70.6)	199 (88.1)	<0.001
RF and ACPA positivity	Yes	335 (43.0)	123 (60.3)	<0.001

vSHS: van der Heijde-modified Sharp score.

Table 4 Final multivariate logistic model estimates, $n=1120$, complete case analysis

Characteristic	Beta	Standard error	OR and 95% CI	*P*
Intercept	−2.8	0.25	—	<0.001
RF positivity				
No	Reference	—		
Yes	0.77	0.17	2.1 [1.5, 3.0]	<0.001
Erosive disease on X-Ray				
No	Reference	—	—	
Yes	0.81	0.16	2.3 [1.7, 3.2]	<0.001
CRP				
< 10	Reference	—	—	
10–30	0.17	0.19	1.2 [0.8, 1.7]	0.39
30 +	0.77	0.18	2.1 [1.5, 3.0]	<0.001
Swollen joint count 28				
< 7	Reference	—	—	
7–10	0.03	0.21	1.03 [0.7, 1.5]	0.89
10 +	0.39	0.19	1.5 [1.01, 2.2]	0.048

OR: odds ratio; 95% CI: CI at a 95% level.

our current study ranges from ± 0.02 minimum to ± 0.08 maximum. This is better than precision level from the ESPOIR matrix (which went up to –0.36 or +0.24) [20]. This is even better than precision level from the SONORA matrix (going up to ± 0.15), even though the modest gain in sample size could have been balanced

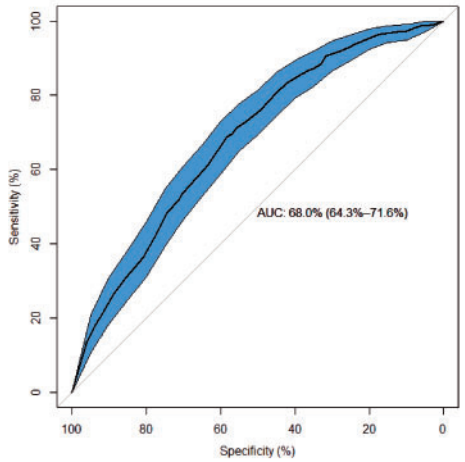
FIG. 1 Calibration and discriminative power of the final multivariate logistic prediction model of rapid radiographic progression

A Hosmer and Lemeshow Goodness-Of-Fit test

Category of Risk	Observed RRP	Expected RRP	Observed/Expected Ratio
[0.05 - 0.09]	7	7.9	0.9
(0.09 - 0.11]	8	12.5	0.6
(0.11 - 0.13]	15	13.3	1.1
(0.13 - 0.16]	26	22.9	1.1
(0.16 - 0.18]	11	12.3	0.9
(0.18 - 0.22]	26	22.4	1.2
(0.22 - 0.25]	40	36.2	1.1
(0.25 - 0.29]	28	28.9	1.0
(0.29 - 0.37]	29	32.2	0.9
(0.37 - 0.47]	43	44.4	1.0
Global	233	232.8	**1, *P*=0.79**

Notes. RRP: Rapid Radiographic Progression, *P* : *P*-value of the test

B Receiver Operating Curve of the final model



Notes. AUC: Area Under the Curve. The are in blue is within the Confidence Interval at a 95% level of the curve. The values in parentheses are the Confidence Interval at a 95% level of the AUC.

A) Hosmer and Lemeshow goodness-of-fit test; **B)** receiving operating curve of the final model.

out by the higher number of probability estimates (from 16 to 36) [18]. A last strength is the robustness of the results using two different strategies for model selection and validation. Nonetheless, there is a decrease in the central tendency of AUC estimate when using the data-splitting strategy. However, the 95% CI of the AUCs from both validation strategies overlap greatly (0.68 [0.64–0.72] *vs* 0.63 [0.55–0.71]). Therefore, this difference may be due to sampling hazard. Or it can be probably explained by the tendency of data-splitting strategy to underestimate the true predictive power of a model, even though overfitting when using 10-fold stratified cross-validation cannot be ruled out [27].

Despite these strengths, our study suffers from some limitations. The main one concerns the discriminatory power of the model. Indeed, the AUC of the model was estimated to be 0.68, which corresponds to a model with moderate discriminatory power. It means that if the predicted probability of developing a RRP was used as a prognostic score with a threshold classifying patients in a binary manner (occurrence of RRP or not), then it would lead to classification errors (false positives and false negatives). If we nevertheless try to search an optimal cutoff by receiving operating curve analysis, then it would lead to a prognostic test with a sensitivity of 0.69, a specificity of 0.58, a negative predictive value of 0.88 and positive predictive value of 0.30. It means this prognostic test would lead to a high number of false-positive. The low positive predictive value can be related to the fact the model seems to not discriminate patients with a very high probability of RRP. This result is consistent with what was previously exhibited with the ASPIRE matrix [20]. According to a recent study exploring the discriminative power of previously published matrices on ESPOIR data, our current model has a

Fig. 2 Prediction matrix of the risk of occurrence of rapid radiographic progression at one year

		Absence of typical RA erosion on radiographs			Presence of typical RA erosions on radiographs		
		SJC < 6	6 ≤ SJC < 10	SJC ≥ 10	SJC < 6	6 ≤ SJC < 10	SJC ≥ 10
RF positivity	CRP ≥ 30	**0.21** [0.14 ; 0.29]	**0.21** [0.15 ; 0.29]	**0.28** [0.21 ; 0.36]	**0.37** [0.28 ; 0.47]	**0.38** [0.29 ; 0.48]	**0.47** [0.39 ; 0.55]
	10 ≤ CRP < 30	**0.13** [0.08 ; 0.18]	**0.13** [0.08 ; 0.19]	**0.18** [0.12 ; 0.24]	**0.25** [0.17 ; 0.33]	**0.25** [0.18 ; 0.33]	**0.33** [0.25 ; 0.41]
	CRP < 10	**0.11** [0.07 ; 0.15]	**0.11** [0.08 ; 0.15]	**0.15** [0.11 ; 0.20]	**0.22** [0.16 ; 0.28]	**0.22** [0.17 ; 0.28]	**0.29** [0.22 ; 0.36]
RF negativity	CRP ≥ 30	**0.11** [0.07 ; 0.17]	**0.11** [0.07 ; 0.17]	**0.15** [0.10 ; 0.21]	**0.22** [0.14 ; 0.31]	**0.22** [0.15 ; 0.31]	**0.29** [0.22 ; 0.37]
	10 ≤ CRP < 30	**0.06** [0.04 ; 0.10]	**0.07** [0.04 ; 0.10]	**0.09** [0.06 ; 0.13]	**0.13** [0.08 ; 0.19]	**0.14** [0.09 ; 0.20]	**0.18** [0.13 ; 0.25]
	CRP < 10	**0.05** [0.03 ; 0.08]	**0.06** [0.03 ; 0.08]	**0.08** [0.05 ; 0.11]	**0.11** [0.07 ; 0.16]	**0.12** [0.08 ; 0.17]	**0.16** [0.11 ; 0.22]

Notes. RRP: Rapid Radiographic Progression, RA : Rheumatoid Arthritis, SJC: Swollen Joint Count, CRP: C-Reactive Protein, RF: Rheumatoid Factor. Estimates are probabilities. Values in brackets are Confidence Intervals at a 95% level.

RRP Risk (average = 0.21)

- **> 1.5 fold average. High risk**
- **> 1 and ≤ 1.5 fold average. Intermediate risk**
- **< 1 and ≥ 0.5 fold average. Low risk**
- **< 0.5 fold average. Very low risk**

better discriminative power than the ASPIRE matrix, but probably inferior to BeST and SONORA matrices (based on the central tendency value of the AUCs) or possibly inferior to comparable (based on the 95% CI of the AUCs) [25]. Nonetheless, even the SONORA model with an AUC at 0.76 would lead to classification errors at a non-negligible level if it was used as a binary prognostic test. At the very least, those considerations reinforce the idea to use this model to express a probability of RRP for each combination of level of characteristics and not as a binary prognostic test. The next step will be to determine an intervention threshold above which the risk of pejorative evolution is perceived too high to prescribe MTX as monotherapy.

Several hypotheses can be made to explain this moderate AUC value. We may have missed some baseline predictors with high predictive power due to lack of data or knowledge. For example, the same analysis restricted to patients' data from which RF and ACPA status were both available, with a variable representing a combination of RF or ACPA positivity instead of RF positivity alone, would have led to a slightly better AUC (0.70 instead of 0.68). Moreover, in the SWEFOT matrix study, smoking status was found to be associated to RRP [22]. Furthermore, the therapeutic trajectory of each patient within the first year after treatment onset probably explains some part of the probability of developing RRP, especially here where there was some heterogeneity in treatment course after 3 months of follow-up between the different cohorts and RCTs. This heterogeneity, while enhancing the generalizability of the results can explain, in part, the moderate AUC value. Nonetheless, accounting for therapeutic trajectory would ideally require longitudinal modeling and would defeat the purpose of proposing a clue for clinicians about the RRP risk at the time of diagnosis.

A last limit is the presence of some missing data, mostly for the outcome (RRP, unavailable for 162 (12.4%) patients). Therefore, the final multivariate model was estimated on 85.8% (1120) of the 1306 patients of the sample. Supplementary Table S9, available at *Rheumatology* online, displays a comparison of the characteristics of the patients used for final model estimation (complete-case analysis) *vs* those excluded for missingness. Patients excluded were more likely to be RF and ACPA positive, had a slight increase in mean SJC28 (10.6 [5.1] *vs* 9.3 [5.2]) and a slight increase in DAS28 (5.9 [1.1] *vs* 5.7 [1.1]) (Supplementary Table S4, available at *Rheumatology* online). Thus, missingness in predictor(s) and/or outcome could be associated with higher values of predictors of RRP. Therefore, we cannot rule out patients excluded were slightly more likely to exhibit RRP, thus the magnitude of the association between the predictors founded and RRP could have been slightly underestimated. It has to be noted we have presented the complete-case analysis as the main analysis and did not perform any imputation techniques because they assume non-informative missing data, which is not the case [35].

Conclusion

An updated matrix estimating the risk of RRP in early RA was built based on the data from various RA populations. It displays an average risk of RRP for 36 combinations of level of four common baseline predictors, with excellent calibration and enhanced precision compared with previously published matrices. Nonetheless, the discriminative power of this model is not perfect and further investigations will be needed to better understand the complexity of predicting RRP in early RA.

Funding: The study has been financially supported by an unrestricted grant from MSD received by B.F. The funding committee had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure statement: N.V. is an employee of MSD. B.F. received consultancy fees from MSD, Pfizer and AbbVie. The other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

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

2 Singh JA, Saag KG, Bridges SL *et al*. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res 2016;68:1–25.

3 Combe B, Landewe R, Daien CI *et al*. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017;76:948–59.

4 Combe B, Landewe R, Lukas C *et al*. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2006;66:34–45.

5 Smolen JS, Breedveld FC, Burmester GR *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 Le Loët X, Berthelot JM, Cantagrel A *et al*. Clinical practice decision tree for the choice of the first disease modifying antirheumatic drug for very early rheumatoid arthritis: a 2004 proposal of the French Society of Rheumatology. Ann Rheum Dis 2006;65:45–50.

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

8 van der Helm-van Mil AHM, le Cessie S, van Dongen H *et al*. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007;56: 433–40.

9 Aletaha D, Neogi T, Silman AJ *et al*. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.

10 Scott DL, Symmons DP, Coulton BL, Popert AJ. Long-term outcome of treating rheumatoid arthritis: results after 20 years. Lancet Lond Engl 1987;1:1108–11.

11 Kiely P, Walsh D, Williams R, Young A. Outcome in rheumatoid arthritis patients with continued conventional therapy for moderate disease activity–the early RA network (ERAN). Rheumatol 2011;50:926–31.

12 Bruynesteyn K, van der Heijde D, Boers M *et al*. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. Arthritis Rheum 2002;46:913–20.

13 Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. Rheumatology 2009;48:1114–21.

14 Sharp JT, Bluhm GB, Brook A *et al*. Reproducibility of multiple-observer scoring of radiologic abnormalities in the hands and wrists of patients with rheumatoid arthritis. Arthritis Rheum 1985;28:16–24.

15 Sharp JT, Wolfe F, Mitchell DM, Bloch DA. The progression of erosion and joint space narrowing scores in rheumatoid arthritis during the first twenty-five years of disease. Arthritis Rheum 1991;34:660–8.

16 Dirven L, van den Broek M, Klarenbeek N *et al*. Eight year results of disease activity steered treatment in a large recent rheumatoid arthritis cohort: clinical and radiological outcomes. Arthritis Rheumatol 2011;63: S157.

17 Tobón G, Saraux A, Lukas C *et al*. First-year radiographic progression as a predictor of further progression in early arthritis: results of a large national French cohort. Arthritis Care Res 2013;65:1907–15.

18 Bombardier C, Li X, van der Heijde D, Chen M. A risk model for the prediction of radiographic progression: results from SONORA study. Arthritis Rheum 2009;60: S276.

19 Visser K, Goekoop-Ruiterman YPM, de Vries-Bouwstra JK *et al*. A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis receiving different dynamic treatment strategies: post hoc analyses from the BeSt study. Ann Rheum Dis 2010;69:1333–7.

20 Durnez A, Vanderschueren G, Lateur L, Westhovens R, Verschueren P. Effectiveness of initial treatment allocation based on expert opinion for prevention of rapid radiographic progression in daily practice of an early RA cohort. Ann Rheum Dis 2011;70:634–7.

21 Fautrel B, Granger B, Combe B *et al*. 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 Saevarsdottir S, Rezaei H, Geborek P *et al*. Current smoking status is a strong predictor of radiographic progression in early rheumatoid arthritis: results from the SWEFOT trial. Ann Rheum Dis 2015;74:1509–14.

23 Conroy RM, Pyörälä K, Fitzgerald AP *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24:987–1003.

24 Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008;19:385–97.

25 Granger B, Combe B, Le Loet X *et al*. Performance of matrices developed to identify patients with early rheumatoid arthritis with rapid radiographic progression despite methotrexate therapy: an external validation study based on the ESPOIR cohort data. RMD Open 2016;2:e000245.

26 De Cock D, Vanderschueren G, Meyfroidt S *et al*. The performance of matrices in daily clinical practice to predict rapid radiologic progression in patients with early RA. Semin Arthritis Rheum 2014;43:627–31.

27 James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer Texts in Statistics, Vol. 103. New York, NY: Springer New York, 2013.

28 Combe B, Benessiano J, Berenbaum F *et al*. The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. Jt Bone Spine Rev Rhum 2007;74:440–5.

29 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. Rheumatol 2008;47:59–64.

30 Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, Allaart CF *et al*. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum 2005;52:3381–90.

31 van Vollenhoven RF, Ernestam S, Geborek P *et al*. Addition of infliximab compared with addition of sulfasalazine and hydroxychloroquine to methotrexate in patients with early rheumatoid arthritis (Swefot trial): 1-year results of a randomised trial. Lancet 2009;374:459–66.

32 St Clair EW, van der Heijde D, Smolen JS *et al*. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. Arthritis Rheum 2004;50:3432–43.

33 Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Int Jt Conf Artif Intell 1995;14:1137–45.

34 R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2016.

35 van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007;16:219–42.