ACR Open Rheumatology
Vol. 3, No. 7, July 2021, pp 457–463
DOI 10.1002/acr2.11266
© 2021 The Authors. *ACR Open Rheumatology* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

American College of Rheumatology
Empowering Rheumatology Professionals

What Is the Persistence to Methotrexate in Rheumatoid Arthritis, and Does Machine Learning Outperform Hypothesis-Based Approaches to Its Prediction?

Helga Westerlind,[1] Mateusz Maciejewski,[2] Thomas Frisell,[1] Scott A Jelinsky,[2] Daniel Ziemek,[2] and Johan Askling[1]

Objective. The objectives of this study were to assess the 1-year persistence to methotrexate (MTX) initiated as the first ever conventional synthetic disease-modifying antirheumatic drug in new-onset rheumatoid arthritis (RA) and to investigate the marginal gains and robustness of the results by increasing the number and nature of covariates and by using data-driven, instead of hypothesis-based, methods to predict this persistence.

Methods. Through the Swedish Rheumatology Quality Register, linked to other data sources, we identified a cohort of 5475 patients with new-onset RA in 2006-2016 who were starting MTX monotherapy as their first disease-modifying antirheumatic drug. Data on phenotype at diagnosis and demographics were combined with increasingly detailed data on medical disease history and medication use in four increasingly complex data sets (48-4162 covariates). We performed manual model building using logistic regression. We also performed five different machine learning (ML) methods and combined the ML results into an ensemble model. We calculated the area under the receiver operating characteristic curve (AUROC) and made calibration plots. We trained on 90% of the data, and tested the models on a holdout data set.

Results. Of the 5475 patients, 3834 (70%) remained on MTX monotherapy 1 year after treatment start. Clinical RA disease activity and baseline characteristics were most strongly associated with the outcome. The best manual model had an AUROC of 0.66 (95% confidence interval [CI] 0.60-0.71). For the ML methods, Lasso regression performed best (AUROC = 0.67; 95% CI 0.62-0.71).

Conclusion. Approximately two thirds of patients with early RA who start MTX remain on this therapy 1 year later. Predicting this persistence remains a challenge, whether using hypothesis-based or ML models, and may yet require additional types of data.

INTRODUCTION

Clinical prognostication of rheumatoid arthritis (RA), at diagnosis as well as at later time points, constitutes one of the biggest challenges in clinical practice. Most RA treatment guidelines list a handful of individual predictors, such as presence of erosions and high systemic inflammation, as markers of poor prognosis or inadequate response to therapy. Few, if any, more-elaborate prediction models are used in clinical practice to inform the choice of treatment or of treatment intensity. As a step toward an individualized approach to RA treatment, identifying baseline predictors to indicate which patients are likely to respond well to and remain on methotrexate (MTX) used as monotherapy would be important because other patients might do better if offered alternative treatments at diagnosis (1,2).

Presented in part at the annual meeting of the American College of Rheumatology, Chicago, IL, October 19-24, 2018.

This report received support from a collaboration between Pfizer and Karolinska Institutet on precision medicine. This work was further supported by the Swedish Research Council, the Stockholm County Council, the Swedish Heart-Lung Foundation, Karolinska Institutet (Strategic Research Area Epidemiology), the Nordic Research Council (Nordforsk), Vinnova, and the Rheumatology Research Foundation, who had no impact on design or interpretation of the study or its results.

[1]Helga Westerlind, PhD, Thomas Frisell, PhD, Johan Askling, MD, PhD: Karolinska Institutet, Solna, Sweden; [2]Mateusz Maciejewski, PhD, Scott A Jelinsky, PhD, Daniel Ziemek, PhD: Pfizer Pharma GmbH, Berlin, Germany.

Karolinska Institutet (Dr. Askling as principal investigator) has or has had research agreements with Abbvie, Astra-Zeneca, Bristol Myers Squibb, Eli Lilly, MSD, Pfizer, Roche, Samsung Bioepis, Sanofi, and UCB, mainly in the context of safety monitoring of biologics via the Swedish Biologics Register (ARTIS). Drs. Maciejewski, Jelinsky and Ziemek are currently employees of Pfizer and own Pfizer stock as part of compensation. No other disclosures relevant to this article were reported.

Address correspondence to Helga Westerlind, Karolinska Universitetssjukhuset, Clinical Epidemiology Division, Eugeniahemmet, T2:01, Solna, Sweden. Email: Helga.Westerlind@ki.se.

Submitted for publication March 30, 2021; accepted in revised form April 1, 2021.

SIGNIFICANCE & INNOVATIONS

- Two thirds of all patients with rheumatoid arthritis (RA) who start treatment with methotrexate monotherapy remain on the therapy 1 year after treatment start.
- Machine learning models were able to derive robust models, performing on par with manual hypothesis-based models, in an automated way.
- Clinical characteristics at RA diagnosis were the most important predictors.

So far, attempts to predict response to MTX have used hypothesis-based approaches (here: limited sets of predictors preselected on the basis of expert opinion) applied to clinical variables, candidate genetic markers (3–6), genome-wide searches for molecular markers (7,8), or targeted searches for epigenetic markers (9). In terms of prediction modeling, these projects have, with few exceptions (6,8,9), rarely yielded an area under the receiver operating characteristic (AUROC) greater than 0.70 (ref. 10). Furthermore, these attempts have mainly targeted prediction of primary treatment response at 3 months, with some extending the time period up to 6 months (5), and the number of individuals included has been small. Longer-term predictions are scarce, as are predictions of persistence.

In the age of big data, more and more data sources to characterize patients are available, such as those offered by linkages of clinical and RA-specific data to health care registers. In hypothesis-based approaches applied to such data, only a fraction of all available information (here: defined as comorbid conditions rather than the information offered by the totality of drug use and health care contacts) will ever be used. Machine learning (ML) methods have been developed to improve prediction modeling of complex data, typically "-omics" data. So far, however, ML has not been widely used in the context of clinical prediction modeling in RA or in the context of the blending of clinical, administrative, and other types of data. Furthermore, it has not been tested if health care data, representing the medical history leading up to RA diagnosis, could offer information to assist prediction of important treatment outcomes.

The main objective of this study was therefore to assess and predict the persistence to MTX as disease-modifying antirheumatic drug (DMARD) monotherapy at 1 year after RA diagnosis based on the majority of available data in the Swedish Rheumatology Quality (SRQ) register, enriched with additional data, using a variety of approaches. Specifically, we aimed to assess the potential gain from using more complex data sets with an increasingly higher resolution of medical history, as well as from using different data-driven analytical approaches, compared with each other and with traditional hypothesis-based epidemiological approaches.

MATERIALS AND METHODS

A schematic overview of the workflow and data sources used in this article is presented in Supplementary Figure A1.

Data sources. The SRQ, started in 1996, is a nationwide register capturing 85% to 90% of prevalent patients with RA in Sweden (11). The SRQ contains clinical data registered at inclusion and at regular visits (including information on the RA diagnosis), disease activity and treatment entered by the treating rheumatologist, and certain patient-reported data from the same return visits. Through the unique personal identification number assigned to Swedish residents, the SRQ was linked to the following nationwide registers: the National Patient Register (NPR) (which contains information on visits to specialty care in the Swedish health care system from 2001 or later), the Prescribed Drug Register (PDR) (all dispensed drugs from 2006 or later), the Total Population Register (socio-demographics), the Longitudinal Integration Database for Health Insurance and Labor Market Studies (sick leave and disability pension), and the Multi-Generation Register (data on first-degree relatives). In brief, health care in Sweden is publicly funded and heavily subsidized; this includes also specialized care of chronic diseases. At the visit, the treating physician assigns a main diagnosis code, which covers the main reason for the health care visit. This can be complemented with up to eight contributory diagnosis codes. Our linkage and the registers covered have previously been described in detail (12).

Study population and MTX treatment. Using the SRQ, we identified a cohort of patients with new-onset RA (diagnosed by a rheumatologist at a rheumatology clinic) who were registered in the SRQ between 2006 and 2016 within 1 year of RA symptoms onset and started on MTX DMARD monotherapy at RA diagnosis as the first ever DMARD. We then followed their subsequent prescriptions of antirheumatic therapies through the SRQ and through the PDR (Supplementary Figure A2). Furthermore, individuals with a visit diagnosis of RA or polyarthritis (*International Classification of Diseases, 10th Revision* codes M05, M06, and M13) in the NPR more than 365 days before the start of their MTX treatment, as recorded in the SRQ, were excluded.

Main and suboutcomes. Our main outcome, persistence on MTX monotherapy, was defined as having a treatment record of MTX in the SRQ spanning 365 days after initiation, regardless of corticosteroid use and without having received any additional DMARD of any type during this period. For contextualization, we further defined five suboutcomes: primary inefficacy (discontinuing MTX within a 60-183-day window after initiation with "no effect" or "decline of achieved effect" listed as reason or starting any other DMARD within the same time period); start of another DMARD (180-365 days after MTX initiation); persistence to MTX monotherapy at 12 months, with no oral steroid use at 9 months after MTX

initiation; persistence to MTX monotherapy, with no oral steroid use, at 36 months after initiation; and persistence to MTX monotherapy, with no oral steroid use, at 36 months after initiation among those who were persistent at 1 year.

Covariates. We created four nested covariate data sets (Supplementary Table A1). All data sets contained clinical data from the SRQ and sociodemographic data (detailed in Supplementary A, Supplementary Materials and Methods). Covariate set A, representing a traditional expert opinion-based set of predictors, additionally contained 20 predefined medical history diagnoses (see Supplementary Table A2 for exact definitions). Covariate set B expanded this information by instead using all primary diagnoses and prescriptions from the linkages to the NPR and PDR (up to 10 years before RA diagnosis). Covariate set C contained the same information as covariate set B but split into time intervals (the year before, 1-5 years before, 5-10 years before [only for *International Classification of Diseases* (ICD) codes]). In covariate sets A to C, we included the main diagnosis code for the visit. However, there is a possibility to enter contributory codes, which also summarize other diagnoses of relevance for the health care contact, although they are more extensively used for inpatient visits. For covariate set D, we included not only the main diagnosis but also the contributory codes. Covariate set A consisted of 48 variables; B, 1313; C, 2033; and D, 4162.

Statistics. We split the data into a training data set consisting of a random selection of 90% of the data, which was used for all modeling, and a holdout data set, the remaining 10%, that was saved for evaluation of model performance. We then applied a filtering threshold of 0.5% for all variables. For all data sets and all outcomes, we ran univariate logistic regressions to assess the association with the outcomes (the main outcome and each of the suboutcomes). Multiple testing was accounted for by applying a false discovery rate (FDR) of 10%. These univariate associations were not used as part of any complex manual or ML-based modeling but served to contextualize the variables included in our covariate data sets.

Hypothesis-based modeling. We made two traditional hypothesis-based models applied to the training data. These models used only the covariates identified by subject-matter considerations (from covariate set A) and variable reparameterizations and selections left to the discretion of a trained epidemiologist (TF). After inspecting distributions and associations of individual covariates and the outcome, the epidemiologist made two models. One was based on manually entering and removing individual variables and testing interaction terms and nonlinear terms for continuous variables (the manual model), informed by the Akaike information criterion values, but also included subjective exclusions based on face validity and the modeler's trust in an observed association. The second model was a simple backward selection logistic regression model that started with covariate set A and was revised at the epidemiologist's discretion by recoding some continuous covariates as categorical or with third-degree polynomial terms and allowing first-degree interactions between several variables (the stepwise model).

ML models. We used the caret package in R (13), with a tune length of 3 (ie, the number of values tested for each hyperparameter in the ML method), and the following methods: least absolute shrinkage and selection operator (Lasso) and elastic net regularization using glmnet (14), support vector machine (SVM) with a linear kernel (R package e1071), random forest (randomForest R package) (15), and extreme gradient boosting (R package XGBoost) (16).

We performed and manually inspected calibration and discrimination plots for all ML methods, data sets and outcomes, and AUROC plots (plotROC package) (17). For all models, we used the AUROC for comparison between methods and models. We also, as a baseline comparator, estimated the AUROC using only sex and age as predictors for all outcomes. Fivefold cross-validation (CV) was applied in all ML models, including the ensemble model, when estimating performance on the training data. We used the caret function varImp to estimate the importance of each predictor and plotted this for the top 10 predictors per model and data set.

After the initial modeling, we estimated, plotted, and visually inspected the correlation between the results of the different methods. The results from four of the ML methods, excluding elastic net for technical reasons (see Supplementary A, Supplementary Materials and Methods), were combined into an ensemble model (R package caretEnsemble*)* by using caretStack, a generalized linear model, and fivefold CV model.

As a final step, all resulting models were run on the holdout data, and the AUROC was estimated for all covariate data sets and outcomes.

RESULTS

Persistence to methotrexate at 1 year. In the analysis, 5475 patients were included: 3737 (68%) were women, and 3555 (65%) were rheumatoid factor–positive. The median age was 61 years (interquartile range: 20). Supplementary Figure A1 shows a flowchart for patient inclusion, and Table 1 shows demographics and distributions of all outcomes.

A total of 3835 (70%) patients remained on MTX DMARD monotherapy 1 year after RA diagnosis. This proportion was similar in the training and holdout data sets (Table 1). A total of 1247 (17%) patients had started another DMARD, and 313 (1%) were not on any DMARD at all at 1 year. At three years, 1808 (34%) of the 3669 patients still observed in the SRQ at this time point remained on MTX DMARD monotherapy.

Univariate association and main outcome. Supplementary Tables T1 to T4 describe the univariate association between the selected covariates surviving FDR correction and the main outcome. Age was the most significantly associated

Table 1. Patient characteristics and distribution of the outcomes in the training and the holdout data sets

	Overall	Training data set	Holdout data set
N	5475	4927	548
Persistent at 1 year, n (%)	3834 (70)	3449 (70)	385 (70)
Primary inefficiency, n (%)	734 (13)	664 (14)	70 (13)
Start of another DMARD, n (%)	723 (13)	654 (13)	69 (13)
Persistent to MTX at 1 year and no steroids at 9 months, n (%)	1690 (31)	1515 (31)	175 (32)
Persistence at 3 years and no steroids, n (%)	1142 (31)[a]	1028 (31)[b]	114 (31)[c]
Persistence at 3 years and no steroids, among those persistent at 1 year,[d] n (%)	1118 (44)[d]	1006 (44)[e]	112 (45)f
Median age (IQR), years	61 (20)	61 (20)	61 (20)
Women, n (%)	3737 (68)	3362 (68)	375 (68)
RF-positive, n (%)	3555 (65)	3178 (65)	368 (67)
Median year of diagnosis (IQR)	2010 (4)	2010 (5)	2010 (4)

Abbreviations: DMARD, disease-modifying antirheumatic drug; IQR, interquartile range; MTX, methotrexate; RF, rheumatoid factor.
[a] Total number of individuals in analysis = 3669.
[b] Total number of individuals in analysis = 3303.
[c] Total number of individuals in analysis = 366.
[d] Total number of individuals in analysis = 2515.
[e] Total number of individuals in analysis = 2267.
[f] Total number of individuals in analysis = 248.

variable in all data sets, with younger age associated to lower persistence to MTX monotherapy. For all data sets, clinical variables at diagnosis were most significantly associated with persistence.

Model performances. For age- and sex-only adjusted models, the AUROC was 0.58 (95% confidence interval [CI] 0.53-0.63) in the holdout data (Table 2). The AUROCs for models resulting from the manual model building approach were 0.66 (95% CI 0.65-0.68) and 0.66 (95% CI 0.64-0.68) for the stepwise model.

Results based on the holdout data sets are presented in Table 2. The highest AUROC was obtained with Lasso regression and elastic net (AUROC = 0.67 [95% CI 0.62-0.71]) for covariate set A. It should be noted, however, that all CIs were overlapping, with the exception of SVM, which performed worse with increased complexity of the data set. The full results are presented in Table 2. Calibration, discrimination, and AUROC plots for the Lasso regression in covariate set A are presented in Figure 1 and for the remaining models in Supplementary B, Supplementary Diagnositics B. The top 10 predictors per algorithm and data set are plotted in Supplementary A, Supplementary Figures B. In general, Lasso and elastic net had top predictors from the ICD and Anatomical Therapeutic Chemical (ATC) classification system codes, whereas the other methods mostly had variables from covariate set A as top predictors. Results from the holdout data set were consistent with those from the training data set (shown in Supplementary Table B1). Plots of variable importance are shown in Supplementary B, Supplementary Figures B.

All results, including associations, diagnostics, and top predictors, for the suboutcomes are reported in Supplementary C, Supplementary Results, Supplementary Tables C1-C2, Supplementary Diagnostics C and Supplementary Figures C. Again, the top predictors differed slightly between Lasso and elastic net compared with the rest of the methods

DISCUSSION

The results of our large study using prospectively recorded data from routine clinical practice enriched with data from national

Table 2. AUROC (95% CI) in the holdout data for all methods and covariate sets, in the manual and machine learning models for persistence to MTX monotherapy 1 year after diagnosis in a cohort of patients with RA diagnosed in 2006-2016 starting MTX in monotherapy

Covariate data set	A	B	C	D
Sex/age adjusted only	0.58 (0.53-0.63)	-	-	-
Backward selection	0.66 (0.60-0.71)	-	-	-
Manual modeling	0.65 (0.60-0.70)	-	-	-
Lasso regression	0.67 (0.62-0.71)	0.67 (0.62-0.72)	0.67 (0.62-0.72)	0.66 (0.62-0.71)
Elastic net	0.67 (0.62-0.71)	0.67 (0.62-0.72)	0.67 (0.62-0.72)	0.66 (0.62-0.71)
Random forest	0.62 (0.57-0.67)	0.61 (0.56-0.66)	0.63 (0.58-0.68)	0.61 (0.56-0.66)
SVM	0.65 (0.60-0.70)	0.58 (0.52-0.63)	0.58 (0.53-0.63)	0.53 (0.48-0.58)
XGBoost	0.61 (0.56-0.66)	0.64 (0.59-0.69)	0.63 (0.58-0.68)	0.63 (0.58-0.68)
Ensemble	0.63 (0.58-0.68)	0.65 (0.60-0.70)	0.65 (0.60-0.70)	0.65 (0.60-0.70)

Note. All starting on MTX monotherapy holdout data.
Abbreviations: AUROC, area under the receiver operating characteristic curve; CI, confidence interval; MTX, methotrexate; RA, rheumatoid arthritis; SVM, support vector machine.

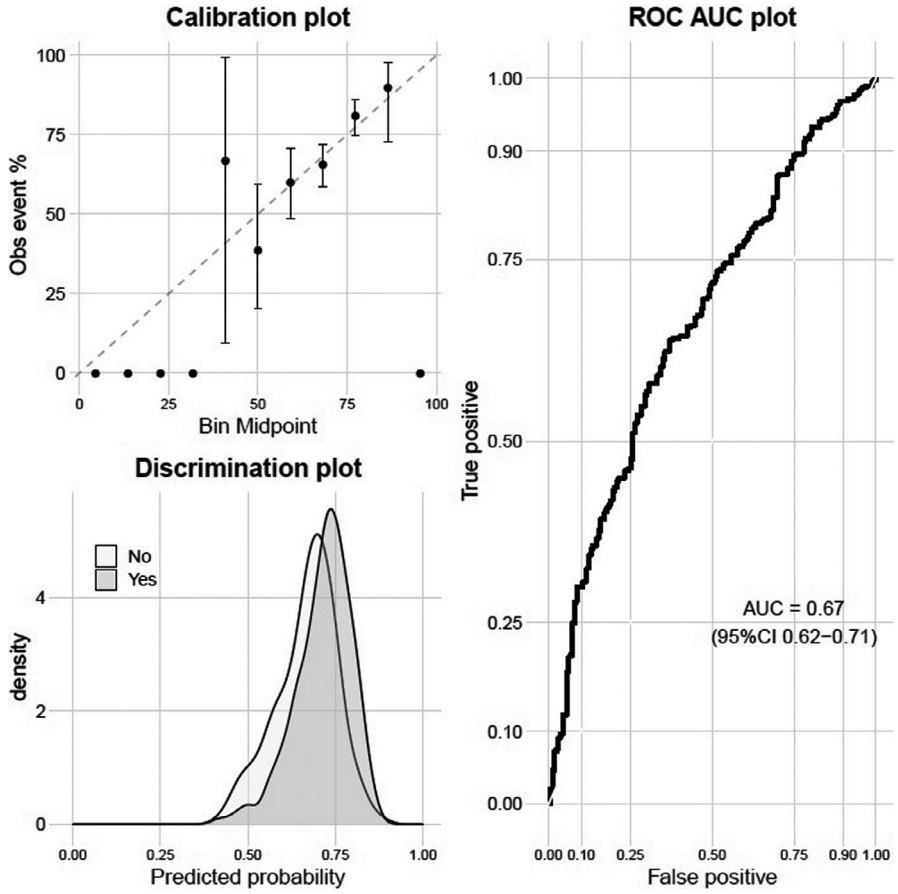

Figure 1. Calibration, discrimination, and area under the receiver operating characteristic curve plots for Lasso regression in covariate set A. AUC, area under the curve; CI, confidence interval; Obs, observation; ROC, receiver operating characteristic.

registers suggest that at 1 year after RA diagnosis, approximately two thirds of all patients who initiated MTX monotherapy remained on this monotherapy. Besides quantifying this clinically interesting proportion, we made a series of important observations: 1) We observed that several factors previously identified as predictors of primary response to MTX (eg, sex, initial RA disease activity, and age) were also associated with persistence to MTX monotherapy at one year. Indeed, of all covariates assessed as individual predictors, clinical variables at diagnosis tended to be most significantly associated with our outcomes. 2) With traditional, hypothesis-based prediction models that made use of preselected variables, we reached moderate predictive performance, as measured by the AUROC. Such manual methods are difficult to scale to thousands of potentially relevant variables. 3) When we assessed the marginal gains of using increasingly more granular information on previous medical history and used five different and more complex ML methods, we found that linear methods, Lasso regression, and elastic net performed the best and also, to a greater extent, had top predictors based on ATC and ICD codes. We also found that for this outcome of interest, the marginal gain over and above traditional and manually assembled hypothesis-based models was small. On the other hand, there was also no loss compared with the manually assembled models, and ML provided robustness in the form of CV as well as potential for a pipeline screening larger data set for potential predictors.

Most previous studies have used clinical baseline characteristics, and most successful results so far have been from studies that have also included genetic data (18). But even through pooling of results by using meta-analyses, no conclusive, continuously replicable results have been found (19). In our study, we added information from several national registers. Although we did not have access to genetic or other molecular data, we instead added information on prior disease history and treatment, which, to our knowledge, has not previously been investigated in the prediction of MTX treatment outcomes. With the increased complexity of our data sets, we were able to increase the granularity of the medical history. However, we could not reach the high AUROCs demonstrated by genetic studies; we instead note that in terms of AUROCs, our results are on a par with those from the study by Plant et al (8), which used clinical variables. Although the model by Plant et al (8) predicted nonresponse at 6 months after initiation of MTX and is thus not directly comparable to our main model of MTX persistence at 12 months, the outcomes are similar to two of our suboutcomes: primary inefficiency and start of another DMARD.

For primary inefficiency (ie, stopping MTX before 6 months for any reason or starting another DMARD within the same time period), our AUROCs were 0.70 (95% CI 0.64-0.77) and 0.61 (95% CI 0.55-0.67) at best, and the most strongly associated covariate for these outcomes was, indeed, clinical presentation at diagnosis.

Regarding associations with individual predictors, our results showed that covariates, such as number of hospitalizations in the years preceding RA diagnosis and number of ICD codes given up to 10 years before diagnosis, were strongly associated with not staying on MTX monotherapy. Taking the association of the clinical variables into account and noticing that a fair amount of the covariates associated with our main outcome are representative of pain and pain conditions, it seems like these patients could present with a higher load of disease at diagnosis, but a further investigation of these relationships was beyond the scope of our study.

The promise and potential pitfalls of big data approaches are linked to the same underlying idea: to consider a large number of potentially relevant variables for prediction of a phenotype. Best practices for ML (eg, CV and holdout data sets) are key to avoiding the reporting of models that might fail to validate in future studies (20). Here, we included up to 4000 additional variables in the ML models compared with the manual modeling approach. For the main outcome, we did not detect any combination of variables that substantially improved global predictive power. For discussion and to illustrate the discriminative ability of our automatically trained model, we plotted back-to-back discrimination plots for two specific results in Figure 2. Here we show our main outcome in comparison to no model at all (horizontal line at 0.5) and a model based on age and sex. Clearly, the discriminative ability improves with the complex model but only slightly so. A much stronger picture is obtained when considering, for example, the suboutcome MTX persistence, with no oral steroid use at 9 months after MTX initiation. In this case, the discriminative ability is substantial. In the latter situation, the predictive gain is driven by steroid use at time of diagnosis (Supplementary Tables T13-T16). Although not a clinically actionable finding, it serves to illustrate the power of finding contrasts of high productivity with minimal manual process. This bodes well for exploration of many additional contrasts in the future.

Strengths of this project include a large, well-characterized cohort with detailed and reliable information on previous disease history and a unique possibility to blend clinical and administrative variables and add increasingly more information from the data sources.

Weaknesses include our approach that aimed at global rather than individual prediction, so we cannot exclude the existence of strong predictors (important for few individuals) hiding inside the data and not discovered by our modeling because of their low prevalence. The application of ML in this article was further naïve in the sense that the methods were treated as a black box, and no modeling of the covariates was done beforehand. It is possible that prediction accuracies could have improved if the latent structure of covariates had been modeled before inclusion in the model. However, such modeling was not in the scope of

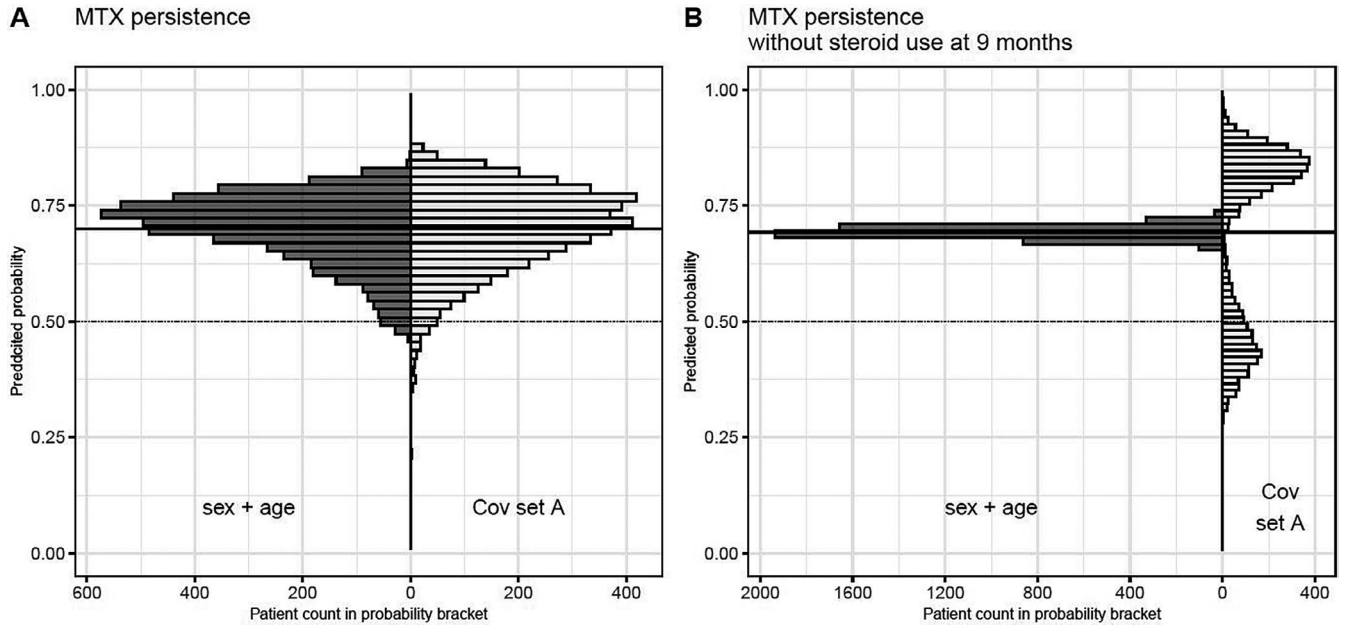

Figure 2. Two back-to-back class probability histograms to demonstrate the impact of modeling of different covariate sets. Each histogram shows the number of patients in a predicted class probability bracket. **A**, Models based on the main contrast, MTX persistence. **B**, Models for the suboutcome, MTX persistence, with no steroid use at 9 months. The left histogram in each panel describes a model based on age and sex alone. The right histogram in each panel depicts a model based on covariate set A (Cov set A). The horizontal line shows prevalence of phenotype. Whereas gains are marginal but notable for the main contrast, prediction becomes markedly better on the basis of the inclusion of covariates beyond age and sex for the suboutcome. MTX, methotrexate.

this project because our intent was to investigate whether prediction accuracy improved simply by using larger volumes of more granular information. Thus, to gauge the clinical usefulness of the models, model calibration, instead of AUROC, could also be of interest. We further only used a linear kernel for the SVM because other kernels consistently failed during training.

In conclusion, despite an increasing number of treatment options in RA, approximately two thirds of all patients who start treatment with MTX monotherapy as their first ever DMARD remain on this drug 1 year later. Although additional big data on comorbidities and medications contain information strongly related to MTX persistence, the contribution of such information to prediction models lead to a marginal gain in the average prediction. At the same time, more complex ML-based approaches performed no worse than manual modeling and can be used for rapid exploration of additional contrasts of interest. Future attempts to model drug persistence in RA should strive at combining not only clinical, comorbidity, and socioeconomic data but also biological and molecular data.

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 published. Dr. Westerlind 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. Westerlind, Maciejewski, Frisell, Jelinsky, Ziemek, Askling.

Acquisition of data. Frisell, Askling.

Analysis and interpretation of data. Westerlind, Maciejewski, Frisell, Jelinsky, Ziemek, Askling.

REFERENCES

1. Farragher TM, Lunt M, Fu B, Bunn D, Symmons DP. Early treatment with, and time receiving, first disease-modifying antirheumatic drug predicts long-term function in patients with inflammatory polyarthritis. Ann Rheum Dis 2010;69:689–95.
2. Scire CA, Verstappen SM, Mirjafari H, Bunn DK, Lunt M, Montecucco C, et al. Reduction of long-term disability in inflammatory polyarthritis by early and persistent suppression of joint inflammation: results from the Norfolk Arthritis Register. Arthritis Care Res (Hoboken) 2011;63:945–52.
3. Hider SL, Silman AJ, Thomson W, Lunt M, Bunn D, Symmons DP. Can clinical factors at presentation be used to predict outcome of treatment with methotrexate in patients with early inflammatory polyarthritis? Ann Rheum Dis 2009;68:57–62.
4. Saevarsdottir S, Wallin H, Seddighzadeh M, Ernestam S, Geborek P, Petersson IF, et al. Predictors of response to methotrexate in early DMARD naive rheumatoid arthritis: results from the initial open-label phase of the SWEFOT trial. Ann Rheum Dis 2011;70:469–75.
5. De Rotte MC, Pluijm SM, de Jong PH, Bulatovic Calasan M, Wulffraat NM, Weel AE, et al. Development and validation of a prognostic multivariable model to predict insufficient clinical response to methotrexate in rheumatoid arthritis. PLoS One 2018;13:e0208534.
6. Wessels JA, van der Kooij SM, le Cessie S, Kievit W, Barerra P, Allaart CF, et al. A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. Arthritis Rheum 2007;56:1765–75.
7. Plant D, Wilson AG, Barton A. Genetic and epigenetic predictors of responsiveness to treatment in RA. Nat Rev Rheumatol 2014;10:329–37.
8. Plant D, Maciejewski M, Smith S, Nair N, Maximising Therapeutic Utility in Rheumatoid Arthritis Consortium, the RAMS Study Group, Hyrich K, et al. Profiling of gene expression biomarkers as a classifier of methotrexate nonresponse in patients with rheumatoid arthritis. Arthritis Rheumatol 2019;71:678–84.
9. Carini C, Hunter E, Scottish Early Rheumatoid Arthritis Inception Cohort Investigators, Ramadass AS, Green J, Akoulitchev A, et al. Chromosome conformation signatures define predictive markers of inadequate response to methotrexate in early rheumatoid arthritis. J Transl Med 2018;16:18.
10. Romao VC, Canhao H, Fonseca JE. Old drugs, old problems: where do we stand in prediction of rheumatoid arthritis responsiveness to methotrexate and other synthetic DMARDs? BMC Med 2013;11:17.
11. SRQ. Visualization and analysis platform (VAP). 2018. URL: https://openvap.compos.se/srq/7_1/
12. Eriksson JK, Askling J, Arkema EV. The Swedish Rheumatology Quality Register: optimisation of rheumatic disease assessments using register-enriched data. Clin Exp Rheumatol 2014;32 Suppl 85:S147–9.
13. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26.
14. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33:1–22.
15. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18–22.
16. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2016; San Francisco. New York: Association for Computing Machinery; 2016. p. 785–94.
17. Sachs MC. plotROC: a tool for plotting ROC curves. J Stat Softw 2017;79:2.
18. Guan Y, Zhang H, Quang D, Wang Z, Parker SC, 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;71:1987–96.
19. Ling S, Bluett J, Barton A. Prediction of response to methotrexate in rheumatoid arthritis. Expert Rev Clin Immunol 2018;14:419–29.
20. Frohlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med 2018;16:150.