Seminars in Arthritis and Rheumatism 56 (2022) 152050

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis

Cynthia Yang [a,*,1], Ross D. Williams [a,1], Joel N. Swerdel [b], João Rafael Almeida [c], Emily S. Brouwer [b], Edward Burn [d,e], Loreto Carmona [f], Katerina Chatzidionysiou [g], Talita Duarte-Salles [e], Walid Fakhouri [h], Antje Hottgenroth [i], Meghna Jani [j], Raivo Kolde [k], Jan A. Kors [a], Lembe Kullamaa [l,m,n], Jennifer Lane [d], Karine Marinier [o], Alexander Michel [p], Henry Morgan Stewart [q], Albert Prats-Uribe [d], Sulev Reisberg [k,r,s], Anthony G. Sena [a,b], Carmen O. Torre [q], Katia Verhamme [a], David Vizcaya [t], James Weaver [b,u], Patrick Ryan [b,u], Daniel Prieto-Alhambra [d], Peter R. Rijnbeek [a]

[a] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
[b] Janssen Research and Development, Titusville, NJ, United States
[c] DETI/IEETA, University of Aveiro, Aveiro, Portugal
[d] Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
[e] Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
[f] Instituto de Salud Musculoesquelética, Madrid, Spain
[g] Department of Medicine, Solna, Rheumatology Unit, Karolinska Institute, Stockholm, Sweden
[h] Eli Lilly and Company, Windlesham, Surrey, United Kingdom
[i] Lilly Deutschland GmbH, Bad Homburg, Germany
[j] Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, United Kingdom
[k] Institute of Computer Science, University of Tartu, Tartu, Estonia
[l] Department of Epidemiology and Biostatistics, National Institute for Health Development, Tallinn, Estonia
[m] Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia
[n] European Patients' Forum, Brussels, Belgium
[o] Servier, Suresnes, France
[p] Epidemiology, Bayer Basel, Basel, Switzerland
[q] Real-World Solutions, IQVIA, Brighton, United Kingdom
[r] STACC, Tartu, Estonia
[s] Quretec, Tartu, Estonia
[t] Bayer Pharmaceuticals, Barcelona, Spain
[u] Observational Health Data Sciences and Informatics, New York, NY, United States

ARTICLE INFO

Keywords:
Rheumatoid arthritis
Methotrexate
Cardiovascular diseases
Infections
Prediction models

ABSTRACT

Background: Identification of rheumatoid arthritis (RA) patients at high risk of adverse health outcomes remains a major challenge. We aimed to develop and validate prediction models for a variety of adverse health outcomes in RA patients initiating first-line methotrexate (MTX) monotherapy.

Methods: Data from 15 claims and electronic health record databases across 9 countries were used. Models were developed and internally validated on Optum® De-identified Clinformatics® Data Mart Database using L1-regularized logistic regression to estimate the risk of adverse health outcomes within 3 months (leukopenia, pancytopenia, infection), 2 years (myocardial infarction (MI) and stroke), and 5 years (cancers [colorectal, breast, uterine] after treatment initiation. Candidate predictors included demographic variables and past medical history. Models were externally validated on all other databases. Performance was assessed using the area under the receiver operator characteristic curve (AUC) and calibration plots.

* Corresponding author.
E-mail address: c.yang@erasmusmc.nl (C. Yang).
[1] Cynthia Yang and Ross D. Williams contributed equally to this paper.

https://doi.org/10.1016/j.semarthrit.2022.152050

Available online 15 June 2022

Findings: Models were developed and internally validated on 21,547 RA patients and externally validated on 131,928 RA patients. Models for serious infection (AUC: internal 0.74, external ranging from 0.62 to 0.83), MI (AUC: internal 0.76, external ranging from 0.56 to 0.82), and stroke (AUC: internal 0.77, external ranging from 0.63 to 0.95), showed good discrimination and adequate calibration. Models for the other outcomes showed modest internal discrimination (AUC < 0.65) and were not externally validated.
Interpretation: We developed and validated prediction models for a variety of adverse health outcomes in RA patients initiating first-line MTX monotherapy. Final models for serious infection, MI, and stroke demonstrated good performance across multiple databases and can be studied for clinical use.
Funding: This activity under the European Health Data & Evidence Network (EHDEN) has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 806968. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Introduction

Compared to the general population, patients with rheumatoid arthritis (RA) have an increased risk of treatment-related adverse events, such as cytopenia and infection, and comorbidities, such as cardiovascular disease (CVD) and cancer [1–3]. Although the management and prognosis of RA has improved in recent decades, identification of RA patients at high risk of adverse health outcomes remains a major challenge [4,5].

The European Alliance of Associations for Rheumatology (EULAR) and the American College of Rheumatology (ACR) recommend initiating methotrexate (MTX) monotherapy (with glucocorticoids) as soon as possible after the diagnosis of RA [6,7], making this the most common treatment for RA worldwide. MTX treatment implies screening or monitoring efficacy and side-effects, as with most disease modifying antirheumatic drugs (DMARDs). Using prediction models to evaluate patient-level risks in RA patients initiating first-line MTX monotherapy could allow clinicians to target those at high risk of adverse health outcomes for increased screening or monitoring throughout the course of treatment.

Few prediction models have been developed for adverse health outcomes in RA patients, with those that have been developed focusing on the risk of either CVDs or serious infection [8–15]. Challenges in the development of RA-specific prediction models have previously been highlighted [10,16]. For example, while existing CVD models estimate 10-year risks, a shorter period may be more appropriate, since most RA patients will change treatments several times during a 10-year period [10]. Additionally, a larger cohort of RA patients would allow for the development of RA-specific prediction models using a larger number of candidate predictors [8]. Finally, most existing models have not been subjected to extensive external validation, which is necessary to understand a model's prediction performance [17]. In this study, we aimed to develop and externally validate prediction models for a variety of adverse health outcomes in RA patients initiating first-line MTX monotherapy, using 15 large-scale claims and electronic health record (EHR) databases across 9 countries and 4 continents.

Materials and methods

This study was conducted within the European Health Data & Evidence Network (EHDEN) project and involved a multidisciplinary team of rheumatologists, clinicians, epidemiologists, data custodians, and data scientists. We developed and validated prediction models using the Patient-Level Prediction framework from the Observational Health Data Sciences and Informatics (OHDSI) initiative [18]. This framework allows for standardized development and extensive validation of prediction models using observational health databases and can be applied to datasets that are mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) [19–21]. The OMOP CDM was developed to transform source data into a common format and enables analytical source code to be shared among researchers. We followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines for reporting [22].

Data sources

We used 15 claims and EHR databases with data mapped to the OMOP CDM from 6 European countries (Spain, Estonia, Netherlands, Germany, France, and the United Kingdom (UK)), the United States of America (USA), Australia, and Japan. The databases are listed in Table 1. Data from the Optum® De-Identified Clinformatics® Data Mart Database, a USA claims database, were used for model development and internal validation. Data from the 14 other databases were used for external validation. Each site obtained institutional review board approval for the study or used de-identified data. Therefore, informed consent was not necessary at any site. Descriptions of the databases are also provided in Table A.1 in Appendix A.

Study population

Adult RA patients (aged 18 years and over) were included in the study population if they had at least 365 days of observation in the database prior to the first drug utilization record of MTX (the index event) and met all of the following inclusion criteria: (1) a diagnosis of RA within 5 years prior to or on index, (2) no drug utilization record of any DMARD any time prior to index, (3) no drug utilization record of any other DMARD on or within 7 days after index, (4) no record indicating any cancer any time prior to or on index, and (5) no record indicating any other inflammatory arthritis (psoriatic arthritis, ankylosing spondylitis, reactive arthritis, any axial spondyloarthropathy) any time prior to or on index.

Detailed definitions of these inclusion criteria, including code lists, are available at http://atlas-demo.ohdsi.org/#/cohortdefinition/1773112.

Outcomes

We investigated outcomes for which RA patients have increased risks compared to the general population and for which RA patients identified at high risk could be targeted for increased screening or monitoring throughout the course of treatment. The first event (binary) of each of the following adverse health outcomes within a period after initiating first-line MTX monotherapy (the index event) was considered: (1) leukopenia, pancytopenia, and infection (serious, opportunistic, all) recorded from 1 day up to 90 days after index, (2) myocardial infarction (MI) and stroke recorded from 1 day up to 2 years after index, (3) cancer (colorectal, breast, uterine) recorded from 1 year up to 5 years after index.

Detailed definitions of these outcomes, including code lists, are available at: https://github.com/ohdsi-studies/EhdenRaPrediction/tree/master/inst/cohorts.

For all outcomes, patients who had any record of the specific outcome within 90 days prior to or at initiation of MTX monotherapy were excluded from the study population. For cancer outcomes, patients who were lost to follow-up within one year of treatment initiation were excluded. For all other outcomes, patients who were lost to follow-up within one day of treatment initiation were excluded. Outcomes for which the final study population contained less than 25 RA patients with an outcome event were omitted from further analysis.

Candidate predictors

Candidate predictors were extracted from data routinely recorded in the database. This included binary indicators of 5-year age groups (i.e., 20−24, 25−29, etc.) and sex, as well as a large set of binary indicators of recorded OMOP CDM concepts for health conditions and drug groups [20]. For health conditions, we considered all data prior to index. For drug groups, we separately considered data from the 30 days prior to index and data from the 365 days prior to index. Finally, three established risk scores (CHA2DS2-VASc, Diabetes Complications Severity Index (DCSI), Charlson Comorbidity Index (CCI) (Romano adaptation)) were calculated using all data prior to index [23–26]. No clinical measurements or sporadically recorded variables were included as candidate predictors to maximize transportability (i.e., ability to apply across databases) of the developed prediction models.

Handling of missing data

In the observational data used in this study, if a candidate predictor was not recorded in a patient's history, we assumed that the candidate predictor was not observed for this patient. Age group and sex are required by the OMOP CDM and were always recorded. For our analyses, if a health condition or drug group was not recorded in a patient's history, we assumed that the health condition or drug group was absent.

Statistical analysis methods

We used logistic regression with predictor selection through L1-regularization [27]. For each outcome, two L1-regularized logistic regression models were developed: (1) one model using all candidate predictors for a data-driven approach of predictor selection, and (2) one model using only age groups and sex as candidate predictors to provide a benchmark.

A random subset of 75% of the patients was used as a training set and the remaining subset of 25% of the patients was used as a test set. First, 3-fold cross-validation was performed on the training set to optimize the regularization parameter, after which the test set was used for internal validation. Discrimination was assessed numerically using the area under the receiver operator characteristic curve (AUC) with a 95% confidence interval (CI). Calibration was assessed graphically by plotting the predicted risks against the observed risks.

To examine model performance across multiple databases, we externally validated the models. Only models with an AUC of at least 0.7 on internal validation were considered good enough to warrant external validation. External validation of each model was limited to those databases within which the corresponding outcome events could be identified. Discrimination and calibration were assessed in the same way as on internal validation.

Software packages containing the analytical source code that was used to develop the models and to externally validate the developed models on databases with data mapped to the OMOP CDM are available at https://github.com/ohdsi-studies/EhdenRaPrediction.

Results

Study population

In the development database (Optum Claims), 21,547 RA patients met the inclusion criteria. For breast cancer and uterine cancer, we only included female patients, resulting in 15,311 RA patients who met the

Table 1
Databases included in the study with data mapped to the OMOP CDM.

Database full name	Database short name	Country	Data type	Patient type	Population size	Data range
Estonian Health Information System	Estonia	Estonia	EHR	All inpatient and outpatient discharge summaries, general population	1.4m	2012-2016
IBM MarketScan® Commercial Database	CCAE	USA	Claims	Privately insured	151m	2000-2020
IBM MarketScan® Medicare Supplemental Database	MDCR	USA	Claims	Retiree supplemental	10m	2000-2020
IBM MarketScan® Multi-State Medicaid Database	MDCD	USA	Claims	Medicaid	30m	2006-2020
Integrated Primary Care Information	IPCI	Netherlands	EHR	Primary care	2.6m	1996-2020
IQVIA Australia EMR	IQVIA Australia	Australia	EHR	Outpatient / General population	6m	2006-2020
IQVIA Disease Analyser Germany EMR	IQVIA Germany	Germany	EHR	Outpatient / General population, public and private insurance	37m	1992-2020
IQVIA Hospital US Charge Master	IQVIA US Hospital	USA	EHR	Inpatient and outpatient hospital encounters, including Emergency Room visits / General population	86m	2007-2020
IQVIA LPD France	IQVIA LPD France	France	EHR	Outpatient / General population	7.8m	1994-2020
IQVIA UK THIN IMRD EMR	IQVIA THIN	UK	EHR	General population / Primary care records with hospitalization / referral information	15m	1989-2020
IQVIA US Ambulatory EMR	IQVIA US Ambulatory	USA	EHR	Outpatient / General population	49m	2006-2020
Japan Medical Data Center	JMDC	Japan	Claims	Society-Managed Health Insurance	12m	2005-2020
Optum® De-Identified Clinformatics® Data Mart Database	Optum Claims	USA	Claims	Privately insured	87m	2001-2020
Optum® De-identified Electronic Health Record Dataset	Optum EHR	USA	EHR	Privately insured	100m	2006-2020
The Information System for Research in Primary Care	SIDIAP-H	Spain	EHR	Primary care linked (partially) to inpatient data	5.8m	2006-2020

inclusion criteria. Table 2 shows the number of RA patients and the number of RA patients with an outcome event in the final study population for each adverse health outcome. An attrition flowchart explaining how we arrived at the number of patients in the final study population for each outcome is available in Fig. B.1 in Appendix B.

Table C.1. in Appendix C shows demographics and baseline characteristics of the final study population for each outcome, based on all data prior to or at initiation of MTX monotherapy. Patients with an outcome event on average had more comorbidities at treatment initiation.

Model specification

A total of 12,724 candidate predictors were extracted from data routinely recorded in the database, of which 18 were binary indicators of 5-year age groups and sex.

The number of predictors in each final model is specified in Table 3. Full lists of candidate predictors and detailed specifications of all final models are available in an interactive R Shiny web application at https://data.ohdsi.org/EhdenRaPrediction/. The candidate predictors can be explored interactively in the 'Model Table' under 'Model'; an overview of the predictors in the final model can also be exported from this tab using 'Download Model'.

Model performance

Model discrimination on internal validation is presented in Table 3, ordered by highest AUC for the data-driven approach. For leukopenia, the AUC was below 0.6, indicating modest discrimination. For pancytopenia, no predictors were identified for both the data-driven approach and the benchmark, and we were therefore unable to develop any prediction model for this outcome. For opportunistic infection, and all infection, the AUCs were 0.6 or lower. For serious infection, MI, and stroke, the data-driven approach resulted in AUCs on internal validation of 0.74 (0.68–0.80), 0.76 (0.72–0.81), and 0.77 (0.73–0.81), respectively, indicating good discrimination. For colorectal cancer and breast cancer, the AUCs were below 0.65. Finally, for uterine cancer, the final study population contained less than 25 RA patients with an outcome event. Therefore, this outcome was omitted from further analysis.

The calibration plots for serious infection, MI, and stroke (Fig. D.1–D.6 in Appendix D) indicated adequate calibration besides good discrimination. We externally validated the models for these three outcomes across the 14 other databases. The AUCs are presented in Table 4, ordered by highest AUC for the data-driven approach. Overall, the models demonstrated good performance across multiple databases. Several 95% CIs were wide due to limited statistical power, making those results difficult to interpret. We therefore focused on the databases within which at least 100 RA patients with an outcome event were identified, which is a recommended minimum for external validation [28]. In these databases, the data-driven approach consistently outperformed the benchmark, with AUCs ranging from 0.62 to 0.76 for serious infection, from 0.65 to 0.75 for MI, and from 0.63 to 0.79 for stroke. The corresponding calibration plots from the data-driven approach are presented in Fig. E.1–E.16 in Appendix E. These plots showed adequate calibration for all three outcomes. As expected, the models tend to underestimate or overestimate risk in databases with a higher or lower incidence, respectively [29]. For example, the model for serious infection underestimated the risk in the IQVIA US Hospital database, where the outcome incidence was more than tenfold higher than the outcome incidence in the development database. To account for this, the models can be recalibrated for use in these databases.

The final data-driven models for serious infection, MI, and stroke, including intercept, coefficients, and OMOP CDM concept IDs, can be found in Tables F.1–F.3 in Appendix F. Several age groups corresponding to older age were selected as predictors. Sex was selected as predictor in the model for stroke. The CHADS2-VASc score was selected as predictor in both the model for MI and the model for stroke. Furthermore, within each model, a large set of binary indicators of conditions and drugs was selected as predictors.

Table 2
Final study population in the Optum Claims database.

Outcome	Number of RA patients	Number of RA patients with outcome event (%)
Leukopenia	21,452	85 (0.4)
Pancytopenia	21,496	30 (0.1)
Serious infection	21,276	316 (1.5)
Opportunistic infection	21,404	161 (0.8)
All infection	19,163	1,957 (10.2)
Myocardial infarction	21,463	417 (1.9)
Stroke	21,425	527 (2.5)
Colorectal cancer	15,584	53 (0.3)
Breast cancer	11,072	100 (0.9)
Uterine cancer	11,104	18 (0.2)

Table 3
Internal discrimination of the models developed on the Optum Claims database.

Outcome	Data-driven approach (age groups, sex, conditions, drugs)		Benchmark (age groups, sex)	
	Number of predictors	AUC (95% CI)	Number of predictors	AUC (95% CI)
Stroke	90	0.77 (0.73–0.81)	16	0.74 (0.70–0.78)
Myocardial infarction	64	0.76 (0.72–0.81)	16	0.72 (0.68–0.76)
Serious infection	87	0.74 (0.68–0.80)	13	0.68 (0.62–0.74)
Opportunistic infection	12	0.60 (0.51–0.68)	1	0.49 (0.42–0.57)
All infection	62	0.59 (0.57–0.62)	6	0.53 (0.50–0.56)
Colorectal cancer	1	0.55 (0.41–0.69)	2	0.64 (0.48–0.79)
Leukopenia	10	0.50 (0.36–0.64)	NA	NA
Breast cancer	NA	NA	4	0.52 (0.42–0.61)

Discussion

In this study, we developed and validated prediction models for a variety of adverse health outcomes in RA patients initiating first-line MTX monotherapy. For serious infection, MI, and stroke, the models demonstrated good performance across multiple databases. Internal validation of these models resulted in AUCs of greater than 0.70 and adequate calibration. External validation of these models resulted in good discrimination, where the data-driven approach of predictor selection (age groups, sex, conditions, drugs) consistently outperformed the benchmark (age groups, sex). This shows that conditions and drugs extracted from routinely recorded data have added value in identifying patients at high risk of serious infection, MI, and stroke. The models showed adequate calibration as well, although for some databases, the models may benefit from recalibration.

For uterine cancer, we were not able to develop models using our data. For this outcome, more data are required. For leukopenia, pancytopenia, opportunistic infection, all infection, colorectal cancer, and breast cancer, we did not externally validate the developed models since they did not discriminate well (AUC < 0.65) on internal validation.

We developed our models using large-scale claims and EHR databases that contain routinely recorded patient information. We chose for a data-driven approach and considered all conditions and drugs in a patient's history as candidate predictors. It is interesting to note that this large set of conditions and drugs in a patient's history also included comedication with nonsteroidal anti-inflammatory drugs and

Table 4
External discrimination of the models for serious infection, myocardial infarction, and stroke.

Outcome	Database	Number of RA patients	Number of RA patients with outcome event (%)	Data-driven approach (age groups, sex, conditions, drugs) - AUC (95% CI)	Benchmark (age groups, sex) - AUC (95% CI)
Serious infection	Estonia	1,441	8 (0.6)	0.83 (0.64–1.00)	0.76 (0.66–0.86)
	SIDIAP-H	2,051	5 (0.2)	0.78 (0.65–0.92)	0.76 (0.58–0.94)
	Optum EHR	29,980	308 (1.0)	0.76 (0.73–0.79)	0.63 (0.60–0.66)
	JMDC	4,871	15 (0.3)	0.72 (0.62–0.82)	0.63 (0.51–0.75)
	MDCR	6,662	154 (2.3)	0.67 (0.62–0.71)	0.59 (0.55–0.64)
	CCAE	29,303	223 (0.8)	0.65 (0.61–0.68)	0.54 (0.51–0.58)
	MDCD	3,793	123 (3.2)	0.63 (0.58–0.68)	0.50 (0.44–0.55)
	IQVIA US Hospital	3,871	776 (20.0)	0.62 (0.59–0.64)	0.56 (0.54–0.58)
Myocardial infarction	IPCI	1,458	24 (1.6)	0.82 (0.72–0.92)	0.78 (0.72–0.84)
	Optum EHR	30,568	466 (1.5)	0.75 (0.73–0.77)	0.71 (0.69–0.73)
	IQVIA LPD France	2,652	5 (0.2)	0.74 (0.60–0.87)	0.50 (0.26–0.75)
	SIDIAP-H	2,057	17 (0.8)	0.72 (0.61–0.84)	0.65 (0.54–0.77)
	IQVIA US Ambulatory	28,129	114 (0.4)	0.72 (0.67–0.77)	0.66 (0.62–0.70)
	MDCD	3,876	88 (2.3)	0.69 (0.63–0.75)	0.61 (0.55–0.67)
	CCAE	29,509	186 (0.6)	0.68 (0.64–0.73)	0.65 (0.61–0.68)
	MDCR	6,739	218 (3.2)	0.67 (0.64–0.71)	0.57 (0.53–0.61)
	IQVIA Germany	8,046	39 (0.5)	0.67 (0.57–0.77)	0.51 (0.41–0.61)
	IQVIA US Hospital	4,331	190 (4.4)	0.65 (0.61–0.69)	0.60 (0.56–0.63)
	Estonia	1,455	18 (1.2)	0.62 (0.48–0.77)	0.71 (0.61–0.81)
	JMDC	4,899	12 (0.2)	0.59 (0.44–0.75)	0.59 (0.40–0.77)
	IQVIA THIN	7,850	51 (0.6)	0.59 (0.50–0.67)	0.64 (0.58–0.70)
	IQVIA Australia	359	7 (1.9)	0.56 (0.29–0.84)	0.62 (0.45–0.80)
Stroke	IPCI	1,467	6 (0.4)	0.95 (0.90–0.99)	0.82 (0.65–0.98)
	Optum EHR	30,506	517 (1.7)	0.79 (0.77–0.81)	0.73 (0.71–0.76)
	IQVIA Germany	8,052	41 (0.5)	0.78 (0.70–0.86)	0.65 (0.57–0.72)
	MDCD	3,864	129 (3.3)	0.78 (0.74–0.82)	0.69 (0.64–0.73)
	Estonia	1,455	23 (1.6)	0.75 (0.67–0.83)	0.72 (0.64–0.81)
	JMDC	4,899	32 (0.7)	0.75 (0.65–0.84)	0.64 (0.55–0.74)
	SIDIAP-H	2,057	23 (1.1)	0.72 (0.63–0.81)	0.71 (0.62–0.79)
	IQVIA US Ambulatory	28,163	115 (0.4)	0.71 (0.67–0.76)	0.68 (0.64–0.72)
	CCAE	29,509	259 (0.9)	0.70 (0.67–0.74)	0.64 (0.61–0.67)
	MDCR	6,734	304 (4.5)	0.67 (0.64–0.70)	0.57 (0.54–0.61)
	IQVIA THIN	7,852	23 (0.3)	0.66 (0.56–0.76)	0.65 (0.55–0.74)
	IQVIA US Hospital	4,306	208 (4.8)	0.63 (0.59–0.67)	0.62 (0.59–0.66)

glucocorticoids, which may capture aspects of RA-specific variables such as disease activity indicators that were not explicitly considered as candidate predictors. It is important to note that our study was focused on prediction and not on evaluating individual predictor associations; it may be misleading to highlight individual predictors as having predictive value by themselves, since this can lead to causal interpretation [30]. It could be interesting to investigate whether explicitly considering certain RA-specific variables as candidate predictors would improve prediction performance. A potential limitation of our study is that we were not able to investigate this using our data.

A potential limitation of our study is that routinely recorded data can be misclassified. If a candidate predictor or an adverse health outcome was not recorded in a patient's history, we assumed but could not be certain that there had not been such an event. Variation in prediction performance across databases may reflect differences in how data are captured. Another potential limitation of our study is that our models were developed for RA patients initiating first-line MTX monotherapy and are therefore only intended for this target population. However, since MTX is the current anchor drug in RA, the developed models are applicable to a large group of RA patients initiating first-line treatment. Furthermore, although it is possible that the models may be applicable to RA patients initiating other treatments, this has not been investigated in our research and would require further validation of the models in those target populations. Finally, a potential limitation of our study is that we did not perform any sensitivity analyses on the inclusion criteria used to obtain our study population. For example, only patients on MTX with no drug utilization record of any other DMARD on or within 7 days after index were included. The 7 days after index offset was chosen to avoid any other DMARDs that occurred at the index date but were entered late in the record. Although less likely, it is possible that a second DMARD that occurred at the index date was entered in the record with a delay longer than 7 days after index.

Our study also has several strengths. To the best of our knowledge, this study is the largest cohort study to date on predicting a variety of adverse health outcomes in RA patients. By developing models using data mapped to the OMOP CDM, we were able to include 14 databases for external validation of the models. The scale of the data allowed us to investigate a variety of adverse health outcomes and validate the models across multiple international databases. Even though more data are still needed for some of the outcomes, our study shows the feasibility of this approach. Additionally, we have provided software packages that contain the analytical source code used to develop the models and to externally validate our developed models on databases with data mapped to the OMOP CDM.

The models developed and internally validated for serious infection, MI, and stroke using a large-scale USA claims database (Optum Claims) showed good performance across multiple claims and EHR databases. We do not believe we have sufficient evidence to recommend the use of the models in clinical practice, but research could be conducted to prove their value as implemented in administrative or EHR software to generate automatic reminders for individuals at high risk of these outcomes. The models are based entirely on routinely recorded data, minimizing the burden to the clinician. However, regulatory approvals would be required before this can be considered, which is beyond the

scope of our current work. The models had particularly good external validation performance on the USA databases and appeared to perform well (with wider confidence intervals) on the international databases too. RA patients identified at high risk of serious infection, MI, or stroke could be targeted for increased screening or monitoring throughout the course of treatment, complementary to current screening or monitoring strategies. In this way, the models may enable clinicians to provide better personalized care to RA patients initiating first-line MTX monotherapy.

Funding

This activity under the European Health Data & Evidence Network (EHDEN) has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 806968. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The sponsor of the study did not have any involvement in the writing of the manuscript or the decision to submit it for publication. The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). DPA is funded by a National Institute for Health Research Senior Research Fellowship. TDS is funded by the Department of Health of the Generalitat de Catalunya under the Strategic Plan for Research and Innovation in Health (PERIS; SLT002/16/00308). RK was supported by the European Social Fund via the IT Academy programme. SR was funded by Estonian Research Council grants PRG1095 and RITA1/02-96-11. The views expressed in this publication are those of the authors and not those of their respective institutions. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit it for publication.

Data sharing statement

The data underlying this article are available at https://data.ohdsi.org/EhdenRaPrediction/.

Declaration of Competing Interest

CY, RDW, JRA, EB, TDS, MJ, RK, JAK, LK, APU, SR, HMS, and COT report no competing interests related to this work. JNS, AGS, JW, and PR are employees of Janssen Research & Development, a pharmaceutical company of Johnson & Johnson, and shareholders of Johnson & Johnson. At the time the study was conducted, ESB was an employee of Janssen Research & Development, a pharmaceutical company of Johnson & Johnson, shareholder of Johnson & Johnson, and shareholder of Takeda Pharmaceuticals. LC reports her institute has been hired for methodological consultancy by AbbVie Spain, S.L.U., Astellas Pharma, SA, Bristol-Myers Squibb, S.A.U. (BMS), Daiichi-Sankyo España, S.A., Dentsply Sirona Iberia, S.A.U., Eisai Farmacéutica, SA, Fresenius Kabi España, S. A. U., Laboratorios Gebro Pharma, SA, Lilly, S.A., Merck Sharp & Dohme España, S.A., Novartis Farmaceutica, SA, Pfizer, S.L.U., Roche Farma, S.A, Sanofi Aventis, UCB Pharma, S.A., outside the submitted work. KC reports consultancy fees from Eli Lilly, AbbVie, and Pfizer, outside the submitted work. WF is an employee and shareholder of Eli Lilly. AHO is an employee of Lilly Deutschland GmbH. JL reports grants from Versus Arthritis, grants from Medical Research Council, outside the submitted work. AM is an employee of Bayer AG. At the time the study was conducted, KM was an employee of Servier. KV works for a research institute which receives/received unconditional research grants from Yamanouchi, Pfizer/Boehringer Ingelheim, Novartis, GSK, UCB, Amgen, Chiesi, none of these are related to the content of this paper. DV reports personal fees from Bayer, during the conduct of the study and outside the submitted work. DPA reports grants and other (DPA's department has received fees for speaker services and advisory board membership) from AMGEN, grants, non-financial support and other (DPA's department has received fees for consultancy services) from UCB Biopharma, grants from Les Laboratoires Servier, outside the submitted work; and Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programmes organised by DPA's department and open for external participants. PRR works for a research institute who receives/received unconditional research grants from Yamanouchi, Pfizer-Boehringer Ingelheim, GSK, Amgen, UCB, Novartis, AstraZeneca, Chiesi, Janssen Research and Development, none of which relate to the content of this work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2022.152050.

References

[1] Listing J, Gerhold K, Zink A. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. Rheumatology (Oxford) 2013;52(1):53–61.
[2] Peters MJ, Symmons DP, McCarey D, Dijkmans BA, Nicola P, Kvien TK, et al. EULAR evidence-based recommendations for cardiovascular risk management in patients with rheumatoid arthritis and other forms of inflammatory arthritis. Ann Rheum Dis 2010;69(2):325–31.
[3] Turesson C, Matteson EL. Malignancy as a comorbidity in rheumatic diseases. Rheumatology (Oxford) 2013;52(1):5–14.
[4] Dougados M, Soubrier M, Antunez A, Balint P, Balsa A, Buch MH, et al. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). Ann Rheum Dis 2014;73(1):62–8.
[5] Turesson C. Comorbidity in rheumatoid arthritis. Swiss Med Wkly 2016;146:w14290.
[6] Singh JA, Saag KG, Bridges Jr SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2016;68(1):1–26.
[7] Smolen JS, Landewe RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis 2020.
[8] Arts EE, Popa CD, Den Broeder AA, Donders R, Sandoo A, Toms T, et al. Prediction of cardiovascular risk in rheumatoid arthritis: performance of original and adapted SCORE algorithms. Ann Rheum Dis 2016;75(4):674–80.
[9] Crowson CS, Hoganson DD, Fitz-Gibbon PD, Matteson EL. Development and validation of a risk score for serious infection in patients with rheumatoid arthritis. Arthritis Rheum 2012;64(9):2847–55.
[10] Crowson CS, Rollefstad S, Kitas GD, van Riel PL, Gabriel SE, Semb AG, et al. Challenges of developing a cardiovascular risk calculator for patients with rheumatoid arthritis. PLoS One 2017;12(3):e0174656.
[11] Curtis JR, Xie F, Chen L, Muntner P, Grijalva CG, Spettell C, et al. Use of a disease risk score to compare serious infections associated with anti-tumor necrosis factor therapy among high- versus lower-risk rheumatoid arthritis patients. Arthritis Care Res (Hoboken) 2012;64(10):1480–9.
[12] Curtis JR, Xie F, Crowson CS, Sasso EH, Hitraya E, Chin CL, et al. Derivation and internal validation of a multi-biomarker-based cardiovascular disease risk prediction score for rheumatoid arthritis patients. Arthritis Res Ther 2020;22(1):282.
[13] Solomon DH, Greenberg J, Curtis JR, Liu M, Farkouh ME, Tsao P, et al. Derivation and internal validation of an expanded cardiovascular risk prediction score for rheumatoid arthritis: a consortium of rheumatology researchers of North America registry study. Arthritis Rheumatol 2015;67(8):1995–2003.
[14] Strangfeld A, Eveslage M, Schneider M, Bergerhausen HJ, Klopsch T, Zink A, et al. Treatment benefit or survival of the fittest: what drives the time-dependent decrease in serious infection rates under TNF inhibition and what does this imply for the individual patient? Ann Rheum Dis 2011;70(11):1914–20.
[15] Wang D, Yeo AL, Dendle C, Morton S, Morand E, Leech M. Severe infections remain common in a real-world rheumatoid arthritis cohort: a simple clinical model to predict infection risk. Eur J Rheumatol 2020.
[16] Jani M, Barton A, Hyrich K. Prediction of infection risk in rheumatoid arthritis patients treated with biologics: are we any closer to risk stratification? Curr Opin Rheumatol 2019;31(3):285–92.
[17] Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? J Am Med Inform Assoc 2019;26(12):1651–4.
[18] Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018;25(8):969–75.
[19] Khalid S, Yang C, Blacketer C, Duarte-Salles T, Fernández-Bertolín S, Kim C, et al. A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. Comput Methods Progr Biomed 2021;211:106394.

[20] Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012;19 (1):54–60.
[21] Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A, et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Med Res Methodol 2020;20(1):102.
[22] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1–73.
[23] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373–83.
[24] Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010;137(2):263–72.
[25] Young BA, Lin E, Von Korff M, Simon G, Ciechanowski P, Ludman EJ, et al. Diabetes complications severity index and risk of mortality, hospitalization, and healthcare utilization. Am J Manag Care 2008;14(1):15–23.
[26] Romano P, Roos L, Jollis J. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46(10): 1075–90.
[27] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc: Ser B (Methodol) 1996;58(1):267–88.
[28] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016; 35(2):214–26.
[29] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic t. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17(1):230.
[30] Ramspek CL, Steyerberg EW, Riley RD, Rosendaal FR, Dekkers OM, Dekker FW, et al. Prediction or causality? A scoping review of their conflation within current observational research. Eur J Epidemiol 2021:1–10.