

RESEARCH

Open Access



Machine learning prediction and explanatory models of serious infections in patients with rheumatoid arthritis treated with tofacitinib

Merete Lund Hetland^{1,2†}, Anja Strangfeld^{3,4†}, Gianluca Bonfanti⁵, Dimitrios Soudis⁶, J. Jasper Deuring^{7,9*†} and Roger A. Edwards^{8†}

Abstract

Background Patients with rheumatoid arthritis (RA) have an increased risk of developing serious infections (SIs) vs. individuals without RA; efforts to predict SIs in this patient group are ongoing. We assessed the ability of different machine learning modeling approaches to predict SIs using baseline data from the tofacitinib RA clinical trials program.

Methods This analysis included data from 19 clinical trials (phase 2, $n = 10$; phase 3, $n = 6$; phase 3b/4, $n = 3$). Patients with RA receiving tofacitinib 5 or 10 mg twice daily (BID) were included in the analysis; patients receiving tofacitinib 11 mg once daily were considered as tofacitinib 5 mg BID. All available patient-level baseline variables were extracted. Statistical and machine learning methods (logistic regression, support vector machines with linear kernel, random forest, extreme gradient boosting trees, and boosted trees) were implemented to assess the association of baseline variables with SI (logistic regression only), and to predict SI using selected baseline variables using 5-fold cross-validation. Missing values were handled individually per prediction model.

Results A total of 8404 patients with RA treated with tofacitinib were eligible for inclusion (15,310 patient-years of total follow-up) of which 473 patients reported SIs. Amongst other baseline factors, age, previous infection, and corticosteroid use were significantly associated with SI. When applying prediction modeling for SI across data from all studies, the area under the receiver operating characteristic (AUROC) curve ranged from 0.656 to 0.739. AUROC values ranged from 0.599 to 0.730 in data from phase 3 and 3b/4 studies, and from 0.563 to 0.643 in data from ORAL Surveillance only.

Conclusions Baseline factors associated with SIs in the tofacitinib RA clinical trial program were similar to established SI risk factors associated with advanced treatments for RA. Furthermore, while model performance in predicting

[†]Merete Lund Hetland and Anja Strangfeld shared first authorship.

[†]J. Jasper Deuring and Roger A. Edwards shared senior authorship.

*Correspondence:

J. Jasper Deuring
Jasper.Deuring@Pfizer.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

SI was similar to other published models, this did not meet the threshold for accurate prediction (AUROC > 0.85). Thus, predicting the occurrence of SIs at baseline remains challenging and may be complicated by the changing disease course of RA over time. Inclusion of other patient-associated and healthcare delivery-related factors and harmonization of the duration of studies included in the models may be required to improve prediction.

Trial registration ClinicalTrials.gov: NCT00147498; NCT00413660; NCT00550446; NCT00603512; NCT00687193; NCT01164579; NCT00976599; NCT01059864; NCT01359150; NCT02147587; NCT00960440; NCT00847613; NCT00814307; NCT00856544; NCT00853385; NCT01039688; NCT02187055; NCT02831855; NCT02092467.

Keywords Machine learning, Prediction models, Rheumatic diseases, Infectious diseases, Janus kinase inhibitor, Treatment safety, Risk stratification, Support vector machines with linear kernel, Random forest, Extreme gradient boosted trees

Background

Risk stratification and assessing causality in disease are founded on two different concepts, i.e., prediction vs. explanation, and thus require different research approaches to address [1]. Finding factors that best predict a current diagnosis or future event is the focus of predictive research, which uses predictive models to identify individuals or populations at risk of disease to allow appropriate interventions [1, 2]. Explanatory research uses models to identify causal factors of an outcome (e.g., risk or protective factors), and thus assesses whether such factors are valid targets of intervention across populations to prevent disease [1, 2]. However, the concepts of prediction and explanation are often conflated in studies attempting to identify “risk factors” in disease [1]. For example, in a systematic review of epidemiological diabetes publications that included “prediction” in their titles, only 745 articles (39%) reported metrics of predictive statistics, while 1165 articles (61%) did not include such metrics [3]. The top reported metrics of actual prediction were area under the receiver operating characteristic (AUROC) curve, sensitivity, and specificity [3]. Furthermore, using simulated data, it was observed that biomarkers with strong statistical association with diabetes can still demonstrate poor predictive validity [3]. Thus, these observations highlight that association is not prediction, though the two are often interchanged.

Like diabetes, rheumatoid arthritis (RA) is a multifactorial, immunological-driven disease with many parameters associated with the disease and treatment outcome [4–7]. In both cases, the development of prediction models based on patient data could improve individualized clinical decision-making so that the right treatment is given to the right patient at the right time [3]. However, it is likely that such models would require a combination of demographic, clinical, biological, and imaging data for accurate patient-level prediction [3, 8].

Patients with RA are at an increased risk of developing serious infections vs. individuals without RA [9]. Additionally, the risk of serious infection in patients with RA

has been observed to vary between some treatments [10, 11]. Tofacitinib is an oral Janus kinase inhibitor for the treatment of RA. Previous studies have explored factors associated with the risk of serious infection in patients receiving advanced treatments for RA, including tofacitinib, which include older age, male gender, previous history of infection, diabetes, and baseline corticosteroid use [11–15]. Furthermore, previous studies have attempted to develop and validate prediction models for a variety of adverse health outcomes in patients with RA from real-world clinical practice data, such as serious infection, myocardial infarction, stroke, and cancer [16–18]. Such prediction models would prove useful in identifying patients at high risk of adverse outcomes and allow for appropriate management of these patients, e.g., increased monitoring throughout treatment. Of the few prediction models that have been developed for RA, most have generally reported moderate to good discriminative power in predicting serious infections [16–18].

The aims of the current post hoc analysis were to apply advanced statistical methodologies and machine learning to confirm generally established association factors for serious infections with advanced treatments for RA, and primarily to generate a predictive model capable of identifying future occurrence of serious infection in patients with RA treated with tofacitinib based on patient-level baseline clinical trial data.

Methods

Patients and study design

This analysis included data from 10 phase 2, six phase 3, and three phase 3b/4 studies (including the ORAL Surveillance study). All studies were randomized clinical trials of tofacitinib in patients with RA and are summarized in Table 1.

Only patients randomized to receive tofacitinib 5 mg twice daily (BID), 10 mg BID, or 11 mg once daily (QD) as monotherapy or in combination with background methotrexate or other conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) were included in the current analysis.

Table 1 Summary of randomized clinical trials included in the analysis

Phase	ClinicalTrials.gov identifier	Protocol number/trial name	Patients included in current analysis, <i>n</i>	Patient population	Tofacitinib dose included in analysis	Study duration	
2	NCT00147498 [19]	A3921019	61	Active RA with inadequate response to, or discontinued therapy due to unacceptable toxicity from, MTX, etanercept, infliximab, or adalimumab	5 mg BID (monotherapy)	6 weeks	
	NCT00413660 [20]	A3921025	145	Active RA with inadequate response to stably dosed MTX	5 and 10 mg BID (with background MTX)	24 weeks	
	NCT00550446 [21]	A3921035	110	Active RA with inadequate response to ≥ 1 DMARD	5 and 10 mg BID (monotherapy)	24 weeks	
	NCT00603512 [22]	A3921039	53	Active RA with inadequate response to stably dosed MTX	5 and 10 mg BID (with background MTX)	12 weeks	
	NCT00687193 [23]	A3921040	105	Active RA with inadequate response to ≥ 1 DMARD	5 and 10 mg BID (monotherapy)	12 weeks	
	NCT01164579 [24]	A3921068	72	Early, active RA and MTX- and bDMARD-naïve	10 mg BID (with or without background MTX)	12 months	
	NCT00976599 [25]	A3921073	15	Active RA with inadequate response to stably dosed MTX	10 mg BID (with background MTX)	4 weeks	
	NCT01059864 [26]	A3921109	111	Active RA	10 mg BID (monotherapy)	12 weeks	
	NCT01359150 [27]	A3921129	112	Active RA	10 mg BID (with or without background MTX)	9 weeks	
	NCT02147587 [28]	A3921237	55	Active RA with inadequate response to stably dosed MTX	5 mg BID (with background MTX)	14 weeks	
	3	NCT00960440 [29]	A3921032/ ORAL Step	267	Active RA with inadequate response to TNFi	5 and 10 mg BID (with background MTX)	6 months
		NCT00847613 [30]	A3921044/ ORAL Scan	637	Active RA with inadequate response to stably dosed MTX	5 and 10 mg BID (with background MTX)	24 months
		NCT00814307 [31]	A3921045/ ORAL Solo	488	Active RA with inadequate response to ≥ 1 DMARD	5 and 10 mg BID (monotherapy)	6 months
		NCT00856544 [32]	A3921046/ ORAL Sync	633	Active RA with inadequate response to ≥ 1 DMARD	5 and 10 mg BID (with background DMARD)	12 months
NCT00853385 [33]		A3921064/ ORAL Standard	405	Active RA with inadequate response to stably dosed MTX	5 and 10 mg BID (with background MTX)	12 months	
NCT01039688 [34]		A3921069/ ORAL Start	770	Active RA and MTX-naïve	5 and 10 mg BID (monotherapy)	24 months	
3b/4	NCT02187055 [35]	A3921187/ ORAL Strategy	760	Active RA with inadequate response to stably dosed MTX	5 mg BID (with or without background MTX)	12 months	
	NCT02831855 [36]	A3921192/ ORAL Shift	694	Active RA with inadequate response to stably dosed MTX	11 mg QD (with or without background MTX)	12 months	
	NCT02092467 [37]	A3921133/ ORAL Surveillance	2911	Active RA with inadequate response to stably dosed MTX	5 and 10 mg BID (with background MTX)	Up to 72 months	

bDMARD biologic disease-modifying antirheumatic drug, *BID* twice daily, *DMARD* disease-modifying antirheumatic drug, *MTX* methotrexate, *QD* once daily, *RA* rheumatoid arthritis, *TNFi* tumor necrosis factor inhibitor

Given the diversity of the patients included in the different studies, analyses were performed on eligible data in three groups: (1) all studies; (2) phase 3 and 3b/4 studies only (i.e., excluding phase 2 studies); and (3) ORAL Surveillance (NCT02092467) only.

All studies were conducted in accordance with the International Council for Harmonisation Good Clinical Practice guidelines, local regulatory requirements, and the ethical principles of the Declaration of Helsinki. Protocols were approved by an Institutional Review Board

or Independent Ethics Committee at each study site. Patients provided written informed consent.

Outcomes

Serious infections were the outcome of interest and were identified from each study data set. A serious infection was defined as any infection (viral, bacterial, and fungal) that required hospitalization for treatment, or parenteral antimicrobial therapy, or resulted in death, was life-threatening (immediate risk of death), resulted in persistent or significant disability/incapacity (substantial

disruption of the ability to conduct normal life functions), or resulted in congenital anomaly/birth defect. A patient who experienced a serious infection was discontinued from the study (no recurrent serious infections were observed), except for ORAL Surveillance in which patients that permanently discontinued study treatment were asked to continue to participate in the trial. All serious infections occurring in a patient from the ORAL Surveillance study were considered. Appropriate laboratory investigations including, but not limited to, cultures were performed to establish the etiology of any serious infection.

Baseline variables

Patient-level, mutual baseline variables were extracted from each of the randomized clinical trials. In total, 129 baseline variables were extracted, which included variables related to demographics, medical history, medication use, disease activity assessments, and laboratory assessments.

The treatment variable included within the current analysis had two possible values: tofacitinib 5 mg BID or tofacitinib 10 mg BID. Therefore, patients randomized to receive either tofacitinib 10 mg BID plus atorvastatin or tofacitinib 10 mg BID plus placebo in NCT01059864 during the double-blind phase were considered as tofacitinib 10 mg BID. For NCT02831855, patients randomized to receive tofacitinib 11 mg QD were considered as tofacitinib 5 mg BID.

Data pre-processing

To handle missing values, the total number of missing values per baseline variable was calculated and a threshold of 70% was applied to retain only those variables where data were available for $\geq 70\%$ of the total number of patients included in the analysis (Additional file 1: Fig. S1). To handle lack of variability, variables where the range of possible values was reduced to a single value (i.e., those for which all patients have the same value) were excluded, for example, “biologic DMARD use at baseline” and “Others combined use at baseline”. Additionally, variables related to the EuroQol-Five Dimensions Health Questionnaire were excluded because they were not measured in six of the 19 clinical studies considered.

For the remaining baseline variables, the handling of missing values for each prediction model varied. For those models incapable of handling missing values, only complete observations were included (i.e., patients with ≥ 1 missing value were excluded). For some of these models, the analysis was also run using maximum likelihood (single or multiple) imputation. For models capable of handling missing values either natively or using the missing incorporated in attribute approach, analysis was run using whole population data (i.e., all patients are included

regardless of missing values). Details on how missing values were handled in each analysis are listed in Table 2.

For each continuous variable, min-max normalization was applied (i.e., the variable was rescaled to 0–1 range) to eliminate any influence of differences in the magnitudes across variables on the final results.

The variables included in the final analysis, following data pre-processing, are summarized in Additional file 1: Table S1.

Multivariate logistic regression analysis

Prior to running any model prediction, logistic regression was applied on the whole patient data set (i.e., without any data splits for cross-validation) to assess the associations of baseline variables with the outcome of interest in a multivariable context. Model performance was assessed based on the Akaike’s information criterion method, a statistical measure for comparative evaluation of models as a trade-off between goodness-of-fit and complexity, in which the smallest value reflects the best model. Step-wise variable selection was performed; this process starts with the null model and at each iteration, variables can be added to or excluded from the model. The process stops when no further improvement in Akaike’s information criterion would be achieved with addition or removal of variables to or from the model, respectively. Regression coefficients and odds ratios (OR) with 95% confidence intervals (CIs) were calculated. Associations were considered significant where the 95% CIs did not include 1.0.

Prediction models

Various machine learning methods (classification models) were implemented to explore how results change by model; these included logistic regression, support vector machines with linear kernel, random forest, extreme gradient boosting trees, and boosted trees. These models are reviewed in detail elsewhere [38]. For each model, the baseline variables were defined as independent variables, while the presence of serious infection was defined as the dependent variable. Continuous variables were included as predictors in the support vector machines algorithm; both continuous and categorical variables were included in all other algorithms.

Statistical analysis methods

Multiple classification algorithms were applied to the data using a repeated k-fold cross-validation (CV) approach [39]. First, the whole data set was randomly split into “k” different partitions, where “k” = 5 in this analysis. Then, in an iterative procedure, the k partitions were removed one at a time, leaving the other partitions (“k-1 partition”) to be considered together. Patients in the removed k partition served as the testing data set, while patients in the k-1 partitions were used as the training

Table 2 Estimated performance metrics

Algorithm	Missing values handling	AUROC	Accuracy ^a , %	Sensitivity ^a , %	Specificity ^a , %	PPV ^a , %	NPV ^a , %
A) All studies (group 1; N=8404)^b							
Logistic regression	Only complete observations	0.705	82.5	37.4	85.5	14.7	95.3
SVM with linear kernel	Only complete observations	0.686–0.691	75.1–75.7	51.0–52.9	76.6–77.2	12.9–13.3	95.9–96.1
Random forest	Only complete observations	0.682–0.733	93.0–93.7	0.0–6.2	98.8–100.0	0.0–30.9	93.7–94.0
Extreme gradient boosting trees ^c	Whole population (no missing value imputation)	0.656–0.739	83.7–93.6	3.8–27.1	87.2–98.9	9.9–20.0	94.5–95.5
Boosted trees ^c	MIA	0.703–0.726	89.6–91.5	11.3–18.4	93.9–96.3	14.6–17.0	94.8–95.1
Logistic regression ^c	ML single imputation	0.693	80.1	40.9	82.5	12.2	95.9
Logistic regression ^c	ML multiple imputation	0.694–0.697	79.8–80.2	40.0–41.5	82.1–82.5	11.9–12.4	95.8–95.9
B) Phase 3 and 3b/4 studies (group 2; N=7565)^b							
Logistic regression	Only complete observations	0.696	81.9	36.3	85.0	14.3	95.1
SVM with linear kernel	Only complete observations	0.680–0.686	74.8–75.5	48.9–51.3	76.6–77.2	12.6–13.4	95.6–95.8
Random forest	Only complete observations	0.673–0.723	92.5–93.5	0.0–5.1	98.6–100.0	0.0–41.7	93.5–93.8
Extreme gradient boosting trees ^c	Whole population (no missing value imputation)	0.599–0.730	87.9–92.9	4.6–22.6	92.2–98.6	11.8–19.9	94.1–94.9
Boosted trees ^c	MIA	0.702–0.720	88.8–90.9	13.1–18.8	93.4–96.0	14.9–17.9	94.4–94.7
Logistic regression ^c	ML single imputation	0.702	82.4	35.7	85.4	13.8	95.3
Logistic regression ^c	ML multiple imputation	0.701–0.704	82.4–82.6	36.4–37.6	85.4–85.6	14.1–14.5	95.4–95.5
C) ORAL Surveillance only (group 3; N=2911)^b							
Logistic regression	Only complete observations	0.611	75.3	32.5	80.9	18.3	90.1
SVM with linear kernel	Only complete observations	0.607–0.610	73.1–73.7	34.7–36.3	78.0–78.8	17.3–17.9	90.1–90.3
Random forest	Only complete observations	0.589–0.635	87.7–88.4	0.0–3.4	98.9–100.0	0.0–63.9	88.3–88.6
Extreme gradient boosting trees ^c	Whole population (no missing value imputation)	0.563–0.643	74.0–87.4	3.9–24.1	80.5–98.3	14.1–27.6	88.6–89.3
Boosted trees ^c	MIA	0.603–0.630	86.3–87.5	3.3–8.0	96.6–98.6	20.1–26.6	88.5–88.8
Logistic regression ^c	ML single imputation	0.624	76.1	35.3	81.5	20.1	90.5
Logistic regression ^c	ML multiple imputation	0.621–0.629	75.9–76.4	34.8–36.3	81.3–81.8	19.8–20.7	90.5–90.7

The AUROC considers the estimated probabilities provided by the models, regardless of any cut-off value, while all other performance measures (i.e., accuracy, sensitivity, specificity, PPV, and NPV) are obtained by applying a cut-off value of 0.5 to the predicted probability obtained (i.e., a patient is classified as having serious infections if their predicted probability is ≥ 0.5)

AUROC area under receiver operating characteristic, MIA missing incorporated in attribute, ML maximum likelihood, N total number of patients included in each group, NPV negative predictive value, PPV positive predictive value, SVM support vector machines

^a Cut-off = 0.5

^b The total number of patients assessed in each model differed according to how missing values were handled by the model

^c Complete patient set. No patients excluded based on missing variables

data set. For each partition, a model is chosen and fitted based on the patients in the training data set and predictions are then made for patients in the test data set. This sequence was iterated for each k partition, enabling predictions for each patient because each patient was in 1 partition and each partition was used once as the test data set. It is important to note that the predictions were provided by k different models (i.e., 1 model for each iteration of the CV). The k-fold CV was repeated multiple times with the random partition changing each time. The values of the hyperparameters for each algorithm were also varied to assess how the results changed. Given that three repetitions of the CV approach were performed in this analysis, the performance metrics of each model are an average of these repetitions. Synthetic Minority Over-sampling Technique (SMOTE) [40] was applied to adjust

for class imbalance between the number of patients with vs. without serious infection.

For evaluating the accuracy of the models to predict serious infection, receiver operating characteristic curves were derived. The AUROC curve was then determined to assess how well each model distinguished between classes (i.e., patients with vs. patients without serious infections); an AUROC value ≤ 0.5 was considered as lacking predictive ability, while a value equal to 1 indicated prediction was always accurate [38]. For the purposes of this analysis, an AUROC value of 0.85 was considered as the threshold for an accurate model.

Results

Patient selection

Demographic and baseline characteristics of the patients included in this analysis are summarized in Table 3. A

Table 3 Demographics and baseline characteristics for all patients included in the analysis

Phase	ClinicalTrials.gov identifier	Total number of patients, N	Total PY	Patients with serious infections, n	IR, per 100 PY	Tofacitinib dose, n	Female, n (%)	Age, mean (SD)	BMI, mean (SD)	Race, n (%)	White	Black	Asian	Other
2	NCT00147498	61	13.6	0	-	61	53 (86.9)	47.9 (10.8)	27.8 (6.8)	42 (68.9)	3 (4.9)	1 (1.6)	15 (24.6)	
	NCT00413660	145	61.7	2	3.2	71	112 (77.2)	53.9 (11.8)	28.0 (5.7)	127 (87.6)	3 (2.1)	-	15 (10.3)	
	NCT00550446	110	48.8	0	-	49	96 (87.3)	52.9 (12.1)	27.3 (5.6)	80 (72.7)	3 (2.7)	11 (10.0)	16 (14.5)	
	NCT00603512	53	11.6	0	-	27	47 (88.7)	50.3 (9.8)	21.5 (3.4)	-	-	53 (100)	-	
	NCT00687193	105	23.8	2	8.4	52	88 (83.8)	53.7 (10.9)	22.1 (3.5)	-	-	105 (100)	-	
	NCT01164579	72	57.8	1	1.7	-	61 (84.7)	49.3 (12.5)	27.6 (5.4)	40 (55.6)	1 (1.4)	-	31 (43.1)	
	NCT00976599	15	1.5	0	-	-	14 (93.3)	53.5 (9.2)	34.8 (8.9)	10 (66.7)	3 (20.0)	1 (6.7)	1 (6.7)	
	NCT01059864	111	24.1	2	8.3	-	99 (89.2)	52.3 (11.5)	27.5 (7.6)	51 (45.9)	5 (4.5)	48 (43.2)	7 (6.3)	
	NCT01359150	112	19.3	1	5.2	-	81 (72.3)	52.7 (10.4)	31.1 (7.3)	99 (88.4)	5 (4.5)	-	8 (7.1)	
	NCT02147587	55	11.8	3	25.4	55	42 (76.4)	61.7 (6.2)	31.4 (7.1)	49 (89.1)	5 (9.1)	1 (1.8)	-	
3	NCT00960440	267	114.7	4	3.5	133	229 (85.8)	55.2 (11.4)	29.3 (7.1)	220 (82.4)	18 (6.7)	19 (7.1)	10 (3.7)	
	NCT00847613	637	1040.3	43	4.1	321	541 (84.9)	52.8 (11.5)	26.2 (6.3)	296 (46.5)	22 (3.5)	266 (41.8)	53 (8.3)	
	NCT00814307	488	233.5	5	2.1	243	423 (86.7)	52.3 (11.6)	27.3 (6.6)	321 (65.8)	22 (4.5)	73 (15.0)	72 (14.8)	
	NCT00856544	633	563.8	10	1.8	315	522 (82.5)	52.3 (11.8)	26.7 (6.5)	347 (54.8)	11 (1.7)	224 (35.4)	51 (8.1)	
	NCT00853385	405	351.5	14	4.0	204	342 (84.4)	52.9 (11.9)	27.3 (6.4)	294 (72.6)	7 (1.7)	59 (14.6)	45 (11.1)	
	NCT01039688	770	1274.1	19	1.5	373	613 (79.6)	49.8 (12.5)	26.7 (5.7)	505 (65.6)	25 (3.2)	131 (17.0)	109 (14.2)	

Table 3 (continued)

Phase	ClinicalTrials.gov identifier	Total number of patients, N	Total PY	Patients with serious infections, n	IR, per 100 PY	Tofacitinib dose, n	Female, n (%)	Age, mean (SD)	BMI, mean (SD)	Race, n (%)	White	Black	Asian	Other
3b/4	NCT02187055	760	655.1	16	2.4	760	630 (82.9)	49.8 (12.8)	28.0 (6.5)	582 (76.6)	30 (3.9)	79 (10.4)	69 (9.1)	
	NCT02831855	694	546.9	12	2.2	694 ^b	532 (76.7)	56.8 (11.8)	28.2 (6.2)	594 (85.6)	33 (4.8)	37 (5.3)	30 (4.3)	
	NCT02092467	2911	10256.4	339	3.3	1455	2293 (78.8)	61.1 (6.9)	29.7 (6.4)	2254 (77.4)	128 (4.4)	121 (4.2)	408 (14.0)	
Total		8404	15310.1	473	3.1	4813	6818 (81.1)	55.5 (11.4)	28.2 (6.5)	5911 (70.3)	324 (3.9)	1229 (14.6)	940 (11.2)	

BID twice daily, *BMI* body mass index, *IR* incidence rate, *PY* patient-years, *QD* once daily, *SD* standard deviation

^a For NCT01059864, patients randomized to receive either tofacitinib 10 mg BID plus atorvastatin (*n*=63) or tofacitinib 10 mg BID plus placebo (*n*=48) were considered as tofacitinib 10 mg BID

^b For NCT02831855, patients randomized to receive tofacitinib 11 mg QD (*n*=694) were considered as tofacitinib 5 mg BID

total of 8404 patients with RA treated with tofacitinib (5 mg BID, *n*=4813; 10 mg BID, *n*=3591) were eligible for inclusion in the study. The total follow-up time for patients was 15,310.13 patient-years. Patients were mostly female (81%) with a mean (standard deviation) age of 55.5 (11.4) years and body mass index of 28.2 (6.5). Of these, serious infections were reported in 473 patients with RA receiving tofacitinib (5 mg BID, *n*=236; 10 mg BID, *n*=237; group 1). The number of patients included in the individual phase 2 studies was generally low (range 15–145).

Across phase 3 and 3b/4 studies, a total of 7565 patients with RA were eligible for inclusion, of which, serious infections were reported in 462 patients (group 2). Data from 2911 patients were analyzed from the ORAL Surveillance study (NCT02092467), with 339 patients reporting serious infections (group 3).

Association analysis

Stepwise logistic regression showed that the largest baseline variable association with serious infection was ethnicity, in which Asian patients were more likely to experience serious infections vs. White patients (OR 2.85; 95% CI 2.09, 3.87; Fig. 1). Other significant associations with serious infections included older age, male gender, and prior and/or current comorbidities (infection, renal and urinary disorders, and chronic obstructive pulmonary disease), as well as concomitant treatments received at baseline (csDMARDs, corticosteroids, psycholeptics, antidepressants, and lipid-lowering agents). The absence of vascular disorders was significantly associated with serious infections. Other significant associations were also observed for some laboratory assessments and disease activity indices, which are shown in Fig. 1.

Prediction of serious infections at baseline performance metrics

An overview of the estimated performance metrics for each of the seven prediction models across all studies (group 1), phase 3 and 3b/4 studies (group 2), and ORAL Surveillance (NCT02092467; group 3), including the various methods for imputation of missing values, is shown in Table 2. When data from all studies were included in the analysis, AUROC ranged from 0.656 to 0.739 (Table 2). Assessment of data from only phase 3 and 3b/4 studies resulted in AUROC ranging from 0.599 to 0.730 (Table 2), while an AUROC between 0.563 and 0.643 was observed when the models were assessed using only ORAL Surveillance study data (Table 2). The sensitivity/specificity and the positive/negative predictive values were generally consistent across all groups and ranged from 0.0–52.9%/76.6–100% and 0.0–63.9%/88.3–96.1%, respectively. Figures S2, S3, and S4 in Additional file 1

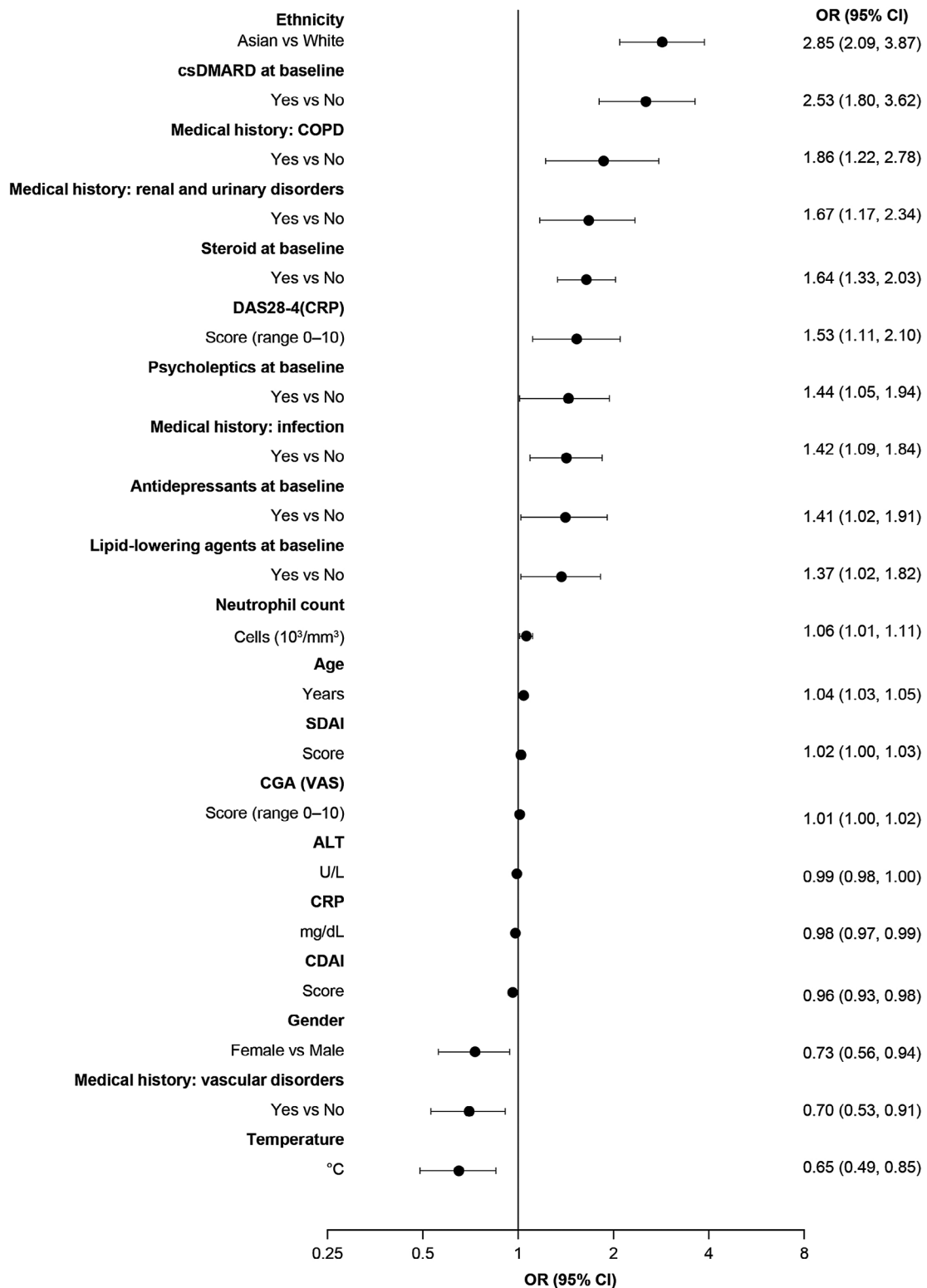


Fig. 1 Baseline variables associated with serious infection (stepwise multivariate logistic regression^a)

^a Stepwise multivariate logistic regression was performed in all patients (N=8404); ORs shown are those for patients with serious infection (n=473). For continuous variables, ORs > 1 indicate a higher risk of serious infection with higher values. ALT, alanine aminotransferase; CDAI, Clinical Disease Activity Index; CGA, Clinician Global Assessment; CI, confidence interval; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; DAS28-4(CRP), Disease Activity Score in 28 joints, C-reactive protein-4; OR, odds ratio; SDAI, Simplified Disease Activity Index; VAS, visual analog scale

show the variable importance for the Extreme Gradient Boosted prediction models.

Discussion

In the current study, we combined baseline patient-level data from 19 randomized clinical trials of patients with RA receiving tofacitinib 5 mg BID (including 11 mg QD) or 10 mg BID and assessed (1) the association of variables with outcomes in a multivariable context and (2) the ability to accurately predict serious infection in individual patients by applying seven independent prediction modeling approaches. Consistent with previously published analyses on the tofacitinib RA clinical development program and those receiving tumor necrosis factor inhibitors (TNFi) or non-TNFi DMARDs [11–15], multivariate logistic regression of the entire data set showed that, amongst other factors, older age, male gender, previous history of infections, and corticosteroid use at baseline are associated with a higher risk of serious infections in patients receiving tofacitinib. These results suggest that the patient composition selected for this analysis seems representative of the previously analyzed cohorts.

By applying seven independent prediction models, we observed that in the selected overall data set (group 1), the threshold for definitive prediction ($AUROC \geq 0.85$) was not achieved; this was consistent when the analysis was focused on data from phase 3 and 3b/4 studies only (group 2) or ORAL Surveillance (group 3). It should be noted that model performance, using data from clinical trial data sets, was generally comparable to that seen in other published models predicting serious infection in RA, using data from clinical practice cohorts, where moderate discriminative power has been reported ($AUROC = 0.68–0.74$) [16–18]. Thus, our findings highlight the continuing challenge to accurately predict serious infections at baseline in patients with RA, which may, in part, arise due to a patient's risk of serious infection varying over time.

By combining data from 19 different clinical trials, 129 different baseline variables could be identified. However, some were excluded from the analysis due to missing values (<70% patient data available) or lack of variability. Variables included in the models still covered a wide range of baseline characteristics, medical history, and concomitant treatments. Nonetheless, many other factors highly associated with infections have been reported from other studies which were not available in the current data set, such as socio-economic and environmental factors [41, 42]. Thus, it is possible that variables not available in the included studies may provide more predictive value.

There were several strengths to the current study; this was a large, well-characterized data set of patients with RA ($N = 8404$; >15,000 patient-years of follow-up) from

multiple randomized clinical trials. All patients included in the analysis had a defined starting point (start of treatment with tofacitinib) and data were collected following a standardized protocol. Furthermore, all serious infections were adjudicated events, the seven independent prediction models assessed are well-established approaches, and model performance was generally consistent across the various groups for predictive modeling. Conversely, there were a number of limitations that should be taken into consideration. The prediction model was limited to the available variables within the included clinical trials, and these were further restricted to only those variables assessed at baseline. The included studies also had different durations, ranging from 6 weeks up to 72 months, which may have affected the number of observed serious infection events in some studies and thereby limited prediction of such events based on baseline variables. Furthermore, clinical trials are highly selective and therefore, the validity of any prediction model based on clinical trial data may be limited in real-world patients. Additionally, there was a large imbalance between the number of patients with serious infections ($n = 473$) vs. those without infection ($n = 7931$), and despite partial correction for this using the SMOTE technique, this imbalance may still have impacted the positive predictive value. As Janus kinase inhibitors, such as tofacitinib, are associated with an increased incidence of herpes zoster [43, 44], prediction of specific bacterial, viral, or fungal infections could be interesting but was not possible in the current study due to the small number of cases of each. It should also be noted that the prediction models were not validated using any internal comparator (e.g., placebo or csDMARD-treated patients without tofacitinib exposure), or any external data source other than the clinical trial data; however, these analyses established a foundation for future research on this topic.

In addition to factors previously suggested, such as patient demographic factors, specific biomarkers, and lifestyle factors [45], collecting information associated with patient-related social, environmental/healthcare delivery-related, and personal demographics may provide further insight. Sample sizes could also be enlarged by including non-serious infections, which may allow broader prediction of infection events and subsequently allow classification by infection subtypes and causative agents, though it should be considered that inclusion of such may introduce larger variability within the data. Although not directly compared, findings from ORAL Surveillance demonstrated that different factors were associated with serious and non-serious infections in patients receiving tofacitinib or TNFi [13], suggesting that factors that increase the risk of serious infections are not similar to those for non-serious infections.

As the disease course of RA varies over time, and can be unique to each individual patient [46], predicting events (e.g., serious infections) with purely baseline-derived variables may not be accurate, particularly with long observation periods, as the patients risk of such events may change over time [15]. Therefore, time-varying variables that consider changes to the patient over time (e.g., for disease activity, glucocorticoid dosage, functional status, and changes in white blood cell composition) may enable more accurate prediction. However, the aim of the current study was to assess the ability to predict the occurrence of serious infections based only on variables at the start of tofacitinib treatment.

Conclusion

Our findings are consistent with previously reported analyses on advanced therapies in RA, including tofacitinib; we observed that older age, male gender, previous history of infections, and corticosteroid use at baseline are associated with a higher risk for serious infections. Furthermore, use of baseline data from a large tofacitinib RA clinical trial program data set with seven independent prediction modeling approaches resulted in a similar model performance to that observed previously. However, this did not meet the threshold for definitive prediction in clinical practice. Thus, prediction of serious infections at baseline using clinical trial data from the tofacitinib RA program is currently challenging, and other patient-associated data, harmonization of study duration, or including time-varying variables may be required to increase the ability to accurately predict serious infection with only baseline characteristics.

Abbreviations

ALT	Alanine aminotransferase
AUROC	Area under receiver operating characteristic
bDMARD	Biologic disease-modifying antirheumatic drug
BID	Twice daily
BMI	Body mass index
CDAI	Clinical Disease Activity Index
CGA	Clinician Global Assessment
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CRP	C-reactive protein
csDMARD	Conventional synthetic disease-modifying antirheumatic drug
CV	Cross-validation
DAS28-4(CRP)	Disease Activity Score in 28 joints, C-reactive protein-4
DMARD	Disease-modifying antirheumatic drug
IR	Incidence rate
MIA	Missing incorporated in attribute
ML	Maximum likelihood
MTX	Methotrexate
NPV	Negative predictive value
OR	Odds ratio
PPV	Positive predictive value
PY	Patient-years
QD	Once daily
RA	Rheumatoid arthritis
SD	Standard deviation
SDAI	Simplified Disease Activity Index

SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support vector machines
TNFi	Tumor necrosis factor inhibitor
VAS	Visual analog scale

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13075-024-03376-9>.

Supplementary Material 1

Acknowledgments

The authors would like to thank the study patients and investigators. The authors also thank Ivana Vranic for her critical review of the manuscript. Medical writing support, under the direction of the authors, was provided by Robert Morgan, PhD, CMC Connect, a division of IPG Health Medical Communications, and was funded by Pfizer, New York, NY, USA, in accordance with Good Publication Practice (GPP 2022) guidelines (Ann Intern Med. 2022;175:1298–304).

Author contributions

MLH and AS contributed to data interpretation. GB contributed to study conception/design and data analysis. DS contributed to study conception/design. JJD and RAE contributed to study conception/design, data analysis, and data interpretation. All authors had access to the data, and reviewed and approved the final manuscript before submission.

Funding

This study was sponsored by Pfizer.

Data availability

Upon request, and subject to review, Pfizer will provide the data that support the findings of this study. Subject to certain criteria, conditions, and exceptions, Pfizer may also provide access to the related individual de-identified participant data. See <https://www.pfizer.com/science/clinical-trials/trial-data-and-results> for more information.

Declaration

Ethics approval and consent to participate

This study was a post hoc analysis of existing data from the tofacitinib rheumatoid arthritis clinical development program. All included studies were approved by an Institutional Review Board or Independent Ethics Committee at each study site. Patients provided written informed consent.

Consent for publication

Not applicable.

Competing interests

MLH has received grants from AbbVie, Biogen, Bristol Myers Squibb, Celltrion, Eli Lilly, Janssen Biologics B.V., Lundbeck Foundation, MSD, Pfizer Inc, Roche, Samsung Bioepis, Sandoz, and Novartis; honoraria from Medac, Pfizer Inc, and Sandoz; is a member of the advisory board for AbbVie; is a co-chair of EuroSpA; and has chaired for DANBIO DRQ. AS is a principal investigator of RABBIT, which is jointly sponsored by a consortium of pharmaceutical manufacturers as follows: AbbVie, Amgen, Bristol Myers Squibb, Celltrion, Eli Lilly, Fresenius-Kabi, Galapagos, Hexal, MSD, Pfizer Inc, Samsung Bioepis, Sanofi-Aventis, UCB, and Viartis; and has received lecture fees from AbbVie, Bristol Myers Squibb, Celltrion, Eli Lilly, MSD, Pfizer Inc, Roche, Sanofi-Aventis, and UCB. GB is a stockholder of Engineering Ingegneria Informatica, which acquired and merged with Fair Dynamics Consulting; and a paid consultant contracted by Health Services Consulting Corporation in connection with this study. Health Services Consulting Corporation was a paid contractor to Pfizer Inc in connection with the formal data analysis. DS and JJD are employees and stockholders of Pfizer Inc. RAE is the owner of Health Services Consulting Corporation and was a paid consultant to Pfizer Inc in connection with data generation.

Description of data

List of baseline variables included in analysis (Table S1); Percentage of available data across all baseline variables extracted from studies in the tofacitinib RA program (Fig. S1); Feature importance of the Extreme Gradient Boosted prediction model for all studies (group 1) (Fig. S2). Feature importance of the Extreme Gradient Boosted prediction model for phase 3 and 3b/4 studies (group 2) (Fig. S3); Feature importance of the Extreme Gradient Boosted prediction model for ORAL Surveillance (NCT02092467; group 3) (Fig. S4).

Author details

¹Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Centre of Head and Orthopaedics, Copenhagen University Hospital Rigshospitalet, Glostrup, Denmark
²Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
³Epidemiology and Health Services Research, German Rheumatism Research Centre (DRFZ), Berlin, Germany
⁴Department of Rheumatology and Clinical Immunology, Charité University Medicine Berlin, Berlin, Germany
⁵Engineering Ingegneria Informatica, Milan, Italy
⁶Pfizer Hellas S.A, Thessaloniki, Greece
⁷Pfizer, Rotterdam, The Netherlands
⁸Health Services Consulting Corporation, Boxborough, MA, USA
⁹Pfizer Netherlands GmbH, Rivium Westlaan, 142 2909 LD Capelle a/d IJssel, Rotterdam, The Netherlands

Received: 18 December 2023 / Accepted: 10 July 2024

Published online: 27 August 2024

References

1. Schooling CM, Jones HE. Clarifying questions about risk factors: predictors versus explanation. *Emerg Themes Epidemiol*. 2018;15:10.
2. Sainani KL. Explanatory versus predictive modeling. *PM R*. 2014;6:841–4.
3. Varga TV, Niss K, Estampador AC, Collin CB, Moseley PL. Association is not prediction: a landscape of confused reporting in diabetes - a systematic review. *Diabetes Res Clin Pract*. 2020;170:108497.
4. Tian Z, McLaughlin J, Verma A, Chinoy H, Heald AH. The relationship between rheumatoid arthritis and diabetes mellitus: a systematic review and meta-analysis. *Cardiovasc Endocrinol Metab*. 2021;10:125–31.
5. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid arthritis. *Nat Rev Dis Primers*. 2018;4:18001.
6. Berbudi A, Rahmadika N, Tjahjadi AI, Ruslami R. Type 2 diabetes and its impact on the immune system. *Curr Diabetes Rev*. 2020;16:442–9.
7. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol*. 2011;11:98–107.
8. Mahler M, Martinez-Prat L, Sparks JA, Deane KD. Precision medicine in the care of rheumatoid arthritis: focus on prediction and prevention of future clinically-apparent disease. *Autoimmun Rev*. 2020;19:102506.
9. Doran MF, Crowson CS, Pond GR, O'Fallon WM, Gabriel SE. Frequency of infection in patients with rheumatoid arthritis compared with controls: a population-based study. *Arthritis Rheum*. 2002;46:2287–93.
10. Pawar A, Desai RJ, Gautam N, Kim SC. Risk of admission to hospital for serious infection after initiating tofacitinib versus biologic DMARDs in patients with rheumatoid arthritis: a multidatabase cohort study. *Lancet Rheumatol*. 2020;2:E84–98.
11. Galloway JB, Hyrich KL, Mercer LK, Dixon WG, Fu B, Ustianowski AP, et al. Anti-TNF therapy is associated with an increased risk of serious infections in patients with rheumatoid arthritis especially in the first 6 months of treatment: updated results from the British Society for Rheumatology Biologics Register with special emphasis on risks in the elderly. *Rheumatology (Oxford)*. 2011;50:124–31.
12. Cohen SB, Tanaka Y, Mariette X, Curtis JR, Lee EB, Nash P, et al. Long-term safety of tofacitinib up to 9.5 years: a comprehensive integrated analysis of the rheumatoid arthritis clinical development programme. *RMD Open*. 2020;6:e001395.
13. Balanescu A, Citera G, Pascual-Ramos V, Bhatt DL, Connell CA, Gold D, et al. Infections in patients with rheumatoid arthritis receiving tofacitinib versus tumour necrosis factor inhibitors: results from the open-label, randomised controlled ORAL Surveillance trial. *Ann Rheum Dis*. 2022;81:1491–503.
14. Salmon JH, Gottenberg JE, Ravaut P, Cantagrel A, Combe B, Flipo RM, et al. Predictive risk factors of serious infections in patients with rheumatoid arthritis treated with abatacept in common practice: results from the Orenca and Rheumatoid Arthritis (ORA) registry. *Ann Rheum Dis*. 2016;75:1108–13.
15. 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;70:1914–20.
16. Yang C, Williams RD, Swerdel JN, Almeida JR, Brouwer ES, Burn E, et al. Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: a multinational real-world cohort analysis. *Semin Arthritis Rheum*. 2022;56:152050.
17. Krabbe S, Grøn KL, Glinthorg B, Nørgaard M, Mehnert F, Jarbøl DE, et al. Risk of serious infections in arthritis patients treated with biological drugs: a matched cohort study and development of prediction model. *Rheumatology (Oxford)*. 2021;60:3834–44.
18. Zink A, Manger B, Kaufmann J, Eisterhues C, Krause A, Listing J, et al. Evaluation of the RABBIT risk score for serious infections. *Ann Rheum Dis*. 2014;73:1673–6.
19. Kremer JM, Bloom BJ, Breedveld FC, Coombs JH, Fletcher MP, Gruben D, et al. The safety and efficacy of a JAK inhibitor in patients with active rheumatoid arthritis: results of a double-blind, placebo-controlled phase IIa trial of three dosage levels of CP-690,550 versus placebo. *Arthritis Rheum*. 2009;60:1895–905.
20. Kremer JM, Cohen S, Wilkinson BE, Connell CA, French JL, Gomez-Reino J, et al. A phase IIb dose-ranging study of the oral JAK inhibitor tofacitinib (CP-690,550) versus placebo in combination with background methotrexate in patients with active rheumatoid arthritis and an inadequate response to methotrexate alone. *Arthritis Rheum*. 2012;64:970–81.
21. Fleischmann R, Cutolo M, Genovese MC, Lee EB, Kanik KS, Sadis S, et al. Phase IIb dose-ranging study of the oral JAK inhibitor tofacitinib (CP-690,550) or adalimumab monotherapy versus placebo in patients with active rheumatoid arthritis with an inadequate response to disease-modifying antirheumatic drugs. *Arthritis Rheum*. 2012;64:17–29.
22. Tanaka Y, Suzuki M, Nakamura H, Toyozumi S, Zwillich SH, Tofacitinib Study Investigators. Phase II study of tofacitinib (CP-690,550) combined with methotrexate in patients with rheumatoid arthritis and an inadequate response to methotrexate. *Arthritis Care Res (Hoboken)*. 2011;63:1150–8.
23. Tanaka Y, Takeuchi T, Yamanaka H, Nakamura H, Toyozumi S, Zwillich S. Efficacy and safety of tofacitinib as monotherapy in Japanese patients with active rheumatoid arthritis: a 12-week, randomized, phase 2 study. *Mod Rheumatol*. 2015;25:514–21.
24. Conaghan PG, Østergaard M, Bowes MA, Wu C, Fuerst T, van der Heijde D, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naïve, early active rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitative and quantitative techniques. *Ann Rheum Dis*. 2016;75:1024–33.
25. Boyle DL, Soma K, Hodge J, Kavanaugh A, Mandel D, Mease P, et al. The JAK inhibitor tofacitinib suppresses synovial JAK1-STAT signalling in rheumatoid arthritis. *Ann Rheum Dis*. 2015;74:1311–6.
26. McInnes IB, Kim HY, Lee SH, Mandel D, Song YW, Connell CA, et al. Open-label tofacitinib and double-blind atorvastatin in rheumatoid arthritis patients: a randomised study. *Ann Rheum Dis*. 2014;73:124–31.
27. Winthrop KL, Silverfield J, Racewicz A, Neal J, Lee EB, Hrycaj P, et al. The effect of tofacitinib on pneumococcal and influenza vaccine responses in rheumatoid arthritis. *Ann Rheum Dis*. 2016;75:687–95.
28. Winthrop KL, Wouters AG, Choy EH, Soma K, Hodge JA, Nduaka CI, et al. The safety and immunogenicity of live zoster vaccination in patients with rheumatoid arthritis before starting tofacitinib: a randomized phase II trial. *Arthritis Rheumatol*. 2017;69:1969–77.
29. Burmester GR, Blanco R, Charles-Schoeman C, Wollenhaupt J, Zerbini C, Benda B, et al. Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: a randomised phase 3 trial. *Lancet*. 2013;381:451–60.
30. van der Heijde D, Tanaka Y, Fleischmann R, Keystone E, Kremer J, Zerbini C, et al. Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: twelve-month data from a twenty-four-month phase III randomized radiographic study. *Arthritis Rheum*. 2013;65:559–70.

31. Fleischmann R, Kremer J, Cush J, Schulze-Koops H, Connell CA, Bradley JD, et al. Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N Engl J Med*. 2012;367:495–507.
32. Kremer J, Li Z-G, Hall S, Fleischmann R, Genovese M, Martin-Mola E, et al. Tofacitinib in combination with nonbiologic disease-modifying antirheumatic drugs in patients with active rheumatoid arthritis: a randomized trial. *Ann Intern Med*. 2013;159:253–61.
33. van Vollenhoven RF, Fleischmann R, Cohen S, Lee EB, García Mejjide JA, Wagner S, et al. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. *N Engl J Med*. 2012;367:508–19.
34. Lee EB, Fleischmann R, Hall S, Wilkinson B, Bradley J, Gruben D, et al. Tofacitinib versus methotrexate in rheumatoid arthritis. *N Engl J Med*. 2014;370:2377–86.
35. Fleischmann R, Mysler E, Hall S, Kivitz AJ, Moots RJ, Luo Z, et al. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. *Lancet*. 2017;390:457–68.
36. Cohen SB, Pope J, Haraoui B, Mysler E, Diehl A, Lukic T, et al. Efficacy and safety of tofacitinib modified-release 11 mg once daily plus methotrexate in adult patients with rheumatoid arthritis: 24-week open-label phase results from a phase 3b/4 methotrexate withdrawal non-inferiority study (ORAL Shift). *RMD Open*. 2021;7:e001673.
37. Ytterberg SR, Bhatt DL, Mikuls TR, Koch GG, Fleischmann R, Rivas JL, et al. Cardiovascular and cancer risk with tofacitinib in rheumatoid arthritis. *N Engl J Med*. 2022;386:316–26.
38. Sande SZ, Seng L, Li J, D'Agostino R. Statistical learning in medical research with decision threshold and accuracy evaluation. *J Data Sci*. 2021;19:634–57.
39. Nakatsu RT. An evaluation of four resampling methods used in machine learning classification. *IEEE Intell Syst*. 2021;36:51–7.
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–57.
41. Hoang U, Liyanage H, Coyle R, Godden C, Jones S, Blair M, et al. Determinants of inter-practice variation in childhood asthma and respiratory infections: cross-sectional study of a national sentinel network. *BMJ Open*. 2019;9:e024372.
42. Smith S, Morbey R, de Lusignan S, Pebody RG, Smith GE, Elliot AJ. Investigating regional variation of respiratory infections in a general practice syndromic surveillance system. *J Public Health (Oxf)*. 2021;43:e153–60.
43. Sunzini F, McInnes I, Siebert S. JAK inhibitors and infections risk: focus on herpes zoster. *Ther Adv Musculoskelet Dis*. 2020;12:1759720x20936059.
44. Redeker J, Albrecht K, Kekow J, Burmester GR, Braun J, Schäfer M, et al. Risk of herpes zoster (shingles) in patients with rheumatoid arthritis under biologic, targeted synthetic and conventional synthetic DMARD treatment: data from the German RABBIT register. *Ann Rheum Dis*. 2022;81:41–7.
45. 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;31:285–92.
46. Scott DL, Steer S. The course of established rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. 2007;21:943–67.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.