RHEUMATOLOGY ADVANCES IN PRACTICE

Rheumatology Advances in Practice 2021;00:1–10
https://doi.org/10.1093/rap/rkab087
Advance Access Publication 13 November 2021

Original article

Prediction of sustained biologic and targeted synthetic DMARD-free remission in rheumatoid arthritis patients

Theresa Burkard [1], Ross D. Williams [2], Enriqueta Vallejo-Yagüe [1], Thomas Hügle[3], Axel Finckh [4], Diego Kyburz [5] and Andrea M. Burden [1]

Abstract

Objectives. The aim was to develop a prediction model of sustained remission after cessation of biologic or targeted synthetic DMARD (b/tsDMARD) in RA.

Methods. We conducted an explorative cohort study among b/tsDMARD RA treatment episode courses stopped owing to remission in the Swiss Clinical Quality Management registry (SCQM; 2008–2019). The outcome was sustained b/tsDMARD-free remission of $\geq$12 months. We applied logistic regression model selection algorithms using stepwise, forward selection, backward selection and penalized regression to identify patient characteristics predictive of sustained b/tsDMARD-free remission. We compared c-statistics corrected for optimism between models. The three models with the highest c-statistics were validated in new SCQM data until 2020 (validation dataset).

Results. We identified 302 eligible episodes, of which 177 episodes (59%) achieved sustained b/tsDMARD-free remission. Two backward and one forward selection model, with eight, four and seven variables, respectively, obtained the highest c-statistics corrected for optimism of $c = 0.72$, $c = 0.70$ and $c = 0.69$, respectively. In the validation dataset (47 eligible episodes), the models performed with $c = 0.99$, $c = 0.80$ and $c = 0.74$, respectively, and excellent calibration. The best model included the following eight variables (measured at b/tsDMARD stop): RA duration, b/tsDMARD duration, other pain/anti-inflammatory drug use, quality of life (EuroQol), DAS28-ESR score, HAQ score, education, and interactions of RA duration and other pain/anti-inflammatory drug use and of b/tsDMARD duration and HAQ score.

Conclusion. Our results suggest that models with up to eight unique variables may predict sustained b/tsDMARD-free remission with good efficiency. External validation is warranted.

Key words: rheumatoid arthritis, biologic DMARD, targeted synthetic DMARD, remission, treatment discontinuation, biologics, Swiss Clinical Quality Management

Key messages

- Suggested prediction models of sustained biologic/targeted synthetic DMARD-free remission can be considered an advancement from the previous studies.
- Suggested prediction models may help clinicians decide who could successfully stop biologic/targeted synthetic DMARDs.
- External validation is warranted before application in clinical practice.

[1]Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, [2]Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands, [3]Department of Rheumatology, Lausanne University Hospital, University of Lausanne, Lausanne, [4]Department of Rheumatology, University Hospitals Geneva, Geneva and [5]Department of Rheumatology, University Hospital Basel, Basel, Switzerland

Submitted 2 July 2021; accepted 28 October 2021

Correspondence to: Andrea M. Burden, Department of Chemistry and Applied Biosciences, Pharmacoepidemiology, Institute of Pharmaceutical Sciences, Federal Institute of Technology (ETH) Zurich, Vladimir-Prelog-Weg 1-5/10, CH-8093 Zurich, Switzerland.
E-mail: andrea.burden@pharma.ethz.ch

CLINICAL SCIENCE

Introduction

The target of RA therapy is clinical remission, commonly defined as a 28-joint DAS (DAS28) <2.6 [1]. If remission cannot be achieved with several conventional synthetic (cs) DMARDs and CSs, the European Alliance of Associations for Rheumatology (EULAR) recommends the use of biological (b) DMARDs or targeted synthetic (ts) DMARDs [2]. EULAR also recommends the tapering of b/tsDMARDs if a patient is in clinical remission but not stopping them. However, several studies report a substantial proportion of b/tsDMARD stops owing to remission (2–18% depending on the cohort) [3–6]. Given that there is no guidance on when, or if, patients may initiate b/tsDMARD drug holidays, but it is often the patient's wish (or the physician's suggestion) to stop b/tsDMARD treatment owing to remission, empirical evidence is needed.

To date, several studies have assessed patient characteristics associated with b/tsDMARD- free remission, including one meta-analysis [7–10]. These studies identified predictors of b/tsDMARD-free remission such as seronegativity or RA duration. However, most of these studies suggested predictors based on univariate analysis or interpretation of each individual patient characteristic in a multivariable model, which is prone to error [11].

The aim of this investigation was to provide insight into who might do well if b/tsDMARDs are stopped owing to remission. Thus, we developed a prediction model of sustained b/tsDMARD-free remission of ≥12 months.

Methods

Study design and data source

We conducted an explorative cohort study among RA patients in the Swiss Clinical Quality Management in Rheumatic Diseases (SCQM). The SCQM was established by the Swiss Society of Rheumatology in 1997, and RA diagnoses were made by a board-certified rheumatologist [12]. Regulatory health authorities in Switzerland have recommended continuous monitoring with the SCQM system for all patients receiving b/tsDMARDs [13]. Patients come from a wide range of settings (i.e. private practices in addition to academic centres) and are usually enrolled before the initiation of a b/tsDMARD to allow its nationwide monitoring [12]. Detailed information on data capture in SCQM can be found in Supplementary Data S1, available at *Rheumatology Advances in Practice* online.

The study was reviewed by the ethics commission of the Canton of Zurich (BASEC-Nr Req-2020-01497). We received pseudo-anonymized data without access to the code key; therefore, a full ethics authorization was waived by the commission. All participants of the SCQM have signed informed consent before enrolment in the analysis, in accordance with the Declaration of Helsinki, and participants may be withdrawn at any time.

Study population

We identified all treatment episodes of bDMARDs (i.e. abatacept, adalimumab, certolizumab, etanercept, golimumab, infliximab, rituximab and tocilizumab) or tsDMARDs (i.e. baricitinib and tofacitinib) that were stopped between 1 August 2008 and 31 July 2018 for ≥31 days. The day of b/tsDMARD stop is termed the index date. We included only episodes that were flagged with the reason for stopping being remission or other reasons according to the physician, which helps to ensure that the treatment stop was intended. Moreover, these reasons for stopping had the longest b/tsDMARD-free periods in a previous descriptive analysis using the SCQM data [10]. Remission was defined by clinicians' assessment of RA disease activity (i.e. DAS28-ESR score <2.6, DAS28-CRP score <2.6) [1] or by the patient-driven RA Disease Activity Index (RADAI) score <1.5) [14]. Finally, only episodes with an RA disease activity measure indicating remission within 4 months before the therapy stop date were eligible (Fig. 1). Patients contributed one or more episodes if the above eligibility criteria were met.

Outcome

We assessed two outcomes in two different analyses. The first outcome was any time restart of a b/tsDMARD after therapy stop (i.e. index date) to assess cumulative incidence of b/tsDMARD restart.

Second, to develop a prediction model of sustained b/tsDMARD-free remission, the outcome was defined as a restart of a b/tsDMARD ≥365 days after therapy stop (i.e. index date) or no restart at all after index date if the patient was followed up for ≥365 days.

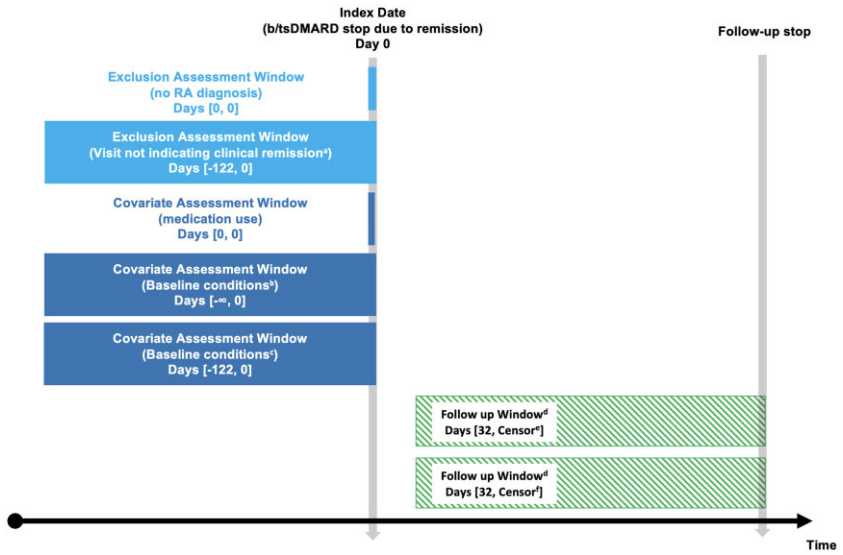
Follow-up

We followed all episodes from the index date until restart of a b/tsDMARD or censoring owing to end of a patient record, whichever happened first. Data were available until 31 July 2019, which allowed every episode a follow-up of ≥365 days.

Covariates

Covariates were measured at the index date and included patient demographics, clinical information (e.g. RA disease activity measures), RA medication use (i.e. DMARDs, other pain/anti-inflammatory drug use, such as analgesics or NSAID use), and certain co-morbidities, such as cardiovascular diseases or musculoskeletal diseases. Missing information was handled using a missing category, because we did not aim to build a prediction model based on imputed values and because missingness also conveys information in prediction modelling according to Sperrin *et al.* [15]. However, to minimize missingness, we carried forward information from the nearest record within defined lookback windows, which can be seen in Supplementary Table S1, available at *Rheumatology Advances in Practice* online.

Fig. 1 Sketch of the study composition



[a]Remission was defined as DAS28-ESR score of <2.6, DAS28-CRP score of <2.6 or RADAI score of <1.5. [b]Time-invariant variables: patient demographics, BMI, smoking, alcohol consumption, activity level, prescriber, co-morbidities, RA advanced treatment duration, RA duration, family history of rheumatic diseases, ACPAs and RF. [c]Highly time-varying variables: RA disease scores and health assessment scores, and other pain/anti-inflammatory drugs. [d]Treatment gaps between b/tsDMARDs of $\leq$31 days were considered as continuous treatment spells. [e]Earliest of b/tsDMARD restart or end of patient record when assessing cumulative incidence of b/tsDMARD restart. [f]Earliest of b/tsDMARD restart or 365 days of follow-up when assessing sustained b/tsDMARD-free remission. bDMARD: biologic DMARD; RADAI: RA disease activity index; tsDMARD: targeted synthetic DMARD.

Given that interactions between variables may improve a prediction model [16], we used the publicly available open source software package InteractionTransformer [17] in Python v.3.8.5 to identify relevant interactions that might improve the model. The InteractionTransformer applies machine learning algorithms to detect interactions associated with the outcome and orders them by so-called SHAP (SHapley Additive ExPlanations) [18] values. We added the top 15 identified interaction terms as additional candidate covariates (Supplementary Data S2, available at *Rheumatology Advances in Practice* online).

Data analysis

We described covariates among the study population at the index date. Furthermore, stratified by b/tsDMARD agent, we estimated Kaplan–Meier curves to assess differences in cumulative incidences of anytime b/tsDMARD restart.

To predict sustained b/tsDMARD-free remission, we performed the following logistic regression selection algorithms, selecting variables based on the highest χ^2 scores, which estimate odds ratios (ORs) with 95% CI of sustained b/tsDMARD-free remission. First, we carried out a stepwise selection, which is a modification of the forward selection, in which effects that enter the model do not necessarily stay. We used a variable entry level of $P \leq 0.5$ and a variable stay level of $P \leq 0.1$ and $P \leq 0.05$ (separately). Second, we carried out a forward selection using a variable entry level of $P \leq 0.5$ and a maximum of 4, 6, 8 and 10 variables allowed (separately). Third, we carried out a backward selection using a variable stay level of $P \leq 0.075$ (instead of $P \leq 0.1$, to obtain a more parsimonious model) and $P \leq 0.05$ (separately). Furthermore, we performed a generalized linear model selection algorithm called least absolute shrinkage and selection operator (LASSO). Model fit was tested using the Hosmer–Lemeshow test and Stukel's test (held for all models). Obtained prediction models were assessed for discrimination properties using c-statistics. To reduce optimism, we performed bootstrapping and estimated optimism-corrected c-statistics for each obtained model [19]. Using the three models with the highest c-statistics corrected for optimism, per model, we additionally compared observed and predicted probabilities to assess calibration [19]. Because cumulative incidences of b/tsDMARD restart differed by agent, we forced all agents into the models to adjust our models for potentially differing patient characteristics by agent. Additionally, we forced the date of remission into the model to adjust for time trends.

In *post hoc* analyses, we used a similar SCQM dataset with follow-up until December 2020. We identified episodes of b/tsDMARD stops between August 2018 and December 2019 (to ensure 365 days of follow-up). Thus, all of the episodes were new episodes, and we

termed this the data validation dataset. In this validation dataset, we tested the same three models with the highest c-statistics corrected for optimism for discrimination and calibration. We performed all statistical analyses using SAS statistical software v.9.4 (SAS Institute, Cary, NC, USA).

Results

Patient characteristics at b/tsDMARD stop

We identified 302 eligible episodes of b/tsDMARD stops owing to remission among 287 patients (95.0% unique patients) between August 2008 and July 2018 (see flow chart in Fig. 2). An overview of the patient characteristics at b/tsDMARD stop is provided in Table 1. The mean age at the index date was 57.2 years, and the majority were women (70.5%). Full information on patient characteristics is available in Supplementary Table S2, available at *Rheumatology Advances in Practice* online.

Cumulative incidence of b/tsDMARD restart stratified by agent

From the cumulative incidence curves, we identified that the median time to any b/tsDMARD restart was 713 days (25th percentile, 182 days; 75th percentile not available because 29% of patients never restarted b/tsDMARDs). The cumulative incidence of b/tsDMARD restart did not differ significantly between individual agents according to Gray's test ($P > 0.05$), but differences were notable upon visual inspection (Supplementary Fig. S1, Supplementary Table S3, available at *Rheumatology Advances in Practice* online).

Prediction of sustained b/tsDMARD-free remission

Among the 302 episodes where b/tsDMARD therapy was stopped owing to remission, 177 episodes (59%) achieved sustained b/tsDMARD-free remission of $\geq$12 months.

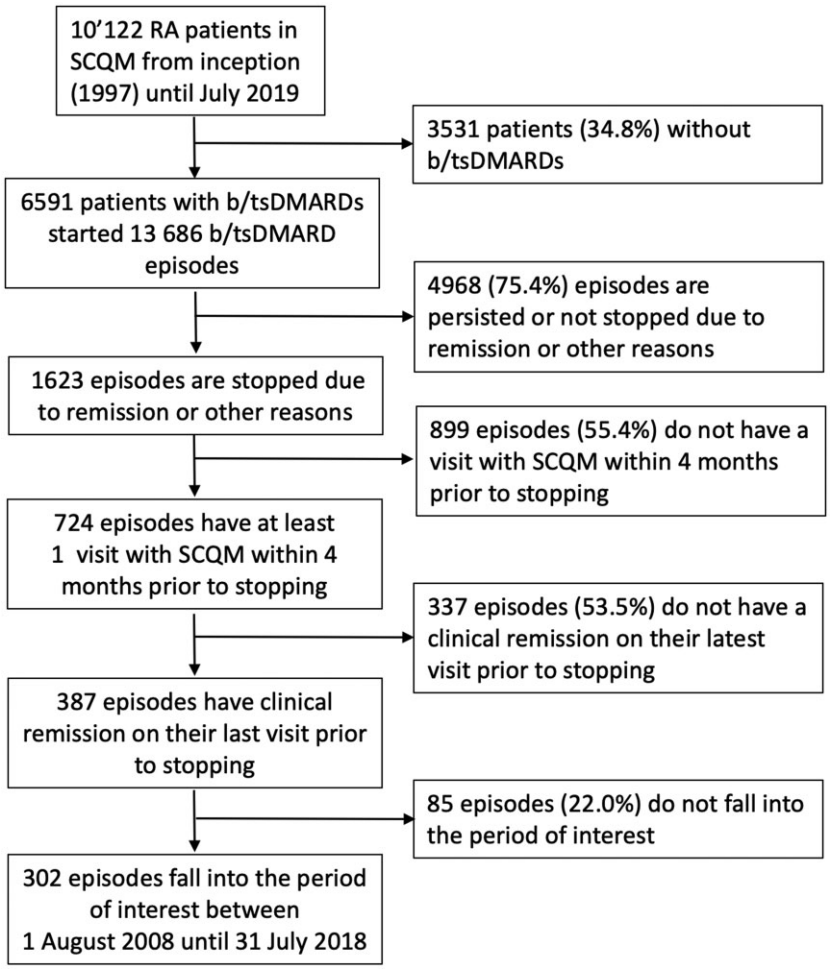
Selected variables according to the model selection procedures with various specifications are listed in Table 2. All model algorithms selected RA duration. Other frequently selected variables were heart disease, education, family history of rheumatic diseases, MTX use, and use of other pain/anti-inflammatory drugs. The most commonly selected interaction term was the interaction between the variables RA duration and other pain/anti-inflammatory drug use. The backward selection algorithm with a variable stay level of $P \leq 0.075$ had the highest naïve and optimism-corrected c-statistics of 0.82 and 0.72, respectively, and contained seven variables. This model contained RA duration, pain/anti-inflammatory drug use and its interaction, family history of rheumatic diseases, duration of b/tsDMARD therapy, HAQ, and the interaction of duration of b/tsDMARD therapy and HAQ, quality of life (EuroQoL), DAS28-ESR and education. The second highest c-statistics, corrected for optimism ($c = 0.70$), was obtained by the backward selection algorithm with a variable stay level of $P \leq 0.05$. This model contained only four variables (RA duration, pain/anti-inflammatory drug use and its interaction, EuroQoL and education). The forward selection with a maximum of 10 variables obtained the third highest c-statistics corrected for optimism ($c = 0.69$). Receiver operating characteristic curves and calibration graphs of these three models with the highest optimism-corrected c-statistics are displayed in Supplementary Figs S2–S7, available at *Rheumatology Advances in Practice* online, respectively.

We subsequently tested these three models for discrimination and calibration using the validation dataset in *post hoc* analyses. In the validation dataset, we identified 47 eligible episodes of b/tsDMARD stop owing to remission between August 2018 and December 2019, of which 28 episodes achieved sustained b/tsDMARD-free remission. Selected patient characteristics are displayed in Supplementary Table S4, available at *Rheumatology Advances in Practice* online; of note, all patients had information on RA duration, but 47% missed information on EuroQol. The backward selection procedure with a variable stay level of $\leq$0.075 obtained a c-statistic of 0.99, the forward selection with a maximum of 10 variables a c-statistic of 0.80, and the backward selection model with a variable stay level of $\leq$0.05 a c-statistic of 0.74. All three models obtained excellent calibration when comparing observed and predicted probabilities (Spiegelhalter test passed) with respective graphs provided in Supplementary Figs S8–S10, available at *Rheumatology Advances in Practice* online. The equations of the three best models are provided as Equations 1–3.

Equations 1–3. Logit functions of sustained b/tsDMARD-free remission with highest discrimination properties: (1) c-statistics in validation dataset of 0.99; (2) c-statistics in validation dataset of 0.80; and (3) c-statistics in validation dataset of 0.74.

1. Log odds of sustained b/tsDMARD-free remission = 2.66 − 0.44 × other pain or anti-inflammatory drug use − 1.98 × (RA duration >4–8 years) − 0.91 × (RA duration >8 years) − 17.22 × missing RA duration − 0.14 × (EuroQol 70–80) + 0.88 × (EuroQol >80) − 1.34 × missing EuroQol − 1.29 × (secondary education level) − 1.54 × (higher education level) − 0.30 × (missing education) + 3.59 × (RA duration >4–8 years and other pain or anti-inflammatory drug use) − 0.28 × (RA duration >8 years and other pain or anti-inflammatory drug use) + 15.64 × (missing RA duration and other pain or anti-inflammatory drug use) − 0.44 × family history of rheumatic diseases + 0.98 × missing family history of rheumatic diseases + 0.50 × (b/tsDMARD duration of >1–3 years) − 1.16 × (b/tsDMARD duration of >3 years) − 0.32 × (HAQ score 0.5–1) − 2.31 × (HAQ score >1) − 0.04 × (missing HAQ score) − 0.34 × (DAS28-ESR score of 2.6–3.2) − 1.43 × (DAS28-ESR score of >3.2) − 0.83 × (missing DAS28-ESR score) − 0.66 × (b/tsDMARD duration of >1–3 years and HAQ score 0.5–1) − 0.44 × (b/tsDMARD duration of >1–3 years and

FIG. 2 Flow chart of the study population



Treatment gaps of ≤31 days were considered as continuous treatment episodes. bDMARD: biologic DMARD; tsDMARD: targeted synthetic DMARD.

HAQ score >1) − 0.15 × (b/tsDMARD duration of >1–3 years and missing HAQ score) + 1.21 × (b/tsDMARD duration of >3 years and HAQ score 0.5–1) + 3.43 × (b/tsDMARD duration of >3 years and HAQ score >1) + 2.16 × (b/tsDMARD duration of >3 years and missing HAQ score).

2. Log odds of sustained b/tsDMARD-free remission = 1.64 − 0.74 × other pain or anti-inflammatory drug use − 1.84 × (RA duration >4–8 years) − 1.17 × (RA duration >8 years) − 16.15 × missing RA duration − 0.40 × family history of rheumatic diseases + 0.66 × missing family history of rheumatic diseases + 0.45 × (b/tsDMARD duration of >1–3 years) − 0.20 × (b/tsDMARD duration of >3 years) + 0.42 × MTX use − 0.60 × heart disease + 0.60 × ever infections + 3.04 × (RA duration >4–8 years and other pain or anti-inflammatory drug use) + 0.52 × (RA duration >8 years and other pain or anti-inflammatory drug use) + 15.58 × (missing RA duration and other pain or anti-inflammatory drug use).
3. Log odds of sustained b/tsDMARD-free remission = 2.32 − 0.75 × other pain or anti-inflammatory drug use − 2.34 × (RA duration >4–8 years) − 1.36 × (RA duration >8 years) − 16.39 × missing RA duration − 0.21 × (EuroQol 70–80) + 0.68 × (EuroQol >80) − 0.44 × missing EuroQol − 1.03 × (secondary education level) − 1.12 × (higher education level) − 0.30 × (missing education) + 3.34 × (RA duration >4–8 years and other pain or anti-inflammatory drug use) + 0.22 × (RA duration >8 years and other pain or anti-inflammatory drug use) + 15.84 × (missing RA duration and other pain or anti-inflammatory drug use)

We suggest that the model should not be used in patients with missing RA duration ($n = 6$) or baricitinib use ($n = 1$) because group sizes were small and there were no patients with missing RA duration in eligible episodes in the validation dataset.

TABLE 1 Selected patient characteristics at recorded stop date of biologic/targeted synthetic DMARD owing to remission

Patient characteristic at index date	Study population ($n = 302$)
Age, mean (S.D.), years	57.2 (14.7)
Women	213 (70.5)
Men	89 (29.5)
Primary education[a]	94 (31.1)
Secondary education[a]	124 (41.1)
Higher education[a]	56 (18.5)
Missing education information	28 (9.3)
BMI, mean (S.D.), kg/m^2	26.0 (5.2)
RA duration, median (IQR), years[b]	7.0 (3.7–12.2)
No family history of rheumatic diseases[c]	202 (66.9)
Family history of rheumatic diseases[c]	89 (29.5)
Missing information on family history	11 (3.6)
RF negative	89 (29.5)
RF positive	202 (66.9)
Missing RF information	11 (3.6)
Index date 2008–2012	103 (34.1)
Index date 2013–2015	96 (31.8)
Index date 2016–2018	103 (34.1)
Duration of b/tsDMARD, median (IQR)	2.2 (1.0–4.3)
TNFi	212 (70.2)
Non-TNFi	81 (26.8)
tsDMARD	9 (3.0)
DAS28-ESR score, mean (S.D.)	2.0 (0.8)
DAS28-CRP score, mean (S.D.)	1.9 (0.6)
RADAI score, median (IQR)	1.4 (0.7–2.8)
HAQ score, median (IQR)	0.25 (0–0.75)
EuroQoL score, mean (S.D.)	79.4 (69–100)
csDMARD use[d]	197 (65.2)
MTX	141 (46.7)
LEF	40 (13.3)
Use of other pain/anti-inflammatory medication[e]	120 (39.7)
Prednisone use	41 (13.6)
Heart diseases	25 (8.3)
Hypertension	102 (33.8)
Hyperlipidaemia (diagnosis or treatment)	63 (20.9)
OA, hip/knee replacement	79 (26.2)
Osteoporosis or fracture (diagnosis or treatment)	70 (23.2)
Prior infections	33 (19.5)

Full information on patient characteristics can be seen in Supplementary Table S2, available at *Rheumatology Advances in Practice* online. Values are expressed as *n* (%) unless otherwise specified. [a]Primary education includes compulsory school, secondary education, vocational training and higher education, high school or university. [b]RA duration assessed from diagnosis until index date; if diagnosis date was not available, we assessed RA duration from first symptoms minus 1 year. [c]Family anamnesis includes RA, ankylosing spondylitis, psoriasis, PsA, chronic IBD and other spondyloarthropathies (e.g. reactive arthritis). [d]csDMARDs used include MTX, LEF, SSZ, chloroquine, AZA, CsA and CYC. [e]Use of other pain/anti-inflammatory medication include coxibs, other analgetics, conventional non-NSAIDs, antidepressants, paracetamol and opiates. bDMARD: biological DMARD; csDMARD: conventional synthetic DMARD; NA: not applicable; RADAI: RA disease activity index; TNFi: TNF-α inhibitor; tsDMARD: targeted synthetic DMARD.

Discussion

This cohort study included 302 episodes of b/tsDMARD being stopped owing to remission between 2008 and 2018, among which 59% of episodes achieved sustained b/tsDMARD-free remission of $\geq$12 months. When predicting sustained b/tsDMARD-free remission, the highest c-statistics corrected for optimism of $c = 0.72$ was achieved by a model containing eight individual predictors and two interaction terms. Moreover, more parsimonious models of only seven individual predictors and one interaction term, and of four individual predictors and the same interaction term obtained a similar high c-statistic corrected for optimism of $c = 0.69$ and $c = 0.70$, respectively. All three models also performed well in new patient episodes in *post hoc* analyses, with $c = 0.99$, $c = 0.80$ and $c = 0.74$, respectively.

TABLE 2 Variables selected by different model selection algorithms predicting sustained biologic/targeted synthetic DMARD-free remission

Parameter	Stepwise selection, entry level $P \leq 0.5$		Forward selection, entry level $P \leq 0.5$				Backward selection		LASSO selection
	Stay level $P \leq 0.1$	Stay level $P \leq 0.05$	≤10 variables	≤8 variables	≤6 variables	≤4 variables	Stay level $P \leq 0.075$	Stay level $P \leq 0.05$	Regularization parameter: 0.49
Naïve c-statistics	0.71	0.68	0.77	0.72	0.71	0.70	0.82	0.77	0.75
Optimism-corrected c-statistics	0.64	0.61	0.694	0.64	0.64	0.63	0.72	0.70	0.686
Heart diseases	x	–	x	x	x	–	–	–	–
MTX use	x	–	x	x	x	x	–	–	x
RA duration	x	x	x	x	x	x	x	x	x
Other pain/anti-inflammatory drug use	–	–	x	x	–	–	x	x	x
Interaction of RA duration and other pain/anti-inflammatory drug use	–	–	x	–	–	–	x	x	x
Family history of RD	–	–	x	x	x	–	x	–	–
Duration of b/tsDMARD use	–	–	x	x	–	–	x	–	x
Prior infections	–	–	x	x	–	–	–	–	–
EuroQol score	–	–	–	–	–	–	x	x	–
DAS28-ESR score	–	–	–	–	–	–	x	–	–
HAQ score	–	–	–	–	–	–	x	–	–
Interaction of HAQ score and duration of b/tsDMARD use	–	–	–	–	–	–	x	–	–
Education	–	–	–	–	–	–	x	x	–

x: variable was selected; –: variable was not selected. b/tsDMARD: biologic or targeted synthetic DMARD; DAS28-ESR: DAS based on 28 joints and ESR; EuroQol: quality of life (EQ-5D); LASSO: least absolute shrinkage and selection operator; RD: rheumatic diseases.

The model with the best optimism-corrected c-statistic in this study ($c = 0.72$) contained seven variables (i.e. RA duration, pain/anti-inflammatory drug use, the interaction between RA duration and pain/anti-inflammatory drug use, family history of rheumatic diseases, duration of b/tsDMARD therapy, HAQ, the interaction b/tsDMARD duration and HAQ, EuroQoL, DAS28-ESR and education). Our finding is consistent with a meta-analysis of 34 studies, which reported that low disease activity assessed through DAS28, lower HAQ and shorter symptom or disease duration were often identified as predictors for the successful discontinuation of biological DMARD [7]. Furthermore, one study in the meta-analysis identified lower RaQol scores [20] (adaptation of the EuroQol questionnaire for RA patients) to be predictive of successful discontinuation of TNF inhibitors [21]. Also, a randomized controlled trial included in the meta-analysis suggested that absence of glucocorticoids or other pain/anti-inflammatory drugs was a predictor of sustained tocilizumab-free remission [22]. Although several of our algorithms selected other pain/anti-inflammatory drug use (e.g. NSAIDs), prednisone use was not selected by any of the algorithms. Furthermore, although family history of rheumatic diseases, education levels, or heart disease and infections were identified as predictors in our best models, these were not suggested as predictors of sustained b/tsDMARD-free remission in previous studies. However, this might be because this information was not available in previous studies. Education is likely to approximate two aspects (income and expectation), both of which might drive the odds of achieving sustained b/tsDMARD-free remission in the same direction: patients with lower education might not restart b/tsDMARDs (owing to 10% co-payments), and patients with a higher education might restart b/tsDMARDs (potentially owing to higher expectations). Thus, our findings add to current evidence regarding who might be taken off b/tsDMARD if in remission, while also taking into account social aspects, but our findings require confirmation in other cohorts.

Although female sex was previously suggested to be negatively associated with bDMARD-free remission [9, 10, 21], a study of a randomized controlled trial assessing sustained infliximab-free remission reported that sex was not a relevant variable after adjusting for other

confounding factors [23]. Moreover, smoking [24] and seronegativity [7, 8] were previously reported to be predictors of sustained bDMARD-free remission. However, sex, smoking or seronegativity were not selected into any of our models. Inconsistencies with previous findings concerning these patient characteristics are probably attributable to confounding, because the treating rheumatologist decides which b/tsDMARD agent to prescribe to which patient. This is a particular problem if sample size is small. Thus, by controlling our results for b/tsDMARD agent, we controlled our analysis for differing characteristics of patients receiving different b/tsDMARD agents and thereby prevented spurious findings.

A recently performed study in the same database (SCQM) by Arnold *et al.* [9] assessed the time to loss of remission defined by either DAS28 $\geq$ 2.6 or b/tsDMARD restart. Arnold *et al.* reported that 76% of patients lost remission within a median time of 9 months, whereas we observed that 59% of patients achieved sustained b/tsDMARD-free remission of $\geq$12 months. Differences are likely to lie in differing definitions of outcomes, because patients in our study might have lost clinical remission but were not considering restarting b/tsDMARDs, which can be considered a harder outcome than disease activity measurements. Despite different analytical approaches, our studies agree that age and seropositivity had no influence, but that disease duration and csDMARD use did matter. Although our study did not select csDMARD use in particular, it selected MTX use, which accounts for most csDMARD use. This observation is consistent with a cohort study, which reported that MTX was observed to help patients maintain low disease activity after discontinuation of adalimumab [25].

The strong decrease from naïve to optimism-corrected c-statistics ($c = 0.82$ to $c = 0.72$) in the model with the highest c-statistics shows that models with many predictors are often overfitted [16]. Nevertheless, our model with the highest optimism-corrected c-statistics performed in an outstanding manner in the validation dataset, with c-statistics of 0.99. However, it is possible that this result is attributable to a low number of patients, and it would benefit from further validation. Moreover, a parsimonious model with fewer variables will be preferable, not only for external validation, but also for application in clinical practice. Thus, the models with the second or third highest c-statistics corrected for optimism ($c = 0.70$ and $c = 0.69$, respectively) containing fewer variables might be preferable models and also valid candidates to be investigated further. Furthermore, the more parsimonious models also performed well when testing the discrimination ($c = 0.74$ and $c = 0.80$, respectively) and calibration in new patient episodes. However, external validation is important, and we invite future researchers to assess the robustness and generalizability of all models.

A strength of our study is the increased predictive power of sustained b/tsDMARD-free remission by the introduction of an interaction term, which was identified through machine learning approaches. None of the previous assessments that identified patient characteristics predictive of b/tsDMARD-free remission considered interaction terms. However, it seemed that the interaction term made the difference concerning discrimination properties of our prediction models, because those which scored highest included at least one interaction term. Furthermore, we adjusted our analysis for confounding by b/tsDMARD agent and index date, which controlled for the patient selection and time trends. Internal validation by bootstrapping in addition to testing of discrimination and calibration in new patient episodes performed well for the three tested models. Moreover, both our overall population and the validation population had an equal balance concerning the number of outcomes, which is beneficial.

However, our analysis is also subject to several limitations. First, external validation for prediction models is needed and was not performed in this study because we did not have a suitable external dataset at hand. However, we invite investigators to assess our model in external cohorts. Second, potential predictors, such as imaging-related variables or laboratory biomarkers such as T-cell frequency, were not available in the dataset. Moreover, we did not assess use of csDMARDs and prednisone use in a time-varying manner, which might have prolonged the duration of b/tsDMARD-free remission. However, this time-varying information is likely to be a proxy for the outcome (i.e. b/tsDMARD restart) and thereby lies on the causal pathway. Therefore, for the development of a prediction model for sustained b/tsDMARD-free remission, it was expedient to use only the information at the index date (i.e. b/tsDMARD stop). Third, the small sample of certain categories and resulting large β coefficients might limit the use of the identified prediction models (e.g. for patients with missing RA duration). However, the patient population in our validation dataset did not have missing RA duration and obtained good discrimination and excellent calibration. Thus, the good performance was not driven by missing data. Fourth, the small sample size of the validation dataset adds some uncertainty to the obtained c-statistics in the validation step in *post hoc* analyses. Fifth, some variables had a high level of missingness (>20%), such as quality of life assessments (SF12, EuroQol and HAQ) and the patient's view on disease activity (i.e. RADAI), which might have skewed our analysis slightly. However, by adjusting for the categories of missing values, we took them into account. Moreover, it seems that prediction models were still suitable for patients with a high level of missing EuroQol information in the validation dataset (47%), because they obtained good discrimination and excellent calibration.

Conclusion

In this study, we developed prediction models to predict b/tsDMARD-free sustained remission of $\geq$12 months while rigorously controlling for confounding and with the

use of interaction terms. Thus, our investigation can be considered an advancement from the previous studies identifying patient characteristics associated with successful b/tsDMARD discontinuation. Suggested prediction models with up to eight variables obtained c-statistics corrected for optimism of $c = 0.74$, $c = 0.80$ and $c = 0.99$ in new patient episodes of the same database. These results are promising, but the accuracy of any prediction algorithm depends on contextual factors. Thus, external validation of our models is warranted before application in a clinical setting.

Acknowledgements

A list of rheumatology offices and hospitals that are contributing to the SCQM registries can be found on www.scqm.ch/institutions. The SCQM is financially supported by pharmaceutical industries and donors. A list of financial supporters can be found on www.scqm.ch/sponsors. The professorship of A.M.B. is partly endowed by the Swiss Pharmacists Association (PharmaSuisse) and the ETH Foundation. We thank Adrian Martinez de la Torre for his statistical advice in this investigation.

Theresa Burkard: substantial contributions to the conception or design of the work; analysis of data for the work, drafting the work, final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ross Williams, Enriqueta Vallejo-Yagüe, Thomas Hügle, Axel Finckh, Diego Kyburz: substantial contributions to the design of the work; interpretation of data for the work; revising the work critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

A.M.B: substantial contributions to the conception or design of the work; the acquisition, interpretation of data for the work; revising the work critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding: No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Disclosure statement: T.H. reports grants or contracts from Pfizer, AbbVie and Novartis, consulting fees and payment or honoraria from Glaxo Smith Kline, Janssen, Pfizer, Novartis, AbbVie, Roche, Merck Sharp Dohme, Sanofi, Bristol Myers Squibb, Eli Lilly and UCB Pharma. A.F. reports grants or contracts from Pfizer, Bristol Myers Squibb and Novartis, consulting fees from AbbVie AB2Bio, Bristol Myers Squibb, Gilead, Pfizer and Viatris, payment or honoraria from Pfizer and Eli-Lilly. D.K. reports grants or contracts from AbbVie and Novartis, consulting fees from AbbVie, Gilead, Novartis, Pfizer and Eli Lilly, payment or honoraria from AbbVie, Gilead, Novartis and Pfizer. The remaining authors have declared no conflicts of interest.

Data availability statement

The data analysed in this study is available from the corresponding author upon reasonable request and after having received approval from the license holder (SCQM). The code used in this study is available from the corresponding author upon reasonable request.

Supplementary data

Supplementary data are available at *Rheumatology Advances in Practice* online.

References

1 Felson DT, Smolen JS, Wells G *et al.* American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Ann Rheum Dis 2011;70:404–13.

2 Smolen JS, Landewé RBM, Bijlsma JWJ *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis 2020;79:685–99.

3 Ebina K, Hashimoto M, Yamamoto W *et al.* Drug tolerability and reasons for discontinuation of seven biologics in 4466 treatment courses of rheumatoid arthritis—the ANSWER cohort study. Arthritis Res Ther 2019;21:91.

4 Youssef P, Marcal B, Button P *et al.* Reasons for biologic and targeted synthetic disease-modifying antirheumatic drug cessation and persistence of second-line treatment in a rheumatoid arthritis dataset. J Rheumatol 2020;47:1174–81.

5 Ebina K, Hashimoto M, Yamamoto W *et al.* Drug tolerability and reasons for discontinuation of seven biologics in elderly patients with rheumatoid arthritis - The ANSWER cohort study. PLoS One 2019;14:e0216624.

6 Strand V, Miller P, Williams SA *et al.* Discontinuation of biologic therapy in rheumatoid arthritis: analysis from the Corrona RA Registry. Rheumatol Ther 2017;4:489–502.

7 Schlager L, Loiskandl M, Aletaha D, Radner H. Predictors of successful discontinuation of biologic and targeted synthetic DMARDs in patients with rheumatoid arthritis in remission or low disease activity: a systematic literature review. Rheumatology (Oxford) 2020;59:324–34.

8 Haschka J, Englbrecht M, Hueber AJ *et al.* Relapse rates in patients with rheumatoid arthritis in stable

remission tapering or stopping antirheumatic therapy: interim results from the prospective randomised controlled RETRO study. Ann Rheum Dis 2016;75:45–51.

9 Arnold S, Jaeger VK, Scherer A *et al.* Discontinuation of biologic DMARDs in a real-world population of patients with rheumatoid arthritis in remission: outcome and risk factors. Rheumatology (Oxford) 2021. doi:10.1093/rheumatology/keab343.

10 Burkard T, Vallejo-Yagüe E, Hügle T, Finckh A, Burden AM. AB0206 description and follow-up of rheumatoid arthritis patients who stopped b/tsDMARD therapy, stratified by cessation reason [abstract]. Ann Rheum Dis 2021;80:1128.

11 Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907–16.

12 Uitz E, Fransen J, Langenegger T, Stucki G. Clinical quality management in rheumatoid arthritis: putting theory into practice. Rheumatology (Oxford) 2000;39:542–9.

13 Finckh A, Simard JF, Duryea J *et al.*; for the Swiss Clinical Quality Management in Rheumatoid Arthritis Project. The effectiveness of anti-tumor necrosis factor therapy in preventing progressive radiographic joint damage in rheumatoid arthritis: a population-based study. Arthritis Rheum 2006;54:54–9.

14 Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology (Oxford) 2000;39:321–7.

15 Sperrin M, Martin GP, Sisk R, Peek N. Missing data should be handled differently for prediction than for description or causal explanation. J Clin Epidemiol 2020;125:183–7.

16 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35:1925–31.

17 Levy JJ, O'Malley AJ. Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. BMC Med Res Methodol 2020;20:171.

18 Lundberg SM, Erion G, Chen H *et al.* From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–7.

19 Steyerberg EW, Harrell FE, Borsboom GJJM *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81.

20 Tijhuis GJ, De Jong Z, Zwinderman AH *et al.* The validity of the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire. Rheumatology (Oxford) 2001;40:1112–9.

21 Saleem B, Keen H, Goeb V *et al.* Patients with RA in remission on TNF blockers: when and in whom can TNF blocker therapy be stopped? Ann Rheum Dis 2010;69:1636–42.

22 Nishimoto N, Amano K, Hirabayashi Y *et al.* Drug free REmission/low disease activity after cessation of tocilizumab (Actemra) Monotherapy (DREAM) study. Mod Rheumatol 2014;24:17–25.

23 Tanaka Y, Takeuchi T, Mimori T *et al.*; for the RRR study investigators. Discontinuation of infliximab after attaining low disease activity in patients with rheumatoid arthritis: RRR (remission induction by Remicade in RA) study. Ann Rheum Dis 2010;69:1286–91.

24 Van Den Broek M, Klarenbeek NB, Dirven L *et al.* Discontinuation of infliximab and potential predictors of persistent low disease activity in patients with early rheumatoid arthritis and disease activity score-steered therapy: Subanalysis of the BeSt study. Ann Rheum Dis 2011;70:1389–94.

25 Tanaka Y, Yamanaka H, Ishiguro N *et al.* Adalimumab discontinuation in patients with early rheumatoid arthritis who were initially treated with methotrexate alone or in combination with adalimumab: 1 year outcomes of the HOPEFUL-2 study. RMD Open 2016;2:e000189.