

Identification of Distinct Disease Activity Trajectories in Methotrexate-Naive Patients With Rheumatoid Arthritis Receiving Tofacitinib Over Twenty-Four Months

Vivian P. Bykerk,¹ Eun Bong Lee,² Ronald van Vollenhoven,³ David C. Gruben,⁴ Lara Fallon,⁵ John C. Woolcott,⁶ and Edward Keystone⁷

Objective. Tofacitinib is an oral JAK inhibitor for the treatment of rheumatoid arthritis (RA). To better understand tofacitinib treatment responses, we used group-based trajectory modeling to investigate distinct disease activity trajectories and associated baseline variables in patients with active RA.

Methods. This post hoc analysis used data from a phase III study of methotrexate-naive patients receiving tofacitinib 5 mg twice daily. Changes in the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate (DAS28-ESR) from baseline to month 24 were used in group-based trajectory modeling to identify distinct disease activity trajectories. Patient and disease characteristics, changes in radiographic progression and patient-reported outcomes, and safety up to month 24 were compared among trajectory groups.

Results. From 346 methotrexate-naive patients, 5 disease trajectory groups, defined by DAS28-ESR scores, were identified, which progressed from high disease activity (HDA) to remission (group 1, n = 28), to low disease activity (LDA) rapidly (group 2, n = 107), to moderate disease activity (group 3, n = 98), to LDA gradually (group 4, n = 46), or remained in HDA (group 5, n = 67), at month 24. At baseline, groups 1 and 2 generally had lower disease activity and more favorable patient-reported outcomes, compared with other groups. Improvements in radiographic progression and patient-reported outcomes over 24 months were generally consistent with DAS28-ESR-predicted disease activity trajectories. Adverse event rates were generally comparable across groups.

Conclusion. Distinct phenotypic subgroups identified heterogeneity in patients with RA normally analyzed as a single population. Trajectory modeling may enable separation of clinically meaningful subsets of patients with RA, and may help optimize treatment outcomes.

INTRODUCTION

Patients with rheumatoid arthritis (RA) exhibit wide variations in disease characteristics, sociodemographic factors, treatment

adherence, and health status, which can affect response to treatment (1). Group-based trajectory modeling is one method to identify groups of patients according to their predicted response to treatment over time, which may be influenced by baseline

ClinicalTrials.gov identifier: NCT01039688.

Supported by Pfizer Inc.

¹Vivian P. Bykerk, BSc, MD, FRCPC: Weill Cornell Medical College, New York, New York; ²Eun Bong Lee, MD, PhD: Seoul National University College of Medicine, Seoul, Republic of Korea; ³Ronald van Vollenhoven, MD, PhD: Amsterdam University Medical Centres, Amsterdam, The Netherlands; ⁴David C. Gruben, PhD: Pfizer Inc, Groton, Connecticut; ⁵Lara Fallon, PhD: Pfizer Canada, Montreal, Quebec, Canada; ⁶John C. Woolcott, PhD: Pfizer Inc, Collegeville, Pennsylvania; ⁷Edward Keystone, MD, FRCPC: Mount Sinai Hospital, Toronto, Ontario, Canada.

Dr. Bykerk has received consulting fees from AbbVie, Amgen, Bristol Myers Squibb, Gilead, Pfizer Inc, Sanofi-Regeneron, and UCB (less than \$10,000 each) and has received grants from Amgen, Bristol-Myers Squibb, Genentech, and UCB. Dr. Lee has received consulting fees from Pfizer Inc (less than \$10,000) and has received grants from GC Pharm and HANDOK Inc. Dr. van Vollenhoven has received consulting fees from AbbVie, AstraZeneca, Biotest, Bristol Myers Squibb, Celgene, Crescendo, Eli Lilly and Company, GlaxoSmithKline, Janssen, Merck, Novartis, Pfizer Inc, Roche, UCB, and Vertex (less than \$10,000 each) and has received grants from AbbVie, Amgen, Bristol Myers Squibb, GlaxoSmithKline, Pfizer Inc, Roche, and UCB. Drs. Gruben, Fallon, and Woolcott are employees and stockholders of Pfizer Inc. Dr. Keystone has received consulting fees from AbbVie, Amgen, Bristol Myers Squibb, Celltrion,

Eli Lilly and Company, Gilead, Janssen, Merck, Myriad Autoimmune, Pfizer Inc, Roche, Sandoz, Sanofi-Genzyme, and Samsung Bioepis (less than \$10,000 each) and has received grants from Amgen, Merck, Pfizer Inc, and PuraPharm. No other disclosures relevant to this article were reported.

Upon request, and subject to certain criteria, conditions, and exceptions (see <https://www.pfizer.com/science/clinical-trials/trial-data-and-results> for more information), Pfizer will provide access to individual de-identified participant data from Pfizer-sponsored global interventional clinical studies conducted for medicines, vaccines, and medical devices (1) for indications that have been approved in the US and/or EU, or (2) in programs that have been terminated (i.e., development for all indications has been discontinued). Pfizer will also consider requests for the protocol, data dictionary, and statistical analysis plan. Data may be requested from Pfizer trials 24 months after study completion. The de-identified participant data will be made available to researchers whose proposals meet the research criteria and other conditions, and for which an exception does not apply, via a secure portal. To gain access, data requestors must enter into a data-access agreement with Pfizer.

Address correspondence to Vivian P. Bykerk, BSc, MD, FRCPC, Weill Cornell Medical College, 525 East 71 Street, New York, NY 10021. Email: bykerkv@hss.edu.

Submitted for publication December 23, 2020; accepted in revised form May 18, 2021.

SIGNIFICANCE & INNOVATIONS

- This group-based trajectory modeling analysis evaluated 346 methotrexate-naive patients in a phase III study. Five disease trajectory groups were identified in patients using the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate. The groups comprised those who progressed from high disease activity (HDA) to remission, low disease activity (LDA) rapidly, moderate disease activity, LDA gradually, or remained in HDA, at month 24.
- Significant differences between trajectory groups in some baseline variables (e.g., sex, disease activity measures, and patient-reported outcomes) were observed.
- Improvements in patient-reported outcomes across trajectory groups over time were generally consistent with improvements in disease activity predicted by group-based trajectory modeling.
- These data demonstrate heterogeneity in patients who are normally analyzed as a single population; further exploration may help to better understand suboptimal treatment responses in rheumatoid arthritis.

characteristics (2–6). Understanding patient characteristics associated with distinct disease activity trajectories may make predicting responses to specific treatments possible at an early stage (2,3).

Tofacitinib is an oral JAK inhibitor for the treatment of RA. The efficacy and safety of tofacitinib 5 and 10 mg twice daily administered as monotherapy or in combination with conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), mainly methotrexate (MTX), in patients with moderately to severely active RA, have been demonstrated in phase II (7–11), phase III (12–18), and phase IIIb/IV (19) studies of up to 24-months duration, and in long-term extension studies with up to 9.5 years of observation (20–22).

The 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate (DAS28-ESR) is a commonly used measure of disease activity status (e.g., remission or low/moderate/high disease activity [LDA/MDA/HDA, respectively]) (23,24). Previously, an analysis of data pooled from 3 phase III trials of patients with RA with a prior inadequate response to csDMARDs receiving tofacitinib 5 mg twice daily for up to 12 months identified distinct disease activity trajectories, characterized by baseline differences in DAS28-ESR and patient-reported outcomes (25).

ORAL Start was a 24-month phase III study of tofacitinib 5 and 10 mg twice daily in MTX-naive patients with active RA (18). This post hoc analysis of tofacitinib 5 mg twice daily data from ORAL Start aimed to identify distinct disease activity trajectories in MTX-naive patients with RA receiving tofacitinib, offering a characterization of baseline variables that could be used as early predictors of response.

PATIENTS AND METHODS

Study design. ORAL Start was a 24-month, randomized, double-blind, phase III study completed in 2013 that compared efficacy and safety of tofacitinib 5 and 10 mg twice daily monotherapy with MTX monotherapy in patients with moderately to severely active RA who were MTX-naive or who had not received a therapeutic dose of MTX (18).

Full study details have been published previously (18). Eligible patients were age ≥ 18 years with a diagnosis of active RA, based on American College of Rheumatology 1987 revised criteria (26–28), had either an ESR of >28 mm/hour or a C-reactive protein level of >7 mg/liter, and had ≥ 3 distinct joint erosions detected on hand/wrist or foot radiographs, or were anti-cyclic citrullinated peptide or rheumatoid factor positive. At baseline, the duration of RA in patients was 2.7–3.4 years. In total, 6.9% of patients had received a nontherapeutic dose of MTX prior to study baseline; the most common non-MTX csDMARDs received by patients prior to study baseline were sulfasalazine and leflunomide (12.9% and 6.3% of patients, respectively). This post hoc analysis included data for patients receiving tofacitinib 5 mg twice daily who were MTX-naive at baseline.

Trajectory analysis. Trajectory groups are understood to be clusters of individuals following similar trajectories of disease response. As with previous analyses (3–5,25), DAS28-ESR scores at baseline and changes in DAS28-ESR over time were used to model predicted trajectories. Disease activity status was defined using DAS28-ESR scores as: HDA >5.1 , MDA ≥ 3.2 to ≤ 5.1 , LDA ≥ 2.6 to <3.2 , and remission <2.6 (23,29).

Outcomes. Patient demographics, baseline disease characteristics and patient-reported outcome scores, changes in radiographic progression (total Sharp score, and erosion and joint space narrowing [JSN] scores, assessed at months 6, 12, and 24), patient-reported outcomes over time (assessed at months 1, 2, 3, 6, 9, 12, 15, 18, 21, and 24), and adverse events (AEs) were compared across predicted disease activity trajectory groups.

Patient-reported outcomes included the Health Assessment Questionnaire disability index (HAQ DI) (30), the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F) (31), the Short Form 36 health survey (SF-36; mental component summary [MCS] and physical component summary [PCS] scores and domain scores: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health) (32,33), and and arthritis pain (visual analog scale [VAS] 0–100 mm) (34,35).

The proportions of patients reporting normative patient-reported outcome scores were also identified and compared; defined for HAQ DI as ≤ 0.25 (36) or <0.5 (functional remission) (37), and for FACIT-F as ≥ 40.1 (31), or more recently, ≥ 43.5 (38). SF-36 MCS and PCS and domain scores were assessed using

age- and sex-matched norms, as per the SF-36 scoring manual (32,33). Improvements in arthritis pain score of $\geq 30\%$ and $\geq 50\%$ from baseline were defined as moderate and substantial clinically important improvements, respectively (34,35). Safety end points were reported through month 24, including AEs, discontinuations due to AEs, and all-cause mortality.

Statistical analysis. For each disease activity trajectory group, predicted DAS28-ESR values and 95% confidence intervals over time, and the proportion (%) of patients within each group, were modeled. Group-based trajectory modeling (6) was applied to DAS28-ESR data to find distinct longitudinal subgroups of patients with similar disease activity changes through month 24. This is a special case of finite mixture models that seeks to classify patients into trajectories using a maximum-likelihood approach, based on the product of the conditional likelihoods for each individual being in the j^{th} group, multiplied by the probability of membership in the j^{th} group ($j = 1, 2, \dots, k$, with k being the number of groups specified).

The modeling algorithm only required a baseline value to allow initial assignment to a trajectory group. Each group was modeled by linear regression of DAS28-ESR versus time (months) added as polynomials (months, months², months³, etc.), and k and the degree of polynomial (p) were specified. For all $k = 1, 2, 3, 4, 5$, and $p = 1, 2, 3, 4, 5$, models were fit, and the Bayesian information criterion (BIC) of each result was calculated; the best BIC chosen from all possibilities was run. The algorithm jointly modeled all groups, using intercept values (month 0, i.e., baseline) as a start for assigning patients. The algorithm then became iterative: at the end of each linear regression, the posterior probability of a patient belonging to a particular group was calculated and patients were reassigned to the group with the highest probability.

This approach continued until no more increase in likelihood was reached, and the algorithm was then stopped. In cases where linear regression was replaced by a generalized linear model, e.g., Poisson or logistic regression, a censored normal distribution was used, where DAS28-ESR was censored to be on the interval 0–10 (6). Modeling was carried out using observed data, with no imputation for missing values (6).

Pair-wise comparisons of demographics and baseline characteristics were performed among predicted disease activity trajectory groups. Equality of mean values of continuous measures were assessed using t -tests, and equality of rates were assessed using chi-square tests. A 2-sided Bonferroni correction for multiple comparisons was applied; consequently, a P value less than or equal to 0.005 indicated statistical significance ($P \leq 0.05$). Missing radiographic data were extrapolated linearly, and patient-reported outcomes were analyzed using a mixed-effects longitudinal model (39), as previously reported (18).

RESULTS

Predicted disease activity trajectory groups. In total, 346 patients with HDA at baseline were included in the analysis. In the trajectory model, the majority of patients (98.8%) had at least 2 values (baseline value plus 1 additional observation post baseline), 84.7% had at least 7 values, and 71.4% had values from each observation. In total, 1.2% of patients in the analysis only had baseline values. In this case, the modeling algorithm placed the patient within the group with the y-intercept closest to the baseline value.

Trajectory modeling found 5 distinct groups of patients with similar predicted disease activity trajectories (Figure 1). Improvement in disease activity (based on DAS28-ESR change) was

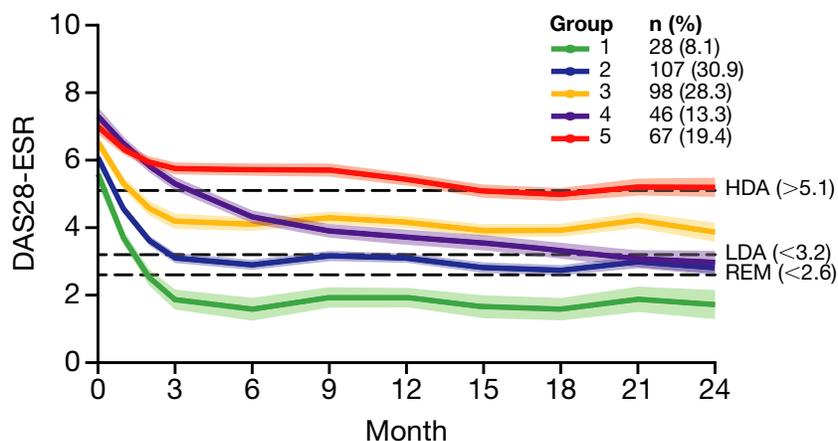


Figure 1. Predicted group trajectories based on the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate (DAS28-ESR), with 95% confidence intervals (95% CIs) identified using group-based trajectory modeling in patients with active rheumatoid arthritis who were methotrexate-naïve and receiving tofacitinib 5 mg twice daily over 24 months. Solid lines represent predicted values; shading indicates the 95% CI. Percentage reflects the proportion of patients in each group; modeling of trajectory groups was based on changes in DAS28-ESR scores over 24 months: high disease activity (HDA) >5.1 , moderate disease activity ≥ 3.2 to ≤ 5.1 , low disease activity (LDA) <3.2 to ≥ 2.6 , and remission (REM) <2.6 (23).

Table 1. Patient demographics and baseline disease characteristics by predicted DAS28-ESR disease activity trajectory groups*

	Group 1: HDA to remission (n = 28)	Group 2: HDA to LDA rapid (n = 107)	Group 3: HDA to MDA (n = 98)	Group 4: HDA to LDA gradual (n = 46)	Group 5: HDA to HDA (n = 67)
Demographics					
Female, no. (%)	11 (39.3) [†]	81 (75.7) [‡]	84 (85.7) [‡]	36 (78.3) [‡]	54 (80.6) [‡]
Age, years	47.0 ± 16.4	52.1 ± 11.3	50.4 ± 12.6	49.9 ± 10.1	48.1 ± 12.6
Body mass index, kg/m ²	25.2 ± 4.5	26.0 ± 4.8	27.2 ± 5.6	27.0 ± 5.8	25.7 ± 6.4
Current smoker, no. (%)	2 (7.1)	23 (21.5)	16 (16.3)	8 (17.4)	13 (19.4)
Geographic region, no. (%)					
US/Canada	11 (39.3)	36 (33.6)	43 (43.9)	20 (43.5)	24 (35.8)
Europe	11 (39.3) [§]	22 (20.6)	15 (15.3)	4 (8.7) [‡]	15 (22.4)
Latin America	5 (17.9)	24 (22.4)	29 (29.6)	9 (19.6)	10 (14.9)
Rest of the world	1 (3.6)	25 (23.4)	11 (11.2)	13 (28.3)	18 (26.9)
Race, no. (%)					
White	24 (85.7) [§]	71 (66.4)	64 (65.3)	23 (50.0) [‡]	38 (56.7)
Other	4 (14.3)	36 (33.6)	34 (34.7)	23 (50.0)	29 (43.3)
Baseline disease characteristics and activity measures					
Rheumatoid arthritis duration, years	1.2 ± 1.9 [¶]	3.2 ± 7.1	2.6 ± 4.4	2.6 ± 4.9	3.7 ± 5.1 [‡]
Day 1 steroid use, no. (%)	10 (35.7)	51 (47.7)	39 (39.8) [§]	30 (65.2) [#]	38 (56.7)
DAS28-ESR score	5.7 ± 0.9 ^{**}	6.1 ± 0.8 ^{**}	6.7 ± 0.8 ^{††}	7.5 ± 0.8 ^{‡‡}	7.1 ± 0.8 ^{§§}
CDAI score	31.2 ± 9.0 ^{**}	33.7 ± 11.1 ^{**}	39.6 ± 11.7 ^{††}	50.6 ± 12.3 ^{¶¶}	43.9 ± 10.5 ^{††}
ESR score	33.1 ± 18.1 [†]	47.9 ± 23.4 ^{##}	57.8 ± 26.8 [‡]	70.8 ± 30.0 ^{§§}	64.2 ± 32.7 ^{§§}
CRP	15.4 ± 16.3	20.3 ± 25.9	22.2 ± 22.0	25.5 ± 31.3	27.8 ± 35.4
Patients with CRP score >7 mg/liter, no. (%)	17 (60.7)	62 (57.9)	72 (73.5)	35 (76.1)	46 (68.7)
Total Sharp score	5.9 ± 10.6 ^{***}	18.7 ± 33.1 [‡]	17.1 ± 41.6	16.4 ± 37.2	29.1 ± 45.2 [‡]
Erosion score	3.3 ± 5.1 ^{***}	9.0 ± 16.4 [‡]	8.8 ± 23.1	8.0 ± 17.7	12.3 ± 19.1 [‡]
Joint space narrowing score	2.6 ± 6.4 ^{***}	9.7 ± 17.7 [‡]	8.2 ± 19.6	8.4 ± 20.4	16.8 ± 27.4 [‡]
Tender joints					
68 count	19.3 ± 10.3 ^{†††}	21.1 ± 12.7 ^{†††}	25.2 ± 11.7 [§]	34.9 ± 15.0 ^{‡‡}	30.2 ± 14.3 ^{§§}
28 count	10.7 ± 4.7 ^{**}	12.6 ± 6.3 ^{**}	15.5 ± 6.2 ^{††}	20.3 ± 6.0 ^{‡‡}	17.8 ± 5.5 ^{§§}
Swollen joints					
66 count	12.8 ± 7.2 [§]	13.6 ± 7.0 ^{†††}	15.9 ± 9.2 [§]	24.5 ± 11.7 ^{¶¶}	17.4 ± 7.9 ^{‡‡‡}
28 count	9.3 ± 3.5 ^{†††}	10.2 ± 4.7 ^{†††}	11.7 ± 5.6 [§]	16.5 ± 6.1 ^{¶¶}	12.5 ± 5.1 ^{†††}
Physician global assessment (VAS 0–100)	59.1 ± 15.3	58.1 ± 16.7 ^{†††}	62.3 ± 15.9	68.2 ± 13.4 ^{§§§}	66.8 ± 16.5 ^{§§§}
Anti-CCP positive, no. (%)	26 (92.9)	93 (86.9)	84 (85.7)	35 (76.1)	59 (88.1)
Rheumatoid factor positive, no. (%)	23 (82.1)	91 (85.0)	78 (79.6)	38 (82.6)	56 (83.6)
Baseline patient-reported outcomes					
HAQ DI score	1.2 ± 0.6 ^{**}	1.3 ± 0.6 ^{**}	1.7 ± 0.6 ^{§§}	1.8 ± 0.7 ^{§§}	1.7 ± 0.6 ^{§§}
FACIT-F total score	33.1 ± 10.6 ^{†††}	32.5 ± 10.6 ^{**}	27.0 ± 9.7 ^{§§§}	22.3 ± 11.3 ^{§§§}	25.3 ± 10.8 ^{§§§}
SF-36 MCS score	47.8 ± 12.0 ^{**}	44.8 ± 11.8 ^{**}	38.5 ± 10.5 ^{§§}	33.7 ± 12.1 ^{§§}	37.2 ± 11.8 ^{§§}
SF-36 PCS score	35.8 ± 6.8 ^{¶¶¶}	34.7 ± 8.1 ^{¶¶¶}	31.5 ± 6.1 ^{§§}	31.9 ± 8.3	30.8 ± 6.5 ^{§§}
Arthritis pain score (VAS 0–100)	50.0 ± 25.4 ^{†††}	49.5 ± 26.0 ^{**}	62.3 ± 22.2 ^{§§§}	71.1 ± 20.6 ^{§§}	67.2 ± 18.0 ^{§§}

(Continued)

Table 1. (Cont'd)

	Group 1: HDA to remission (n = 28)	Group 2: HDA to LDA rapid (n = 107)	Group 3: HDA to MDA (n = 98)	Group 4: HDA to LDA gradual (n = 46)	Group 5: HDA to HDA (n = 67)
Patients with arthritis pain score >40, no. (%)	17 (60.7)†††	68 (63.6)**	83 (84.7)§§§	43 (93.5)§§	63 (94.0)§§
Patient global assessment (VAS 0–100)	51.1 ± 27.9†††	50.4 ± 26.8**	64.4 ± 23.0§§§	69.7 ± 21.6§§	68.8 ± 18.2§§

* Values are the mean ± SD unless indicated otherwise. The number of patients assessed for each characteristic may be lower than the total number. Ranges based on the Disease Activity Score in 28 joints using the erythrocyte sedimentation rate (DAS28-ESR): high disease activity (HDA) >5.1, moderate disease activity (MDA) ≥3.2 to ≤5.1, low disease activity (LDA) <3.2 to ≥2.6, remission <2.6 (ref. 23). A 2-sided Bonferroni correction for multiple comparisons was applied; $P \leq 0.05$ indicated statistical significance. CCP = cyclic citrullinated peptide; CDAI = Clinical Disease Activity Index; CRP = C-reactive protein; FACIT-F = Functional Assessment of Chronic Illness Therapy–Fatigue; HAQ DI = Health Assessment Questionnaire disability index; MCS = mental component summary; PCS = physical component summary; SF-36 = Short Form 36 health survey; VAS = visual analog scale.

† $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 3. $P \leq 0.05$ versus group 4. $P \leq 0.05$ versus group 5.

‡ $P \leq 0.05$ versus group 1.

§ $P \leq 0.05$ versus group 4.

¶ $P \leq 0.05$ versus group 5.

$P \leq 0.05$ versus group 3.

** $P \leq 0.05$ versus group 3. $P \leq 0.05$ versus group 4. $P \leq 0.05$ versus group 5.

†† $P \leq 0.05$ versus group 1. $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 4.

‡‡ $P \leq 0.05$ versus group 1. $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 3.

§§ $P \leq 0.05$ versus group 1. $P \leq 0.05$ versus group 2.

¶¶ $P \leq 0.05$ versus group 1. $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 3. $P \leq 0.05$ versus group 5.

$P \leq 0.05$ versus group 1. $P \leq 0.05$ versus group 4. $P \leq 0.05$ versus group 5.

**** $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 5.

††† $P \leq 0.05$ versus group 4. $P \leq 0.05$ versus group 5.

‡‡‡ $P \leq 0.05$ versus group 2. $P \leq 0.05$ versus group 4.

§§§ $P \leq 0.05$ versus group 2.

¶¶¶ $P \leq 0.05$ versus group 3. $P \leq 0.05$ versus group 5.

greatest during the first 3 months of treatment for group 1 (n = 28), group 2 (n = 107), and group 3 (n = 98), which improved from HDA into remission, LDA, and MDA, respectively, by month 3, and remained there through month 24. Group 4 (n = 46) showed continued gradual improvement through months 3–24 to LDA, while group 5 (n = 67) showed minimal improvement, with patients in HDA at baseline and month 24.

Demographics and baseline characteristics of trajectory groups. Baseline patient demographic information, disease characteristics, disease activity measures, and patient-reported outcomes by predicted disease activity trajectory group are shown in Table 1. Group 1 had a significantly lower proportion of female patients compared with all other groups. Age, body mass index (BMI), and geographic location were similar across trajectory groups, except group 1, which had a significantly higher proportion of patients from Europe, and patients of White race, compared with group 4. Group 1 also had the shortest duration of RA, which was significantly shorter, compared with group 5.

Groups 1 and 2 had significantly lower mean DAS28-ESR scores at baseline than groups 3–5. Group 4 had the highest mean DAS28-ESR score at baseline. Mean ESR level was significantly lower in group 1, compared with other groups, and mean tender joint count scores in 68 and 28 joints were significantly lower in groups 1 and 2, compared with groups 3 (tender joint count

scores in 28 joints only), 4, and 5. Group 1 also had the lowest baseline total Sharp score, and erosion and JSN scores, which were significantly lower than in groups 2 and 5. Differences in baseline patient-reported outcomes among groups were generally consistent with differences seen in clinical measures.

Changes in radiographic progression. The total Sharp score increased over time in groups 3 and 5, and change from baseline was highest in group 5 at month 24, followed by group 3. Total Sharp score increased from baseline to month 12 in group 1, and between months 6 and 12 in group 2, with no further increases in total Sharp score in either group from months 12 to 24. There were minimal changes in total Sharp score through month 24 in group 4 (Figure 2A).

In group 5, the erosion score increased from baseline to month 24, while erosion scores in groups 1, 3, and 4 increased from baseline to month 6, with minimal further changes observed from months 6 to 24. In group 2, the erosion score fell from baseline to month 24. At month 24, the highest increase from baseline erosion score was observed in group 5, followed by group 1 (Figure 2B).

JSN scores increased up to month 24 in groups 3 and 5, and from baseline to months 6 and 12 in groups 1 and 2, respectively, with no or minimal subsequent changes to month 24. In group 4, the JSN score fell from baseline to month 12, and this response was maintained to month 24 (Figure 2C).

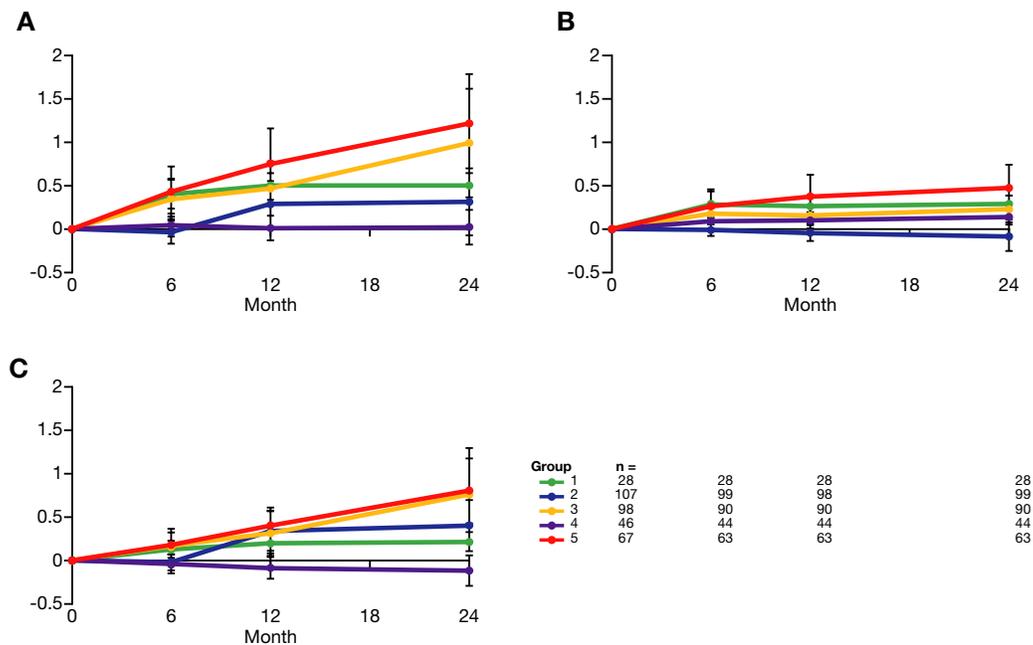


Figure 2. Mean change from baseline over time across the disease activity trajectory groups based on the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate (DAS28-ESR). **A**, Total Sharp score, mean (SE); **B**, Erosion score, mean (SE); **C**, Joint space narrowing score, mean (SE). For **A–C**, the columns of group numbers in the legend correspond to the 0, 6-month, 12-month, and 24-month time points, respectively. Group 1: high disease activity (HDA) to remission; group 2: HDA to low disease activity (LDA) rapid; group 3: HDA to moderate disease activity (MDA); group 4: HDA to LDA gradual; group 5: HDA to HDA. HDA >5.1, MDA ≥3.2 to ≤5.1, LDA <3.2 to ≥2.6, remission <2.6 (23). Missing radiographic data were extrapolated linearly.

Changes in patient-reported outcomes. The mean changes from baseline in HAQ DI, FACIT-F total score, and SF-36 MCS and PCS scores at month 24 are shown in Table 2, and absolute scores from baseline to month 24 are shown in Figure 3. Group 4 had the largest numerical improvement in HAQ DI score at month 24, followed by group 1; improvements in HAQ DI score

were similar in groups 2–5. The proportion of patients reporting normative HAQ DI scores was numerically highest in group 1 and lowest in groups 3 and 5 (Table 2). Proportions reporting HAQ DI functional remission were closely aligned with normative HAQ DI scores.

Group 4 had the largest numerical improvement in FACIT-F total score at month 24, followed by group 1. Improvements in

Table 2. Mean change in patient-reported outcome scores and proportion of patients reporting scores ≥ normative values at 24 months across DAS28-ESR disease activity trajectory groups*

Patient-reported outcome	Group 1: HDA to remission (n = 28)	Group 2: HDA to LDA rapid (n = 107)	Group 3: HDA to MDA (n = 98)	Group 4: HDA to LDA gradual (n = 46)	Group 5: HDA to HDA (n = 67)
HAQ DI score, mean change ± SD	-1.1 ± 0.7	-0.9 ± 0.8	-0.8 ± 0.7	-1.4 ± 0.8	-0.8 ± 0.7
Scores ≥ normative values (≤0.25)	21/23 (91.3)	54/81 (66.7)	20/67 (29.9)	19/37 (51.4)	7/39 (17.9)
Functional remission (<0.5)	21/23 (91.3)	57/81 (70.4)	21/67 (31.3)	24/37 (64.9)	9/39 (23.1)
FACIT-F total score, mean change ± SD	10.0 ± 11.7	7.8 ± 11.3	8.6 ± 10.6	15.1 ± 12.3	8.1 ± 10.3
Scores ≥ normative values (≥40.1)	18/23 (78.3)	46/81 (56.8)	19/67 (38.4)	13/37 (35.1)	13/39 (33.3)
SF-36 MCS score, mean change ± SD	3.1 ± 14.5	4.5 ± 13.1	5.6 ± 12.6	11.8 ± 13.0	8.0 ± 12.7
Scores ≥ normative values†	14/23 (60.9)	46/81 (56.8)	21/67 (31.3)	13/37 (35.1)	13/39 (33.3)
SF-36 PCS score, mean change ± SD	15.2 ± 9.5	12.9 ± 9.8	10.9 ± 8.9	15.7 ± 10.2	8.6 ± 9.8
Scores ≥ normative values†	15/23 (65.2)	44/81 (54.3)	15/67 (22.4)	18/37 (48.6)	4/39 (10.3)
Arthritis pain (VAS 0–100), mean change ± SD	-35.0 ± 23.8	-29.0 ± 31.5	-33.1 ± 30.8	-56.6 ± 24.5	-32.5 ± 27.5

* Values are the number of patients included in the analysis/the number of patients evaluated at month 24 (%), unless indicated otherwise. Ranges are based on the Disease Activity Score in 28 joints using the erythrocyte sedimentation rate (DAS28-ESR): high disease activity (HDA) >5.1, moderate disease activity (MDA) ≥3.2 to ≤5.1, low disease activity (LDA) <3.2 to ≥2.6, remission <2.6 (ref. 23). Patient-reported outcomes were analyzed using a mixed-effects longitudinal model. FACIT-F = Functional Assessment of Chronic Illness Therapy–Fatigue; HAQ DI = Health Assessment Questionnaire disability index; MCS = mental component summary; PCS = physical component summary; SF-36 = Short Form 36 health survey; VAS = visual analog scale.

† SF-36 normative MCS and PCS scores were based on age- and sex-matched norm scores (ref. 32).

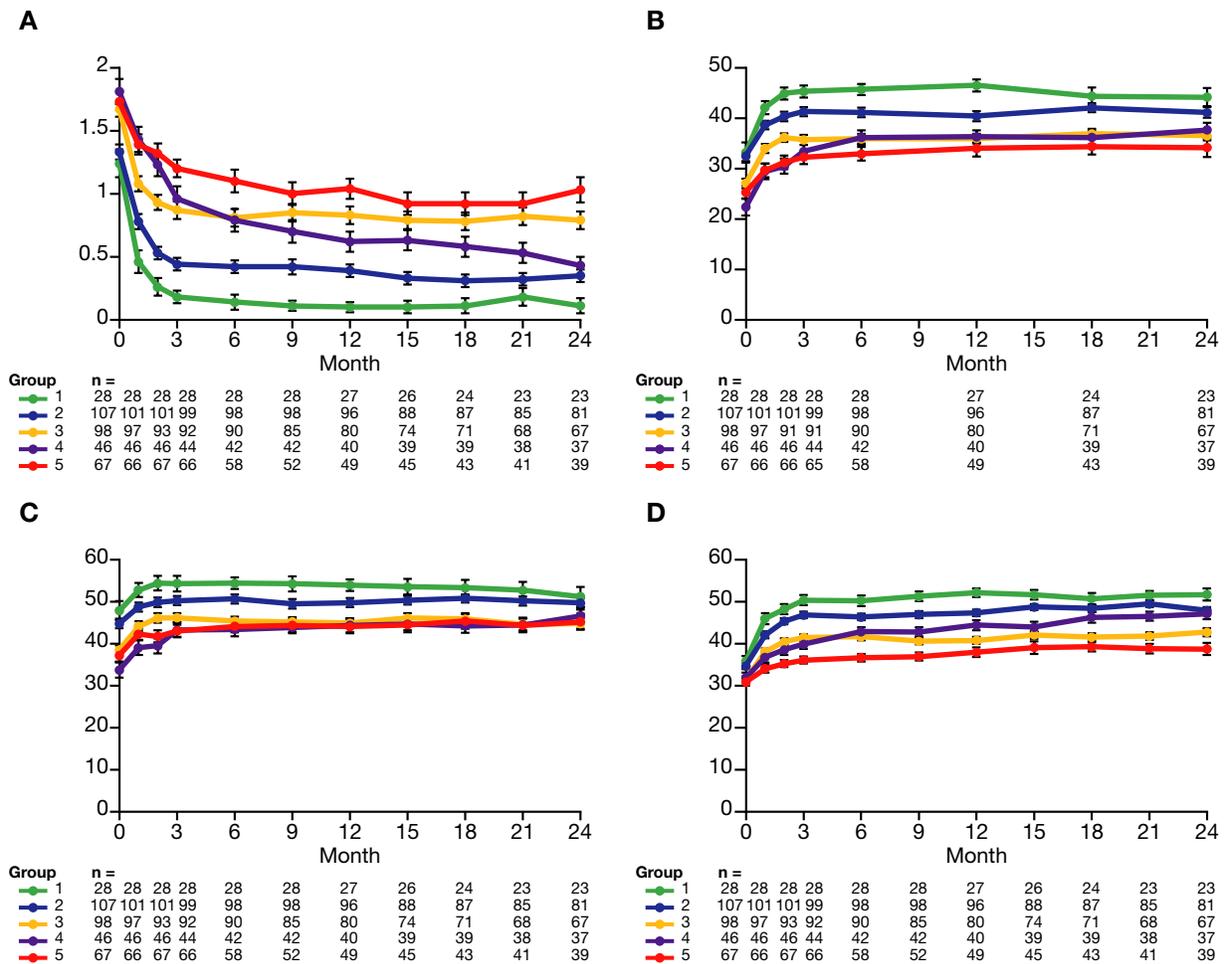


Figure 3. Mean absolute score over time in any disease activity trajectory group based on the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate (DAS28-ESR). **A**, Health Assessment Questionnaire disability index score, mean (SE); **B**, Functional Assessment of Chronic Illness Therapy–Fatigue total score, mean (SE); **C**, Short Form 36 health survey (SF-36) mental component summary score, mean (SE); **D**, SF-36 physical component summary score, mean (SE). **A–D**, The first and fourth columns of group numbers in the legend correspond to the 0 and 3-month time points, respectively. Group 1: high disease activity (HDA) to remission; group 2: HDA to low disease activity (LDA) rapid; group 3: HDA to moderate disease activity (MDA); group 4: HDA to LDA gradual; group 5: HDA to HDA. HDA >5.1, MDA ≥3.2 to <5.1, LDA <3.2 to ≥2.6, remission <2.6 (23). Patient-reported outcomes were analyzed using a mixed-effects longitudinal model.

FACIT-F total score were generally similar in groups 2, 3, and 5. The proportions of patients reporting normative FACIT-F total scores were numerically highest in group 1 and lowest in groups 3 and 5 (Table 2).

Numerical improvements in SF-36 MCS and PCS scores were highest in group 4. Groups 1 and 2 had the smallest improvements in SF-36 MCS score, while groups 3 and 5 had the smallest improvements in SF-36 PCS score (Table 2). The proportions of patients reporting normative SF-36 MCS and PCS scores were numerically highest in groups 1 and 2, and lowest in groups 3 and 5. The proportions of patients reporting normative values in SF-36 domain scores were generally consistent with those reporting normative SF-36 MCS and PCS scores, with the exception of the bodily pain domain, where groups 1 and 4 had the highest proportions reporting normative scores (see Supplementary Table 1, available on the *Arthritis Care & Research*

website at <http://onlinelibrary.wiley.com/doi/10.1002/acr.24709/abstract>).

At month 24, group 4 had the largest mean change in arthritis pain score (Table 2), and the proportions reporting ≥30%/≥50% improvements in arthritis pain score were highest in groups 1 and 4 and lowest in group 3 (see Supplementary Figure 1, available on the *Arthritis Care & Research* website at <http://onlinelibrary.wiley.com/doi/10.1002/acr.24709/abstract>).

AEs across trajectory groups. Discontinuation rates were numerically lower in groups 1, 2, and 4, compared with groups 3 and 5 (Figure 4). Discontinuations due to AEs were lowest in groups 2 and 4 and highest in group 5 (see Supplementary Table 2, available on the *Arthritis Care & Research* website at <http://onlinelibrary.wiley.com/doi/10.1002/acr.24709/abstract>). There were 2 deaths (1 each in groups 2 and 3) (Figure 4).

	Group 1 HDA to REM (n = 28)	Group 2 HDA to LDA rapid (n = 107)	Group 3 HDA to MDA (n = 98)	Group 4 HDA to LDA gradual (n = 46)	Group 5 HDA to HDA (n = 67)
Patients with discontinuations	5 (17.9)	24 (22.4)	31 (31.6)	9 (19.6)	29 (43.3)
Deaths	0	1 (0.9)	1 (1.0)	0	0
Adverse events					
Blood and lymphatic system disorders	2 (7.1)	5 (4.7)	5 (5.1)	2 (4.4)	7 (10.4)
Gastrointestinal disorders	8 (28.6)	30 (28.0)	30 (30.6)	9 (19.6)	18 (26.9)
General disorders and administration site conditions	3 (10.7)	10 (9.4)	9 (9.2)	3 (6.5)	9 (13.4)
Infections and infestations	12 (42.9)	48 (44.9)	44 (44.9)	14 (30.4)	28 (41.8)
Injury, poisoning, and procedural complications	3 (10.7)	12 (11.2)	9 (9.2)	3 (6.5)	3 (4.5)
Investigations	5 (17.9)	24 (22.4)	18 (18.4)	13 (28.3)	16 (23.9)
Metabolism and nutrition disorders	1 (3.6)	5 (4.7)	9 (9.2)	5 (10.9)	10 (14.9)
Musculoskeletal and connective tissue disorders	6 (21.4)	25 (23.4)	15 (15.3)	5 (10.9)	21 (31.3)
Nervous system disorders	5 (17.9)	12 (11.2)	13 (13.3)	5 (10.9)	12 (17.9)
Respiratory, thoracic, and mediastinal disorders	4 (14.3)	9 (8.4)	3 (3.1)	4 (8.7)	9 (13.4)
Skin and subcutaneous tissue disorders	3 (10.7)	12 (11.2)	10 (10.2)	4 (8.7)	7 (10.4)
Vascular disorders	1 (3.6)	7 (6.5)	10 (10.2)	10 (21.7)	5 (7.5)

Heatmap color key (rates expressed as percentages)						
0	>0–5	>5–10	>10–15	>15–20	>20–25	>25–30
>30–35	>35–40	>40–45	>45–50	>50–55	>55–60	>60–65
>65–70	>70–75	>75–80	>80–85	>85–90	>90–95	>95–100

Figure 4. Heatmap of the incidence of adverse events (AEs) in any disease activity trajectory group based on the 4-variable Disease Activity Score in 28 joints, using the erythrocyte sedimentation rate. Values are the number (%), and data are presented for AEs by system organ class with an incidence of $\geq 10\%$ in at least 1 trajectory group. High disease activity (HDA) >5.1 , moderate disease activity (MDA) ≥ 3.2 to ≤ 5.1 , low disease activity (LDA) <3.2 to ≥ 2.6 , remission (REM) <2.6 (ref. 23).

Analysis of AEs indicated that incidences were generally comparable across groups (Figure 4). Group 4 had the numerically lowest proportion of patients with blood and lymphatic disorders, gastrointestinal disorders, general disorders and administration site conditions, infections and infestations, musculoskeletal and connective tissue disorders, nervous system disorders, and skin and subcutaneous tissue disorders, compared with other groups. In contrast, rates of investigations and vascular disorders were highest in group 4. In group 5, a numerically higher proportion of patients experienced AEs in several system organ classes, compared with other groups, most notably musculoskeletal and connective tissue disorders.

Consistent with previous analyses, across all trajectory groups, the most common AEs were nasopharyngitis and upper respiratory tract infection, followed by nausea, headache, and hypertension (see Supplementary Table 2, available on the *Arthritis Care & Research* website at <http://onlinelibrary.wiley.com/doi/10.1002/acr.24709/abstract>).

DISCUSSION

Identification of distinct latent trajectory groups among patients could inform treatment optimization and decision-making regarding subsequent lines of therapy. This post hoc analysis of

data from ORAL Start is the first trajectory modeling analysis of MTX-naïve patients with RA receiving tofacitinib. Based on the DAS28-ESR response to month 24, we identified 5 distinct predicted disease activity trajectories in patients receiving tofacitinib 5 mg twice daily. Groups 1–3 improved from HDA at baseline to remission, LDA, and MDA, respectively, over 3 months, and disease activity generally plateaued thereafter (i.e., improvements were maintained to month 24). Group 4 gradually improved from HDA to LDA over 24 months, while patients in group 5 remained in HDA at month 24.

There were significant differences in baseline characteristics between groups, including sex, disease activity measures, joint damage, and patient-reported outcomes. Group 1, which had the greatest improvement in disease activity at month 24 also had the lowest disease activity and most favorable radiographic and patient-reported outcome scores, as well as the shortest RA duration, at baseline. In contrast, group 5, which remained in HDA at month 24 had the longest RA duration at baseline and the highest total Sharp score. Significant differences in disease duration and radiographic evaluation at baseline were observed in group 1 versus group 5 only. Possibly other baseline characteristics, or interactions between characteristics not explored in the current model, may contribute to overall drug efficacy or influence the attainment of remission versus LDA.

At baseline, DAS28-ESR and Clinical Disease Activity Index (CDAI) scores were lowest in groups 1 and 2, followed by groups 3 and 5, and were highest in group 4. This finding suggests that baseline DAS28-ESR or CDAI scores may be predictive of short-term improvements in disease activity, such as those observed in groups 1, 2, and 3, which had the greatest improvement in disease activity during the first 3 months of treatment, but may be less predictive of the long-term improvements observed in group 4 over 24 months.

Improvements in patient-reported outcomes were generally consistent with predicted DAS28-ESR trajectories and plateaued after 3 months, suggesting that early patient-reported outcome data may be useful in informing treatment strategies. At month 24, the proportions of patients reporting HAQ DI scores \geq normative values, and functional remission in HAQ DI, were generally consistent with predicted DAS28-ESR trajectories.

Discontinuations due to AEs were numerically higher in groups 3 and 5, compared with groups 1, 2, and 4, and group 5 had a relatively poorer safety profile compared with other groups. No consistent pattern could be identified between disease activity trajectories and incidence of AEs across groups, and careful monitoring of safety is required for all patients, irrespective of predicted disease trajectory.

While groups 4 and 5 had the highest baseline disease activity, radiographic scores, impaired quality of life, and fatigue (as assessed by DAS28-ESR, CDAI, swollen joint count in 66 joints, FACIT-F, arthritis pain score, and patient global assessment), the trajectories of these groups diverged over time. Group 5 also had greater radiographic progression over time, compared with group 4. Group 4 experienced generally greater improvements in patient-reported outcomes, notably at month 24, than group 5. In addition, similarities between groups 3 and 5 in outcome measures through month 24, and the differential disease activity trajectories observed in these patients, compared with those in group 4, merits further discussion. A higher proportion of patients in groups 3 and 5 were female, and group 3, followed by group 4, had the highest baseline BMI (associated with a poorer prognosis), and group 3 also had the highest proportion of patients from Latin America (which may have implications for socioeconomic factors that affect outcomes). Patients in group 5 had the highest mean total Sharp score at baseline, which may be indicative of previously undertreated disease. However, while a numerical difference in the proportion of female patients was observed in groups 3 and 5 versus group 4, no significant differences in baseline factors between groups 3–5 were identified.

At baseline, the mean total Sharp score and mean erosion and JSN scores were lowest in group 1 and highest in group 5, while groups 2, 3, and 4 were generally numerically similar. Only minimal changes in radiographic scores were observed through month 24, which were unlikely to be clinically relevant; with the exception of group 5, baseline total Sharp score and erosion and JSN scores were not predictive of disease activity at month 24.

Previously, distinct RA disease activity trajectories, characterized by baseline differences in disease activity and patient-reported outcomes, were identified over 12 months in patients receiving tofacitinib 5 mg twice daily who had an inadequate response to csDMARDs and were biologic DMARD (bDMARD)–naïve (25). Similar to the current analysis, 5 disease activity trajectories were identified that improved from HDA to remission, LDA, and MDA (2 groups: based on rapid or gradual improvement), or remained in HDA at month 12; patients with higher disease activity at baseline were generally less likely to achieve improvements at month 12 (25). This result is consistent with the findings of the present analysis, where, with the exception of group 4, baseline disease activity was predictive of disease status at month 24.

Disease trajectories have also been identified in patients with RA receiving other treatments. An observational analysis in patients with early RA receiving combination csDMARDs identified 3 disease activity trajectories (good, moderate, and poor), demonstrating an association between persistence with initial csDMARD therapy and lower long-term disease activity (5). Furthermore, another analysis in patients with early RA in an observational cohort study found that baseline physician global assessment score was highest in those who improved from HDA to remission (equivalent to group 1 in the current analysis), and numerically lower in patients who improved from HDA to LDA or MDA (equivalent to groups 2 and 3, respectively), while patient global assessment scores were similar in all 3 groups (2). This finding contrasts with the results of the current analysis, where groups 1 and 2 generally had lower baseline disease activity, higher quality of life, lower fatigue, and numerically lower baseline physician global assessment and patient global assessment scores, compared with groups 3–5. These discrepancies may be due to differences in disease duration and severity between the populations evaluated in the previous and current analyses; only patients with ≤ 12 months of symptoms were included in the previous analysis, and the majority (51%) had MDA at baseline, whereas patients in the current analysis had a mean disease duration of 1.2–3.7 years, and all had HDA at baseline.

An analysis of patients with early RA following a treat-to-target strategy (using an escalating csDMARD to csDMARD + bDMARD treatment regimen) over 12 months, identified 3 response trajectories (fast response, slow response, and poor outcome); clinical outcomes and patient-reported outcomes over time were greatest in the fast response group (3). However, unlike the current analysis, the fast response group (82.6% of patients) were in MDA at baseline. Likewise, a pooled analysis of registry data for patients with established RA receiving abatacept identified 3 response trajectories (rapid, gradual, and inadequate). Time to discontinuation due to lack of efficacy was shorter in the group with the poorest response over time; however, again, the majority (91.7%) were in MDA at baseline (4).

It should be noted that trajectory groups identified by modeling should not be considered permanent, but instead

represent summaries of disease progression. Building on this analysis, future investigation into the heterogeneity of treatment responses could examine which clinical variables cluster together in similar-behaving trajectory groups. In particular, further exploration of clustering of disease and patient characteristics associated with more severe disease (e.g., longer disease duration, higher levels of initial structural damage, current smoking status, higher BMI, initial steroid use, greater pain sensitization) would be of potential interest.

A strength of this post hoc analysis was the use of data from a clinical trial, which enrolled a unique patient population for which tofacitinib is not indicated. Further validation of the model, through trajectory analysis of registry and/or real-world data, would strengthen the interpretation of any results. This was a descriptive analysis, which identified and characterized trajectory groups based on disease activity. The analysis was limited by small patient numbers in some trajectory groups. Also, a possible result of increasing the number of possible groups (k) is that a group containing a relatively large proportion of the analysis population could be spuriously separated into 2 new groups, containing the lower and higher proportions of the original population, without offering any new insight into the underlying trajectories. The algorithm may also return a group with no members; hence, user input is required to compare and interpret competing sets of trajectory results to select the best model.

In conclusion, this post hoc analysis identified phenotypic subgroups with distinct disease activity trajectories in MTX-naive patients treated with tofacitinib, reflecting heterogeneity in patients normally analyzed as a single group. More thorough exploration of the heterogeneity of any given patient population, in terms of a preplanned cluster analysis subsequent to the presentation of clinical trial outcomes, may help practitioners identify which patients are more likely to respond to treatment and provide a means of matching the right patient with the right treatment. Identification of distinct latent trajectory groups of patients enrolled in clinical trials could provide a better understanding of the characteristics of particular patient cohorts, give further insight into the impact of treatments under investigation, inform future trial development, and ultimately optimize outcomes. Future analyses to investigate potential effect modifiers that may predispose a patient to a specific response trajectory are warranted.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Bykerk had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Bykerk, Lee, van Vollenhoven, Gruben, Fallon, Woolcott, Keystone.

Acquisition of data. Gruben.

Analysis and interpretation of data. Bykerk, Lee, van Vollenhoven, Gruben, Fallon, Woolcott, Keystone.

ROLE OF THE STUDY SPONSOR

Pfizer Inc were involved in the study design and in data collection. All authors, including those employed by Pfizer Inc, had a role in data analysis, data interpretation, and writing the manuscript. Medical writing support, under the guidance of the authors, was provided by Anthony McCluskey, PhD, CMC Connect, McCann Health Medical Communications, and was funded by Pfizer Inc, New York, in accordance with Good Publication Practice (GPP3) guidelines (Ann Intern Med 2015;163:461–464). Publication of this article was not contingent upon approval by Pfizer Inc.

REFERENCES

- Goetz I, Carter GC, Lucero M, Zarotsky V, Alatorre CI, Cantrell RA, et al. Review of treatment response in rheumatoid arthritis: assessment of heterogeneity. *Curr Med Res Opin* 2011;27:697–711.
- Barnabe C, Sun Y, Boire G, Hitchon CA, Haraoui B, Thorne JC, et al. Heterogeneous disease trajectories explain variable radiographic, function and quality of life outcomes in the Canadian early Arthritis CoHort (CATCH). *PLoS One* 2015;10:e0135327.
- Siemons L, ten Klooster PM, Vonkeman HE, Glas CA, van de Laar MA. Distinct trajectories of disease activity over the first year in early rheumatoid arthritis patients following a treat-to-target strategy. *Arthritis Care Res (Hoboken)* 2014;66:625–30.
- Courvoisier DS, Alpizar-Rodriguez D, Gottenberg JE, Hernandez MV, Iannone F, Lie E, et al. Rheumatoid arthritis patients after initiation of a new biologic agent: trajectories of disease activity in a large multinational cohort study. *EBioMedicine* 2016;11:302–6.
- Wabe N, Wojciechowski J, Wechalekar MD, Cleland LG, McWilliams L, Lee A, et al. Disease activity trajectories in early rheumatoid arthritis following intensive DMARD therapy over 3 years: association with persistence to therapy. *Int J Rheum Dis* 2017;20:1447–56.
- Nagin D. *Group-based modeling of development*. Cambridge (MA): Harvard University Press; 2005.
- 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;64:617–29.
- 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.
- 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.
- Tanaka Y, Suzuki M, Nakamura H, Toyozumi S, Zwillich SH, and the 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.
- 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.
- Van der Heijde D, Strand V, Tanaka Y, Keystone E, Kremer J, Zerbini CA, et al. Tofacitinib in combination with methotrexate in patients with rheumatoid arthritis: clinical efficacy, radiographic, and safety outcomes from a twenty-four-month, phase III study. *Arthritis Rheumatol* 2019;71:878–91.

13. Van Vollenhoven RF, Fleischmann R, Cohen S, Lee EB, García Meijide JA, Wagner S, et al. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. *N Engl J Med* 2012;367:508–19.
14. 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.
15. 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.
16. 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.
17. Kremer J, Li ZG, 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.
18. 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.
19. 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.
20. Wollenhaupt J, Silverfield J, Lee EB, Curtis JR, Wood SP, Soma K, et al. Safety and efficacy of tofacitinib, an oral Janus kinase inhibitor, for the treatment of rheumatoid arthritis in open-label, long-term extension studies. *J Rheumatol* 2014;41:837–52.
21. Yamanaka H, Tanaka Y, Takeuchi T, Sugiyama N, Yuasa H, Toyozumi S, et al. Tofacitinib, an oral Janus kinase inhibitor, as monotherapy or with background methotrexate, in Japanese patients with rheumatoid arthritis: an open-label, long-term extension study. *Arthritis Res Ther* 2016;18:34.
22. Wollenhaupt J, Lee EB, Curtis JR, Silverfield J, Terry K, Soma K, et al. Safety and efficacy of tofacitinib for up to 9.5 years in the treatment of rheumatoid arthritis: final results of a global, open-label, long-term extension study. *Arthritis Res Ther* 2019;21:89.
23. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res (Hoboken)* 2012;64:640–7.
24. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 update of the American College of Rheumatology recommended rheumatoid arthritis disease activity measures. *Arthritis Care Res (Hoboken)* 2019;71:1540–55.
25. Bykerk VP, Van Vollenhoven R, Connell C, Gruben D, Fallon L, Woolcott J, et al. SAT0217 identification of distinct disease activity trajectories in patients with rheumatoid arthritis receiving tofacitinib over 12 months [abstract]. *Ann Rheum Dis* 2018;77 Suppl 2:969.
26. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729–40.
27. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35.
28. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
29. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1–26.
30. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982;9:789–93.
31. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79.
32. Maruish ME. User's manual for the SF-36v2® health survey. 3rd ed. Lincoln (RI): QualityMetric Incorporated; 2011.
33. Ware JE, Kosinski M, Dewey JE. How to score version two of the SF-36 health survey. Lincoln (RI): QualityMetric Incorporated; 2000.
34. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105–21.
35. Busse JW, Bartlett SJ, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, et al. Optimal strategies for reporting pain in clinical trials and systematic reviews: recommendations from an OMERACT 12 workshop. *J Rheumatol* 2015;42:1962–70.
36. Krishnan E, Sokka T, Häkkinen A, Hubert H, Hannonen P. Normative values for the Health Assessment Questionnaire disability index: benchmarking disability in the general population. *Arthritis Rheum* 2004;50:953–60.
37. Keystone EC, van der Heijde D, Kavanaugh A, Kupper H, Liu S, Guérette B, et al. Clinical, functional, and radiographic benefits of long-term adalimumab plus methotrexate: final 10-year data in longstanding rheumatoid arthritis. *J Rheumatol* 2013;40:1487–97.
38. Montan I, Löwe B, Cella D, Mehnert A, Hinz A. General population norms for the Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue scale. *Value Health* 2018;21:1313–21.
39. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. 2nd ed. Hoboken (NJ): John Wiley; 2011.