ACR Open Rheumatology
Vol. 4, No. 9, September 2022, pp 811–818
DOI 10.1002/acr2.11468

American College of Rheumatology
Empowering Rheumatology Professionals

Identifying Factors Associated With Treatment Response in Rheumatoid Arthritis Clinical Trials

Anthony J. Cordisco,[1] Marianna Olave,[2] Michael D. George,[1] and Joshua F. Baker[3]

Objective. Despite a wealth of studies evaluating rheumatoid arthritis (RA) therapies, it remains difficult to compare efficacies across trials due to heterogeneous study populations. We sought to identify patient/trial characteristics associated with clinical response to enable fairer comparisons.

Methods. We reviewed 565 disease-modifying antirheumatic drug studies compiled for American College of Rheumatology (ACR) management guidelines. Seventy-two articles on randomized controlled phase II/III trials from 1995 to 2018 reporting the proportion of patients achieving 20%, 50% or 70% improvement in the ACR's RA disease score (ACR20/50/70) or Disease Activity Score-28 with erythrocyte sedimentation rate or C-reactive protein (DAS28-ESR or DAS28-CRP) with follow-up more than 3 months were included. We explored associations between 34 patient/trial characteristics and ACR responses. We constructed multivariable models using these factors to compute expected response rates and to compare observed with expected response rates across therapies.

Results. Among eligible clinical trials, later publication year, baseline DAS28-CRP score, methotrexate/biologic naivety, baseline ESR, follow-up of 52 weeks or more, number of subjects enrolled, and anticitrullinated peptide antibody seropositivity were associated with greater ACR response. Greater age, longer disease duration, higher baseline Sharp score, and steroid use were associated with lower response rates. Predictive models incorporating these factors explained 29%, 37%, and 53% of variance in ACR20, ACR50, and ACR70, respectively. Overall, comparing observed versus expected rates of response across trials more closely approximated results of head-to-head trials. For example, although observed responses numerically favored adalimumab to tofacitinib, comparison of observed versus expected results across trials more closely approximated the results from a head-to-head trial ("Oral Rheumatoid Arthritis triaL [ORAL] Strategy").

Conclusion. We identified factors associated with ACR response in RA trials. Adjusting for expected outcomes yielded therapy comparisons somewhat more similar to head-to-head trials. These findings could inform other across-trial comparisons, particularly when head-to-head trials are lacking.

INTRODUCTION

In recent years, the number of disease-modifying treatment options for rheumatoid arthritis (RA) has dramatically expanded beyond the conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) thanks to the introduction of new drugs against diverse therapeutic targets. This includes tumor necrosis factor α (TNFα) inhibitors, interleukin-6 inhibitors, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) co-stimulation modulators, Janus kinase inhibitors, and anti-CD20 antibodies. The efficacy of these drugs both individually and in combination with csDMARDs, especially methotrexate (MTX), is well-established based on randomized controlled clinical trials, marking welcome progress in a prevalent and morbid disease.

Comparing response rates of these drugs with one another is challenging because of the significant heterogeneity between study populations recruited for clinical trials. It has also not been

[Correction added on 19 July 2022, after first online publication: The article title was corrected in this version.]

Mr. Cordisco is supported by a Rheumatology Research Foundation Medical and Graduate Student Preceptorship. Dr. Baker is supported by a Veterans Affairs (VA) Clinical Science Research and Development Merit Award (I01 CX001703) as well as by the VA Rehabilitation Research & Development (I21 CX003157; I01 CX003644). Dr. George is supported by NIH grant K23AR073931 and funding from GSK. The contents of this work do not represent the views of the Department of the VA or the United States Government.

[1]Anthony J. Cordisco, MA, Michael D. George, MD, MSCE: Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [2]Marianna Olave, BA: Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [3]Joshua F. Baker, MD, MSCE: Philadelphia VA Medical Center, Philadelphia, Pennsylvania, USA.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr2.11468&file=acr211468-sup-0001-Disclosureform.pdf.

Address correspondence to Anthony J. Cordisco, MA, University of Pennsylvania, Jordan Medical Education Center, Floor 6, 3400 Civic Center Boulevard, Building 421, Philadelphia, PA 19104. Email: anthony.cordisco@pennmedicine.upenn.edu.

Submitted for publication December 30, 2021; accepted in revised form May 6, 2022.

SIGNIFICANCE & INNOVATIONS

- A number of patient and trial characteristics were associated with clinical responses in clinical trials of pharmacologic therapies for rheumatoid arthritis.
- Differences in patient and trial characteristics explained 37% and 53% of the variance in response rates across treatment arms for the proportion of patients achieving 50% or 70% improvment in the ACR's RA disease score (ACR50 and ACR70).
- Consideration of patient and trial characteristics facilitated somewhat more accurate cross-trial comparison of therapies compared with published head-to-head trials.

practicable to conduct head-to-head trials directly comparing the efficacy of all the different combinations of available therapies. Observational studies are influenced by channeling and confounding by indication, which limit the interpretability of comparative data. We hypothesized that certain patient characteristics and features of clinical trial design would have an important effect on response rates. Such factors might include, among others, patient age, sex, duration of RA, rheumatoid factor (RF) and/or anticitrullinated peptide antibody (ACPA) serostatus, baseline disease activity, number of prior drug failures, and whether or not a clinical trial was sponsored by a pharmaceutical company. Understanding the role of patient and trial factors in predicting treatment response would help facilitate across-trial comparisons and add to methods such as network meta-analysis (1).

Which patient baseline characteristics and trial features are associated with clinical response rates and to what degree clinical trial response rates can be predicted from these factors has not been established. Successful identification of predictive factors would facilitate more accurate trial-to-trial comparisons of efficacy through population adjustment. We sought to identify patient and study factors associated with the proportion of patients achieving 20%, 50%, or 70% improvement in the American College of Rheumatology's (ACR) RA disease score (ACR20/50/70) by using published data from a large number of phase II/III clinical trials. We constructed multivariable models incorporating these factors to predict ACR responses and facilitate more accurate direct comparison of therapies.

MATERIALS AND METHODS

Identification of relevant trials and data extraction. We identified relevant published clinical trials through multiple strategies. We leveraged the database of clinical trials compiled for the ACR's efforts to update its management guidelines for RA as a reference. We also performed literature searches through PubMed to identify additional trials. As a secondary analysis of previously published data, this study was not considered human subjects research and therefore was exempt from Institutional Review Board review.

Table 1. Treatment arms by therapy and therapy class

Therapy class	No. of treatment arms	Therapy	No. of treatment arms
csDMARDs	58	MTX	43
		Triple therapy	4
		Other csDMARDs	11
TNFα inhibitor mono	14	Adalimumab mono	7
		Infliximab mono	0
		Etanercept mono	4
		Golimumab mono	2
		Certolizumab pegol mono	1
IL-6 inhibitor mono	5	Sarilumab mono	1
		Tocilizumab mono	4
JAK inhibitor mono	9	Tofacitinib mono	8
		Baricitinib mono	1
Rituximab mono	1	Rituximab mono	1
Abatacept mono	1	Abatacept mono	1
TNFα inhibitor + MTX	51	Adalimumab + MTX	14
		Infliximab + MTX	12
		Etanercept + MTX	4
		Golimumab + MTX	11
		Certolizumab pegol + MTX	10
IL-6 inhibitor + MTX	15	Tocilizumab + MTX	11
		Sarilumab + MTX	4
Abatacept + MTX	10	Abatacept + MTX	10
Rituximab + MTX	5	Rituximab + MTX	10
JAK inhibitor + MTX	15	Tofacitinib + MTX	8
		Baricitinib + MTX	7

Abbreviations: csDMARD, conventional synthetic disease-modifying antirheumatic drug; IL-6, interleukin-6; JAK, Janus kinase; MTX, methotrexate; TNFα, tumor necrosis factor alpha.

Inclusion criteria were as follows: articles had to report on randomized controlled phase II/III trials of now-approved disease-modifying drugs for RA published between 1995 and 2018 with outcomes assessed at time-points greater than 3 months. Although phase II studies were not excluded, nearly all the included studies were phase III. Safety trials, cost-effectiveness trials, meta-analyses, long-term open-label extension studies, studies with crossover or treat-to-target designs, and studies that used exclusively functional, quality of life, or patient-reported outcomes were not included. Studies that reported neither ACR20/50/70 responses nor Disease Activity Score-28 (DAS28)-ESR/CRP as endpoints were excluded. Nearly all the trials were anonymized although this was not strictly required for inclusion. A very small number of trials did not anonymize patients due to difficulty concealing particular drugs' delivery mechanisms. However, in these cases, evaluators remained anonymized. In situations in which multiple articles reported on clinical trial results from multiple follow-up times, we used data from the longest follow-up time reported (before patients entered an open-label extension period). When numerical values for ACR response were not provided, the numbers were interpolated from graphs. A total of 565 articles were initially screened based on their titles of which 410 were reviewed in more detail. Ultimately, 72 articles totaling 185 treatment arms were included; the complete list and their references are provided in Supplementary Table 1.

Key trial characteristics and study outcomes. The primary treatment endpoints of interest were the ACR20/50/70 response rates. The rates of ACR response were extracted for each of the treatment arms to which patients were randomized. Planned secondary endpoints included the proportion of patients achieving DAS28-ESR or DAS28-CRP remission, although these endpoints were reported much less frequently and were not analyzed.

Treatment arms were assigned a numerical identifier based on the therapy that the patients in that arm were randomized to receive. This included an identifier for the specific therapy (eg, adalimumab) and an identifier for the drug class (eg, TNFα inhibitor). Treatment arms for biologic therapies were also coded based on whether the biologic was given concurrently with a csDMARD (eg, tocilizumab + MTX). Separate variables also specified whether or not the patients in the treatment arm were MTX naïve and/or biologic naïve. The MTX-naïve variable was used to distinguish whether MTX use represented a new, active therapy initiation versus continuation of previous background therapy. The specific therapies, drug classes, and drug combinations that were evaluated in the study are defined in Table 1. As we sought specifically to create models that could inform treatment comparisons, therapies themselves were not included in predictive models.

We extracted 34 baseline patient or trial factors that were hypothesized to be potentially important predictors of clinical response rates. These included trial characteristics (eg, whether or not it was primarily sponsored by a pharmaceutical company) and characteristics of participants randomized to a given treatment arm (eg, average disease duration prior to enrollment). In general, many variables were not reported in every eligible trial arm and, indeed, some variables of theoretical importance were found to be rarely reported, such as patient smoking status. The baseline characteristics and response endpoints extracted, and the proportion of treatment arms reporting them are detailed in Supplementary Tables 2A and B.

Statistical analysis. An initial series of univariate regressions was performed to describe associations between baseline patient and trial characteristics (independent variables) and ACR20, 50, and 70 response rates (dependent variables) among all trial arms with available data. These analyses then informed multivariable predictive models.

Table 2. Full multivariable predictive models for ACR50 response

	β (95% CI)
Model 1	
Age, y*	−1.00 (−1.88 to −0.12)
Disease duration, y	−0.53 (−1.18 to 0.12)
Calendar year published*	0.40 (0.00 to 0.80)
Early escape permitted**	−7.83 (−12.3 to −3.35)
Number of patients**	0.027 (0.01 to 0.045)
Follow up ≥52 wk	2.47 (−2.38 to 7.32)
Constant*	−714 (−1520 to 96.4)
Model 2	
Age, y	1.25 (−0.62 to 3.12)
Disease duration, y*	−1.54 (−3.17 to 0.09)
Calendar year published	−0.14 (−1.60 to 1.31)
Early escape permitted	−0.14 (−7.17 to 6.89)
Number of patients	−0.016 (−0.048 to 0.015)
Follow up ≥52 wk	7.12 (−1.53 to 15.8)
DAS28-CRP**	17.6 (8.05 to 27.2)
Constant	176 (−2750 to 3100)
Model 3	
Age, y	−0.39 (−1.61 to 0.83)
Disease duration, y*	−0.99 (−1.99 to 0.013)
Calendar year published*	0.54 (−0.073 to 1.16)
Early escape permitted	−6.51 (−13.0 to −0.01)
Number of patients	0.014 (−0.013 to 0.042)
Follow up ≥52 wk	4.05 (−3.26 to 11.4)
ESR, mm/h*	0.27 (−0.06 to 0.59)
Constant	−1040 (−2300 to 203)
Model 4	
Age, y	4.14 (−1.82 to 10.1)
Disease duration, y	2.82 (−5.36 to 11.0)
Calendar year published	1.87 (−6.53 to 10.3)
Early escape permitted	−18.5 (−60.7 to 23.7)
Number of patients	0.012 (−0.15 to 0.18)
Follow ≥52 wk	−0.51 (−28.0 to 27.0)
DAS28-CRP*	38.0 (3.72 to 72.4)
ESR, mm/h*	−0.22 (−2.28 to 1.85)
Constant	−4170 (−21,300 to 13,000)

Abbreviations: ACR50, percentage of patients achieving at least 50% improvement in American College of Rheumatology rheumatoid arthritis disease score; CI, confidence interval; DAS28-CRP, Disease Activity Score-28-C-reactive protein; ESR, erythrocyte sedimentation rate.
* $P < 0.10$.
** $P < 0.01$.
*** $P < 0.001$.

Predictive models were developed in a multistage process. Variables that were associated with ACR response in univariate analyses ($P < 0.10$) were included in multivariable models in a multistage process to avoid dropping studies that did not report certain characteristics. An initial model was constructed from variables that were associated with response in univariate analyses and that were reported in every study. We then performed sequential regressions in which other variables of interest were added to this base model one at a time and tested in the subset of studies for which those data were available. Variables that were independently associated with ACR response were then included in a group of final predictive models. In a sensitivity analysis, we also explored weighting the predictive models based on the size (n) of the trial arm.

Predictive models were used to determine an expected ACR20, ACR50, and ACR70 score for each treatment arm based on available trial and patient characteristics. From combinatorics, predicting ACR responses for every treatment arm required the total number of models contained within a group to be 2^n, where n was the number of variables associated with response that were not reported in every treatment arm. There were 2 models for ACR20, 4 for ACR50, and 16 for ACR70; the 4 models for ACR50 are detailed as an example in Table 2. The model selected for a given treatment arm was selected based on the information available from that study (algorithm for ACR50 shown in Figure 1).

We then generated the observed minus the expected response for each treatment arm, thereby defining residual responses over and beyond what the predictive models attributed to patient and trial characteristics. We then compared the observed and residual response rates of therapies with one another to determine how the apparent differences in efficacy changed when accounting for the trial and patient characteristics. Both the nominal and residual response rates for a given therapy were determined using meta-analysis across all treatment arms using that therapy in order to account for varying sample sizes. This was done for abatacept versus adalimumab and tofacitinib versus adalimumab (all combined with MTX) because the results could be compared with those of two published head-to-head trials: the "Abatacept versus adaliMumab comParison in bioLogic-naïvE RA subjects with background methotrexate (AMPLE)" trial of abatacept + MTX versus adalimumab + MTX and the "ORAL Strategy" trial of tofacitinib + MTX versus adalimumab + MTX (2,3).

All statistical analysis was performed using Stata 14.2 (StataCorp, LP, College Station, TX).

RESULTS

Association of trial characteristics with ACR response rates. Trial characteristics associated with ACR response rates are shown in Table 3. A number of trial features and population baseline characteristics were associated with ACR responses. Later calendar year of publication and higher baseline DAS28-CRP score at enrollment were associated with greater likelihood of achieving all degrees of ACR response. Response rates were also higher in trials with a greater proportion

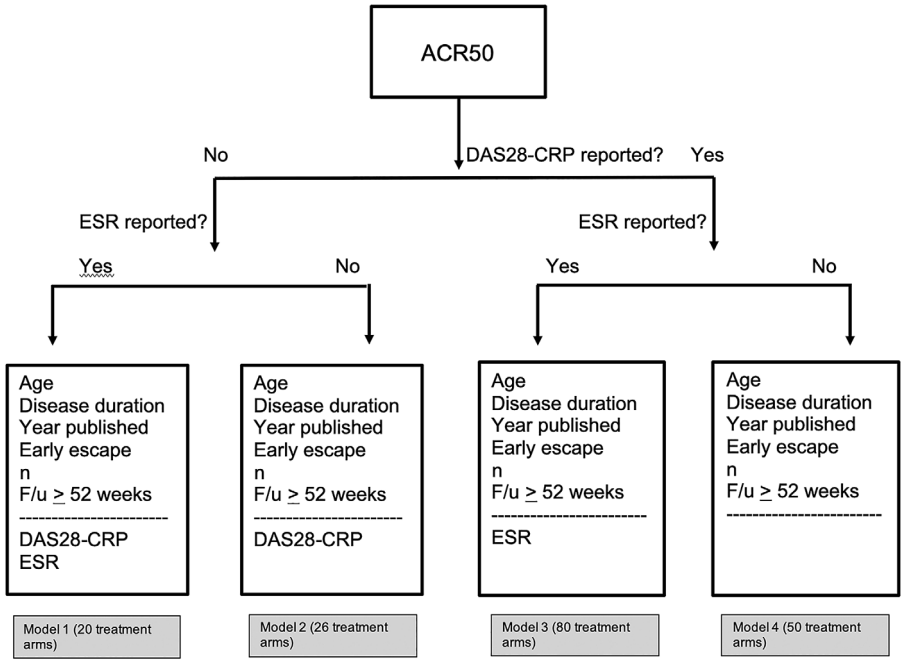

Figure 1. Algorithm for choosing a regression model for a treatment arm to predict the expected outcome for ACR50 response. Variables above the dashed lined were reported in all treatment arms extracted and are included in every model, albeit with varying coefficients. ACR50, percentage of patients achieving at least 50% improvement in American College of Rheumatology rheumatoid arthritis disease score; DAS28-CRP, Disease Activity Score-28-C-reactive protein; ESR, erythrocyte sedimentation rate; F/u, follow-up.

Table 3. Single variable regressions for ACR20/50/70 with baseline patient/trial characteristics

Baseline characteristic	ACR20 (95% CI)	ACR50 (95% CI)	ACR70 (95% CI)
Calendar year published (N = 174)	0.52* (0.07 to 0.97)	0.61** (0.22 to 1.00)	0.55*** (0.26 to 0.84)
DAS28-ESR (N = 114)	0.30 (−7.61 to 8.20)	1.65 (−5.26 to 8.56)	2.16 (−3.14 to 7.46)
DAS28-CRP (N = 46)	13.3** (4.88 to 21.7)	13.6** (6.21 to 21.0)	8.42* (1.96 to 14.89)
Age, y (N = 174)	−1.66** (−2.65 to −0.67)	−1.90*** (−2.76 to −1.04)	−1.63*** (−2.26 to −1.00)
Female, % (N = 174)	−0.05 (−0.47 to 0.36)	−0.12 (−0.48 to 0.25)	−0.13 (−0.41 to 0.15)
White, % (N = 91)	0.04 (−0.06 to 0.15)	0.06 (−0.04 to 0.16)	0.05 (−0.02 to 0.13)
Smokers, % (N = 5)	−1.43 (−3.10 to 0.25)	−1.03 (−2.37 to 0.31)	−0.66 (−1.55 to 0.23)
BMI, kg/m^2 (N = 19)	−1.06 (−5.87 to 3.75)	−0.62 (−4.37 to 3.13)	−0.53 (−2.94 to 1.88)
HAQ (N = 157)	−4.97 (−17.4 to 7.48)	0.74 (−10.4 to 11.9)	3.22 (−5.44 to 11.9)
SJC (28 joints) (N = 24)	−1.46 (−3.58 to 0.65)	−1.32 (−3.37 to 0.73)	−1.19 (−2.92 to 0.54)
TJC (28 joints) (N = 24)	−1.06 (−2.22 to 0.09)	−0.97 (−2.00 to 0.05)	−0.89 (−1.74 to −0.03)
SJC (66 joints) (N = 148)	−0.45 (−1.22 to 0.32)	−0.29 (−0.97 to 0.38)	−0.20 (−0.70 to 0.30)
TJC (68 joints) (N = 148)	−0.39 (−0.91 to 0.13)	−0.23 (−0.69 to 0.22)	−0.09 (−0.43 to 0.25)
Patient Pain (VAS) (N = 101)	0.26 (−0.33 to 0.86)	0.22 (−0.30 to 0.75)	0.16 (−0.23 to 0.55)
Patient Global Score (N = 105)	−0.11 (−0.57 to 0.35)	−0.13 (−0.54 to 0.28)	−0.03 (−0.33 to 0.27)
Evaluator Global Score (N = 105)	−0.05 (−0.65 to 0.55)	−0.17 (−0.71 to 0.36)	−0.14 (−0.53 to 0.25)
Sharp score (N = 80)	−0.17 (−0.35 to 0.01)	−0.22** (−0.37 to −0.07)	−0.21*** (−0.33 to −0.096)
Using corticosteroids, % (N = 125)	−0.01 (−0.21 to 0.19)	−0.10 (−0.27 to 0.08)	−0.14* (−0.27 to −0.02)
Disease duration, y (N = 174)	−1.14** (−1.82 to −0.45)	−1.45*** (−2.03 to −0.87)	−1.44*** (−1.85 to −1.03)
RF+, % (N = 134)	0.14 (−0.15 to 0.42)	0.18 (−0.06 to 0.42)	0.16 (−0.02 to 0.35)
ACPA+, % (N = 53)	−0.27 (−0.71 to 0.18)	0.17 (−0.25 to 0.60)	0.36[a] (0.01 to 0.71)
ESR, mm/h (N = 98)	0.30 (−0.11 to 0.71)	0.34* (0.00 to 0.68)	0.29* (0.05 to 0.52)
CRP, mg/L (N = 154)	−0.01 (−0.28 to 0.26)	−0.03 (−0.27 to 0.20)	−0.03 (−0.21 to 0.15)
MTX naïve, % (N = 162)	0.07* (0.01 to 0.13)	0.10*** (0.05 to 0.15)	0.10*** (0.07 to 0.14)
Biologic naïve, % (N = 106)	0.11* (0.01 to 0.22)	0.10* (0.01 to 0.19)	0.06 (0.00 to 0.13)
Sponsored by industry (N = 174)	−0.72 (−7.97 to 6.53)	0.67 (−5.55 to 6.89)	1.96 (−2.70 to 6.62)
F/u ≥ 52 wk (N = 174)	4.47 (−0.72 to 9.67)	9.27*** (4.88 to 13.7)	9.86*** (6.73 to 13.0)
Early escape permitted (N = 174)	−9.28*** (−14.3 to −4.22)	−7.50** (−12.0 to −2.99)	−5.63** (−9.01 to −2.25)
Non-Western study (N = 174)	1.67 (−5.15 to 8.49)	−0.17 (−6.22 to 5.87)	−1.41 (−5.94 to 3.13)
Number of patients (N = 174)	0.029** (0.01 to 0.049)	0.034*** (0.016 to 0.051)	0.031*** (0.018 to 0.044)

Note: N is the number of observations for a given association between one of the ACR responses and the baseline variable of interest.
Abbreviations: ACPA, anticitrullinated peptide antibody; ACR20/50/70, percentage of patients achieving at least 20%, 50%, or 70% improvement in American College of Rheumatology rheumatoid arthritis disease score; BMI, body mass index; CI, confidence interval; CRP, C-reactive protein; DAS28, Disease Activity Score-28; ESR, erythrocyte sedimentation rate; F/u, follow-up; HAQ, health assessment questionnaire; MTX, methotrexate; RF, rheumatoid factor; SJC, swollen joint count; TJC, tender joint count; VAS, visual analog score.
* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

of subjects who were either MTX or biologic naïve and in trials enrolling a greater number of patients. More advanced age, longer disease duration at enrollment, prior MTX or biologic failure, and a trial design permitting early escape were associated with a lower likelihood of achieving all degrees of ACR response. Trials with patients with a higher baseline ESR and a follow-up of 52 weeks or more also had higher rates of achieving ACR50 and ACR70 response, whereas a higher baseline radiographic damage score (Sharp score) was associated with a lower likelihood of response. ACR70 response was also more likely in trials with a greater percentage of patients who were seropositive for ACPA and less likely in those with a greater percentage of patients using corticosteroids at baseline.

Notable variables that were not significantly associated with ACR response rates included sex, swollen/tender joint counts, patient and evaluator global scores, RF serostatus, whether or not a trial was primarily sponsored by a pharmaceutical company as opposed to an academic or government institution, and the geographic location of the trial (using the first author's institution location as a proxy). Although the DAS28-CRP score was associated with treatment response, there was no significant association for baseline DAS28-ESR or CRP itself among the studies that reported these quantities. In exploratory models weighting based on trial size, patient pain visual analog score and smoking were also found to be associated with ACR20 response (not shown), although other predictors were similar.

Predictive models. The full groups of models used for ACR50 are shown in Table 2. Pearson correlations between the predicted ACR20, ACR50, and ACR70 response rates and the actual response rates yielded correlation coefficients of 0.53, 0.62, and 0.73, respectively, suggesting that predictive models accounted for 29%, 37%, and 53% of the variance in the response rates, respectively.

Without accounting for patient and trial factors, the average ACR20, ACR50, and ACR70 response rates appeared

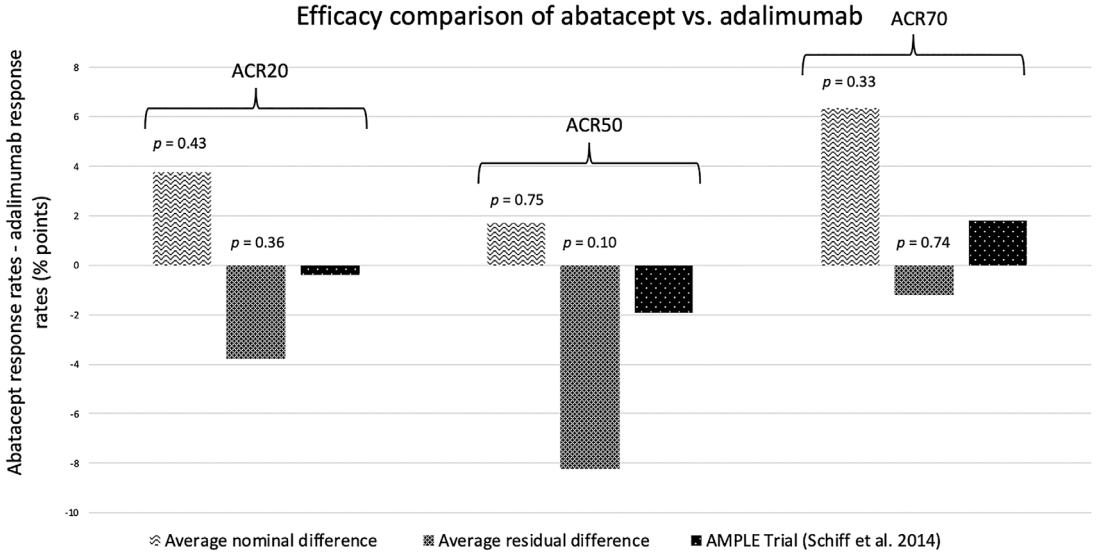

Figure 2. Efficacy comparison of abatacept versus adalimumab (both with background MTX). ACR20/50/70, percentage of patients achieving at least 20%, 50%, or 70% improvement in American College of Rheumatology rheumatoid arthritis disease score; AMPLE, Abatacept versus adaliMumab comParison in bioLogic-naïvE RA subjects with background methotrexate; MTX, methotrexate.

numerically higher for abatacept compared with adalimumab (both combined with background MTX) by 3.8% (P = 0.43), 1.7% (P = 0.75), and 6.3% (P = 0.33), respectively (Figure 2). The differences in residual response were not statistically significant but numerically favored adalimumab: –3.6% (P = 0.39), –7.0% (P = 0.11), and –5.2% (P = 0.10). In the head-to-head AMPLE trial, there were no significant advantages for abatacept compared with adalimumab in ACR20, ACR50, and ACR70: –0.4%, –1.9%, and 1.8%, respectively (2).

Without accounting for patient and trial factors, average ACR20, ACR50, and ACR70 response rates numerically favored adalimumab compared with tofacitinib (both combined with MTX) by 6.5% (P = 0.04), 8.0% (P = 0.04), and 5.9% (P = 0.19), respectively. However, the differences in the residual responses (adjusted for trial factors) of tofacitinib versus adalimumab were smaller and not statistically significant: 3.6% (P = 0.39), 7.0% (P = 0.11), and 5.2% (P = 0.10), respectively (Figure 3). In the head-to-head "ORAL" Strategy trial, tofacitinib + MTX response

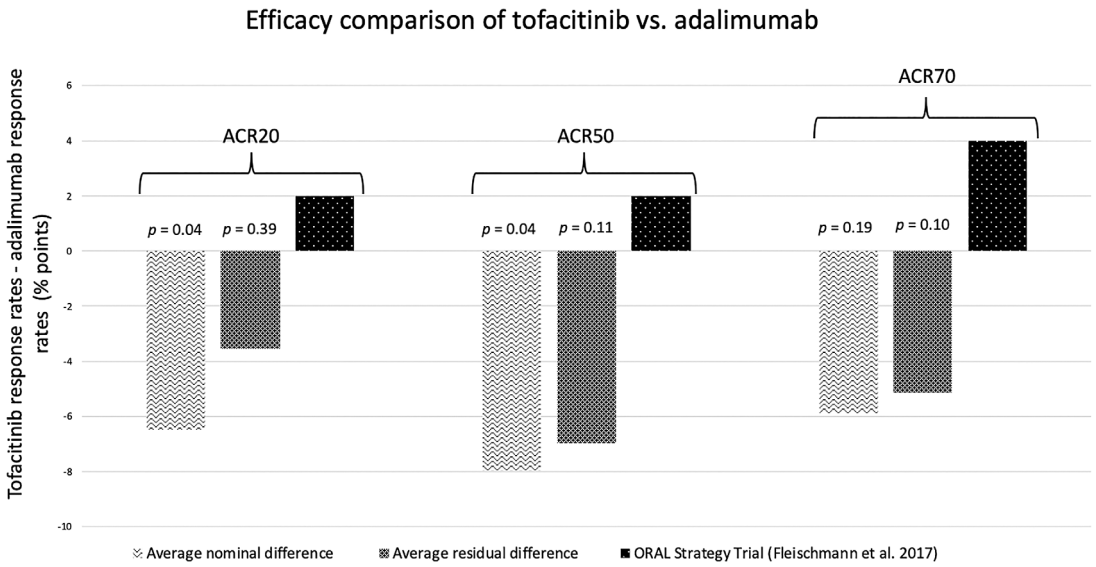

Figure 3. Efficacy comparison of tofacitinib versus adalimumab (both with background MTX). ACR20/50/70, percentage of patients achieving at least 20%, 50%, or 70% improvement in American College of Rheumatology rheumatoid arthritis disease score; MTX, methotrexate; ORAL, Oral Rheumatoid Arthritis triaL.

rates were not significantly different but were numerically higher than for adalimumab + MTX by 2%, 2%, and 4%, respectively.

DISCUSSION

We identified multiple patient-specific and trial-specific factors associated with ACR response rates among clinical trials of pharmacologic therapies for RA. Of note, we identified that more recent year of publication, higher baseline DAS28-CRP score, higher proportion of MTX- or biologic-naïve subjects, higher ESR, ACPA positivity, larger sample size, and a follow-up time of 52 weeks or more were each associated with greater ACR response. Meanwhile, longer disease duration, more advanced age, higher radiographic damage score, corticosteroid use at baseline, and trial design allowing early escape were all associated with lower ACR response rates. All of these variables were commonly reported baseline characteristics of clinical trials. Awareness of their association with ACR response may allow for a more nuanced interpretation of response rates from individual trials and may inform future trial design.

The positive association between higher baseline DAS28-CRP score and ACR response may reflect the fact that ACR20, ACR50, and ACR70 describe proportional improvements in disease severity. Patients who enroll in a clinical trial with higher baseline disease activity have greater opportunity to achieve greater proportional improvement. Higher baseline disease activity has previously been noted to be associated with greater ACR response. In a Swedish cohort study using patients on TNF inhibitors, patients with greater baseline DAS28 score were more likely to achieve an ACR20 response (odds ratio [OR]: 1.44, 95% confidence interval [CI]: 1.19-1.75) (4). Thus, it should be expected that trials that recruit patients with higher disease activity will have higher ACR response rates.

We also observed that older age, greater disease duration, greater joint damage, and prior treatment failure were all associated with lower response rates in clinical trials. Symptoms that are due to accumulated tissue damage might be less likely to respond to immunomodulatory therapies aimed at reducing the activity of the inflammatory disease. Our observations support prior studies that have suggested that longer disease duration and prior treatment failure predict poorer response to therapy (5,6). The Swedish cohort study of TNF inhibitors also found that earlier age at the start of therapy was significantly associated with ACR50 response (4).

There are likely several reasons why studies using an early escape design reported lower response rates. In many studies, patients who are allowed early escape are considered treatment non-responders from that point forward. Also, studies with an early escape design had a significantly lower percentage of patients who were MTX naïve (13.7% vs. 41.0%, $P = 0.001$) and trend toward a lower percentage being biologic naïve (78.7% vs. 90.5%, $P = 0.06$). It is possible that studies aimed at enrolling patients who have had RA for a long time and/or failed prior therapies are more likely to build early escape into their design in anticipation of some patients doing poorly. Ultimately, early escape was a more useful variable than treatment naïveté for predictive models because, as a feature of clinical trial design, it was reliably reported. In contrast, failure of prior therapies was not universally described.

We found that there was an association between the number of subjects enrolled in clinical trials and response rates, but this is unlikely to be causal. The number of subjects enrolled is moderately positively correlated with the calendar year of publication (r [174] = 0.26, $P = 0.0004$), and more recent publication date is itself associated with response. Additionally, trials with follow-up times of 1 year or greater had significantly more patients enrolled on average than those with follow-up times under 1 year: 236 versus 155, $P < 0.001$.

Prior studies have identified factors associated with response to therapy that we did not identify in our study. One systematic review evaluated randomized controlled trials and observational studies (case-control, cohort, and case series designs) (7). The investigators found that female sex was associated with lower response rates, largely based on a study by Anderson et al in 2000 (5). Another meta-analysis of randomized controlled trials of DMARDs found that placebo-arm ACR20 response was lower in clinical trials performed largely in non-Western countries, but we found no such association (8). Previously, smoking has been found to be a negative predictor of ACR response in cohort studies. In one such study, it was associated with lower likelihood of response in infliximab patients with an OR of 0.77 (95% CI: 0.60-0.99) (9). Unfortunately, smoking status was rarely reported in clinical trials assessed in our study.

Using the factors associated with response, we constructed multivariable models that accounted for a substantial proportion of the variance in ACR response, particularly for ACR70. These models were used to calculate expected ACR responses and the residuals (observed minus the expected). Comparison of the residual responses for two DMARD comparisons—abatacept versus adalimumab and tofacitinib versus adalimumab—yielded differences that, for some levels of ACR response, were more like those seen in published head-to-head trials. However, this effect was modest.

Investigators have aimed to overcome barriers in comparing data from different trials through network meta-analysis (10). These studies compare therapies indirectly by taking advantage of similar comparison arms. However, network meta-analyses require no violation of the transitivity assumption and rarely consider differences in patient and trial characteristics, which may influence treatment effects across trials and affect validity. Population adjustment for both prognostic variables and effect modifiers may add value to traditional network meta-analysis. Although the data presented in this study demonstrate improvements in the ability to compare therapies across different trials

through population adjustment, the reduction in bias is generally modest, and residual bias remains a concern. Use of population adjustment in combination with network meta-analysis may be of value.

A significant limitation of this study is that many of the variables shown to be associated with ACR response were not consistently reported in clinical trial articles. For example, although baseline DAS28-CRP score was strongly associated with ACR response, it was only reported in 31% of the treatment arms extracted and primarily after 2008. Similarly, ACPA was associated with ACR response but was uncommonly reported (31% of articles) and also primarily after 2008. Other variables of theoretical interest, such as smoking, were reported so infrequently that no informative analysis could be done.

We circumvented the problem of missing data by building groups of models for ACR20, ACR50, and ACR70 that used different sets of independent variables. However, missing data for key variables likely limited the variance these models could explain. Unmeasured baseline variables may yet explain a significant proportion in the remaining total variance in ACR response. The results of the current study importantly suggest that greater efforts to characterize trial populations in publications of clinical trials will be of value. This will help to better understand how individual trial populations compare with others regarding important patient factors, and to understand how trial population might have influenced response rates. A machine learning approach might have improved the predictive capacity of the model; however, we did not use this approach because no validation sample was available. Finally, the factors we have identified should not be interpreted as having a causal effect on response to therapies, although some of the associations identified in this work may warrant further investigation.

In conclusion, head-to-head trials of contemporary biologic DMARDs are uncommon, and it is tempting to compare efficacies of different therapies by comparing results from clinical trials conducted under distinct circumstances. This study characterized patient and trial factors associated with ACR response rates and used these factors to adjust across-trial comparisons. The models constructed in this work produce comparisons of DMARDs more similar to known head-to-head studies, although the effect is modest and may suffer from residual confounding. This type of approach may be useful for comparing drugs and drug classes across clinical studies, especially when no head-to-head trials exist, and when other approaches such as network meta-analysis are not appropriate or possible.

ACKNOWLEDGMENTS

Mr. Cordisco would like to acknowledge the support of a Rheumatology Research Foundation Medical and Graduate Student Preceptorship. He would also like to acknowledge constructive feedback on this project courtesy of Alexis R. Ogdie-Beatty, MD, MSCE. Dr. Baker would like to acknowledge the support of a Veterans Affairs Clinical Science Research & Development Merit Award (I01 CX001703) and a Rehabilitation Research & Development Merit Award (I01 RX003644). The contents of this work do not represent the views of the Department of Veterans Affairs or the United States government. The authors would like to thank the ACR Rheumatoid Arthritis Management Guideline Committee for providing access to the literature review.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. All authors have agreed to be accountable for all aspects of the work, had full access to all of the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Cordisco, Baker.
Acquisition of data. Cordisco, Olave, Baker.
Analysis and interpretation of data. Cordisco, George, Baker.

REFERENCES

1. Leahy J, Walsh C. Assessing the impact of a matching-adjusted indirect comparison in a Bayesian network meta-analysis. Res Synth Methods 2019;10:546–68.
2. Schiff M, Weinblatt ME, Valente R, van der Heijde D, Citera G, Elegbe A, et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis: two-year efficacy and safety findings from AMPLE trial. Ann Rheum Dis 2014;73:86–94.
3. 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.
4. Karlsson JA, Kristensen LE, Kapetanovic MC, Gülfe A, Saxne T, Geborek P. Treatment response to a second or third TNF-inhibitor in RA: results from the South Swedish Arthritis Treatment Group Register. Rheumatology (Oxford) 2008;47:507–13.
5. Anderson JJ, Wells G, Verhoeven AC, Felson DT. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum 2000;43:22–9.
6. Sanmarti R, Gomez A, Ercilla G, Gratacos J, Larrosa M, Suris X, et al. Radiological progression in early rheumatoid arthritis after DMARDS: a one-year follow-up study in a clinical setting. Rheumatology (Oxford) 2003;42:1044–9.
7. 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.
8. Nagai K, Matsubayashi K, Ide K, Seto K, Kawasaki Y, Kawakami K. Factors influencing placebo responses in rheumatoid arthritis clinical trials: a meta-analysis of randomized, double-blind, placebo-controlled studies. Clin Drug Investig 2020;40:197–209.
9. Hyrich KL, Watson KD, Silman AJ, Symmons DP, British Society for Rheumatology Biologics Register. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology (Oxford) 2006;45:1558–65.
10. Orme ME, Macgilchrist KS, Mitchell S, Spurden D, Bird A. Systematic review and network meta-analysis of combination and monotherapy treatments in disease-modifying antirheumatic drug-experienced patients with rheumatoid arthritis: analysis of American College of Rheumatology criteria scores 20, 50, and 70. Biologics 2012;6:429–64.