

SUPPLEMENT SECTION

METHODS

Patient inclusion and exclusion criteria

Patients included in the study had active disease at baseline, defined by at least 8 of 68 tender joints and 6 of 66 swollen joints, and with CRP >10 mg/L. All patients received stable MTX dose (allowed range 10 to 25 mg/week) for a minimum of 6 weeks prior to the screening visit; patients within the Asia-Pacific Region (Taiwan, South Korea, Malaysia, Philippines, Thailand, and India) were allowed a minimum dose of 6 mg/week. Patients must have continued the stable dose of MTX for the duration of the study. No MTX dose increase was allowed in the study from 6 weeks prior to screening to EOT visit. All patients received folic acid according to country regulations to prevent MTX toxicity.

Patients below age 18 or above age 75 were excluded from the study. Patients were also excluded if they had autoimmune disease other than RA, or systemic involvement of significance (such as vasculitis, pulmonary fibrosis, Felty's syndrome), received treatment with DMARDs other than MTX within 4 weeks or 12 weeks (depending on the DMARD half-life) prior to screening. Use of parenteral glucocorticoids or intraarticular prednisone within 4 weeks prior to screening was excluded, as was use of oral glucocorticoid greater than 10 mg/day or equivalent/day, or a change in dosage within 4 weeks prior to the baseline visit. Patients with past history of nonresponse to anti TNF therapy, and those with a past history of non-response to treatment with biologic agents for RA or being treated with biologic agents within 3 months of screening were also excluded.

Patient assessments

At the randomisation/baseline visit (week 0) and at weeks 2, 4, 8, and 12, safety, vital sign and efficacy assessments (that included a complete joint examination for tender joint count and swollen joint count) were performed. Adverse events and SAEs were assessed weekly until end of treatment (EOT). Patients who discontinued treatment before week 12 were assessed at the EOT visit with complete clinical and laboratory evaluation. Only patients who completed the study period and chose not to enroll in the open-label, long-term extension study had a post-treatment safety follow-up clinic visit on day 127 (week 18), 6 weeks after the EOT visit.

Assessment of TJC and SJC were performed by an assessor independent from the investigator and blinded from the patient's data. Laboratory tests were performed at baseline and throughout the study. Throughout the study, the investigator, sponsor, and patient were blinded to CRP and serum IL-6 levels. All data were collected during scheduled clinic and home visits.

Study treatment

Treatment was administered as a single 2 ml SC injection in the abdomen on day 1 (Week 0) at the investigational site; patients were trained to self-administer treatment on non-clinic visit days. Clinic visits occurred every other week to day 85 (Week 12). Home visits were carried out on days 43 (Week 6) and 71 (Week 10).

Statistical analysis

Analysis of primary efficacy endpoint:

Correction for the multiplicity that arose from testing multiple doses of sarilumab against placebo was addressed using the Hommel procedure with the Hommel adjusted $p < 0.05$ considered statistically significant.^{41,42}

Patients who discontinued treatment due to lack of efficacy or used rescue medication were considered as non-responders for all time points beyond the time they discontinued or started rescue medication. For patients who discontinued due to reasons other than lack of efficacy, a last observation carried forward (LOCF) procedure was applied to missing data for all 7 ACR components from the point of treatment discontinuation or rescue.

To demonstrate the robustness of the conclusions, a statistical sensitivity analysis was performed for the ACR20 response. The sensitivity analysis used the LOCF approach to impute missing tender and swollen joint counts (TJC, SJC), and considered all patients 'non-responders' for all subsequent analysis time points after treatment discontinuation (for any reason) or rescue medication use. The remaining 5 ACR components were not imputed.

Analysis of secondary efficacy endpoint

Descriptive statistics including number of subjects, mean, standard error and least-square means (LS means) were provided. In addition, difference in LS means, the corresponding 95% CI and the p-value were provided for comparisons of each sarilumab dose against placebo. An ANOVA model, including terms for treatment, prior biologic use and region, was used to assess treatment differences in ACR (defined as the lowest percentage improvement from baseline of three measures: TJC, SJC, and median improved score of the five remaining ACR components) at Week 12. Descriptive

statistics including number of subjects, mean, standard error and LS means were provided. In addition, difference in LS means, the corresponding 95% CI and the p-value were provided for comparisons of each sarilumab dose against placebo. The multiplicity in the secondary efficacy endpoints of ACR50 and ACR70 was corrected *post hoc* using simple Bonferroni adjustment with the unadjusted $p < 0.01$ considered statistically significant.

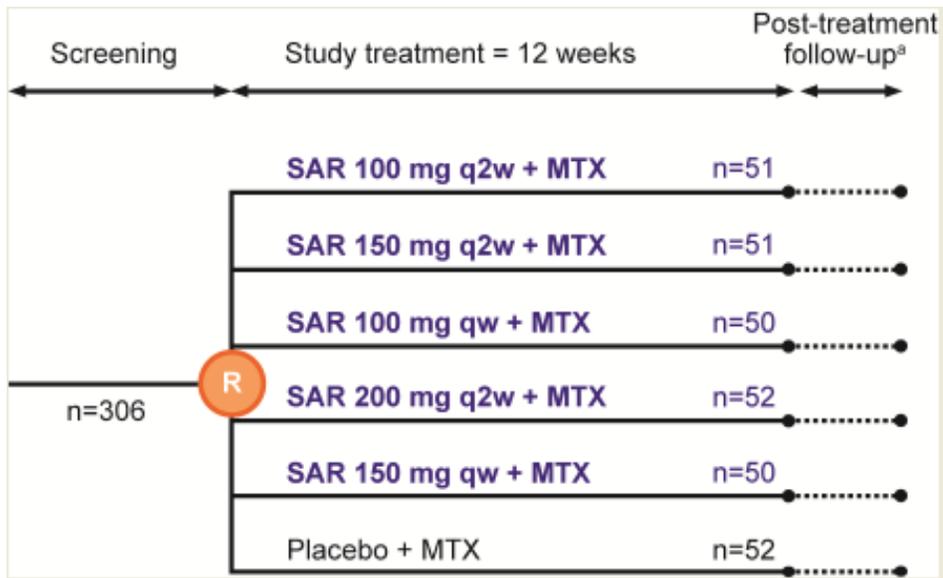
RESULTS

Safety

Autoimmune or lupus-like syndrome, gastrointestinal ulceration or confirmed diverticulitis, hypersensitivity reactions or anaphylaxis, as well as neurological disorders and opportunistic infections were either absent or occurred with low frequency (between 2% and 4%) across all treatment groups.

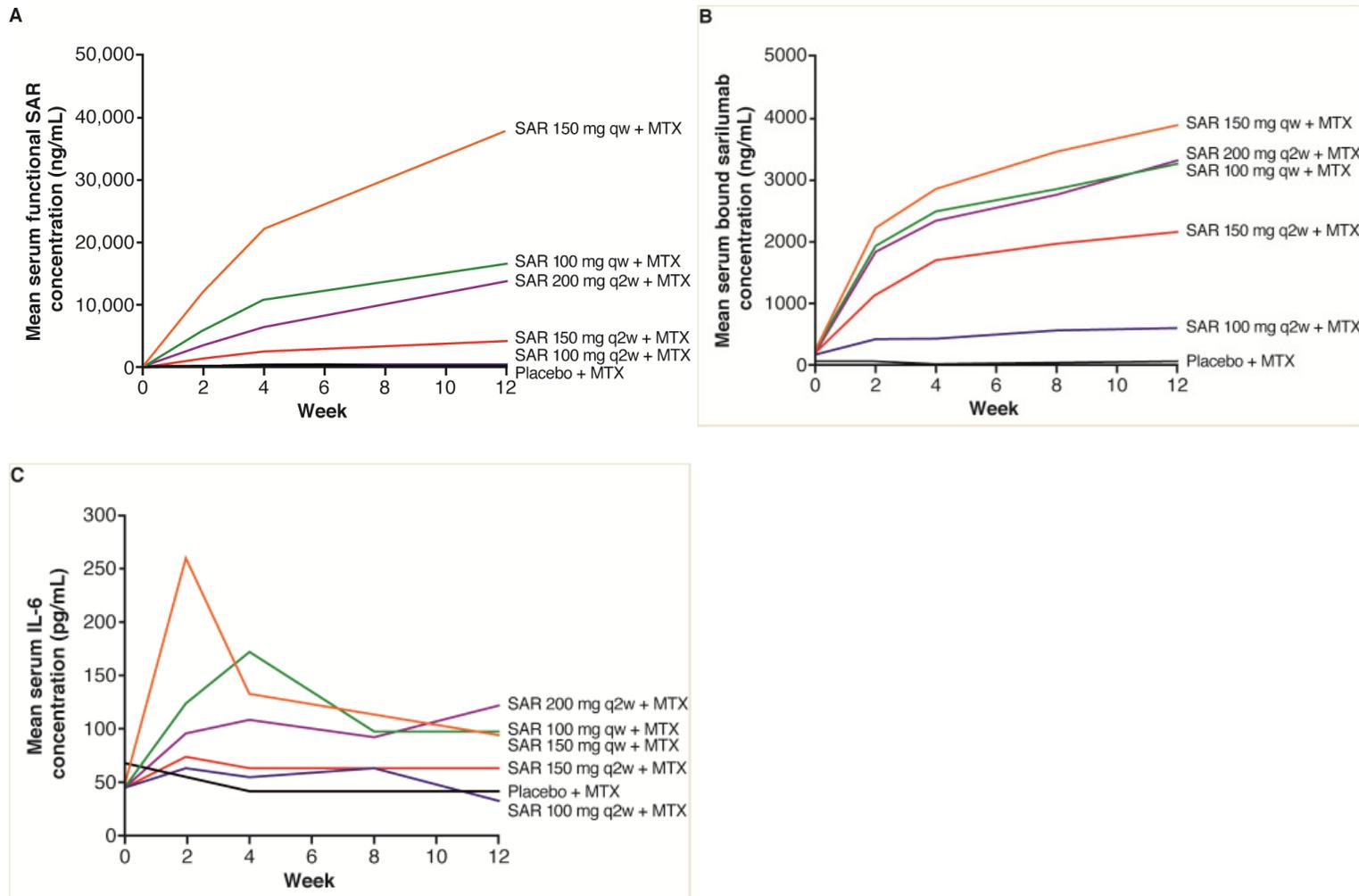
Drug-induced liver injury, defined as ALT $> 3 \times$ ULN and total bilirubin $> 2 \times$ ULN (with no evidence of biliary obstruction), was not reported in any patient with an elevated ALT. Changes from baseline in other liver function parameters, glucose, haemoglobin A1c, triglycerides and creatine phosphokinase were infrequent and did not require medical intervention.

Overall, changes in haematologic parameters from baseline at week 12 were not considered to be clinically significant and did not require medical intervention. There were no abnormalities noted in physical presentation, vital signs, urine pH, electrolyte and renal function parameters. The proportion of patients experiencing altered ECG parameters was low and similar across all groups; no dose response was noted with respect to QT prolongation by either the Bazett or the Fridericia calculation.^{43,44}



Supplement figure 1 Study design. ^aPatients could enter a long-term open-label extension; 6 weeks of follow up for those who did not enter the extension.

SAR=sarilumab, MTX=methotrexate 10 to 25 mg/week, q2w=every 2 weeks, qw=every week, SC=subcutaneous, R=randomization, n=number of patients enrolled.



Supplement figure 2 SAR pharmacokinetic and pharmacodynamic properties over time. A) Mean serum functional SAR PK trough concentrations, B) mean serum bound SAR PK trough concentrations, and C) mean serum IL-6 concentrations. SAR=sarilumab, MTX=methotrexate, q2w=every 2 weeks, qw=every week, IL-6=Interleukin 6, sIL-6R α =soluble interleukin-6 receptor-alpha.

Supplement Table 1 CDAI scores at baseline, week 12, and change from baseline □ ITT population

CDAI	Placebo (N=52)	SAR 100 mg q2w (N=51)	SAR 150 mg q2w (N=51)	SAR 100 mg qw (N=50)	SAR 200 mg q2w (N=52)	SAR 150 mg qw (N=50)
Baseline						
Number	51	51	51	50	51	50
Mean (SD)	40.63 (12.85)	44.74 (13.53)	41.41 (13.31)	40.32 (10.82)	40.37 (12.32)	40.48 (10.22)
Week 12						
Number	52	51	51	49	50	50
Mean (SD)	25.99 (16.09)	26.94 (17.03)	18.79 (13.17)	18.38 (11.82)	16.90 (11.78)	19.85 (15.34)
P value for change from baseline*	□	0.2494	0.0056	0.0122	0.0025	0.0361

CDAI=clinical disease activity index

*Using ANOVA (type 3) with factors: Planned Arm Number. Records with missing values for factors or responses were excluded from statistical analyses.

Supplement Table 2 Incidence of ACR20 response at week 12 □ treatment-by-subgroup interaction testing □ ITT population

Subgroup	P-value for interaction*
Gender (male vs female)	0.5402
Race (Caucasian/white vs all other races)	0.9524
Region (Western countries, South America, rest of the world) [†]	0.2262
Age group (<65, ≥65 and <75, ≥75 years)	0.4174
Baseline weight (<50, ≥50 and <100, ≥100 kg)	0.9202
BMI (<25, ≥25 and < 30, ≥30 kg/m ²)	0.3428
Prior biological use (yes, no) [‡]	0.6949
Rheumatoid factor (positive, negative)	0.5486
Anti-CCP antibody	0.9022
Baseline CRP (≤1.5 mg/dL, >1.5 mg/dL)	0.2666
Duration of RA (≤ median, > median in years)	0.8895
Number of prior DMARDs (none, 1, 2, ≥3)	1.000
Smoking history (yes, no)	0.4746

*Logistic regression model with terms of treatment, prior biologic use, region, subgroup, treatment-by-subgroup

[†]Logistic regression model with terms of treatment, prior biologic use, region, treatment-by-region

[‡]Logistic regression model with terms of treatment, prior biologic use, region, treatment-by-prior biologic use

Supplement Table 3 Incidence of ACR20 response at week 12 – patients with and without prior biologic use

ACR20 at Week 12 n (%)	Placebo (N=52)	SAR 100 mg q2w (N=51)	SAR 150 mg q2w (N=51)	SAR 100 mg qw (N=50)	SAR 200 mg q2w (N=52)	SAR 150 mg qw (N=50)
Prior biologic use						
Number	12	13	12	12	14	12
Responders	4 (33.3)	7 (53.8)	8 (66.7)	5 (41.7)	9 (64.3)	7 (58.3)
Non-responders	8 (66.7)	6 (46.2)	4 (33.3)	7 (58.3)	5 (35.7)	5 (41.7)
OR (95% CI) vs placebo*	□	3.14 (0.54, 18.40)	3.69 (0.71, 19.08)	1.45 (0.27, 7.98)	3.96 (0.74, 21.23)	2.67 (0.54, 13.28)
No prior biologic use						
Number	40	38	39	38	38	38
Responders	20 (50.0)	18 (47.4)	26 (66.7)	26 (68.4)	25 (65.8)	29 (76.3)
Non-responders	20 (50.0)	20 (52.6)	13 (33.3)	12 (31.6)	13 (34.2)	9 (23.7)
OR (95% CI) vs placebo*	□	0.88 (0.35, 2.21)	2.08 (0.82, 5.27)	2.20 (0.83, 5.88)	1.98 (0.77, 5.06)	4.52 (1.46, 14.03)

Reason for discontinuation of a prior biologic were: costs (n=8), lack of efficacy (n=2), other (n= 65). Per protocol exclusion criteria, patients could not enter the trial if they had experienced an inadequate response to a prior biologic.

*Mantel-Haenszel estimate: model stratified by region.