CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

761037Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

SECONDARY STATISTICAL MEMORANDUM

CLINICAL STUDIES

NDA/BLA #:	BLA 761037
Drug Name:	Kevzara (sarilumab)
Indication:	Treatment of adult patients with moderately to severely active rheumatoid arthritis who have had an inadequate response or intolerance to one or more disease-modifying antirheumatic drugs (DMARDs)
Applicant:	Sanofi
PDUFA Date:	May 22, 2017
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Gregory Levin
Concurring Reviewers:	Yongman Kim
	Thomas Permutt
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Suzette Peng, Janet Maynard
Project Manager:	Christine Ford

Keywords: effect size, intent-to-treat, labeling, missing data

BACKGROUND

The efficacy of Kevzara (sarilumab) was evaluated in detail in the primary statistical review by Dr. Yongman Kim, and I agree with the key conclusions of Dr. Kim's review. In particular, there is convincing statistical evidence that sarilumab is effective for treatment of patients with moderately to severely active rheumatoid arthritis (RA) who have had an inadequate response or intolerance to one or more disease-modifying antirheumatic drugs. The reader is referred to Dr. Kim's review dated September 2, 2016 for additional details on the design and results of the phase 3 clinical trials supporting the effectiveness of sarilumab.

This memorandum elaborates on a specific aspect of the statistical review: the determination of the most appropriate statistical methodology for evaluating and presenting radiographic progression results in labeling. Drug effects on progression of structural damage, as assessed by x-ray and the modified total Sharp score (mTSS), are expected to predict drug effects on important long-term patient outcomes, such as decline in function and risk of disability. The mTSS is calculated by summing a radiologist's quantification of the extent of joint space width at 42 joints and bone erosion at 44 joints. The score has a range of 0–448.

The effects of sarilumab 150 mg and 200 mg, administered subcutaneously every other week, on radiographic progression were evaluated in a single study: Study 11072 (Part B Cohort 2) was a randomized, double-blind, placebo-controlled, parallel-group, 52-week, phase 3 clinical trial. Starting at Week 16, patients on placebo who had less than 20% improvement in either their swollen or tender joint counts at two consecutive visits or any other clear lack of efficacy based on investigator judgment were proposed to be rescued with open-label sarilumab 200 mg. Escape from placebo to sarilumab occurred in 29% and 39% of patients randomized to placebo by Week 24 and Week 52, respectively. X-rays were taken in patients at baseline, Week 24, and Week 52, as well as at the time point of rescue in the subset of patients who escaped, unless an x-ray had been taken in the previous three months.

ESTIMAND OF INTEREST

Before considering the choice of statistical methodology, it is critical to discuss and identify the estimand of interest, i.e., the specific measure of drug effect on radiographic progression that is of interest. During the sarilumab review, we considered two potential estimands: (1) the *de facto* or *treatment policy* estimand, i.e., the difference in mean change in mTSS over 52 weeks between all patients assigned to sarilumab and all patients assigned to placebo *regardless of escape*; and (2) the difference in mean change in mTSS over 52 weeks between all patients assigned to sarilumab and all patients assigned to placebo *in a setting where patients on placebo do not receive biologic escape therapy.*

There are pros and cons of each of these potential estimands. Despite the inclusion of data after placebo escape (cross-over) to sarilumab, evaluation of the treatment policy estimand (#1) is expected to be sensitive to drug effects, given that structural damage is generally understood to be irreversible, i.e., any joint space or erosion changes occurring on placebo in the first four months of the trial are not expected to go away after escape to sarilumab. This is in contrast to

evaluations of symptomatic endpoints, such as joint counts and functional assessments, which may show considerable improvement toward baseline values within a few weeks of biologic treatment. Furthermore, the treatment policy estimand (#1) reflects a real-world effect of assignment to sarilumab versus assignment to placebo as an add-on to methotrexate (MTX) in MTX inadequate responders. However, the public health relevance of this comparison is unclear, given that the control arm may be receiving a sub-standard-of-care treatment—inadequate responders to MTX would typically receive treatment with a biologic such as a TNF inhibitor or IL-6 inhibitor (or a small-molecule JAK inhibitor) in clinical practice.

Estimand #2, the difference between treatment groups *in the absence of escape therapy on placebo*, has appeal in that it is not impacted by cross-over between treatment arms and thus its evaluation may be more sensitive to drug effects than the evaluation of estimand #1. However, its relevance from a public health perspective is also unclear, given that it involves a comparison against a control arm (placebo + MTX, without biologic escape, in inadequate responders to MTX) that is a hypothetical rather than real-world treatment regimen. Furthermore, any evaluation of this estimand needs to rely on unverifiable assumptions given current ethical considerations for clinical trials in RA.

We note that considerations about the estimand(s) of interest in a specific clinical trial setting are greatly impacted by the study design and in particular, the choice of control group.[1] For example, as discussed above, the relevance of the treatment policy estimand might be questioned in the sarilumab phase 3 trial due to the comparison against a sub-standard-of-care treatment policy (placebo + MTX, in inadequate responders to MTX). However, if the control arm instead receives a reasonable representation of standard of care, the treatment policy estimand compares patient outcomes between two potential real-world treatment regimens and is of clear interest from a public health perspective. For example, a trial could compare a new biologic to an active biologic control in MTX inadequate responders, or could compare a new biologic to MTX in MTX-naïve patients. The evaluation of the treatment policy estimand in trials with these designs would provide information relevant to actual treatment decisions being made in clinical practice. Consideration should therefore be given to such alternative designs for generating evidence of and evaluating the extent of drug effects on radiographic progression.

STATISTICAL METHODOLOGY

The pre-specified statistical analysis of the effect of sarilumab on radiographic progression utilized a linear regression model and an approach often termed *linear extrapolation* to handle missing and post-escape data. The linear extrapolation approach, which has been used in previous RA trials, imputes a single Week 52 value in patients who escape or withdraw from the study prior to Week 52. In the applicant's analysis, patient data after escape (+14 days) were considered missing. Then, the applicant fit a line through the baseline score and the last observed radiographic score before escape (+14 days) and used that line to assign a Week 52 value to the patient. If the interest is in estimand #2, the linear extrapolation approach requires the assumption that placebo patients' scores on average would, in the absence of escape,

[1] Ideally, the discussion about the estimand of interest would happen *before* the discussion about the design and choice of control group.

continue to change at the same linear rate as was observed through the time of escape. This assumption is strong and unverifiable, and may tend to overstate true progression on placebo. In addition, the linear extrapolation approach is a single-imputation method that does not appropriately take into account the statistical uncertainty in the imputation process. This leads to underestimates of the variability and overestimates of the degree of evidence of a treatment effect. Alternatively, one can interpret the linear extrapolation approach as providing an estimate of the difference between treatment groups in the on-treatment slope. However, this estimand is likely not a meaningful surrogate for long-term benefit in all patients. For example, patients on sarilumab who show no early progression but cannot tolerate or adhere to the therapy will show benefit (no progression) according to the on-treatment slope despite the fact that they may have progressed after treatment discontinuation.

An alternative approach includes in the analysis all observed Week 52 x-ray data, including data collected after treatment discontinuation or escape, with patients analyzed according to their randomized treatment group. This analysis reliably targets estimand #1, the treatment policy estimand. This analysis might also be expected to conservatively target estimand #2, given that patients on placebo who meet escape criteria would be expected to have less future progression in the absence of escape to an effective biologic therapy.

During the review cycle, some concerns were raised about the analysis based on all observed data because of the inclusion of data after cross-over from placebo to sarilumab. Therefore, we also considered alternative analyses to more reliably evaluate estimand #2 than the pre-specified linear extrapolation approach. In particular, we sent an information request to the applicant requesting an additional analysis comparing slopes of progression between the treatment arms. The analysis utilized a linear mixed effects model and included all radiographic data observed prior to escape (+14 days), including such data collected at any time point during the 52-week double-blind period. Patient data on the placebo arm after escape (+14 days) were considered missing. Observed data on both sarilumab arms after escape were included in the analysis. The model included the following as covariates: time (study day of x-ray / 365.25), treatment, treatment-by-time interaction, region, and prior biologic use. The treatment-by-time interaction coefficients for the two sarilumab dosing regimens represent differences in slopes (differences in mean changes per year) versus placebo and were of primary interest. This analysis still relies on strong and unverifiable assumptions, e.g., that progression is on average roughly linear over time and that missing values after escape in placebo patients who escape are similar to values over time among placebo patients with observed data, conditional on a linear model of the baseline covariates and the time of the x-ray, and the observed value prior to escape. However, the analysis more appropriately handles statistical uncertainty (presuming the assumptions hold) than the single-imputation linear extrapolation approach. We note that there are a number of alternative methodological approaches that could be considered for evaluating estimand #2—additional research regarding the most appropriate analysis is warranted.

The applicant expressed concerns during the resubmission review cycle about the possibility of including results from a *post hoc* analysis in labeling. We agree that analysis pre-specification is a critical aspect of clinical trial design, and that unplanned, data-driven analyses conducted after completion of a clinical trial to try to identify results supporting drug effects can be biased and

should be interpreted with caution.[2] However, the *post hoc* analyses conducted as part of this review are fundamentally different. We agree with the applicant that there is convincing evidence of an effect of sarilumab on radiographic progression, and agree that the labeling should include a claim of this effect. The question of interest is not *whether there is evidence of a treatment effect* but rather *what analysis provides the best of the treatment effect* that we all agree exists. Furthermore, our discussions about the most appropriate choice of analysis have been based on scientific and statistical considerations rather than based on data-driven considerations about what analysis looks best, or looks worst, or whether results differ greatly between methods. Therefore, the *post hoc* nature of the analyses does not introduce bias in results, and such results should be included in labeling if there is an agreement that they are more appropriate scientifically than those produced with the pre-specified methodology.

RECOMMENDATIONS

During the review of the original BLA submission, the statistical and clinical review teams had numerous discussions with the applicant regarding the most appropriate presentation of radiographic results in labeling. Because of concerns with the linear extrapolation approach, the statistics review team recommended the inclusion of results based on the analysis including all observed data, including data collected after escape. The applicant ultimately agreed to our recommendation—the observed data results are in the version of the label submitted by the applicant in the resubmission in response to the complete response. However, during the review of the applicant's resubmission, we had additional internal discussions about the choice of estimand and methodology. Discussions with clinical colleagues identified an interest in estimand #2, the difference between treatment groups in the absence of biologic escape on placebo. Therefore, we explored alternative methodologies to more reliably evaluate this estimand than the previously implemented linear extrapolation approach. Ultimately, we requested that the applicant carry out a mixed effects model approach to compare slopes of progression, as described above.

We believe that either the analysis including all observed data or the analysis comparing slopes of progression based on a mixed effects model would be a more appropriate choice for deriving and presenting results in labeling than the linear extrapolation approach. Our recommendation is based on the following:

(1) We have concerns with the reliability of results based on linear extrapolation because such results rely on strong and unverifiable scientific assumptions and the use of inappropriate statistical methodology, and more appropriate alternative statistical approaches are available.

[2] Certain *post hoc* analyses conducted to try to identify *lack of evidence of drug effects* would also be problematic. For example, if the primary analysis is valid, it is not appropriate for FDA to conduct a variety of alternative primary analyses with similar assumptions as the primary analysis to try to identify analyses that do not generate evidence of an effect. Such an approach does not evaluate sensitivity to violations in assumptions and could increase the chance of failing to approve an effective therapy. Instead, sensitivity analyses should evaluate whether findings persist under plausible alternative assumptions to those of the primary analysis—these analyses might need to be conducted *post hoc* if not adequately pre-specified.

(2) There is convincing evidence of a treatment effect with those more appropriate alternative statistical approaches, such that a conclusion that sarilumab delays radiographic progression does not need to rely on the linear extrapolation–based results. Results for all three approaches are shown in Table 1 below.
(3) Use of either of the alternative analyses would be more consistent with the recommendations in the 2010 National Research Council Report report *The Prevention and Treatment of Missing Data in Clinical Trials*. In particular, our considerations are based on the goals of evaluating the estimand of interest with minimal and plausible missing data assumptions and ensuring that results are convincing even if those assumptions are violated. The observed data analysis reliably targets one estimand of interest and conservatively targets an alternative estimand of interest, with minimal assumptions. The mixed effects model analysis excluding post-escape data on placebo and comparing slopes is also considered reasonable in this setting—it targets a clear estimand of interest, and although underlying assumptions are unverifiable, sensitivity analyses establish that results are convincing under plausible, alternative assumptions.

Table 1. Evaluation of Effects of Sarilumab on Radiographic Progression using Three Alternative Statistical Methodologies

	Placebo + MTX (N = 398)	**Sarilumab 150 mg + MTX (N = 400)**	**Sarilumab 200 mg + MTX (N = 399)**
Linear Regression Analysis based on All Observed Data			
Mean Change at 52 Weeks	2.02	0.58	0.15
Difference from Placebo + MTX (95% CI)		-1.43 (-2.01, -0.85)	-1.86 (-2.45, -1.28)
Linear Regression Analysis based on Linear Extrapolation			
Mean Change at 52 Weeks	2.53	0.65	0.01
Difference from Placebo + MTX (95% CI)		-1.88 (-2.74, -1.01)	-2.52 (-3.38, -1.66)
Mixed Effects Model Analysis to Compare Slopes			
Rate of Change per year over 52 Weeks	2.24	0.64	0.23
Difference from Placebo + MTX (95% CI)		-1.60 (-2.20, -1.00)	-2.01 (-2.61, -1.41)

Sources: Reviewer analyses and applicant May 8, 2017 response to information request
Abbreviations: MTX=methotrexate; CI=confidence interval

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

GREGORY P LEVIN
05/18/2017

YONGMAN KIM
05/18/2017

THOMAS J PERMUTT
05/18/2017
I concur.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	BLA 761,037/0000
Drug Name:	Kevzara™ (sarilumab) 200 mg s.c.
Indication(s):	Treatment of Rheumatoid Arthritis (RA)
Applicant:	Sanofi-Aventis
Date(s):	Submitted: October 30, 2015
	PDUFA: October 30, 2016
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yongman Kim, Ph.D.
Concurring Reviewers:	Gregory Levin, Ph.D.
	Thomas Permutt, Ph.D.
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Suzette Peng, M.D.
	Janet Maynard, M.D.
Project Manager:	Christine Ford

Keywords: BLA, clinical studies, early escape, missing data

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Sanofi-Aventis has proposed Kevzara™ (sarilumab) for the treatment of adult patients with moderate to severe rheumatoid arthritis (RA) who have had an inadequate response to previous therapy with one or more disease-modifying antirheumatic drugs (DMARDs) (b) (4) The applicant submitted the results from two phase 3 clinical trials, EFC11072 and EFC10832 (11072 and 10832 for short hereafter), to support the efficacy of sarilumab for the treatment of RA. The applicant claims that the results from these trials provide substantial evidence of efficacy by the predefined primary efficacy endpoint ACR20 at Week 24.

Based on my review of the data from the two phase 3 studies, 11072 and 10832, there is sufficient evidence to support the efficacy of sarilumab 150 mg and 200 mg in treating patients with RA. The analysis of the primary efficacy endpoint, ACR20 at Week 24, was statistically significant in the two studies reviewed. In Study 11072, the ACR20 response rates were 58%, 66%, and 33% for the sarilumab 150 mg, 200 mg, and placebo arms, respectively. In Study 10832, the response rates were 56%, 61%, and 34% for the sarilumab 150 mg, 200 mg, and placebo arms, respectively. This evidence was further supported by the analyses of secondary endpoints, including HAQ-DI at Week 16, and Major Clinical Response and mTSS at Week 52 in Study 11072, and HAQ-DI at Week 12, and DAS28-CRP and SF-36 PCS at Week 24 in Study 10832. Therefore, from a statistical perspective, the overall package provided substantial evidence of sarilumab's efficacy benefit.

2 INTRODUCTION

2.1 Overview

This application (BLA #761,037) was submitted on October 30, 2015 in support of sarilumab 150 mg and 200 mg doses for the treatment of patients with rheumatoid arthritis as an original Biological License Application.

Sarilumab is a human IgG1 monoclonal antibody that binds specifically to both soluble and membrane-bound interleukin-6 (IL-6) receptors (sIL-6Rα and mIL-6Rα), and has been shown to inhibit IL-6 mediated signaling through these receptors. Overproduction of IL-6 has been suggested to play pathological roles in rheumatoid arthritis and there is precedent that targeting the receptor with monoclonal antibodies confers therapeutic benefits in this indication. As the first drug in the IL-6 inhibitor class, Acetmra (tocilizumab) was approved for RA by FDA in 2010. Kevzava™ (sarilumab) is proposed to be available in 150 mg/1.14 mL and 200 mg/1.14 mL pre-filled syringes. The applicant recommends the subcutaneous (SC) dose regimen of sarilumab 200 mg every two weeks (q2w) with an option to reduce dose to 150 mg q2w to manage those patients who experience decreased neutrophil counts, decreased platelet counts or elevated liver transaminases.

The submission included the results from two phase 3, randomized, double-blind, placebo-controlled studies, 11072 and 10832, that were similar in design. The objective of the

phase 3 studies was to evaluate the efficacy and safety of sarilumab compared with placebo in patients with RA. Patients were to receive randomized, double-blind study treatment for 52 weeks in Study 1072 and for 24 weeks in Study 10832. The primary efficacy outcome variable was the response rate of ACR20 at Week 24.

History of Drug Development and Regulatory Interactions

The sarilumab clinical development program for RA was introduced to the Division of Pulmonary, Allergy, and Rheumatology Products under IND 100,632. Communication with the applicant regarding their development plan is documented under this IND. Pertinent parts of the statistical portion of those communications are summarized herein.

In September 2009, the applicant had a type C meeting with the Division, where input was received regarding the proposed phase 2/3 study 11072:

- *Agency agreed to Study EFC11072 with an operationally seamless design (proposed Phase 2/3 study). Also, the Agency indicated that patients in cohort B1 should not be included in efficacy analyses, but should be used for safety.*

In September 2011, the applicant had an End-of-Phase 2 meeting with the Division. The Division provided the following statistical comments on the proposed protocols and analysis plan:

- *The proposed analysis plan for the radiographic endpoint is generally acceptable. You should continue to obtain radiographs on patients who escape or drop out of the study in order to be able to provide a sensitivity analysis based on these data (i.e., a "retrieved dropout" analysis). If sensitivity analyses are not consistent with the primary analysis using linear extrapolation, then it may not be possible to draw a definitive conclusion regarding treatment effect on radiographic outcomes.*
- *The Division responded (to the Applicant's question regarding non-inferiority design with active comparator - tocilizumab) that the purpose of the comparison is not for a marketing claim, nor to establish superiority to another product, but rather to gather data to aid in the assessment of sarilumab's toxicity profile compared to that of other products in the same drug class.*

In April 2012, the Division provided the following responses to the applicant's request for advice and comments:

- *The Agency agreed with the modification of the terminology for the primary objectives and co-primary endpoints while maintaining the previously proposed hierarchical testing procedure.*
- *The Agency advised on approaches to handling data for treatment-related dropouts. Measures should be taken to protect the reliability of the radiographic endpoint.*
- *The Agency agreed that it would be appropriate to claim statistical significance of the treatment effect on the ACR20 endpoint even in the absence of a significant effect on the other 2 co-primary endpoints. However, the Agency may not consider this sufficient for regulatory purposes. A benefit in terms of signs and symptoms only may or may not be sufficient to balance the benefit-risk profile for the product and would be a review issue.*

In May 2013, the Division provided the following responses to the applicant's request for advice and comments:

- *The Agency agreed with the proposed analyses for the radiographic data, in particular the adequacy of the proposed sensitivity analyses for Study EFC11072.*
- *The Agency agreed with the Sponsor's approach to the analysis of the HAQ-DI endpoint.*

- *The Agency agreed with the Sponsor's proposed approached to handle missing data for the primary analysis of ACR20 and modified van der Heijde total Sharp score; however, the Agency did not agree with the approach to handle missing data for the HAQ-DI.*

In July 2014, the Division provided the following responses to the applicant's request for advice and comments:

- *The Sponsor proposed analysis of the co-primary endpoint for physical function (HAQ-DI) is in accordance with the Agency's previous advice on the SAP for Study EFC11072 received on 15 May 2013 and 21 August 2013.*

In October 2014, the applicant had a pre-BLA meeting with the Division. The Division provided the following statistical comments on the proposed protocols and analysis plan:

- *For all analyses (of safety), report estimated differences and confidence intervals for all pairwise treatment arm comparisons. In addition, for all integrated analyses, you should appropriately account for study differences, either by adjusting for study in a model or carrying out meta-analyses of within-study results.*

In December 2015, after the filing meeting, the Division sent the following statistical Information Request (IR) to the Applicant to help explore the potential effect of missing data on the reliability of efficacy results:

1. *For all endpoints proposed for inclusion on the product label, we request additional supportive analyses that include all observed data, including any outcomes collected after escape or discontinuation of study medication.*
2. *You have not provided sensitivity analyses that sufficiently evaluate the potential impact of missing data on the reliability of efficacy results. For co-primary endpoints, examine the potential effects of missing data and rescue on your results using tipping point sensitivity analyses.*

In February 2016, the Division sent the following statistical Information Requests (IRs) to the Applicant to clarify discrepancy between escape criteria and actual escapes:

1. *We calculate that, among patients in part B cohort 2 who were eligible for escape at Week 16 according to the swollen and tender joint count criteria, 75/140 (54%) of patients on placebo initiated rescue therapy, as compared to 19/79 (24%) and 20/71 (28%) of patients on sarilumab 150 mg and 200 mg, respectively. We calculated relatively similar results in Study EFC10832, with smaller differences between treatment arms in the proportions. Please clarify why such low proportions of patients on all treatment arms who met the escape criteria actually initiated escape therapy. In addition, while we recognize that the comparison between treatment arms is no longer a randomized comparison, provide any insight you may have into why a greater proportion of patients on placebo than sarilumab who met the rescue criteria actually initiated rescue therapy.*

2.1.1 Specific Studies Reviewed

The focus of this review is on the efficacy data from two phase 3 efficacy studies, 11072 Part B Cohort 2 and 10832. The design of the two studies is described in Table 1.

Table 1. Clinical Trials Reviewed

Trial No.	Phase	Design	Treatment Arms	Number of Patients	Dates*
11072	3	52-week,	sarilumab 150 mg	400	03/2011-

Part B Cohort 2		randomized, double-blind, parallel-group, placebo-controlled	sarilumab 200 mg	399	10/2013
			Placebo	398	
10832	3	24-week, randomized, double-blind, parallel-group, placebo-controlled	sarilumab 150 mg	181	10/2012-03/2015
			sarilumab 200 mg	184	
			Placebo	181	

Source: Reviewer
*Dates correspond to the start and the end of the study.

2.2 Data Sources

Data were submitted by the applicant to the CDER electronic data room in SAS transport format. Protocols, correspondence, data listings, program code, and study reports were accessed under the network path \\cdsesub1\evsprod\bla761037\761037.enx.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

In general, the submitted efficacy data were acceptable in terms of quality and integrity. I was able to derive the primary and secondary efficacy endpoints for the studies reviewed. No noticeable deviations between the raw datasets and analysis datasets relevant to primary and secondary endpoints were identified. The statistical analyses of my derived endpoints were consistent with the applicant's analyses.

Based on the information provided in this submission, each study seemed to be conducted properly and was consistent with the history of regulatory interactions and protocol revisions/amendments.

3.2 Evaluation of Efficacy

The applicant conducted two phase 3, randomized, double-blind, placebo-controlled international studies, 11072 and 10832. In nearly all study design features (eligibility criteria, study schedule, primary efficacy outcome variable and analysis, secondary and exploratory efficacy outcome measures and analyses), the studies were similar. The two studies differed mainly in the duration of the trial – 52 weeks in study 11072 and 24 weeks in study 10832 and the treatment prior to the studies – MTX in study 11072 and TNFα in study 10832. Also, the radiographic assessment of the structural damage progression was conducted only in study 11072.

3.2.1 Study 11072

The objective of the study was to evaluate the efficacy and safety of sarilumab 150 mg every 2 weeks and 200 mg every 2 weeks compared with placebo in patients with RA who were inadequate responders to or intolerant of methotrexate. Patients were to receive randomized study treatment in a double-blind manner for 52 weeks.

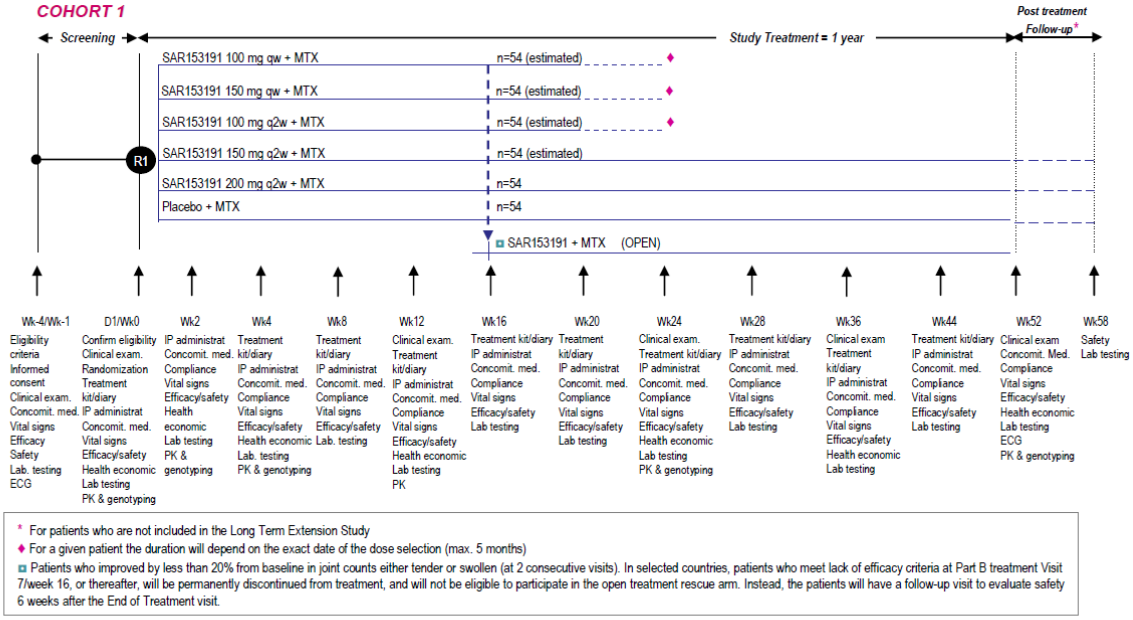
Study Design and Endpoints
This study had an operationally seamless design that used a Phase 2 design to select the relevant dose regimens (Part A, dose-ranging), followed by a Phase 3 design to test the selected dose regimens (Part B, pivotal), without interruption of patient recruitment. The design ensured complete protection of the double-blind.

Part A was a 12-week, 6-arm, dose-ranging study, intended to select the 2 dose regimens for further evaluation of sarilumab, based on efficacy (reduction in signs and symptoms) and safety. Patients were randomly assigned to receive either placebo once a week (qw), sarilumab 100 mg qw, sarilumab 150 mg qw, sarilumab 100 mg once every 2 weeks (q2w), sarilumab 150 mg q2w, or sarilumab 200 mg q2w, in a ratio of 1:1:1:1:1:1. The maximum duration of the study per patient was 22 weeks (up to 4 weeks for screening, 12 weeks for treatment, and 6 weeks for follow-up). Patients from Part A did not participate in Part B of the study. Patients who completed Part A and were eligible could enter an open-label, long-term, extension study (LTS11210). The results from Part A were analyzed after the last patient completed Part A and were presented in a separate study report.

Part B was a 52-week, 6-arm study with 2 cohorts, intended to confirm the efficacy and safety of the 2 dose regimens (150 mg q2w and 200 mg q2w) selected from Part A. Patients in Part B Cohort 1 were randomly assigned to receive either placebo qw, sarilumab 100 mg qw, sarilumab 150 mg qw, sarilumab 100 mg q2w, sarilumab 150 mg q2w, or sarilumab 200 mg q2w, in a ratio of 1:1:1:1:1:1, as in Part A (Figure 1). Once the results from Part A were known and the doses for further evaluation in Part B were selected, patients in Cohort 1 who were taking the selected doses of sarilumab 150 mg q2w or 200 mg q2w, or placebo, continued in this study. After dose selection, blinding of the patients in Cohort 1 who were receiving the selected doses (or placebo) was maintained. Patients in the "non-selected" sarilumab dose groups in Cohort 1 were discontinued from Study EFC11072 and could enter Study LTS11210. Patients for Part B Cohort 2 were recruited after dose selection from Part A and were randomly assigned to receive either placebo q2w, sarilumab 150 mg q2w, or sarilumab 200 mg q2w, in a ratio of 1:1:1 (Figure 2). Randomization was stratified by region and prior biologic use.

From Week 16, patients with lack of efficacy defined as less than 20% improvement from baseline in either swollen joint count (SJC) or tender joint count (TJC) for 2 consecutive visits, or any other clear lack of efficacy based on investigator judgment were proposed to be rescued with open-label sarilumab at the highest available dose at the time of transfer into the rescue treatment arm, and continued in the study according to their planned visit schedule. Patients that were not rescued were discontinued from the study. The maximum duration of the study per patient was 62 weeks (up to 4 weeks for screening, 52 weeks for treatment, and 6 weeks for follow-up). All patients who completed Part B and were eligible could enter Study LTS11210.

Figure 1. Study Schema for Study 11072 Part B Cohort 1

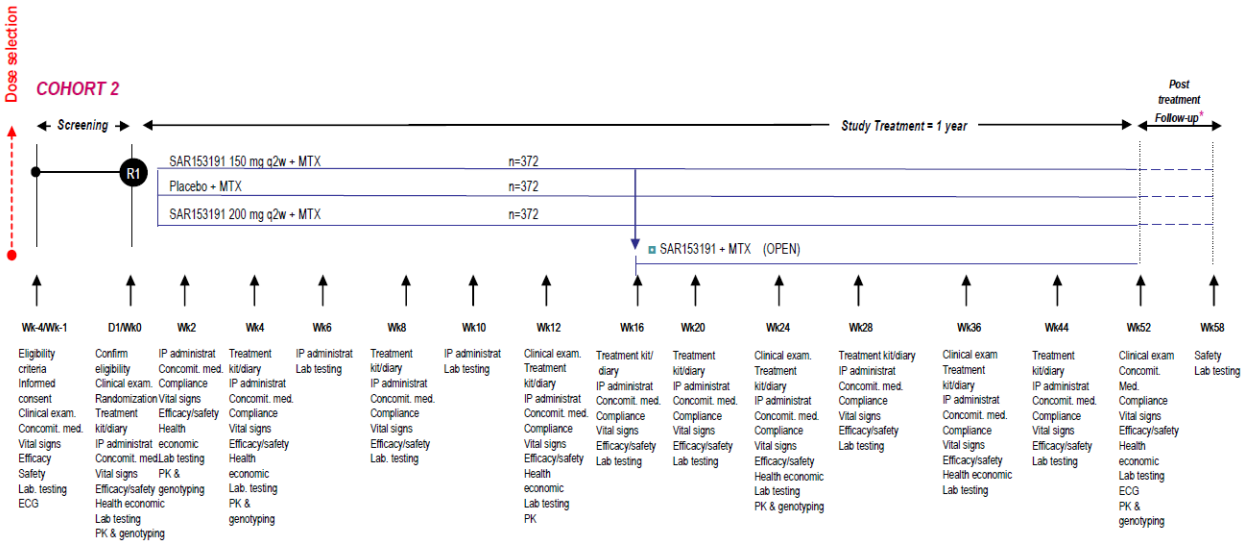


Open-label sarilumab was administered at a dose of 150 mg qw until the site was approved to enroll patients in Cohort 2.

Source: Excerpted from the Clinical Study Report for Study 11072 (page 24).

Figure 2. Study Schema for Study 11072 Part B Cohort 2

Open-label sarilumab was administered at a dose of 200 mg q2w.



Source: Excerpted from the Clinical Study Report for Study 11072 (page 25).

The design of this study had been discussed and agreed upon by the FDA. The operationally seamless design of this study required that patient enrollment in Part B start after the last patient was randomized in Part A, without waiting for the dose selection based on the results from Part A. Thus, to ensure complete protection of the double-blind, Part B patients belonged to 2 distinct cohorts according to the time of their enrollment.

The population enrolled in the study consisted of adults with active RA with an inadequate response to MTX. Following were some of the inclusion criteria:

- Diagnosis of rheumatoid arthritis (RA) as defined by the 1987 revised American College of Rheumatology (ACR) criteria with disease duration of no less than 3 months and ACR class I-III
- Patients were to have been treated with, and tolerated, a minimum of 12 weeks of treatment with MTX prior to the randomization visit
- Patient with moderate to severe active disease defined as:
    - At least 8 out of 68 joints assessed as painful or tender on motion at both screening and baseline visits, and
    - At least 6 out of 66 joints assessed as swollen at both screening and baseline visits, and
    - hs C-reactive protein >6 mg/L (>0.6 mg/dL) at the screening visit

Following were some of the exclusion criteria:

- Prior therapy with a TNF antagonist or any other biologic agents within 3 months prior to randomization
- Autoimmune disease other than RA or significant systemic involvement (vasculitis, pulmonary fibrosis, Felty's syndrome)
- Current treatment with DMARDs/immunosuppressive agents other than MTX

Patients who withdrew were to be assessed using the procedure normally planned for the end of treatment (EOT) (Week 52: Visit 13) and post-treatment safety follow-up (Visit 14) visits. The Investigator was to make the best effort to contact the patients who were lost to follow-up to identify the reason why they had failed to attend the visit and determine their health status, at least their vital status.

The three key efficacy endpoints in Part B Cohort 2 were

- The event of achieving an ACR20 response at Week 24 (primary),
- Change from baseline in HAQ-DI at Week 16 (key secondary),
- Change from baseline in the modified Van der Heijde total Sharp score at Week 52 (key secondary).

Baseline was defined as the last available value prior to the first dose of the study medication.

A patient was defined as an ACR20 responder if, and only if, the following three conditions were met:

1. they had a ≥ 20% improvement in the number of tender joints (based on 68 joints)
2. they had a ≥ 20% improvement in the number of swollen joints (based on 66 joints)
3. they had a ≥ 20% improvement in three of the following five domains
    - Patient's Global Assessment (measured on a VAS scale, 0-100)
    - Physician's Global Assessment (measured on a VAS scale, 0-100)
    - Patient's assessment of pain (measured on a VAS scale, 0-100)
    - Patient's Assessment of Physical function (HAQ-DI score, 0-3)
    - Acute phase reactant measured by CRP (mg/dL, >0)

The HAQ-DI is a standardized questionnaire developed for use in RA. The HAQ-DI, with the past week as the time frame, focuses on whether the respondent "is able to…" do the activity and covers eight categories in 20 items: dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities, for which there are at least 2 questions by category. The four responses for the HAQ-DI questions are graded as follows: without any difficulty = 0; with some difficulty = 1; with much difficulty = 2; and unable to do = 3.

The degree of joint damage was assessed using the van der Heijde modified total Sharp score (mTSS). This methodology quantifies the extent of bone erosions for 44 joints and joint space narrowing for 42 joints, with higher scores representing greater damage. The van der Heijde mTSS at a time point is the sum of the scores from both the erosion score and the joint space narrowing score, for a maximum score of 448.

Another key secondary efficacy endpoint was major clinical response. Major clinical response is defined as the event of maintaining an improvement as assessed by the ACR70 for at least 24 consecutive weeks during the 52-week period.

Exploratory efficacy endpoints included ACR50 at Week 24, ACR70 at Week 24, DAS28-CRP at Week 24, radiographic progression of mTSS at Week 52, and SF-36 PCS at Week 24.

The SF-36 is a generic questionnaire measuring general health status (quality of life) in the last 4 weeks before completing the questionnaire. The SF-36 is a 36 item questionnaire that measures eight multi-item dimensions of health: physical functioning (10 items) social functioning (2 items) role limitations due to physical problems (4 items), role limitations due to emotional problems (3 items), mental health (5 items), energy/vitality (4 items), pain (2 items), and general health perception (5 items). For each dimension, item scores are coded, summed, and transformed on to a scale from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). Two standardized summary scores can also be calculated from the SF-36; the physical component summary (PCS) and the mental health component summary (MCS).

Statistical Methodologies

The primary analysis population was the ITT population for Part B Cohort 2. The binary ACR20 response at Week 24 was analyzed with a two-sided CMH test stratified by prior biologic use and region. Separate pairwise comparisons of the response rates between each dose regimen of sarilumab and placebo were derived. The MH estimate of the odds ratio and the corresponding 95% CI were derived by testing each active dose group versus placebo separately.

Patients who discontinued treatment, received rescue with open-label sarilumab, or who had missing data were considered non-responders in the primary analysis. Therefore, the binary endpoint should in fact be considered a composite response endpoint defined by: (1) not receiving rescue (generally through achieving at least 20% improvement in both the swollen and tender joint counts at Week 16); (2) remaining on treatment and in the study through the time point of interest (e.g., Week 24); and (3) achieving a response in the outcome of interest at the time point of interest (e.g., ACR20 at Week 24).

As a sensitivity analysis, the data collected after treatment discontinuation or rescue were set to missing, then a last observation carried forward (LOCF) procedure from the point of treatment discontinuation or rescue was applied to impute missing data for all seven ACR components for all visits post that point. Responder status was determined using the imputed data. The LOCF analysis was considered inappropriate since it is based on strong and unverifiable assumptions about the missing data mechanism. Furthermore, as a single-imputation approach, it does not take into account the uncertainty in the imputation process. Therefore, I conducted an intent-to-treat analysis with post-escape observed data and non-responder imputation for dropouts to evaluate an intention-to-treat or de facto estimand, i.e., the difference in outcomes in all randomized patients regardless of adherence or use of ancillary therapies. I carried out similar sensitivity analyses incorporating observed post-escape data for secondary endpoints.

In addition, after the filing meeting, based on concerns about the handling of non-responders at Week 16, we sent an information request for additional sensitivity analyses including tipping point analyses for the primary endpoint, and the applicant submitted sensitivity analyses as per the IR.

The continuous HAQ-DI change from baseline at Week 16 was analyzed with a mixed model for repeated measures (MMRM). The repeated-measures analysis was based on the restricted maximum likelihood method assuming an unstructured covariance structure to model the within-subject errors. The model, including treatment, region, prior biologic use, visit (all visits from week 2 to week 16), and treatment-by-visit interaction as fixed effects and baseline as a covariate, was used to test the difference between each active treatment group versus placebo in the change from baseline in HAQ-DI at Week 16. The data collected after treatment discontinuation or rescue were set to missing. Therefore, the MMRM analysis assumed a missing-at-random (MAR) mechanism for missing data due to dropout and post-rescue data.

As a sensitivity analysis, data collected after treatment discontinuation or rescue were set to missing, then an LOCF procedure from the point of treatment discontinuation or rescue was used to impute any missing HAQ-DI values for all visits post that point. The change from baseline in HAQ-DI at Week 16 was assessed by fitting an ANCOVA model with the baseline covariate and factors for treatment, region and prior biologic use. In addition to the above protocol-specified analysis, because I consider the LOCF analysis not useful by the same reason as above, I conducted continuous responder analysis using available post-rescue data with missing data due to dropout considered as the worst outcomes.

The van der Heijde mTSS change from baseline at Week 52 was analyzed with a two-sided rank-based analysis of covariance (rank ANCOVA) model adjusted for baseline, prior biologic use and region. The linear extrapolation method was the primary method used to impute missing or post-rescue Week 52 modified total Sharp score, erosion score, or joint space narrowing score. The data collected after treatment discontinuation + 14 days or rescue + 14 days were set to missing before linear extrapolation.

The sensitivity approaches to handle the missing or post-rescue Week 52 modified total Sharp

score are described as follows:

- Approach 1 – Mean rank imputation
- Approach 2 – LOCF with data post treatment discontinuation or rescue + 14 days set to missing and imputed by LOCF
- Approach 3 – As-observed cases (excluding post treatment discontinuation or rescue data)
- Approach 4 – Observed cases (including post treatment discontinuation or rescue data)
- Approach 5 – Linear extrapolation including both post treatment discontinuation and rescue data.

In addition to the above protocol-specified analyses, I conducted an analysis using all observed data, including available post-rescue data, to evaluate mTSS at Week 24 rather than Week 52; there was less missing data and rescue at this earlier time point, and 24 weeks still might be a sufficient treatment duration to assess radiographic progression of structural damage.

The primary and key secondary efficacy endpoints were tested for each sarilumab dose versus placebo in a testing strategy designed to protect the family-wise type 1 error rate at α=5% (two-sided). The applicant proposed a hierarchical testing procedure with a Bonferroni correction to adjust for the multiple doses and endpoints. Thus, a hierarchical testing procedure was used for the multiple endpoints at $\alpha = 0.025$ for each dose regimen separately. The hierarchy was:

1. Incidence of ACR20 response at Week 24,
2. Change from baseline in HAQ-DI at Week 16,
3. Change from baseline in the modified total Sharp score at Week 52,
4. Incidence of achieving major clinical response during the 52-week period.

Sample Size Calculation

The study power was based on the change in modified Total Sharp Score (mTSS). The estimate was calculated based on the Wilcoxon/Mann Whitney rank-sum test in the nQuery Advisor 6.01 software. Computations were based on the main scenario of 2 dose regimens selected for Part B Cohort 2. Assuming an alpha of 0.025 to address the multiplicity across the 2 active dose regimens, 90% power, a Week 52 mean change of 1.10 and 0.35 in the placebo and active groups, respectively, an associated standard deviation (SD) of 2.6 and a missing data rate of 15%, there was a requirement of 372 patients per group. The assumed mean changes and SD are based on results from the tocilizumab program (Lithe study[1]). The SD was estimated as the average SD between the active and placebo treatment groups, increased slightly to be conservative. Likewise, the mean difference was decreased slightly.

Changes in the statistical analysis plan

There were seven amendments to the original protocol (October 19, 2009): Amendment 1 (February 9, 2010), Amendment 2 (June 1, 2010), Amendment 3 (April 4, 2011), Amendment 4 (November 4, 2011), Amendment 5 (August 8, 2012), Amendment 6 (October 29, 2012), and Amendment 7 (October 8, 2013). The applicant claimed that these amendments were made prior to unblinding and analyses of the efficacy data. The changes included the following:

> • In the primary method of Part B Cohort 2 missing data handling, the data collected after treatment discontinuation or rescue will be set to missing. No imputation of missing post-baseline values will be performed. Responder status will be determined if possible. With these rules patients will automatically become non-responders for all time points beyond the time point they started rescue medication or

discontinued study treatment. (They initially proposed LOCF imputation for treatment discontinuation or rescue.)
• Change from baseline in HAQ-DI will be analyzed with a mixed model for repeated measures (MMRM). (They initially proposed an analysis of covariance (ANCOVA) model.)
• Modified missing data handling rules in the sensitivity approaches for the modified Total Sharp Score so the protocol statistical section is consistent with the current SAP.
• As a potential sensitivity analysis for missing data handling, the data collected after treatment discontinuation or rescue will be set to missing, then an LOCF procedure from the point of treatment discontinuation or rescue will be applied to impute missing data for each continuous secondary efficacy variable for all visits post that point.(They initially proposed that the average of change was to be imputed by the average of change from baseline in HAQ-DI before Week 8 for treatment discontinuation or rescue.)

Patient Disposition, Demographic and Baseline Characteristics

A total of 1197 patients were randomized to Cohort 2. Among them, 3 patients (011072-032-008-230, 011072-076-012-202, and 011072-158-001-203) were not treated. The majority (81%) of patients completed 52 weeks and 90% of patients completed 24 weeks (Table 2). The number of patients who withdrew from the study prior to Week 24 and Week 52 was comparable among treatment groups. The most common reasons for discontinuation prior to Week 52 were adverse event and lack of efficacy with comparable rates in the sarilumab treatment groups but lower rates in the placebo group.

Of the 398 (100%) patients in the placebo group, 95 (24%), 117 (29%), and 156 (39%) patients were rescued to sarilumab 200 mg by Week 16, 24, and 52, respectively. On the other hand, of the 400 (100%) patients in the sarilumab 150 mg group, 30 (8%), 45 (11%), and 55 (14%) patients were rescued to the sarilumab 200 mg by Week 16, 24, and 52, respectively. And of the 399 (100%) patients in the sarilumab 200 mg group, 28 (7%), 32 (8%), and 46 (12%) patients were rescued to the sarilumab 200 mg by Week 16, 24, and 52, respectively. At all three time points, there was considerably greater rescue on placebo than the two sarilumab arms; this imbalance is itself suggestive of efficacy. At Week 16, among those who met swollen/tender joint count rescue criteria on each arm, a greater proportion were actually rescued on the placebo group compared to the sarilumab groups. There was also a slightly greater proportion of patients rescued due to investigator judgment on the placebo arm. Acknowledging that these comparisons are based on post-randomization subgroups, I requested the applicant to clarify the reason for this imbalance. The applicant argued that investigators implemented rescue therapy based not only on changes in tender joint count and swollen joint count, but also on their clinical assessments of changes in patient-reported outcomes evident during the weeks prior to rescue. More patients on placebo showed worsening in these additional outcomes, resulting in the observed imbalance. Also the applicant explained that patients who satisfied protocol-specified rescue criteria but were not rescued reported greater improvements in self-assessments as well as reductions in active joint counts during the weeks preceding the decision to initiate rescue than patients who were rescued. I think that the applicant answered my question reasonably. See the appendix for the explanation by the applicant.

Table 2. Patients' Accountability, N (%) (All Randomized Patients)

	SAR 150 mg n (%)	SAR 200 mg n (%)	Placebo n (%)

Randomized	400 (100)	399 (100)	398 (100)
Completed Wk16	369 (92)	369 (93)	383 (96)
Met rescue criteria	51 (13)	46 (12)	127 (32)
Rescued by criteria	19 (5)	20 (5)	75 (19)
Rescued by investigator judgment	11 (3)	8 (2)	20 (5)
Rescued	30 (8)	28 (7)	95 (24)
Completed Wk24	359 (90)	353 (89)	371 (93)
Rescued	45(11)	32 (8)	117 (29)
Completed Wk52	320 (80)	316 (79)	332 (83)
Rescued	55 (14)	46 (12)	156 (39)
Discontinued Wk52	80 (20)	83 (21)	66 (17)
Adverse event	50 (13)	57 (14)	12 (3)
Lack of efficacy	5 (1)	6 (2)	4 (1)
Other	25 (6)	20 (5)	50 (13)

Note: SAR stands for sarilumab.
Source: Reviewer & Clinical Study Report CSR-EFC11072-16.2.1-EN (page 6)

The demographic and baseline disease characteristics were generally balanced and comparable between the treatment groups (Table 3). Overall, the average patient in the study was 51 years old, 74 kg in weight, and had 8 years of duration of RA. The majority of patients were Caucasian and approximately 82% of patients were female. About 70% of patients were naïve to TNF alpha inhibitors and about 60% of patients used methotrexate at baseline.

Table 3. Patients' Demographic and Baseline Characteristics by Treatment

	SAR 150 mg N=400	**SAR 200 mg N=399**	**Placebo N=398**
Age (years)			
N	400	399	398
Mean	50	51	51
SD	12	12	11
Median	51	52	52
Min-Max	18-74	19-75	19-75
Gender, n (%)			
Female	319 (80)	337 (85)	321 (81)
Male	81 (20)	62 (15)	77 (19)
Race, n (%)			
White	345 (86)	343 (86)	343 (86)
Black	10 (3)	8 (2)	10 (3)
Asian	33 (8)	33 (8)	32 (8)
Other	12 (3)	15 (4)	13 (3)
Weight (kg)			
N	398	398	398
Mean	74	75	74
SD	19	20	17
Median	71	72	72
Min-Max	32-151	37-168	42-165
BMI (kg/m2)**			
N	396	398	397
Mean	28	29	28
SD	7	7	6
Median	27	28	27
Min-Max	15-65	16-55	16-54
DAS28-CRP (>5.1)			
N	399	399	398
Mean	6.0	6.0	5.9
SD	0.9	0.9	0.9
Median	5.9	5.9	5.9
Min-Max	2.8-8.3	3.4-8.0	3.5-8.1
Duration of RA (years)			

N	400	399	398
Mean	10	9	9
SD	9	7	8
Median	7	7	7
Min-Max	1-45	1-34	1-44
Prior biologic use, n (%)			
Yes	108 (27)	110 (28)	109 (27)
No	292 (73)	289 (72)	289 (73)
Rheumatoid factor, n (%)			
Yes	345 (87)	328 (83)	336 (84)
No	51 (13)	69 (17)	62 (16)
Tender joint count (0-68)			
N	400	399	398
Mean	27	27	27
SD	14	15	14
Median	24	23	24
Min-Max	8-68	3-68	5-68
Swollen joint count (0-66)			
N	400	399	398
Mean	17	17	17
SD	9	10	9
Median	14	14	15
Min-Max	2-56	3-66	3-56
CRP (mg/L)			
N	400	399	398
Mean	23	22	20
SD	23	24	23
Median	15	16	13
Min-Max	1-209	1-203	1-203
HAQ-DI (0-3)			
N	400	399	398
Mean	1.6	1.7	1.6
SD	0.6	0.6	0.7
Median	1.8	1.8	1.7
Min-Max	0-3	0-3	0-3

Source: Excerpted from the Clinical Study Report for Study EFC11072 – 15.1 (pages 3- 32).

Results and Conclusions

Primary Efficacy Endpoint:

1. *ACR20 at Week 24*

The analysis of the primary endpoint showed statistically significantly greater ACR20 responses at Week 24 for both sarilumab dosing regimens compared to placebo (Table 4). As pre-specified in the protocol, all dropouts prior to Week 24 were treated as non-responders.

There was some dropout prior to Week 24 (11% of active and 7% of placebo) on both arms, in addition to a substantial proportion of patients being rescued by meeting escape criteria or investigator's judgment prior to Week 24 (10% of active and 29% of placebo), as dictated by the design, with disproportionately more patients being rescued (and thus being considered non-responders in the primary analysis) on placebo. As a result, the treatment effect in the primary analysis might be to a large degree driven by an effect on tender and swollen joint counts at Weeks 12, 16, and 20 rather than on ACR20 response at Week 24. Therefore, I consider the observed data sensitivity analysis that includes outcomes collected after rescue to be important, as this analysis attempts to evaluate the effect on ACR20 at Week 24 regardless of whether subjects met the escape criteria at Weeks 16 or 20, were rescued by investigator's judgment, discontinued study treatment, or dropped out of the study.

The applicant's primary and sensitivity analyses appeared to support efficacy of both sarilumab dosing regimens – statistically significant difference in ACR20 responses at Week 24 between each sarilumab dosing regimen and placebo.

Table 4. Applicant's analyses of ACR20 response at Week 24

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Primary analysis with NRI	SAR 150mg (N=400)	232/400 (58)	vs. Placebo	2.8	(2.1, 3.7)	<0.0001
	SAR 200mg (N=399)	265/399 (66)	vs. Placebo	4.0	(3.0, 5.3)	<0.0001
	Placebo (N=398)	133/398 (33)				
Sensitivity analysis with LOCF imputation	SAR 150mg (N=400)	256/400 (64)	vs. Placebo	3.2	(2.4, 4.3)	<0.0001
	SAR 200mg (N=399)	285/399 (71)	vs. Placebo	4.5	(3.3, 6.1)	<0.0001
	Placebo (N=398)	142/398 (36)				

Note: p-value based on CMH test stratified by prior biologic use and region.
Source: Excerpted from the Clinical Study Report for Study EFC11072 (page 103) & 15.2 (page 4).

I conducted an analysis based on all observed data, including outcomes post-rescue to sarilumab. My analysis considered patients who dropped out of the study to be non-responders and supported the conclusion of efficacy of the sarilumab 150 mg and 200 mg dosing regimens over placebo. The estimated effects using the observed post-escape data were smaller than in the applicant's primary and sensitivity analyses, as might be expected due to the considerable number of placebo patients who crossed over to sarilumab prior to Week 24 (Table 5, upper rows).

Since there were patients who met the pre-specified rescue criteria but were not actually rescued, and there was an imbalance in the proportions of such patients among treatment groups, in order to assess any bias in actual rescue, I also conducted an analysis treating patients who met rescue criteria as non-responders regardless of whether they were actually rescued. This analysis also considered patients who dropped out of the study to be non-responders and supported the conclusion of efficacy of the sarilumab 150 mg and 200 mg dosing regimens over placebo (Table 5, lower rows).

Table 5. Reviewer's analyses of ACR20 response at Week 24

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Sensitivity analysis with observed data for subjects who were rescued	SAR 150mg (N=400)	256/400 (64)	vs. Placebo	1.9	(1.4, 2.5)	<0.0001
	SAR 200mg (N=399)	275/399 (69)	vs. Placebo	2.4	(1.8, 3.1)	<0.0001
	Placebo (N=398)	193/398 (48)				
Sensitivity analysis treating patients who met rescue criteria as non-responders	SAR 150mg (N=400)	227/400 (57)	vs. Placebo	3.0	(2.2, 4.0)	<0.0001
	SAR 200mg (N=399)	275/399 (69)	vs. Placebo	4.1	(3.0, 5.5)	<0.0001
	Placebo (N=398)	122/398 (31)				

Note: p-value based on CMH test stratified by prior biologic use and region.
Source: Reviewer

I also conducted a Week 16 analysis prior to the start of rescue, treating patients who dropped out prior to Week 16 as non-responders. This analysis also supported the conclusion of efficacy of the sarilumab 150 mg and 200 mg dosing regimens over placebo (Table 6).

Table 6. Reviewer's analysis of ACR20 response at Week 16

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Sensitivity analysis with observed data for subjects who were rescued	SAR 150mg (N=400)	240/400 (60)	vs. Placebo	2.9	(2.2, 3.9)	<0.0001
	SAR 200mg (N=399)	266/399 (67)	vs. Placebo	3.8	(2.9, 5.2)	<0.0001
	Placebo (N=398)	137/398 (34)				

Note: p-value based on CMH test stratified by prior biologic use and region.
Source: Reviewer

In addition, in the response to FDA's IR after the filing meeting, the applicant submitted the following tipping point analysis results for the primary endpoint:

> For the signs and symptoms primary endpoint, ACR20 incidence, a tipping point did not exist in Study EFC11072; even if all patients with missing data randomized to placebo were assumed to be responders, and those randomized to sarilumab were assumed to be non-responders.

In the tipping point analysis by the applicant, observed data following rescue (originally classified as non-response) were included as observed and assumptions on missing data due to study discontinuation were varied to see if there were any tipping points. To assess the impact of missing data due to both dropout, as well as the impact of rescue, I also conducted a tipping point analysis varying assumptions about the missing data as follows:

– Distribution of ACR20 status at Week 24

Actual Data	
Sarilumab 150 mg	232 observed response 41 non-response due to dropout 45 non-response due to rescue
Sarilumab 200 mg	265 observed response 46 non-response due to dropout 32 non-response due to rescue
Placebo	133 observed response 27 non-response due to dropout 121 non-response due to rescue

– Counts in tipping point analysis

	Sarilumab 150 mg	Sarilumab 200 mg	Placebo
Response Non-response	232 + J 82 + (86 – J)	265 + J 56 + (78 – J)	133 + K 117 + (148 – K)

For a comparison of an sarilumab regimen versus placebo, the following notations are made.
J: Number of responders in uncertain cases from patients randomized to a sarilumab dose
K: Number of responders in uncertain cases from patients randomized to placebo
Between-treatment comparisons were performed using a chi-square test comparing each randomized sarilumab dose versus placebo, for each possible combinations of J and K. The table shows the counts for the comparisons, where J takes value from 0 to 86 for sarilumab 150 mg and from 0 to 78 for sarilumab 2000mg, and K from 0 to 148.

– Sarilumab 150 mg vs. Placebo

Sarilumab 150 mg	P-values less than 0.025 (*) # responders in Placebo patients with missing ACR20														
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140
0	*	*	*	*	*	*	*	*							
10	*	*	*	*	*	*	*	*	*						
20	*	*	*	*	*	*	*	*	*	*					
30	*	*	*	*	*	*	*	*	*	*	*				
40	*	*	*	*	*	*	*	*	*	*	*	*			
50	*	*	*	*	*	*	*	*	*	*	*	*	*		
60	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
70	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
80	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

– Sarilumab 200 mg vs. Placebo

Sarilumab 200 mg	P-values less than 0.025 (*) # responders in Placebo patients with missing ACR20														
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140
0	*	*	*	*	*	*	*	*	*	*	*				
10	*	*	*	*	*	*	*	*	*	*	*	*			
20	*	*	*	*	*	*	*	*	*	*	*	*	*		
30	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
40	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
50	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
60	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
70	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
80	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

My analyses indicate that results were tipped from statistically significant to not statistically significant when placebo patients who dropped out or were rescued had ACR20 response rates of at least 80%, which is likely implausible. In my opinion, the applicant's interpretation of the tipping point analysis results appears reasonable and my additional tipping point analysis resolves our concern with the handling of patients who dropped out or crossed over to the sarilumab treatment from Week 16.

2. *Components of ACR20 response at Week 24*

I was able to confirm the results of the applicant's analyses of the components of the primary endpoint, ACR20 response at Week 24. Analyses of all the components of ACR were statistically significant in favor of both sarilumab doses and there was no single component driving the efficacy in terms of ACR20 response (Table 7). A key limitation of these analyses at Week 24 is a significantly reduced subset of placebo patients remaining at Week 24 (e.g., 251 out of 398 randomized patients for tender joint count), as many of placebo patients met rescue criteria or were judged as non-responders by the investigator and crossed over to sarilumab at Week 16. The considerable rescue destroys the integrity of randomization, although it is likely that the subset of patients remaining on placebo at Week 24 represents a healthy subset of the randomized population, thus leading to conservative inference in comparisons against the

sarilumab arms. Because of the considerable proportion of patients either having crossed over from placebo to sarilumab or having missing data at Week 24, I carried out additional analyses of the ACR20 components and key continuous secondary endpoints at Week 16 (prior to rescue). Results are relatively similar to those at Week 24 and are presented in the Appendix.

Table 7. Applicant's analysis of ACR20 components at Week 24

	Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
TJC	SAR 150mg (N=400)	312	-16.9 (0.7)	vs. Placebo	-6.8	(-8.6, -4.9)	<0.0001
	SAR 200mg (N=399)	320	-17.4 (0.7)	vs. Placebo	-7.3	(-9.1, -5.5)	<0.0001
	Placebo (N=398)	251	-10.1 (0.7)				
SJC	SAR 150mg (N=400)	312	-10.6 (0.4)	vs. Placebo	-4.0	(-5.1, -2.8)	<0.0001
	SAR 200mg (N=399)	320	-11.3 (0.4)	vs. Placebo	-4.6	(-5.8, -3.5)	<0.0001
	Placebo (N=398)	251	-6.6 (0.4)				
Pain VAS	SAR 150mg (N=400)	313	-28.5 (1.3)	vs. Placebo	-13.1	(-16.8, -9.3)	<0.0001
	SAR 200mg (N=399)	321	-31.8 (1.3)	vs. Placebo	-16.4	(-20.2, -12.7)	<0.0001
	Placebo (N=398)	253	-15.4 (1.4)				
Physician global VAS	SAR 150mg (N=400)	313	-42.5 (1.1)	vs. Placebo	-13.1	(-16.2, -9.9)	<0.0001
	SAR 200mg (N=399)	321	-44.3 (1.1)	vs. Placebo	-14.9	(-18.0, -11.8)	<0.0001
	Placebo (N=398)	253	-29.4 (1.2)				
Patient global VAS	SAR 150mg (N=400)	312	-28.3 (1.3)	vs. Placebo	-12.5	(-16.1, -8.9)	<0.0001
	SAR 200mg (N=399)	319	-32.9 (1.3)	vs. Placebo	-17.1	(-20.1, -13.6)	<0.0001
	Placebo (N=398)	253	-15.7 (1.4)				
HAQ-DI	SAR 150mg (N=400)	313	-0.56 (0.03)	vs. Placebo	-0.24	(-0.33, -0.16)	<0.0001
	SAR 200mg (N=399)	316	-0.57 (0.03)	vs. Placebo	-0.25	(-0.34, -0.17)	<0.0001
	Placebo (N=398)	253	-0.32 (0.03)				
CRP (mg/L)	SAR 150mg (N=400)	311	-12.6 (1.2)	vs. Placebo	-12.5	(-15.9, -9.2)	<0.0001
	SAR 200mg (N=399)	317	-17.0 (1.2)	vs. Placebo	-16.9	(-20.2, -13.6)	<0.0001
	Placebo (N=398)	251	-0.1 (1.3)				

Note: No imputation for missing data is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, prior biological use, region, visit, and treatment-by-visit interaction.
Source: Reviewer & the Clinical Study Report for Study EFC11072-16.2.6-EN (pages 237, 252, 267, 297, 312, 8, 282).

In summary, the study showed statistically significant evidence in favor of the sarilumab 150 mg and 200 mg dosing regimens on the ACR20 response at Week 24 (primary efficacy endpoint). Several sensitivity analyses were conducted to assess the robustness of the primary analysis. The conclusions from these analyses were consistent in general.

Key Secondary Efficacy Endpoints:

I was able to confirm the results of the applicant's analyses of the key secondary and selected exploratory efficacy endpoints. I also conducted sensitivity analyses to assess the impact of early escape and missing data due to dropout from the study. All p-values for the secondary endpoints presented here are nominal.

3. *Change from baseline in HAQ-DI at Week 16*

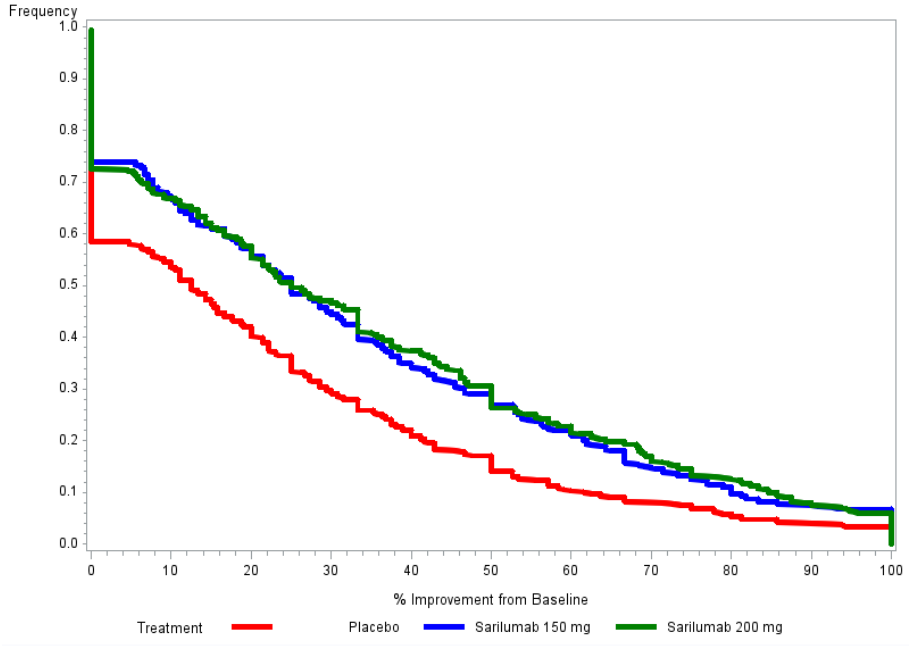
The mean change in HAQ-DI at Week 16 (just prior to escape) in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 8). Also my cumulative responder curves with worst score imputation for missing data showed separation of the curves between the sarilumab dosing regimens and placebo (Figure 3).

Table 8. Applicant's analysis of change from baseline in HAQ-DI at Week 16

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=400)	362	-0.53 (0.03)	vs. Placebo	-0.24	(-0.31, -0.16)	<0.0001
SAR 200mg (N=399)	365	-0.55 (0.03)	vs. Placebo	-0.26	(-0.34, -0.18)	<0.0001
Placebo (N=398)	378	-0.29 (0.03)				

Note: All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, prior biological use, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC11072 (page 104).

Figure 3. Cumulative responder curves based on change from baseline in HAQ-DI at Week 16



Note: Missing data were imputed with worst score.
Source: Reviewer

4. *Change from baseline in mTSS at Week 52*

The mean change in mTSS at Week 52 in patients treated with the sarilumab dosing regimens was statistically significantly less compared to patients treated with placebo (Table 9). The applicant's sensitivity analysis with the same model, using the postrescue data for patients who crossed over to sarilumab from Week 16, was consistent with results from the pre-specified analysis with linearly extrapolated (LE) data for escapers or for other missing data (Table 10). The applicant's analysis of the rates of no progression also supported the pre-specified analysis (Table 11). I also carried out an analysis of mean change in mTSS at Week 24 with the same model as the applicant's analysis, using the postrescue data up to 24 weeks for patients who crossed over to sarilumab from Week 16. There was less missing data at Week 24 than Week 52, and results also provided evidence of effects for the sarilumab arms, although effect sizes were smaller at Week 24 than Week 52 (Table 12). The cumulative distribution curves with worst score imputation for missing data showed some separation of the curves between sarilumab dosing regimens and placebo. Approximately 25 percent of patients showed no change (Figure 4). In addition, the applicant conducted a tipping point analysis which supported the efficacy on mTSS, as assumptions about the missing data under which the conclusions changed were extreme and therefore considered implausible (see Appendix for results and methods of the tipping point analyses.)

Table 9. Applicant's analysis of change from baseline in mTSS at Week 52

Treatment Group	n	Mean Change (SD)	Comparison	Mean Difference	p-value
SAR 150mg (N=400)	352	0.90 (4.66)	vs. Placebo	-1.88	<0.0001
SAR 200mg (N=399)	359	0.25 (4.61)	vs. Placebo	-2.53	<0.0001
Placebo (N=398)	352	2.78 (7.73)			

Note: Rank ANCOVA model stratified by prior biologic use and region with linear extrapolation for missing data due to dropout or escape to rescue.
Source: Excerpted from the Clinical Study Report for Study EFC11072 (page 105).

Table 10. Applicant's sensitivity analysis of change from baseline in mTSS at Week 52

Treatment Group	n	LS Mean Change (SD)	Comparison	Mean Difference	p-value
SAR 150mg (N=400)	318	0.60 (3.56)	vs. Placebo	-1.44	<0.0001
SAR 200mg (N=399)	316	0.17 (2.97)	vs. Placebo	-1.87	<0.0001
Placebo (N=398)	325	2.04 (4.52)			

Note: Rank ANCOVA model stratified by prior biologic use and region with postrescue data. The missing data due to dropout were not imputed.
Source: Excerpted from the Clinical Study Report for Study EFC11072-15.2 (page 16).

Table 11. Applicant's supportive analysis of rates of no progression from baseline to Week 52 in mTSS

Treatment Group	No progression n (%)	Comparison	Odds Ratio (95% CI)	p-value
SAR 150mg (N=400)	191 (48)	vs. Placebo	1.5 (1.1, 1.9)	0.0094
SAR 200mg (N=399)	222 (56)	vs. Placebo	2.0 (1.5, 2.7)	<0.0001
Placebo (N=398)	154 (39)			

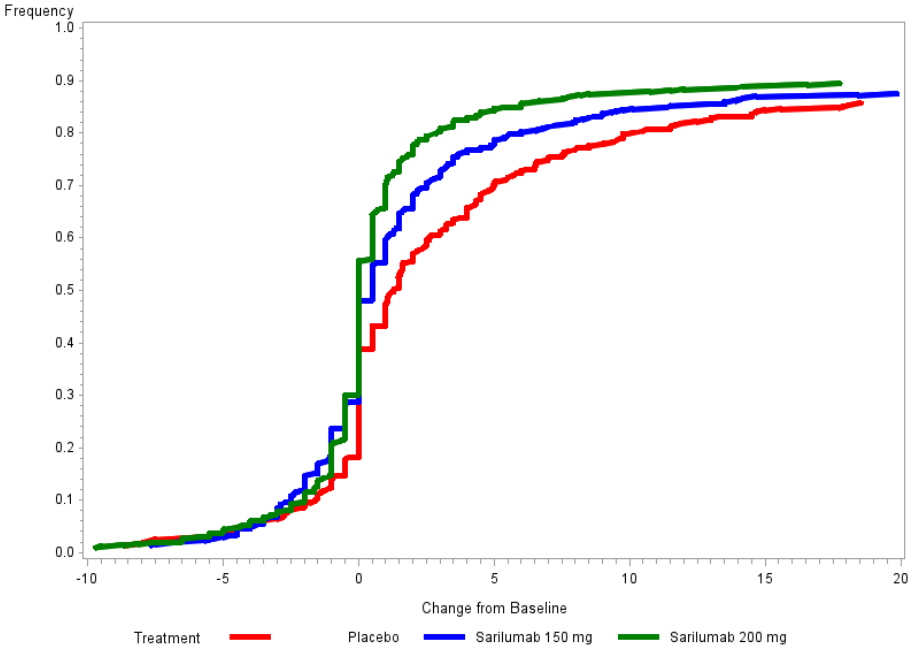
Note: Radiographic progression of the mTSS is defined as a change from baseline in the mTSS >0. The linear extrapolation method is used to impute missing or postrescue Week 52 modified total Sharp scores.
p-value based on CMH test stratified by prior biologic use and region.
Source: Excerpted from the Clinical Study Report for Study EFC11072 (page 110).

Table 12. Reviewer's supportive analysis of change from baseline in mTSS at Week 24

Treatment Group	n	LS Mean Change	Comparison	Mean Difference (SE)	p-value
SAR 150mg (N=400)	340	0.43 (2.71)	vs. Placebo	-0.72	0.0018
SAR 200mg (N=399)	343	0.14 (2.25)	vs. Placebo	-1.01	<0.0001
Placebo (N=398)	348	1.15 (3.44)			

Note: Rank ANCOVA model stratified by prior biologic use and region with postrescue data.
Source: Reviewer.

Figure 4. Cumulative distribution of change from baseline in mTSS at Week 52



Note: Missing data were imputed with the worst score.
Source: Reviewer

5. *Major clinical response at Week 52*

Major clinical response (MCR) was defined as the event of achieving and maintaining ACR70 for at least 24 consecutive weeks during the 52-week period. A statistically significantly larger proportion of patients in the sarilumab dosing groups achieved major clinical response compared to the placebo group (Table 13).

Table 13. Applicant's analysis of major clinical response at Week 52

Treatment Group	Response n (%)	Comparison	Odds Ratio (95% CI)	p-value
SAR 150mg (N=400)	51 (13)	vs. Placebo	4.7 (2.5, 8.9)	<0.0001
SAR 200mg (N=399)	59 (15)	vs. Placebo	5.6 (2.9, 10.5)	<0.0001
Placebo (N=398)	12 (3)			

Note: Major clinical response = Achieving ACR70 for at least 24 consecutive weeks during the 52-week period. Patients are considered ACR70 non-responders from the time they started rescue medication or discontinued study medication. P-value based on CMH test stratified by prior biologic use and region.
Source: Excerpted from the Clinical Study Report for Study EFC11072 (page 113).

Exploratory Endpoints:

6. *Change from baseline in DAS28-CRP at Week 24*

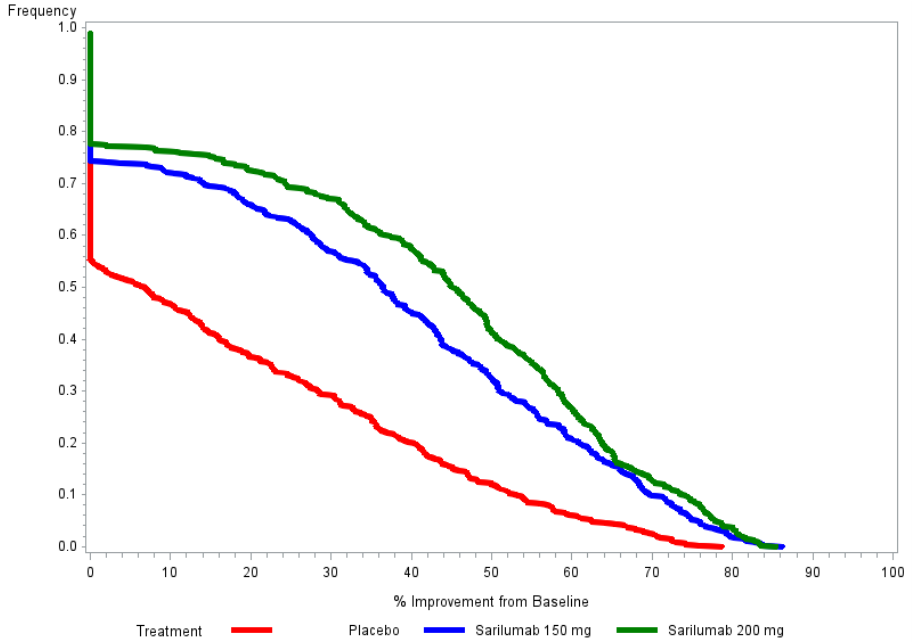
The mean reduction in DAS28-CRP at Week 24 in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 14). Also my cumulative responder curves that used worst score imputation for missing data showed separation of the curves between the sarilumab dosing regimens and placebo (Figure 5).

Table 14. Applicant's analysis of change from baseline in DAS28-CRP at Week 24

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=400)	308	-2.45 (0.08)	vs. Placebo	-1.29	(-1.50, -1.07)	<0.0001
SAR 200mg (N=399)	314	-2.82 (0.08)	vs. Placebo	-1.65	(-1.87, -1.44)	<0.0001
Placebo (N=398)	249	-1.17 (0.08)				

Note: DAS28-CRP =0.56 x sqrt(28TJC) + 0.28 x sqrt(28SJC) + 0.36 x Log(CRP+1) + 0.014 x Patient global VAS + 0.96.
All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, prior biological use, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC11072-16.2.6-EN (page 327).

Figure 5. Cumulative responder curves based on change from baseline in DAS28-CRP at Week 24



Note: Missing data due to dropout and escape to rescue were imputed with worst score.
Source: Reviewer

7. *Change from baseline in SF36-PCS at Week 24*

The mean change in SF36-PCS at Week 24 in patients treated with the sarilumab dosing

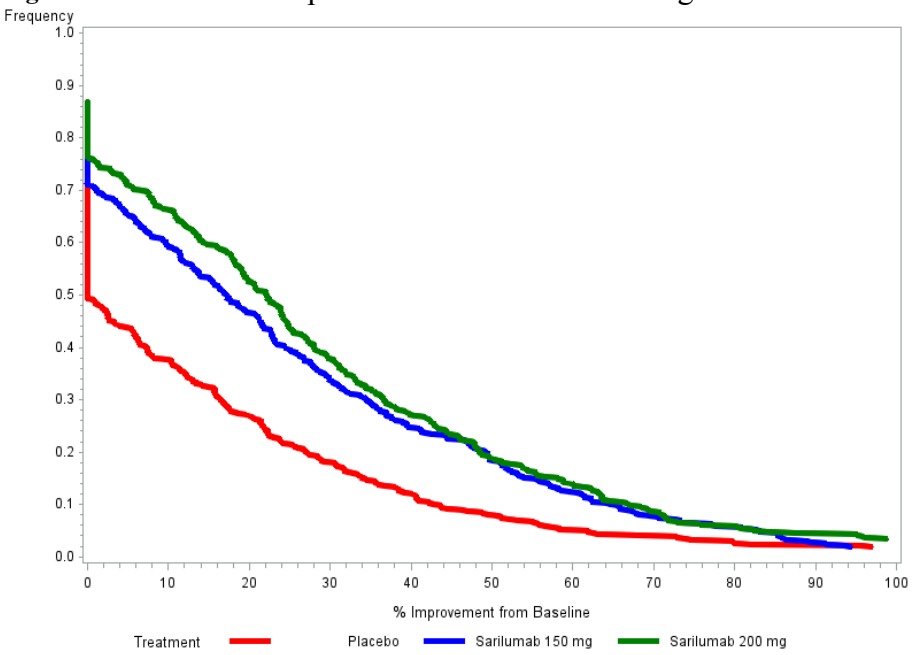
regimens was statistically significantly greater compared to patients treated with placebo (Table 15). Also my cumulative responder curves with worst score imputation for missing data showed separation of the curves between the sarilumab dosing regimens and placebo (Figure 6). Results for both the physical component and mental component summary scores and all eight domains of SF-36 are presented in the Appendix.

Table 15. Applicant's analysis of change from baseline in SF36-PCS at Week 24

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=400)	299	8.0 (0.4)	vs. Placebo	2.8	(1.6, 4.1)	<0.0001
SAR 200mg (N=399)	309	8.4 (0.4)	vs. Placebo	3.2	(2.0, 4.4)	<0.0001
Placebo (N=398)	246	5.2 (0.5)				

Note: All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, prior biological use, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC11072-16.2.6-EN (page 413).

Figure 6. Cumulative responder curves based on change from baseline in SF36-PCS at Week 24



Note: Missing data due to dropout and escape to rescue were imputed with worst score.
Source: Reviewer

Summary of Results

In summary, study data demonstrated that both doses of sarilumab, 150 mg and 200 mg, were superior compared to placebo with respect to the primary endpoint of ACR20 at Week 24, and key secondary endpoints included in the multiplicity adjustment hierarchy of HAQ-DI change from baseline at Week 16, mTSS change from baseline at Week 52 and MCR at Week 52. Analyses of all the endpoints remained statistically significant in sensitivity analyses using different approaches to handle early escape and missing data due to dropout. Additional analyses of exploratory endpoints also provided supportive evidence of efficacy.

3.2.2 Study 10832

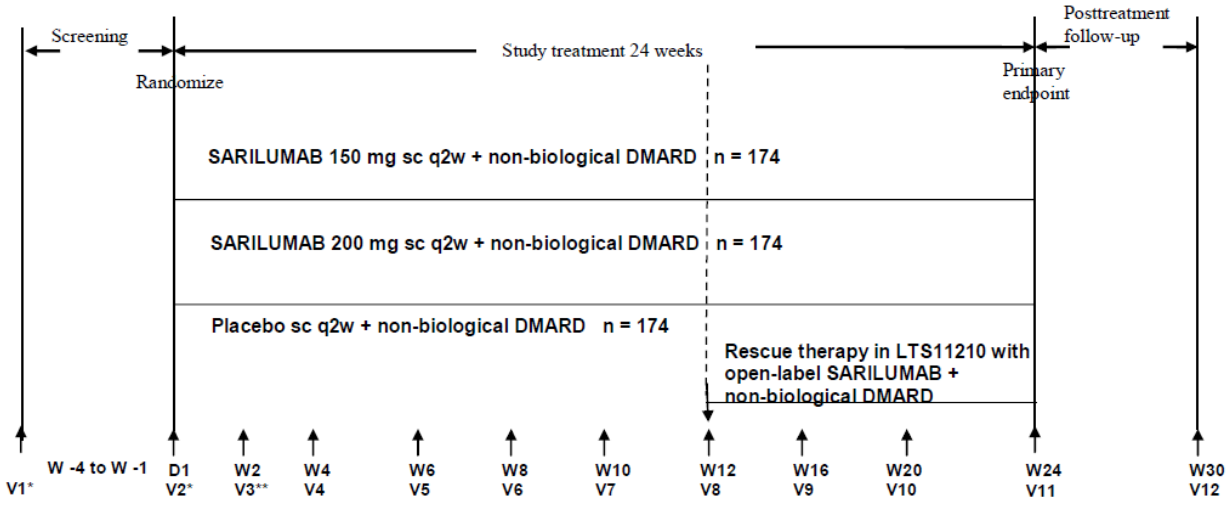
The objective of the study was to evaluate the efficacy and safety of sarilumab 150 mg every 2 weeks and 200 mg every 2 weeks compared with placebo in patients with RA who were inadequate responders to or intolerant of TNF-α antagonists. Patients were to receive randomized study treatment in a double-blind manner for 24 weeks.

Study Design and Endpoints

This was a 3-arm, multi-center, randomized, double-blind, parallel-group, placebo-controlled 24-week Phase 3 study. Patients and investigators were blinded to the allocation to active or placebo treatment. Patients were stratified by region at study entry and number of previous anti-TNFs (1 versus >1). Patients were randomized in a 1:1:1 ratio to receive SC injections of sarilumab 150 mg every 2 weeks (q2w) or sarilumab 200 mg q2w or placebo q2w.

From Week 12 onwards, patients with a lack of efficacy defined as less than 20% improvement from baseline in either swollen joint count (SJC) or tender joint count (TJC) for two joint assessments that were at least 4 weeks apart were allowed to be rescued with open-label sarilumab in the ongoing long-term safety study LTS11210.

Figure 7. Study Schema for Study 10832



Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 20).

The population enrolled in the study consisted of adults with active RA with an inadequate response or intolerance to TNF-α antagonists. Following were some of the inclusion criteria:

- Diagnosis of RA ≥6 months duration, according to the American College of Rheumatology (ACR)/ EULAR 2010 RA Classification Criteria
- Anti-TNF-α therapy failures, defined as patients with an inadequate clinical response defined by the investigator, after being treated for at least 3 consecutive months, and/or intolerance to at least 1 anti-TNF-α blocker(s), resulting in or requiring their discontinuation
    - TNF-α-blockers may include, but are not limited to: etanercept, infliximab,

adalimumab, golimumab and/or certolizumab

- Continuous treatment with 1 or a combination of non-biologic DMARDs (except for simultaneous combination use of LEF and MTX) for at least 12 consecutive weeks prior to randomization and on a stable dose(s) for at least 6 consecutive weeks prior to screening:
    - Methotrexate – 10 to 25 mg/week PO or intra muscular (or per local labeling requirements for the treatment of RA if the dose range differs)
    - Leflunomide – 10 to 20 mg PO daily
    - Sulfasalazine – 1000 to 3000 mg PO daily
    - Hydroxychloroquine – 200 to 400 mg PO daily
- Moderate-to-severely active RA, defined as:
    - at least 8 of 68 tender joints and 6 of 66 swollen joints at screening and baseline visits and
    - Hypersensitive CRP (hs-CRP) ≥8 mg/L at screening

Following were some of the exclusion criteria:

- Past history of, or current, autoimmune or inflammatory systemic or localized joint disease(s) other than RA
- Severe active systemic RA, including but not limited to vasculitis, pulmonary fibrosis, and/or Felty's syndrome
- Treatment with previous RA-directed biologic agents with other than TNF-α antagonist mechanisms:
    - Anakinra: within 28 days prior to randomization
    - Abatacept: within 42 days prior to randomization
    - Rituximab or other cell depleting agent: Within 6 months prior to randomization or until total lymphocyte count and CD-19+ lymphocyte count are normalized, whichever is longer

Patients who withdrew were to be assessed using the procedure normally planned for the end of treatment (EOT) (Week 24: Visit 11) and post-treatment safety follow-up (Visit 12) visits. The Investigator was to make the best effort to contact the patients who were lost to follow-up to identify the reason why they had failed to attend the visit and determine their health status, at least their vital status.

The two key efficacy endpoints were

- The event of achieving an ACR20 response at Week 24 (primary),
- Change from baseline in HAQ-DI at Week 12 (key secondary),

Baseline was defined as the last available value prior to the first dose of the study medication.

Other key secondary efficacy variables were DAS28-CRP, ACR50, ACR70, DAS28-CRP<2.6, Clinical disease activity index (CDAI), HAQ-DI, SF-36 PCS, SF-36 MCS, The functional assessment of chronic illness therapy fatigue scale (FACIT-Fatigue), Morning Stiffness VAS, The rheumatoid arthritis-work productivity survey (WPS-RA), Rheumatoid arthritis impact of disease (RAID), and EuroQoL (EQ-5D-3L) at Week 24.

The CDAI includes four components: SJC (28 joints), TJC (28 joints), patient's global disease activity (in cm), and physician's global assessment (in cm). The CDAI is a simple numerical summation of these 4 individual components, and ranges from 0 to 76.

The FACIT-Fatigue is a 13-item questionnaire rated 0 to 4 originally developed to measure fatigue in patients with cancer and widely used validated in RA patients. The patient was asked to answer 13 questions on a scale of 0 to 4 (0=not at all, 1=a little bit, 2=somewhat, 3=quite a bit, 4=very much). The total score ranges from 0 to 52. High scores represent more fatigue.

The WPS-RA is a questionnaire that evaluates productivity limitations within work and within home associated with RA over the previous month. The questionnaire was interviewer-administered and is based on patient self-report. It contains 9 questions addressing employment status (1 item), productivity at work (3 items), and productivity within and outside the home (5 items).

The RA impact of disease (RAID) score is a composite measure of the impact of RA on patients that takes into account 7 domains: pain, functional disability, fatigue, physical and emotional well-being, quality of sleep, and coping. The RAID is calculated based on 7 numerical rating scale (NRS) questions. Each NRS is assessed as a number between 0 and 10, which correspond to the domains mentioned above. The values for each of these domains were weighted by patient assessment of relative importance and combined in a single score.

The EQ-5D-3L comprises 5 questions on mobility, self-care, pain, usual activities, and psychological status with 3 possible answers for each item (1=no problem, 2=moderate problems, 3=severe problems) and a vertical visual analog scale that allows the patients to indicate their health state today that can range from 0 (worst imaginable) to 100 (best imaginable).

Statistical Methodologies

The statistical methods including analysis set, models, handling data for subjects who escaped to rescue from Week 12, and handling missing data due to dropout were the same as in Study 11072.

After the filing meeting, we sent an information request for additional sensitivity analyses including tipping point analyses for the primary endpoint, and the applicant submitted sensitivity analyses as per the IR.

In order to protect the family-wise type 1 error rate at α=5% (two-sided), the applicant proposed a similar hierarchical testing procedure with a Bonferroni correction to adjust for the multiple doses and endpoints as in Study 11072. A hierarchical testing procedure was used for the multiple endpoints at $\alpha = 0.025$ for each dose regimen separately. The hierarchy was:

1. Incidence of ACR20 response at Week 24,
2. Change from baseline in HAQ-DI at Week 12,
3. Change from baseline in DAS28-CRP at Week 24,
4. Incidence of ACR50 response at Week 24,

5. Incidence of ACR70 response at Week 24,
6. Incidence of DAS28-CRP < 2.6 response at Week 24,
7. Change from baseline in CDAI at Week 24,
8. Change from baseline in HAQ-DI at Week 24,
9. Change from baseline in SF36-PCS at Week 24,
10. Change from baseline in SF36-MCS at Week 24,
11. Change from baseline in FACIT-Fatigue at Week 24,
12. Change from baseline in Morning Stiffness VAS at Week 24,
13. Change from baseline in WPS-RA at Week 24,
14. Change from baseline in RAID at Week 24,
15. Change from baseline in EQ-5D-3L at Week 24

Sample Size Calculation

The study power was based on the change in HAQ-DI at Week 24. Computations were based on the data from Study 11072 Part B: SD = 0.52 and treatment difference equal to 0.2 in the low dose group and equal to 0.28 in the high dose group at Week 12. Assuming an alpha of 0.025 to address the multiplicity across the 2 active dose regimens, at least 90% power was to be achieved with 174 patients per treatment group, resulting in a total of 522 patients.

Changes in the statistical analysis plan

There were three amendments to the original protocol (June 27, 2012): Amendment 1 (September 6, 2012), Amendment 2 (April 3, 2013), and Amendment 3 (October 29, 2014). The applicant claimed that these amendments were made prior to unblinding and analyses of the efficacy data. The changes in SAP occurred at Amendment 3 only and included the following:

> • To modify the analyses for the co-primary endpoint related to change in physical function as measured by the HAQ-DI and replace the co-primary endpoint related to change in physical function as measured by the HAQ-DI from the "average of change from baseline in the HAQ-DI from week 8 to Week 24" to the co-primary endpoint of "the change from baseline in the HAQ-DI at Week 12."

Patient Disposition, Demographic and Baseline Characteristics

A total of 546 patients were randomized and treated. These patients represent the ITT/efficacy population and the safety population. Approximately two-thirds (66%) of patients completed 24 weeks and 91% of patients completed 12 weeks (Table 2). The number of patients who discontinued study treatment prior to Week 12 was comparable among the treatment groups. However, more patients in the placebo group (44%) discontinued study treatment prior to Week 24 compared to the sarilumab groups (29%), largely due to a difference in the proportions of patients who met escape criteria and were rescued with open-label sarilumab. The most common reasons for discontinuation prior to Week 24 other than escape to open label rescue in the extension study were adverse event and lack of efficacy. More patients in the sarilumab treatment groups (9.6%) discontinued treatment due to adverse events than in the placebo group (5.0%). But more patients in the placebo group (2.8%) discontinued treatment due to lack of efficacy than in the sarilumab treatment groups (1.6%).

Of the 181 (100%) patients in the placebo group, 63 (35%) patients were rescued to open-label sarilumab 200 mg by Week 24. On the other hand, of the 181 (100%) patients in the sarilumab 150 mg group, 25 (14%) patients were rescued to sarilumab 200 mg by Week 24. And of the 184 (100%) patients in the sarilumab 200 mg group, 26 (14%) patients were rescued to open-label

sarilumab 200 mg by Week 24. At Week 12, relatively more patients in the placebo group were rescued compared to the sarilumab groups among the patients who met the non-response criteria (Table 16). I requested the applicant to clarify the reason for this imbalance and the applicant answered my question reasonably. See the appendix for the explanation by the applicant.

Table 16. Patients' Accountability, N (%) (All Randomized Patients)

	SAR 150 mg n (%)	SAR 200 mg n (%)	Placebo n (%)
Randomized	181 (100)	184 (100)	181 (100)
Completed Wk12	165 (91)	168 (91)	165 (91)
Met rescue criteria	38 (21)	41 (22)	66 (36)
Rescued	20 (11)	22 (12)	45 (25)
Completed Wk24	125 (69)	133 (72)	101 (56)
Rescued	25(14)	26 (14)	63 (35)
Entered LTS	146 (81)	153 (83)	157 (87)
Discontinued Wk24 (not entering LTS)	31 (17)	25 (14)	17 (9)
Adverse event	18 (10)	17 (9)	9 (5)
Lack of efficacy	4 (2)	2 (1)	5 (3)
Other	9 (5)	6 (4)	3 (1)

Note: SAR stands for sarilumab.
Source: Reviewer & Clinical Study Report CSR-EFC10832 (page 65)

The demographic and baseline disease characteristics were generally balanced and comparable between the treatment groups (Table 17). Overall, the average patient in the study was 53 years old, 78 kg in weight, and had 12 years of duration of RA. The majority of patients were Caucasian and approximately 82% of patients were female. About 23% of patients used more than 1 anti-TNF at baseline.

Table 17. Patients' Demographic and Baseline Characteristics by Treatment

	SAR 150 mg N=181	SAR 200 mg N=184	Placebo N=181
Age (years)			
N	181	184	181
Mean	54	53	52
SD	12	13	12
Median	54	54	53
Min-Max	23-88	19-87	24-79
Gender, n (%)			
Female	142 (79)	151 (82)	154 (85)
Male	39 (21)	33 (18)	27 (15)
Race, n (%)			
White	134 (74)	130 (71)	124 (68)
Black	8 (4)	5 (3)	7 (4)
Asian	3 (2)	1 (1)	1 (1)
Other	36 (20)	48 (25)	49 (27)
Weight (kg)			
N	181	184	181
Mean	79	77	79
SD	22	21	21
Median	74	73	76
Min-Max	40-183	46-146	45-150
BMI (kg/m2)**			
N	181	184	181

Mean	29	29	30
SD	7	7	8
Median	28	28	28
Min-Max	16-63	19-52	18-58
DAS28-CRP			
N	181	184	181
Mean	6.1	6.3	6.2
SD	0.9	1.0	0.9
Median	6.1	6.3	6.1
Min-Max	3.3-8.0	3.9-8.3	4.4-8.1
Duration of RA (years)			
N	181	184	181
Mean	12	13	12
SD	9	10	10
Median	10	10	10
Min-Max	1-46	1-46	1-54
Anti CCP antibody, n (%)			
Yes	135 (27)	137 (76)	150 (83)
No	45 (25)	43 (24)	30 (17)
Rheumatoid factor, n (%)			
Yes	135 (75)	132 (73)	142 (79)
No	46 (25)	49 (27)	38 (21)
Tender joint count (0-68)			
N	181	184	181
Mean	28	30	29
SD	16	16	15
Median	24	27	29
Min-Max	5-68	4-68	8-68
Swollen joint count (0-66)			
N	181	184	181
Mean	20	20	20
SD	11	12	11
Median	16	17	17
Min-Max	6-66	3-66	6-60
CRP (mg/L)			
N	181	184	181
Mean	24	31	26
SD	23	28	25
Median	17	22	17
Min-Max	0-148	0-142	1-147
HAQ-DI (0-3)			
N	181	184	181
Mean	1.7	1.8	1.8
SD	0.6	0.6	0.6
Median	1.8	1.9	1.9
Min-Max	0-3	0-3	0-3

Source: Excerpted from the Clinical Study Report for Study EFC10832 (pages 72- 75).

Results and Conclusions

Primary Efficacy Endpoint:

1. *ACR20 at Week 24*

The analysis of the primary endpoint showed statistically significantly greater ACR20 responses at Week 24 for both sarilumab dosing regimens compared to placebo. As pre-specified in the protocol, all dropouts prior to Week 24 were treated as non-responders (Table 18).

There was a substantial number of patients who discontinued treatment or escaped to rescue prior to Week 24 (29% of active and 44% of placebo). There were disproportionately more patients

being rescued (and thus being considered non-responders in the primary analysis) on placebo (14% of active and 35% of placebo). As a result, the treatment effect in the primary analysis might be to a large degree driven by an effect on tender and swollen joint counts at Weeks 8, 12, 16, and 20 rather than on ACR20 response at Week 24. Therefore, I consider the observed data sensitivity analysis that includes outcomes collected after rescue to be important, as this analysis attempts to evaluate the effect on ACR20 at Week 24 regardless of whether subjects met the escape criteria at Weeks 16 or 20, discontinued study treatment, or dropped out of the study.

The applicant's primary and sensitivity analyses appeared to support efficacy of both sarilumab dosing regimens – there was a statistically significant difference in ACR20 responses at Week 24 between each sarilumab dosing regimen and placebo.

Table 18. Applicant's analyses of ACR20 response at Week 24

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Primary analysis with NRI	SAR 150mg (N=181)	101/181 (56)	vs. Placebo	2.7	(1.7, 4.2)	<0.0001
	SAR 200mg (N=184)	112/184 (61)	vs. Placebo	3.3	(2.1, 5.1)	<0.0001
	Placebo (N=181)	61/181 (34)				
Sensitivity analysis with LOCF imputation	SAR 150mg (N=181)	109/181 (60)	vs. Placebo	2.8	(1.8, 4.4)	<0.0001
	SAR 200mg (N=184)	123/184 (67)	vs. Placebo	3.7	(2.4, 5.8)	<0.0001
	Placebo (N=181)	67/181 (37)				

Note: CMH test stratified by number of previous anti-TNFs and region.
Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 81) & 15.2 (page 83).

I conducted an analysis based on all observed data, including outcomes post-rescue to sarilumab. My analysis considered patients who dropped out of the study to be non-responders. The estimated effects using the observed post-escape data were much smaller than in the applicant's primary and sensitivity analyses, and there was only statistical significance for the sarilumab 200 mg dosing regimen (Table 19). This happened mainly due to the fact that considerably more placebo patients (35%) were crossed over to sarilumab compared to the active groups (14%), and their signs and symptoms were improved after the cross over. This trend toward improved outcomes after cross-over and subsequently, smaller estimated effect sizes when including post cross-over data, is not inconsistent with a finding that sarilumab is efficacious.

Table 19. Reviewer's analyses of ACR20 response at Week 24

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Sensitivity analysis with observed data for subjects who were rescued	SAR 150mg (N=181)	111/181 (61)	vs. Placebo	1.4	(0.9, 2.1)	0.157
	SAR 200mg (N=184)	120/184 (65)	vs. Placebo	1.6	(1.6, 2.5)	0.029
	Placebo (N=181)	98/181 (54)				
Sensitivity analysis treating patients who met rescue criteria as non-responders	SAR 150mg (N=181)	98/181 (54)	vs. Placebo	2.7	(1.7, 4.3)	<0.0001
	SAR 200mg (N=184)	108/184 (59)	vs. Placebo	3.3	(2.1, 5.1)	<0.0001
	Placebo (N=181)	58/181 (32)				

Note: CMH test stratified by number of previous anti-TNFs and region.
Source: Reviewer

I also conducted a Week 12 analysis prior to the start of rescue, treating patients who dropped out prior to Week 12 as non-responders. This analysis also supported the conclusion of efficacy of the sarilumab 150 mg and 200 mg dosing regimens over placebo (Table 20).

Table 20. Reviewer's analysis of ACR20 response at Week 12

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
Sensitivity analysis with observed data for subjects who were rescued	SAR 150mg (N=181)	98/181 (54)	vs. Placebo	2.0	(1.3, 3.1)	0.0013
	SAR 200mg (N=184)	115/184 (63)	vs. Placebo	3.0	(1.9, 4.6)	<0.0001
	Placebo (N=181)	68/181 (38)				

Note: CMH test stratified by number of previous anti-TNFs and region.
Source: Reviewer

In addition, in the response to FDA's IR after the filing meeting, the applicant submitted the following tipping point analysis results for the primary endpoint:

> The results for Study EFC10832 lose significance after inclusion of post-rescue data at Week 24; approximately one-third of the placebo patients were receiving sarilumab at this time. Consequently the results of the tipping point analysis speak more to the efficacy of sarilumab than to the effect of missing data post-rescue.

In the tipping point analysis by the applicant, assumptions on missing data due to study discontinuation were varied to see if there were any tipping points. To assess the impact of rescue and missing data due to dropout, I conducted an additional tipping point analysis varying assumptions on the missing data as follows:

– Distribution of ACR20 status at Week 24

Actual Data	
Sarilumab 150 mg	101 observed response 31 non-response due to dropout 25 non-response due to rescue
Sarilumab 200 mg	112 observed response 25 non-response due to dropout 26 non-response due to rescue
Placebo	61 observed response 17 non-response due to dropout 63 non-response due to rescue

– Counts in tipping point analysis

	Sarilumab 150 mg	Sarilumab 200 mg	Placebo
Response	101 + J	112 + J	61 + K
Non-response	24 + (56 – J)	21 + (51 – J)	40 + (80 – K)

For a comparison of an sarilumab regimen versus placebo, the following notations are made.
J: Number of responders in uncertain cases from patients randomized to a sarilumab dose
K: Number of responders in uncertain cases from patients randomized to placebo
Between-treatment comparisons were performed using a chi-square test comparing each randomized sarilumab dose versus placebo, for each possible combinations of J and K. The table shows the counts for the comparisons, where J takes value from 0 to 56 for sarilumab 150 mg and from 0 to 51 for sarilumab 2000mg, and K from 0 to 80.

– Sarilumab 150 mg vs. Placebo

Sarilumab 150 mg	P-values less than 0.025 (*) # responders in Placebo patients with missing ACR20								
	0	10	20	30	40	50	60	70	80
0	*	*	*						
10	*	*	*	*					
20	*	*	*	*	*				
30	*	*	*	*	*	*			
40	*	*	*	*	*	*	*		
50	*	*	*	*	*	*	*	*	

– Sarilumab 200 mg vs. Placebo

Sarilumab 200 mg	P-values less than 0.025 (*) # responders in Placebo patients with missing ACR20								
	0	10	20	30	40	50	60	70	80
0	*	*	*	*					
10	*	*	*	*	*				
20	*	*	*	*	*	*			
30	*	*	*	*	*	*	*		
40	*	*	*	*	*	*	*	*	
50	*	*	*	*	*	*	*	*	*

My analyses indicates that the statistically significant results were tipped under the assumptions that placebo patients who dropped out or were rescued had ACR20 response rates of at least 30% and all such sarilumab patients were non-responders. Assuming a response rate of around 30% in the active group, 60% - 70% of placebo patients who dropped out or were rescued would have to be responders to tip the statistical significance, which is likely implausible. Therefore, in my opinion, the applicant's interpretation of the tipping point analysis results appears reasonable and my additional tipping point analysis resolves our concern with the handling of patients who dropped out or crossed over to the sarilumab treatment from Week 12.

2. *Components of ACR20 response at Week 24*

I was able to confirm the results of the applicant's analyses of the components of the primary endpoint, ACR20 response at Week 24. Analyses of all the components of ACR were statistically significant in favor of both sarilumab doses and there was no single component

driving the efficacy in terms of ACR20 response (Table 21). As in Study 11072, a key limitation of these analyses at Week 24 is a significantly reduced subset of placebo patients remaining at Week 24, as many of placebo patients met rescue criteria and crossed over to sarilumab at Week 12. The considerable rescue destroys the integrity of randomization, although it is likely that the subset of patients remaining on placebo at Week 24 represents a healthy subset of the randomized population, thus leading to conservative inference in comparisons against the sarilumab arms. Because of the considerable proportion of patient either having crossed over from placebo to sarilumab or having missing data at Week 24, I carried out additional analyses of the ACR20 components and key continuous secondary endpoints at Week 12 (prior to rescue). Results are relatively similar to those at Week 24 and are presented in the Appendix.

Table 21. Applicant's analysis of ACR20 components at Week 24

	Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
TJC	SAR 150mg (N=181)	127	-14.4 (1.0)	vs. Placebo	-3.9	(-6.7, -1.1)	0.0065
	SAR 200mg (N=184)	137	-17.0 (1.0)	vs. Placebo	-6.4	(-9.2, -3.6)	<0.0001
	Placebo (N=181)	101	-10.6 (1.1)				
SJC	SAR 150mg (N=181)	127	-11.6 (0.7)	vs. Placebo	-3.4	(-5.3, -1.5)	0.0005
	SAR 200mg (N=184)	137	-11.9 (0.7)	vs. Placebo	-3.8	(-5.6, -1.9)	<0.0001
	Placebo (N=181)	101	-8.2 (0.7)				
Pain VAS	SAR 150mg (N=181)	127	-31.9 (2.1)	vs. Placebo	-10.6	(-16.5, -4.8)	0.0004
	SAR 200mg (N=184)	135	-33.7 (2.0)	vs. Placebo	-12.4	(-18.2, -6.6)	<0.0001
	Placebo (N=181)	98	-21.3 (2.3)				
Physician global VAS	SAR 150mg (N=181)	127	-40.7 (1.7)	vs. Placebo	-12.1	(-16.8, -7.4)	<0.0001
	SAR 200mg (N=184)	137	-43.2 (1.6)	vs. Placebo	-14.7	(-19.3, -10.0)	<0.0001
	Placebo (N=181)	101	-28.6 (1.8)				
Patient global VAS	SAR 150mg (N=181)	127	-29.6 (2.0)	vs. Placebo	-9.8	(-15.5, -4.1)	0.0008
	SAR 200mg (N=184)	137	-31.3 (2.0)	vs. Placebo	-11.5	(-17.2, -5.9)	<0.0001
	Placebo (N=181)	100	-19.8 (2.2)				
HAQ-DI	SAR 150mg (N=181)	127	-0.52 (0.05)	vs. Placebo	-0.18	(-0.32, -0.05)	0.0078
	SAR 200mg (N=184)	136	-0.58 (0.05)	vs. Placebo	-0.24	(-0.38, -0.11)	0.0004
	Placebo (N=181)	101	-0.34 (0.05)				
CRP (mg/L)	SAR 150mg (N=181)	126	-15.2 (1.5)	vs. Placebo	-11.6	(-15.7, -7.6)	<0.0001
	SAR 200mg (N=184)	137	-23.3 (1.4)	vs. Placebo	-19.7	(-23.6, -15.7)	<0.0001
	Placebo (N=181)	100	-3.6 (1.6)				

Source: Reviewer & the Clinical Study Report for Study EFC10832-16.2.6-EN (pages 127, 138, 149, 171, 182, 46, 160).

In summary, the study showed statistically significant evidence in favor of the sarilumab 150 mg

and 200 mg dosing regimens on the ACR20 response at Week 24 (primary efficacy endpoint). Several sensitivity analyses were conducted to assess the robustness of the primary analysis. The conclusions from these analyses were consistent in general.

Key Secondary Efficacy Endpoints:

I was able to confirm the results of the applicant's analyses of the key secondary efficacy endpoints. I also conducted sensitivity analyses to assess the impact of missing data due to early escape and dropout from the study. All p-values for the secondary endpoints presented here are nominal.

1. *Change from baseline in HAQ-DI at Week 12*

The mean change in HAQ-DI at Week 12 (just prior to escape) in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 22). The applicant's sensitivity analysis based on multiple imputation assuming missing data were missing at random (MAR) was consistent with the primary analysis results (Table 23). However, the sensitivity analysis appeared to not address the issue in handling escape since it treated the post escape data as missing prior to doing multiple imputation. Also my cumulative responder curves with worst score imputation for missing data showed some separation of the curves between the sarilumab dosing regimens and placebo (clearer separation with sarilumab 200 mg than with sarilumab 150 mg versus placebo) (Figure 8).

Table 22. Applicant's analysis of change from baseline in HAQ-DI at Week 12

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=181)	165	-0.46 (0.04)	vs. Placebo	-0.20	(-0.32, -0.09)	0.0007
SAR 200mg (N=184)	171	-0.47 (0.04)	vs. Placebo	-0.21	(-0.33, -0.10)	0.0004
Placebo (N=181)	170	-0.26 (0.04)				

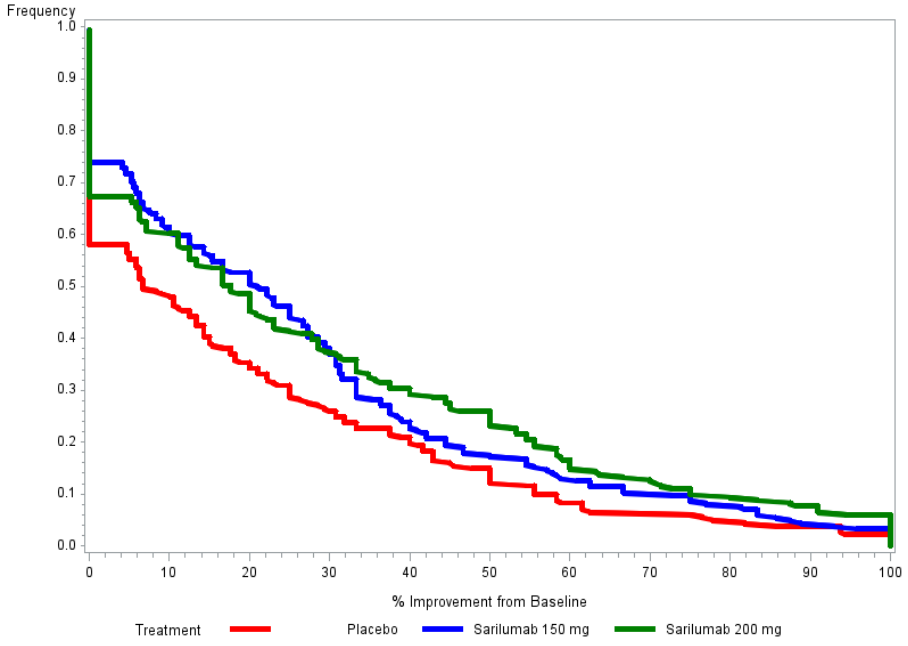
Note: All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, number of previous anti-TNFs, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 82).

Table 23. Applicant's sensitivity analysis of change from baseline in HAQ-DI at Week 12

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=181)	165	-0.46 (0.04)	vs. Placebo	-0.20	(-0.31, -0.08)	0.0009
SAR 200mg (N=184)	171	-0.47 (0.04)	vs. Placebo	-0.21	(-0.33, -0.09)	0.0004
Placebo (N=181)	170	-0.26 (0.04)				

Note: All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. Multiple imputation is used to handle the missing HAQ-DI measurements MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, number of previous anti-TNFs, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 88).

Figure 8. Cumulative responder curves based on change from baseline in HAQ-DI at Week 12



Note: Missing data were imputed with worst score.
Source: Reviewer

2. *Change from baseline in DAS28-CRP at Week 24*

The mean reduction in DAS28-CRP at Week 24 in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 24). The rates of response defined as DAS28-CRP < 2.6 were statistically significantly different between groups favoring sarilumab over placebo (Table 25). Also my cumulative responder curves that used worst score imputation for missing data showed separation of the curves between the sarilumab dosing regimens and placebo (Figure 9).

Table 24. Applicant's analysis of change from baseline in DAS28-CRP at Week 24

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=181)	126	-2.35 (0.11)	vs. Placebo	-0.97	(-1.28, -0.66)	<0.0001
SAR 200mg (N=184)	136	-2.82 (0.11)	vs. Placebo	-1.44	(-1.75, -1.13)	<0.0001
Placebo (N=181)	99	-1.38 (0.12)				

Note: DAS28-CRP =0.56 x sqrt(28TJC) + 0.28 x sqrt(28SJC) + 0.36 x Log(CRP+1) + 0.014 x Patient global VAS + 0.96.
All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. P-values were based on MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, number of previous anti-TNFs, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC10832-16.2.6-EN (page 193).

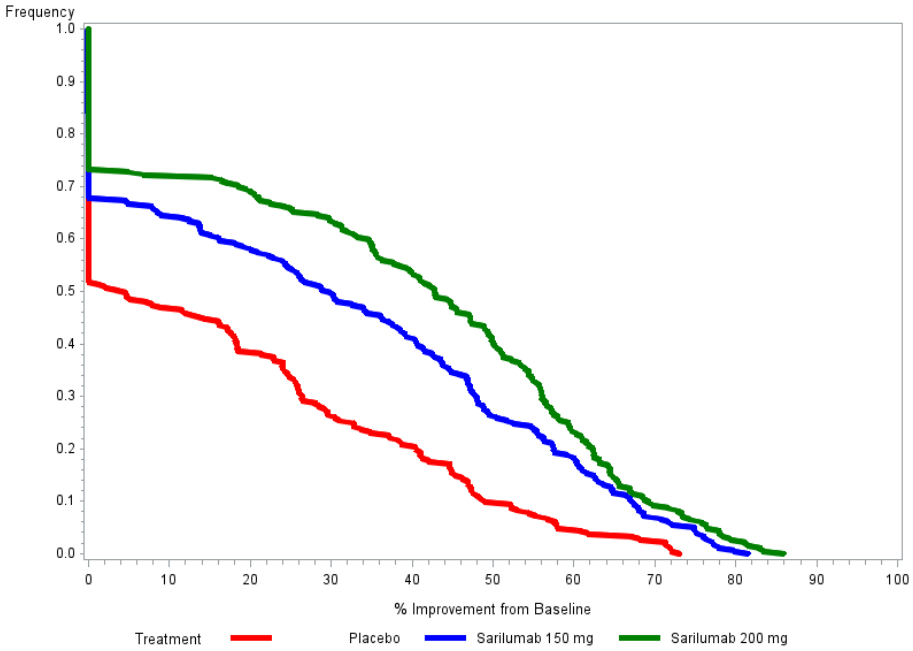
Table 25. Applicant's analyses of DAS28-CRP < 2.6 response at Week 24

Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
SAR 150mg (N=181)	45/181 (25)	vs. Placebo	4.6	(2.3, 9.1)	<0.0001
SAR 200mg (N=184)	53/184 (29)	vs. Placebo	5.8	(2.9, 11.4)	<0.0001
Placebo (N=181)	13/181 (7)				

Note: Patients are considered to be not < 2.6 from the time they started rescue medication or discontinued study medication.
CMH test stratified by number of previous anti-TNFs and region.

Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 101).

Figure 9. Cumulative responder curves based on change from baseline in DAS28-CRP at Week 24



Note: Missing data due to dropout and escape to rescue were imputed with worst score.
Source: Reviewer

3. *ACR50 and ACR70 at Week 24*

The rates of ACR50 and ACR70 response were statistically significantly different between groups favoring sarilumab over placebo (Table 26).

Table 26. Applicant's analyses of ACR50 and ACR70 response at Week 24

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence Interval	p-value
ACR50	SAR 150mg (N=181)	67/181 (37)	vs. Placebo	3.0	(1.8, 5.0)	<0.0001
	SAR 200mg (N=184)	75/184 (41)	vs. Placebo	3.4	(2.0, 5.6)	<0.0001
	Placebo (N=181)	33/181 (18)				
ACR70	SAR 150mg (N=181)	36/181 (20)	vs. Placebo	3.6	(1.8, 7.3)	0.0002
	SAR 200mg (N=184)	30/184 (16)	vs. Placebo	2.7	(1.3, 5.4)	0.0056
	Placebo (N=181)	13/181 (7)				

Note: Patients were considered non-responders from the time they started rescue medication or discontinued study medication.
P-values were based on CMH test stratified by number of previous anti-TNFs and region.
Source: Excerpted from the Clinical Study Report for Study EFC10832-16.2.6-EN (pages 102 & 111).

4. *Change from baseline in Clinical Disease Activity Index (CDAI) at Week 24*

The mean reduction in CDAI at Week 24 in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 27). Also my cumulative responder curves that used worst score imputation for missing data showed separation of the curves between the sarilumab dosing regimens and placebo

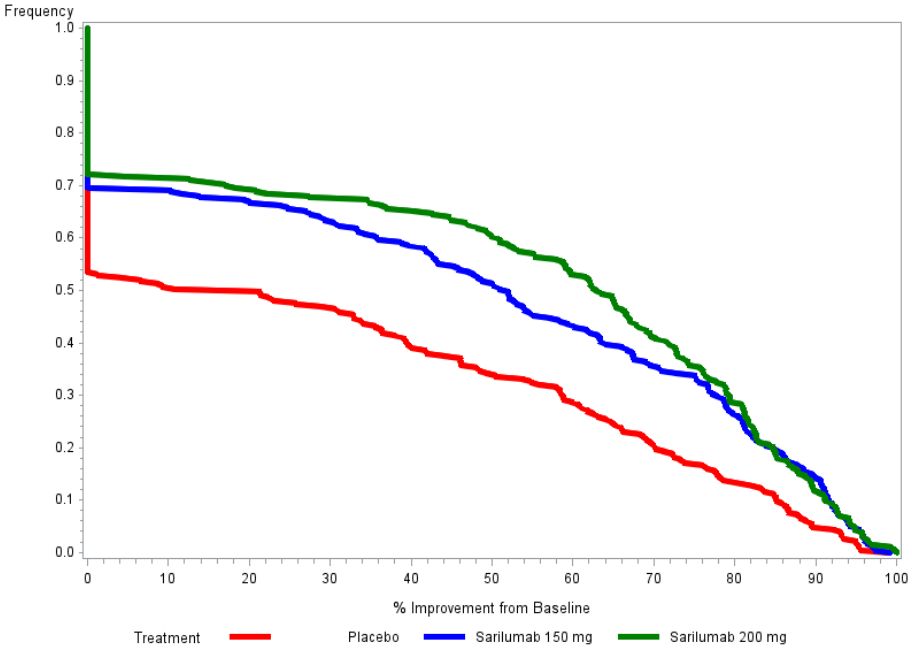
(Figure 10).

Table 27. Applicant's analysis of change from baseline in CDAI at Week 24

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=181)	127	-23.7 (1.1)	vs. Placebo	-7.3	(-10.4, -4.2)	<0.0001
SAR 200mg (N=184)	136	-26.1 (1.1)	vs. Placebo	-9.7	(-12.8, -6.6)	<0.0001
Placebo (N=181)	100	-16.4 (1.2)				

Note: CDAI = 28TJC + 28SJC + Patient global VAS + Physician global VAS.
All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. P-values were based on MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, number of previous anti-TNFs, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC10832-16.2.6-EN (page 222).

Figure 10. Cumulative responder curves based on change from baseline in CDAI at Week 24



Note: Missing data due to dropout and escape to rescue were imputed with worst score.

5. *Change from baseline in SF36-PCS at Week 24*

The mean change in SF36-PCS at Week 24 in patients treated with the sarilumab dosing regimens was statistically significantly greater compared to patients treated with placebo (Table 28). On the other hand, my cumulative responder curves with worst score imputation for missing data showed that only sarilumab 200 mg was separated from placebo over the whole range of response cut points and sarilumab 150 mg was separated from placebo below response cut points of roughly 50% improvement (Figure 11).

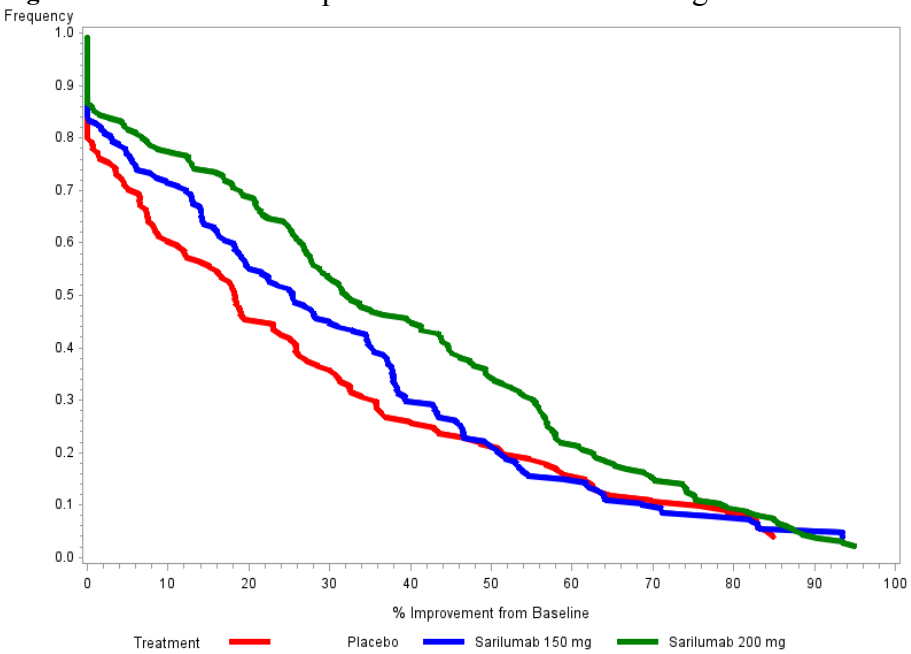
Table 28. Applicant's analysis of change from baseline in SF36-PCS at Week 24

Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
SAR 150mg (N=181)	123	7.7 (0.7)	vs. Placebo	3.3	(1.5, 5.0)	0.0004
SAR 200mg	134	8.5 (0.6)	vs. Placebo	4.1	(2.3, 5.8)	<0.0001

(N=184)		
Placebo (N=181)	99	4.4 (0.7)

Note: All assessments are set to missing from the time a patient receives rescue medication or discontinues study medication early. No imputation is performed. MMRM with PROC MIXED assuming an unstructured covariance structure: model = baseline, treatment, number of previous anti-TNFs, region, visit, and treatment-by-visit interaction.
Source: Excerpted from the Clinical Study Report for Study EFC10832-16.2.6-EN (page 253).

Figure 11. Cumulative responder curves based on change from baseline in SF36-PCS at Week 24



Note: Missing data due to dropout and escape to rescue were imputed with worst score.
Source: Reviewer

However, SF-36 MCS at Week 24 did not show statistically significant differences between both sarilumab doses and placebo. Results for the two summary components and all eight domains of SF-36 are presented in the Appendix. Therefore, the sequential testing procedure stopped for the subsequent secondary efficacy endpoints of FACIT-Fatigue, Morning Stiffness VAS, WPS-RA, RAID, and EQ-5D-3L at Week 24 declaring no statistical significance although nominal p-values for those endpoints were smaller than 0.025 (see Appendix for the results of these endpoints).

Summary of Results

In summary, study data demonstrated that both doses of sarilumab, 150 mg and 200 mg, were superior compared to placebo with respect to the primary endpoint of ACR20 at Week 24, and key secondary endpoint of HAQ-DI change from baseline at Week 12. The other secondary efficacy variables in the multiplicity hierarchy with statistically significant differences between sarilumab and placebo were DAS28-CRP, ACR50, ACR70, DAS28-CRP<2.6, CDAI, HAQ-DI, SF-36 PCS at Week 24. Analyses of all the endpoints remained statistically significant in sensitivity analyses using different approaches to handle missing data due to early escape or dropout.

3.3 Evaluation of Safety

The assessment of the safety of the study drug was mainly conducted by the reviewing medical team. The reader is referred to Dr. Suzette Peng's review for information regarding the safety profile of the drug.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The following are tabular and graphical presentations of results from subgroup analyses by demographics, region, and baseline disease characteristics in terms of ACR20 response at Week 24, based on integrated data from Studies 11072 and 10832, for the sarilumab 200 mg dose. The subgroup analyses were largely consistent with the results from the overall population in terms of ACR20 response, although there was some evidence of quantitative interactions between treatment and subgroups such as rheumatoid factor and anti CCP status – the rheumatoid factor negative and anti CCP negative subgroups showed smaller, though still statistically significant, effects of sarilumab treatment relative to other complementary subgroups (Table 29 & Figure 12). The subgroup analysis results for the sarilumab 150 mg dose were largely similar to those for the 200 mg dose (Table 34 & Figure 13 in Appendix).

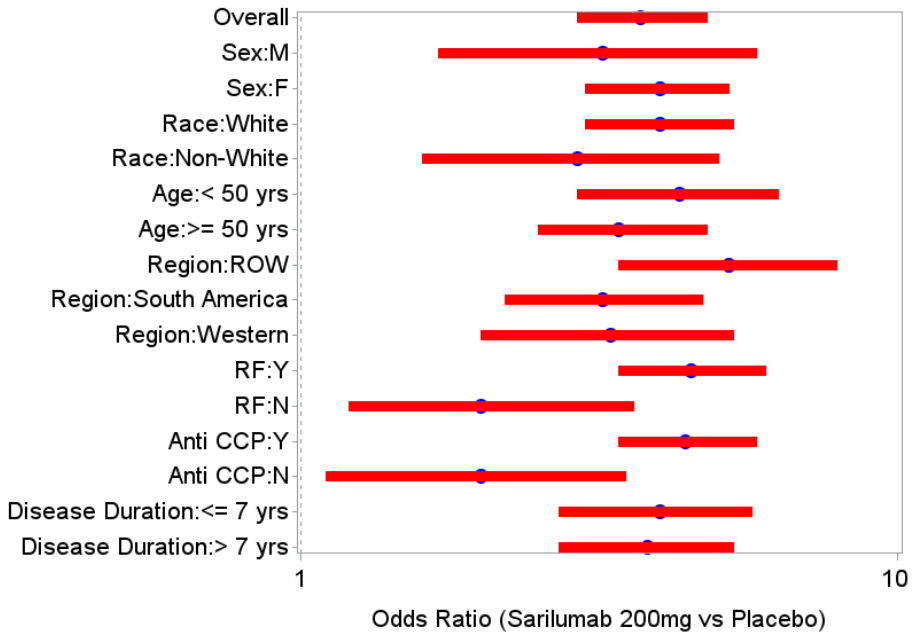
Table 29. Reviewer's Subgroup Analyses on ACR20 response at Week 24 – Studies 11072 part B cohort 2 and 10832 pooled

	SAR 200mg		Placebo		
	N	n(%)	N	n(%)	Odds ratio (95% CI)
Overall (p<0.001) [a]					
	583	377(65)	579	194(34)	3.7 (2.9, 4.8)
Sex (p=0.43) [b]					
Male	95	57(60)	104	36(35)	3.2 (1.7, 5.8)
Female	488	320(66)	475	158(33)	4.0 (3.0, 5.2)
Age (p=0.43) [b]					
≤50 yrs	256	178(70)	260	96(37)	4.3 (2.9, 6.3)
>50 yrs	327	199(61)	319	98(31)	3.4 (2.5, 4.8)
Region (p=0.21) [b]					
ROW [c]	200	131(66)	199	54(27)	5.2 (3.4, 7.9)
SM [c]	229	162(71)	229	99(43)	3.2 (2.2, 4.7)
WEST [c]	154	84(55)	151	41(27)	3.3 (2.0, 5.3)
Race (p=0.30) [b]					
White	473	309(65)	467	153(33)	4.0 (3.0, 5.3)
N-White	110	68(62)	112	41(37)	2.9 (1.6, 5.0)
RF (P=0.01) [b]					
Y	460	310(67)	478	153(32)	4.5 (3.4, 6.0)
N	118	66(56)	100	41(41)	2.0 (1.2, 3.6)

Anti CCP (P=0.02) [b]					
Y	474	312(66)	490	153(31)	4.4 (3.4, 5.8)
N	103	63(61)	88	40(45)	2.0 (1.1, 3.5)
Time Since RA Diagnosis (P=0.97) [b]					
≤7 yrs	251	171(68)	271	101(37)	4.0 (2.7, 5.7)
>7 yrs	332	206(62)	308	93(30)	3.8 (2.7, 5.3)

Source: Reviewer
[a] Logistic regression model with same covariates as primary analysis, also adjusting for study, comparing sarilumab 200 mg to placebo.
[b] Logistic regression model with same terms in [a] and with interaction between treatment arm and subgroup. P-value is for the interaction.
[c] ROW includes South Korea, Lithuania, Poland, Russia, Taiwan, Turkey, Ukraine. SM includes Argentina, Brazil, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru. WEST includes Australia, Canada, Czech Republic, Germany, Greece, Hungary, Israel, Italy, New Zealand, Portugal, Spain, USA.

Figure 12. Reviewer's Subgroup Analyses on ACR20 response at Week 24 – Studies 11072 part B cohort 2 and 10832 pooled



Note. X-axis is on logarithmic scale with base 10.

4.2 Dose Comparison

To help evaluate and compare the benefit-risk profile of the two sarilumab doses, I compared the two doses based on key efficacy endpoint data from the placebo-controlled pre-rescue period from the two studies (16 weeks for study 11072 and 12 weeks for study 10832). Dose comparison regarding the radiographic structural damage (mTSS at Week 24) was based on a single study 11072. Odd ratios or mean differences between arms with 95% confidence intervals are presented. There were some trends of a dose response for ACR responses, mTSS, and DAS28-CRP, with trends toward slightly greater responses on sarilumab 200 mg than 150 mg (Table 30).

Table 30. Comparative efficacy analyses on two sarilumab doses based on integrated data from studies 11072 and 10832

	Treatment Group	n/N (%)	Comparison	Odds Ratio	95% Confidence

		/LSMean (SE)		/LSMean Difference	Interval
ACR20	SAR 150mg (N=581)	338/581 (58)			
	SAR 200mg (N=583)	381/583 (65)	vs. SAR 150mg	1.4	(1.1, 1.7)
ACR50	SAR 150mg (N=581)	175/581 (30)			
	SAR 200mg (N=583)	225/583 (39)	vs. SAR 150mg	1.5	(1.2, 1.9)
ACR70	SAR 150mg (N=581)	89/581 (15)			
	SAR 200mg (N=583)	106/583 (18)	vs. SAR 150mg	1.2	(0.9, 1.7)
HAQ-DI	SAR 150mg (N=581)	-0.52 (0.03)			
	SAR 200mg (N=583)	-0.54 (0.03)	vs. SAR 150mg	-0.02	(-0.09, 0.06)
DAS28-CRP	SAR 150mg (N=581)	-2.17 (0.06)			
	SAR 200mg (N=583)	-2.55 (0.06)	vs. SAR 150mg	-0.38	(- 0.54,-0.23)
mTSS	SAR 150mg (N=400)	0.43 (2.71)			
	SAR 200mg (N=399)	0.14 (2.25)	vs. SAR 150mg	-0.28	(-0.70, 0.15)

Note: 95% confidence intervals based on logistic regression for binary endpoints and ANVOVA for continuous endpoints, adjusted for prior MTX/biologic use, regionand study. mTSS results come from only study 11072
Source: Reviewer

4.3 Comparison with tocilizumab

Again to help evaluate the benefit-risk profile, I compared the two sarilumab doses with the approved active control tocilizumab based on exploratory efficacy endpoints from the safety study 11370. The study was a randomized, double-blind, double-dummy, parallel group, 3-arm, 24-week, active comparator controlled study. Approximately 200 patients with active, moderate to severe RA for ≥3 months, treated with non-biologic DMARDs, were randomized in a 2:1:1 ratio to tocilizumab 20 mg/mL (a 60 minutes single IV infusion) q4w, sarilumab 150 mg q2w, or sarilumab 200 mg q2w. Odd ratios or mean differences with 95% confidence interval are presented. There was a slight numerical favorable trend for tocilizumab compared to both sarilumab doses with regard to ACR responses and HAQ-DI, but not for DAS28-CRP (Table 31).

Table 31. Comparative efficacy analyses on two sarilumab doses versus tocilizumab

	Treatment Group	n/N (%) /LSMean (SE)	Comparison	Odds Ratio /LSMean Difference	95% Confidence Interval
ACR20	SAR 150mg (N=49)	31/49 (63)	vs. TCZ	0.6	(0.3, 1.3)
	SAR 200mg (N=51)	35/51 (69)	vs. TCZ	0.7	(0.3, 1.4)
	TCZ (N=102)	77/102 (75)			
ACR50	SAR 150mg	18/49 (37)	vs. TCZ	0.8	(0.4, 1.7)

	(N=49)				
	SAR 200mg (N=51)	21/51 (41)	vs. TCZ	1.0	(0.5, 2.0)
	TCZ (N=102)	42/102 (41)			
ACR70	SAR 150mg (N=49)	9/49 (18)	vs. TCZ	0.8	(0.3, 2.0)
	SAR 200mg (N=51)	7/51 (14)	vs. TCZ	0.5	(0.2, 1.3)
	TCZ (N=102)	23/102 (23)			
HAQ-DI	SAR 150mg (N=49)	-0.56 (0.09)	vs. TCZ	0.07	(-0.14, 0.28)
	SAR 200mg (N=51)	-0.58 (0.09)	vs. TCZ	0.05	(-0.16, 0.26)
	TCZ (N=102)	-0.63 (0.06)			
DAS28-CRP	SAR 150mg (N=49)	-2.86 (0.18)	vs. TCZ	-0.06	(-0.47, 0.36)
	SAR 200mg (N=51)	-2.98 (0.18)	vs. TCZ	-0.18	(-0.59, 0.24)
	TCZ (N=102)	-2.80 (0.12)			

Note: 95% confidence intervals based on logistic regression for binary endpoints and ANVOVA for continuous endpoints, adjusted for region and baseline ANC level. TCZ stands for tocilizumab.
Source: Reviewer

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

During my review of the application, several potential statistical issues were identified:

- Potential Effect of Escape and Missing Data on Reliability of Efficacy Results
- Adequacy of Radiographic Data to Support a Claim
- Evidence to Support Proposed Dosing Recommendations

The first issue was the potential effect of escape and missing data. For the analysis of the primary endpoint, the applicant pre-specified an approach that considered patients who early escaped to sarilumab or who dropped out to be non-responders. The placebo group had disproportionately more non-responders at Week 16 in Study 11072 (at Week 12 in Study 10832) compared to the active group, prior to the primary analysis time point of Week 24. Therefore, the treatment effect in the primary analysis might be driven entirely by an effect on early escape at Week 16 in Study 11072 (at Week 12 in Study 10832) rather than an effect on ACR20 at Week 24. The applicant conducted a sensitivity analysis with last observation carried forward imputation and a tipping point analyses. I also conducted an additional sensitivity analysis with observed post-escape data that considered dropouts to be non-responders. Results from the sensitivity analyses suggested that sarilumab is efficacious notwithstanding the early escape and missing data.

The second issue was the adequacy of the radiographic data to support an efficacy claim in the Clinical Studies section of labeling. The analysis of radiographic data from study 11072 for inhibition of structural damage showed statistical significance in favor of sarilumab. The mean changes from baseline at Week 52 in mTSS were 2.78, 0.90, and 0.25 for the placebo, sarilumab 150 mg dose, and sarilumab 200 mg dose arms, respectively. The estimated mean differences in

mTSS (95% confidence interval; p-value) were -1.88 (-2.74, -1.01; p<0.001) between sarilumab 150 mg dose and placebo and -2.52 (-3.38, -1.66; p<0.001) between sarilumab 200 mg dose and placebo, respectively. The main analysis used linear extrapolation for early escape and missing data due to dropout. However, in order to provide a reliable assessment of the intention-to-treat estimand (the difference in progression at Week 52 in all randomized patients regardless of adherence), the linear extrapolation approach requires the strong and unverifiable assumption that patients' scores would have continued to change at exactly the same linear rate that was observed through the time of withdrawal. In addition, linear extrapolation is a single-imputation approach that does not appropriately take into account the uncertainty in the imputation process. Therefore, I conducted several sensitivity analyses without the linear extrapolation. Based on the applicant's and my various sensitivity analyses including analysis with post rescue data, there was statistical evidence showing inhibition of structural damage progression with both doses of sarilumab. Although there was only a single study assessing radiographic structural damage progression, I conclude that the evidence is sufficient due to the highly significant p-values, the consistency of results across the two doses and when including post-rescue data, and the similar findings for another drug in the same class (tocilizumab).

Finally, I believe that there is sufficient evidence of efficacy to support both the 150 mg and 200 mg doses of sarilumab. The applicant's proposed labeling recommends that patients start treatment with the 200 mg dose, with potential reduction of the dose to 150 mg for management of neutropenia, thrombocytopenia and elevated liver enzymes. Both doses of sarilumab 200 mg and 150 mg demonstrated benefit with respect to the primary endpoint ACR20 and several important secondary endpoints including HAQ-DI in two independent placebo-controlled clinical trials. Furthermore, there were trends toward slightly greater efficacy on sarilumab 200 mg than 150 mg based on analyses of integrated data from the two phase 3 studies. A determination of the adequacy of the applicant's proposed dosing recommendation will be made in collaboration with the clinical team based on a benefit-risk comparison of the two doses.

5.2 Collective Evidence

In the two phase 3 studies 11072 and 10832 that I reviewed, the analysis of the predefined primary efficacy endpoint, ACR20 at Week 24, was statistically significant. In Study 11072, the ACR20 response rates were 58%, 66%, and 33% for the sarilumab 150 mg, 200 mg, and placebo arms, respectively. In Study 10832, the response rates were 56%, 61%, and 34% for the sarilumab 150 mg, 200 mg, and placebo arms, respectively.

More specifically, the efficacy data from Study 11072 provided statistical evidence of efficacy for the sarilumab 150 mg and 200 mg doses for treatment of RA based on ACR20 at Week 24, the primary endpoint, and all the key secondary endpoints, including HAQ-DI at Week 16, mTSS at Week 52, and Major Clinical Response at Week 52. The efficacy data from Study 10832 provided statistical evidence of efficacy for the sarilumab 150 mg and 200 mg doses based on the primary endpoint of ACR20 response at Week 24 and most of the key secondary endpoints – HAQ-DI at Week 12, DAS28-CRP, ACR50, ACR70, DAS28-CRP < 2.6 response, CDAI, HAQ-DI, and SF36-PCS at Week 24. Some of the secondary endpoints in the multiplicity hierarchy – SF36-MCS, FACIT-Fatigue, Morning Stiffness VAS, WPS-RA, RAID, and EQ-5D-

3L at Week 24 - were not shown to be statistically significant, although trends were in favor of the sarilumab arms.

In summary:

- There was evidence of efficacy for the primary and most secondary endpoints for the 150 mg and 200 mg doses in both studies.
- There was evidence of efficacy for the radiographic endpoint for the 150 mg and 200 mg dose in one study.

Therefore, the overall package provides substantial evidence of efficacy for the proposed 150 and 200 mg doses for treating RA.

5.3 Labeling Recommendations

The following is an excerpt from the relevant clinical studies section in the proposed label. I generally agree with the study description and primary analysis results and their interpretation. However, I recommend that Week 16 (Week 12) (b) (4) results are included for continuous endpoints such as the ACR components and HAQ-DI for Study 11072 (Study 10832). I would also recommend describing the escape rules in the design description and how these patients were treated in the primary analysis (i.e., as non-responders). Estimated effects at Week 16/12 are more reliable because of considerably less escape and missing data at Week 16/12 (b) (4) In addition, I recommend clarification on the statements (b) (4) . The best approach to reliably present radiographic results, given the amount of escape and missing data prior to Weeks 24/52, and the use of linear extrapolation, will also be discussed further by the review team. I also recommend the removal of results highlighted in yellow (b) (4) except for endpoints agreed as clinically important, such as claims of (b) (4) achievement of a clinically relevant improvement in HAQ-DI (change from baseline of ≥ 0.3 units).

(b) (4)

9 Pages of Draft Labeling have been Withheld in Full as B4 (CCI/TS) immediately following this page

APPENDIX

Table 32. Reviewer's analysis of ACR20 components at Week 16 (Study 11072)

	Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
TJC	SAR 150mg (N=400)	365	-15.2 (0.6)	vs. Placebo	-5.1	(-6.8, -3.4)	<0.0001
	SAR 200mg (N=399)	367	-16.1 (0.6)	vs. Placebo	-6.0	(-7.7, -4.3)	<0.0001
	Placebo (N=398)	382	-10.1 (0.6)				
SJC	SAR 150mg (N=400)	365	-9.7 (0.4)	vs. Placebo	-3.3	(-4.4, -2.2)	<0.0001
	SAR 200mg (N=399)	367	-10.4 (0.4)	vs. Placebo	-4.1	(-5.2, -3.0)	<0.0001
	Placebo (N=398)	382	-6.4 (0.4)				
Pain VAS	SAR 150mg (N=400)	364	-26.6 (1.2)	vs. Placebo	-11.5	(-14.9, -8.2)	<0.0001
	SAR 200mg (N=399)	365	-30.5 (1.2)	vs. Placebo	-16.4	(-18.7, -12.1)	<0.0001
	Placebo (N=398)	382	-15.1 (1.2)				
Physician global VAS	SAR 150mg (N=400)	363	-35.3 (1.1)	vs. Placebo	-11.9	(-15.0, -8.8)	<0.0001
	SAR 200mg (N=399)	367	-38.1 (1.1)	vs. Placebo	-14.9	(-17.7, -11.6)	<0.0001
	Placebo (N=398)	379	-23.4 (1.2)				
Patient global VAS	SAR 150mg (N=400)	364	-26.1 (1.2)	vs. Placebo	-10.4	(-13.6, -7.1)	<0.0001
	SAR 200mg (N=399)	364	-29.9 (1.2)	vs. Placebo	-14.2	(-17.5, -11.0)	<0.0001
	Placebo (N=398)	381	-15.7 (1.2)				
HAQ-DI	SAR 150mg (N=400)	363	-0.53 (0.03)	vs. Placebo	-0.24	(-0.31, -0.16)	<0.0001
	SAR 200mg (N=399)	365	-0.55 (0.03)	vs. Placebo	-0.26	(-0.33, -0.18)	<0.0001
	Placebo (N=398)	380	-0.29 (0.03)				
CRP (mg/L)	SAR 150mg (N=400)	361	-14.7 (0.9)	vs. Placebo	-13.7	(-16.0, -11.4)	<0.0001
	SAR 200mg (N=399)	357	-18.6 (0.9)	vs. Placebo	-17.6	(-19.9, -15.3)	<0.0001
	Placebo (N=398)	378	-1.0 (0.8)				

Source: Reviewer

Table 33. Reviewer's analysis of ACR20 components at Week 12 (Study 10832)

	Treatment Group	n	LS Mean Change (SE)	Comparison	Mean Difference	95% Confidence Interval	p-value
TJC	SAR 150mg (N=181)	165	-13.7 (1.0)	vs. Placebo	-5.2	(-7.8, -2.6)	<0.0001
	SAR 200mg (N=184)	172	-14.9 (1.0)	vs. Placebo	-6.3	(-8.9, -3.8)	<0.0001
	Placebo (N=181)	172	-8.6 (1.0)				
SJC	SAR 150mg (N=181)	165	-10.5 (0.7)	vs. Placebo	-3.8	(-5.6, -1.9)	<0.0001

	SAR 200mg (N=184)	172	-10.6 (0.7)	vs. Placebo	-3.8	(-5.7, -2.0)	<0.0001
	Placebo (N=181)	172	-6.8 (0.7)				
Pain VAS	SAR 150mg (N=181)	166	-26.9 (1.9)	vs. Placebo	-11.8	(-16.9, -6.7)	<0.0001
	SAR 200mg (N=184)	171	-30.6 (1.9)	vs. Placebo	-15.4	(-20.5, -10.3)	<0.0001
	Placebo (N=181)	171	-15.1 (1.9)				
Physician global VAS	SAR 150mg (N=181)	165	-33.6 (1.8)	vs. Placebo	-10.9	(-15.6, -6.2)	<0.0001
	SAR 200mg (N=184)	171	-35.4 (1.7)	vs. Placebo	-12.7	(-17.4, -8.0)	<0.0001
	Placebo (N=181)	172	-22.7 (1.7)				
Patient global VAS	SAR 150mg (N=181)	165	-25.3 (1.8)	vs. Placebo	-11.5	(-16.4, -6.7)	<0.0001
	SAR 200mg (N=184)	171	-27.4 (1.8)	vs. Placebo	-13.6	(-18.5, -8.8)	<0.0001
	Placebo (N=181)	172	-13.8 (1.8)				
HAQ-DI	SAR 150mg (N=181)	165	-0.46 (0.04)	vs. Placebo	-0.20	(-0.32, -0.09)	0.0007
	SAR 200mg (N=184)	171	-0.47 (0.04)	vs. Placebo	-0.21	(-0.33, -0.09)	0.0004
	Placebo (N=181)	170	-0.26 (0.04)				
CRP (mg/L)	SAR 150mg (N=181)	165	-16.1 (1.5)	vs. Placebo	-11.4	(-15.2, -7.7)	<0.0001
	SAR 200mg (N=184)	170	-23.0 (1.4)	vs. Placebo	-19.3	(-23.1, -15.6)	<0.0001
	Placebo (N=181)	168	-3.6 (1.4)				

Source: Reviewer

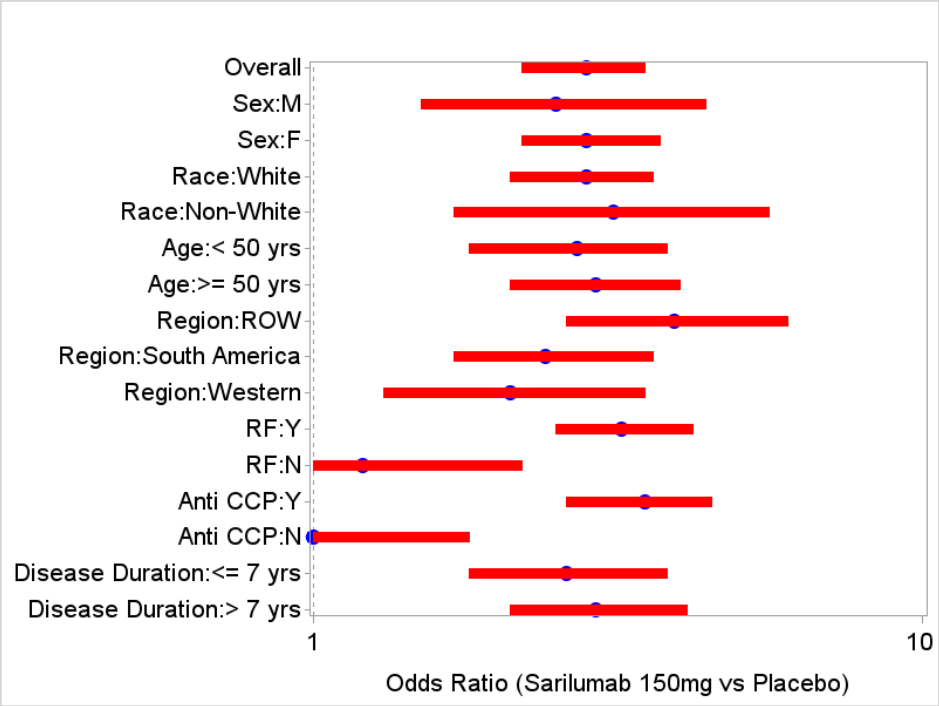
Table 34. Reviewer's Subgroup Analyses on ACR20 response at Week 24 – Studies 11072 part B cohort 2 and 10832 pooled

	SAR 150mg		**Placebo**		
	N	n(%)	N	n(%)	Odds ratio (95% CI)
Overall (p<0.001) [a]					
	581	333(57)	579	194(34)	2.8 (2.2, 3.5)
Sex (p=0.72) [b]					
Male	120	67(56)	104	36(35)	2.5 (1.5, 4.4)
Female	461	266(58)	475	158(33)	2.8 (2.2, 3.7)
Age (p=0.68) [b]					
≤50 yrs	254	149(59)	260	96(37)	2.7 (1.8, 3.8)
>50 yrs	327	184(56)	319	98(31)	2.9 (2.1, 4.0)
Region (p=0.13) [b]					
ROW [c]	200	118(59)	199	54(27)	3.9 (2.6, 6.0)
SM [c]	229	148(65)	229	99(43)	2.4 (1.7, 3.6)
WEST [c]	152	67(44)	151	41(27)	2.1 (1.3, 3.5)
Race (p=0.92) [b]					
White	479	274(57)	467	153(33)	2.8 (2.1, 3.6)

N-White	102	59(58)	112	41(37)	3.1 (1.7, 5.6)
RF (P=0.01) [b]					
Y	480	289(60)	478	153(32)	3.2 (2.5, 4.2)
N	97	42(43)	100	41(41)	1.2 (0.7, 2.2)
Anti CCP (P=0.01) [b]					
Y	494	297(60)	490	153(31)	3.5 (2.6, 4.5)
N	84	36(43)	88	40(45)	1.0 (0.5, 1.8)
Time Since RA Diagnosis (P=0.55) [b]					
≤7 yrs	264	156(59)	271	101(37)	2.6 (1.8, 3.8)
>7 yrs	317	177(56)	308	93(30)	2.9 (2.1, 4.1)

Source: Reviewer
[a] Logistic regression model with same covariates as primary analysis, also adjusting for study, comparing sarilumab 200 mg to placebo.
[b] Logistic regression model with same terms in [a] and with interaction between treatment arm and subgroup. P-value is for the interaction.
[c] ROW includes South Korea, Lithuania, Poland, Russia, Taiwan, Turkey, Ukraine. SM includes Argentina, Brazil, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru. WEST includes Australia, Canada, Czech Republic, Germany, Greece, Hungary, Israel, Italy, New Zealand, Portugal, Spain, USA.

Figure 13. Reviewer's Subgroup Analyses on ACR20 response at Week 24 – Studies 11072 part B cohort 2 and 10832 pooled



Note. X-axis is on logarithmic scale with base 10.

Table 35. Reviewer's analyses on 8 subdomains of SF-36

	Study 11072 (Week 24) n LS mean change from baseline (SE) LS mean diff, 95% CI	Study 10832 (Week 12) n LS mean change from baseline (SE) LS mean diff, 95% CI

	p-value vs PBO			p-value vs PBO		
	PBO (N=398)	SAR150 (N=400)	SAR200 (N=399)	PBO (N=181)	SAR150 (N=181)	SAR200 (N=184)
SF36-PCS	361 5.3 (0.4)	352 7.8 (0.4) 2.5 (1.4, 3.6) <0.0001	345 7.9 (0.4) 2.6 (1.6, 3.7) <0.0001	169 3.7 (0.6)	160 6.9 (0.6) 3.2 (1.6, 4.7) <0.0001	165 6.8 (0.6) 3.1 (1.6, 4.6) <0.0001
Physical functioning	369 11.5 (1.2)	359 16.7 (1.2) 5.2 (2.0, 8.4) 0.0016	350 16.4 (0.6) 4.9 (1.7, 8.1) 0.0030	170 6.7 (1.7)	165 14.7 (1.7) 8.0 (3.6, 12.4) 0.0004	171 14.7 (1.6) 8.0 (3.7, 12.3) 0.0003
Role-physical	366 12.9 (1.2)	357 18.4 (1.2) 5.6 (2.6, 8.6) 0.0009	350 19.3 (1.2) 6.4 (3.4, 9.5) <0.0001	170 10.3 (1.7)	162 16.8 (1.7) 6.5 (2.1, 10.9) 0.0040	169 16.3 (1.7) 6.0 (1.6, 10.3) 0.0078
Bodily pain	367 16.9 (1.1)	358 24.7 (1.1) 7.7 (4.9, 10.6) <0.0001	353 26.6 (1.1) 9.7 (6.8, 12.5) <0.0001	170 11.6 (1.5)	164 22.0 (1.6) 10.4 (6.3, 14.5) <0.0001	170 24.3 (1.5) 12.7 (8.6, 16.8) <0.0001
General health	365 7.6 (0.9)	356 12.1 (0.9) 4.4 (2.1, 6.8) 0.0002	352 13.9 (0.9) 6.3 (4.0, 8.7) <0.0001	170 6.4 (1.3)	164 8.8 (1.3) 2.4 (-1.0, 5.8) 0.1600	169 10.9 (1.3) 4.6 (1.2, 7.9) 0.0082
SF36-MCS	361 4.2 (0.5)	352 5.5 (0.5) 1.3 (0.1, 2.7) 0.0628	345 7.5 (0.5) 4.3 (2.8, 5.8) <0.0001	169 3.5 (0.7)	160 5.1 (0.8) 1.6 (-0.3, 3.6) 0.1005	165 6.5 (0.7) 3.0 (1.0, 4.9) 0.0028
Vitality	367 11.1 (1.0)	358 13.6 (1.0)	352 17.1 (1.0)	170 8.5 (1.4)	165 13.1 (1.5)	171 15.1 (1.4)

		2.5 (0.2, 5.1) 0.0646	6.0 (3.3, 8.6) <0.0001		4.6 (0.9, 8.4) 0.0163	6.6 (2.8, 10.3) 0.0007
Social functioning	368 9.9 (1.1)	359 16.1 (1.2) 6.2 (3.1, 9.2) <0.0001	353 18.9 (1.2) 9.0 (6.0, 12.0) <0.0001	170 9.1 (1.7)	165 17.2 (1.7) 8.1 (3.6, 12.5) 0.0004	171 16.2 (1.7) 7.1 (2.7, 11.5) 0.0018
Role-emotional	366 10.2 (1.3)	357 14.2 (1.3) 4.0 (0.7, 7.3) 0.0186	350 16.2 (1.3) 6.0 (2.6, 9.3) 0.0005	169 8.2 (1.9)	161 12.6 (1.9) 4.3 (-0.7, 9.3) 0.0884	168 13.6 (1.9) 5.4 (0.4, 10.3) 0.0338
Mental health	367 8.1 (0.9)	358 10.1 (0.9) 2.1 (0.3, 4.5) 0.0923	352 13.1 (0.9) 5.1 (2.6, 7.5) <0.0001	170 5.3 (1.3)	165 7.8 (1.3) 2.5 (-0.9, 5.9) 0.1565	171 12.1 (1.3) 6.7 (3.3, 10.1) 0.0001

Note: Analyses in Study 11072 used post rescue data at Week 24 for early escaped patients.

AGENCY QUESTION / REQUEST FOR INFORMATION ITEM NO. 1:

Your protocol for Study EFC11072 indicates that "from Week 16, patients with lack of efficacy defined as less than 20% improvement from baseline in either SJC or TJC for two consecutive visits, or any other clear lack of efficacy based on investigator judgment will be proposed to be rescued with open-label SAR153191 highest available dose…"

We calculate that, among patients in part B cohort 2 who were eligible for escape at Week 16 according to the swollen and tender joint count criteria, 75/140 (54%) of patients on placebo initiated rescue therapy, as compared to 19/79 (24%) and 20/71 (28%) of patients on sarilumab 150 mg and 200 mg, respectively. We calculated relatively similar results in Study EFC10832, with smaller differences between treatment arms in the proportions.

Please clarify why such low proportions of patients on all treatment arms who met the escape criteria actually initiated escape therapy. In addition, while we recognize that the comparison between treatment arms is no longer a randomized comparison, provide any insight you may have into why a greater proportion of patients on placebo than sarilumab who met the rescue criteria actually initiated rescue therapy. (We note that a similar trend was observed for the proportions who initiated rescue by investigator judgment). The manner in which rescue is implemented is important because patients who initiated rescue were considered non-responders in the primary analysis of ACR20 at Week 24.

Sanofi response:

The Sponsor concludes that Investigators implemented rescue therapy based on changes in tender joint count (TJC), swollen joint count (SJC), and their clinical assessments of changes in patient reported pain, disease activity, and physical function evident during the weeks prior to rescue, resulting in a higher proportion of patients randomized to placebo exhibiting minimal improvement or worsening in clinical factors other than TJC or SJC. Patients who satisfied protocol-specified rescue criteria but were not rescued reported greater improvements in self-assessments of pain, disease activity and physical function, as well as reductions in active joint counts during the weeks preceding the decision to initiate rescue than patients who were rescued. Tender joint count, SJC, and physician's global assessment of disease activity were significant covariates in a logistic regression model of the probability of rescue. Elimination of these factors resulted in no differences between treatment groups, further validating that the decision to rescue was based on the Investigators' global impression of lack of improvement or worsening during the preceding weeks.

Applicant's tipping point analyses on mTSS:

(excerpted from Response to Agency Request: Information Request on 12-Jan-2016)

Results:

For mean change in mTSS at Week 52, the results in Study EFC11072 remained significant even under conservative assumptions; ie, that the imputed average response for patients with missing data in the placebo group equaled the observed sarilumab response, and the imputed average response for patients with missing data in the sarilumab groups equaled the observed placebo response. The tipping point at which significance was lost required assumptions of no change in the placebo group and mean increases of 3 points or higher in the sarilumab groups.

Methods:

For Study EFC11072, the tipping-point for the co-primary endpoint of change from baseline in mTSS at Week 52 was assessed using the following analyses (Analysis 1 through 3).
In the entire analysis, all collected mTSS data were used including data collected during the double-blind period, during the post-rescue period, and during the post-study period if the patient prematurely discontinued the study treatment. Note that the best efforts were made to perform Xray assessments at the scheduled time points for all patients, including patients who discontinued prematurely from the study.
Analysis 1. Pattern of missingness.
First, the pattern of missingness was assessed for each of the 3 treatments (placebo, 150 mg q2w sarilumab, and 200 mg q2w sarilumab).
In this analysis, only the baseline and Week 52 data were included.

Analysis 2. A conservative approach using multiple imputation.
Second, an imputation approach was performed using a multiple imputation procedure to assess the effect of the missing data. The multiple imputation used the following conservative imputation procedure:
• missing data in the placebo group at Week 52 was imputed using the empirical distribution of the sarilumab group at Week 52,
• missing data in the sarilumab group at Week 52 was imputed using the empirical distribution of the placebo group at Week 52,
• the multiple imputation was repeated 100 times.

Patients with missing baseline mTSS were excluded from the analysis.
To facilitate such an imputation procedure, the analyses were carried out in 2 different models: the first model consisted of placebo and 150 mg q2w sarilumab, and the second model consisted of placebo and 200 mg q2w sarilumab. The parametric analysis of covariance (ANCOVA) used the covariate baseline to analyze the effect of each sarilumab versus placebo at Week 52.
Note that as a sensitivity analysis, the parametric ANCOVA model was used instead of the nonparametric rank ANCOVA model because the multiple imputation analysis for nonparametric inference is not well established. Results reported in the appendix of the Clinical Summary of Efficacy confirm that the parametric ANCOVA model and the nonparametric rank ANCOVA model give similar results in Study EFC11072.

Analysis 3. Tipping-point analyses.
Third, since the conservative approach in the above step still led to statistically significant results (p-value <0.025 for each dose), further tipping-point analyses were conducted. In this analysis, since smaller changes in baseline in mTSS indicates greater efficacy, the following analyses were carried out:
• missing data in the placebo group at Week 52 was imputed using the empirical distribution of the sarilumab group at Week 52 + Eff1 (Eff1 is called "Shift in Placebo") which results in an imputation representing greater efficacy than sarilumab, and
• missing data in the sarilumab group at Week 52 was imputed using the empirial distribution of the placebo group at Week 52 + Eff2 (Eff2 is called "Shift in Sarilumab") which results in an imputation representing less efficacy than placebo,
• where Eff1 varies from 0 to -3 by 0.5 and Eff2 varies from 0 to 3 by 0.5 to provide different scenarios: the tipping-point will be the Eff1 and Eff2 that makes the p-value cross the significance level (0.025 in the current setting).

Summary of the secondary efficacy endpoints (Study 10832)

APPEARS THIS WAY ON ORIGINAL

Table 25 - Hierarchical order for the secondary efficacy endpoints

	Placebo + MTX (N = 181)	Sarilumab 150mg q2w + DMARD (N = 181)		Sarilumab 200mg q2w + DMARD (N = 184)	
Parameter[a]		**Estimate**[b]	**P-value**[c]	**Estimate**[b]	**P-value**[c]
Primary endpoints					
ACR20 – Week 24	61(33.7%)	101(55.8%)	**< 0.0001**	112(60.9%)	**< 0.0001**
HAQ-DI – Week 12	-0.26(0.043)	-0.46(0.04)	**0.0007**	-0.47(0.043)	**0.0004**
Secondary endpoints					
DAS28-CRP – Week 24	-1.38(0.119)	-2.35(0.111)	**< 0.0001**	-2.82(0.108)	**< 0.0001**
ACR50 – Week 24	33 (18.2%)	67 (37.0%)	**< 0.0001**	75 (40.8%)	**< 0.0001**
ACR70 – Week 24	13 (7.2%)	36 (19.9%)	**0.0002**	30 (16.3%)	**0.0056**
DAS28-CRP<2.6 – Week 24	13 (7.2%)	45 (24.9%)	**< 0.0001**	53 (28.8%)	**< 0.0001**
CDAI – Week 24	-16.35(1.195)	-23.65(1.136)	**< 0.0001**	-26.08(1.109)	**< 0.0001**
HAQ-DI – Week 24	-0.34(0.051)	-0.52(0.049)	**0.0078**	-0.58(0.048)	**0.0004**
SF-36 Physical – Week 24	4.40(0.692)	7.65(0.653)	**0.0004**	8.48(0.630)	**< 0.0001**
SF-36 Mental – Week 24	4.74(0.902)	6.26(0.848)	0.2026	6.76(0.817)	0.0854
FACIT – Fatigue – Week 24	6.82(0.863)	9.86(0.802)	0.0078	10.06(0.778)	0.0040
Morning Stiffness – Week 24	-21.66(2.390)	-32.30(2.231)	0.0008	-33.79(2.148)	0.0001
WPS-RA– Week 24			0.0004		0.0003
RAID – Week 24	-1.8(0.203)	-2.55(0.189)	0.0057	-2.80(0.183)	0.0002
EQ-5D-3L – Week 24	0.19(0.024)	0.29(0.023)	0.0034	0.34(0.022)	< 0.0001

a For further details of the endpoint definition and analysis method see the SAP (16-1-9-sap).
b Values presented are number and percent of responders for binary variables and LS mean change from baseline with standard error for continuous variables
c Nominal p-values. All values in bold font are significant according to the hierarchical testing procedure.

Source: Excerpted from the Clinical Study Report for Study EFC10832 (page 94).

References

1. Hoffmann-LaRoche Inc. Briefing document for tocilizumab biologic license application 125276 for FDA arthritis advisory committee. July 2008.

APPEARS THIS WAY ON ORIGINAL

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

YONGMAN KIM
09/02/2016

GREGORY P LEVIN
09/02/2016

THOMAS J PERMUTT
09/02/2016
I concur

STATISTICAL FILING REVIEW FOR A NEW NDA/BLA

NDA/BLA Number: BLA761037
NDA/BLA Type: Standard
Stamp Date: 10/30/2015
Applicant: Sanofi
Drug Name: (b) (4) (sarilumab)
Indication: Treatment of adult patients with moderately to severely active rheumatoid arthritis (RA) who have had an inadequate response or intolerance to one or more Disease-Modifying Anti-Rheumatic Drugs (DMARDs).
Statistical Team: Yongman Kim PhD & Gregory Levin PhD
Clinical Team: Suzette Peng MD & Janet Maynard MD
Medical Division: Division of Pulmonary, Allergy, and Rheumatology Products
Project Manager: Christine Ford

Introduction:

This submission is for an original BLA of sarilumab for RA indication.
Sarilumab is a recombinant human immunoglobulin (IgG)1 monoclonal antibody that binds specifically to both soluble and membrane-bound interleukin-6 (IL-6) receptors (sIL-6Rα and mIL-6Rα) and inhibits IL-6-mediated signaling.
As of cut-off date of 31 July 2015, the sarilumab clinical development program for RA included 9 Phase 2 and 3 studies of which 4 have been completed (EFC11072 Parts A and B [Phase 2 and 3, respectively], EFC10832 [Phase 3], EFC13752 [Phase 3], and SFY13370 [Phase 3]). See the table below for studies conducted in the clinical development program.

APPEARS THIS WAY ON ORIGINAL

Study	Background therapy	Study Population	Study duration	Treatment (Number of patients)	Rescue[a]	Eligible for LTS11210
EFC11072 Part A (Phase 2)	MTX	RA, MTX-IR	12 weeks	Placebo (52) Sarilumab: 100 mg qw (50); 150 mg qw (50); 100 mg q2w (51); 150 mg q2w (51); 200 mg q2w (52)	No	Yes
EFC11072 Part B (Phase 3)	MTX	RA, MTX-IR	52 weeks	Cohort 1[b] Placebo (30) Sarilumab: 100 mg qw (29); 150 mg qw (27); 100 mg q2w (28); 150 mg q2w (30); 200 mg q2w (28) Cohort 2[b]: Placebo (398) Sarilumab: 150 mg q2w (400); 200 mg q2w (399)	From Week 16	Yes
EFC10832 (Phase 3)	DMARD	RA, TNF-IR or TNF-α intolerant	24 weeks	Placebo (181) Sarilumab: 150 mg q2w (181); 200 mg q2w (184)	From Week 12	Yes
SFY13370 (Phase 3)	DMARD	RA, TNF-IR or TNF-α intolerant	24 weeks	Sarilumab: 150 mg q2w (49); 200 mg q2w (51) Tocilizumab: 4/8 mg/kg q4w (102)	No	Yes
EFC13752 (Phase 3)	None (monotherapy)	RA, non-biologic DMARD-IR or DMARD intolerant	24 weeks	OL sarilumab: 150 mg q2w (65); 200 mg q2w (67)	No	Yes
LTS11210 (Phase 3)[c,d]	DMARD (EFC11072, EFC10832, SFY13370 or ACT11575) None (EFC13752)	See initial study criteria	5 years	Sarilumab 150 mg qw prior to Phase 3 dose selection (316) Sarilumab 200 mg q2w after Phase 3 dose selection (1682)	NA	NA
MSC12665 (Phase 3)[c]	DMARD	RA	Main: 12 weeks Extension: 52 weeks	Main study: OL sarilumab PFS: 150 mg q2w (53); 200 mg q2w (56) OL sarilumab AI: 150 mg q2w (56); 200 mg q2w (52) Extension: OL sarilumab PFS: 150 mg q2w (192)	No	No

Among them, EFC11072 Part B, Cohort 2 and EFC10832 are pivotal efficacy studies that I will review for substantial evidence of efficacy for approval.
The key measures of efficacy assessing clinical response in the sarilumab clinical development program for RA are American College of Rheumatology response criteria, Disease Activity Score for 28 joints, Clinical Disease Activity Index, physical function as measured by HAQ-DI, and radiographic progression as quantified by mTSS.

The following are key elements of statistics-related interactions between the applicant and the FDA:

- Type C meeting (9/29/2009):
  - Agency agreed to Study EFC11072 with an operationally seamless design (proposed Phase 2/3 study). Also, the Agency indicated that patients in cohort B1 should not be included in efficacy analyses, but should be used for safety.
- EOP2 meeting (9/15/2011):
  - Agency agreed that the RA program was acceptable to support the proposed indication. The Agency recommended the Sponsor incorporate tocilizumab into a Phase 3 study for use as a benchmark for safety comparison.
- Email response from Agency regarding information provided in Serial No. 0171 (4/10/2012):
  - The Agency agreed with the modification of the terminology for the primary objectives and co-primary endpoints while maintaining the previously proposed hierarchical testing procedure.
  - Report results for the HAQ-DI at Week 24 as an important secondary analysis. The Agency advised on approaches to handling data for treatment-related dropouts. Measures should be taken to protect the reliability of the radiographic endpoint.

    - The Agency agreed that it would be appropriate to claim statistical significance of the treatment effect on the ACR20 endpoint even in the absence of a significant effect on the other 2 co-primary endpoints.
- Sponsor response to FDA Comments on SAP provided in Serial No. 0297 (6/3/2013):
    - The Sponsor provided proposed revisions to the SAP for Study EFC11072 to address Agency comments regarding the analysis of the HAQ-DI endpoint.
- Email response from Agency regarding information provided in Serial No. 0297 (8/21/2013):
    - The Agency acknowledged the Sponsor's efforts to define an analysis of the HAQ-DI endpoint that appropriately handles missing or escaped patients.
- Comments on SAP for Study EFC10832 (9/23/2014)
    - The Agency noted that the approach to consider patients who discontinue treatment to be non-responders in analyses of ACR20 is reasonable, but that the primary outcome is therefore a composite measure of treatment success.
    - The Agency indicated that Week 12 HAQ-DI data collected after patients stop treatment should be included in the analysis.
    - The Agency indicated that the proposed sensitivity analyses are not sufficient and recommended multiple imputation approaches.
- Pre-BLA Meeting (10/22/2014)
    - The Agency recommended additional safety analyses that compare treatment groups and more appropriately account for study differences and the study designs.

My statistical review will confirm the applicant's key analyses on RA signs and symptoms and radiographic structural damage and conduct sensitivity analyses to check robustness of efficacy data regarding assumptions on missing data mainly due to discontinuation of study treatment and rescue treatment.

Filing Checklist:

On **initial** overview of the NDA/BLA application for refuse-to-file (RTF):

	Content Parameter	**Yes**	**No**	**NA**	**Comments**
1	Index is sufficient to locate necessary reports, tables, data, etc.	x			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	x			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	x			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	x			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? ___Yes_____

Potential Review Issues:

Content Parameter	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	**x**			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	**x**			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			**x**	
Appropriate references for novel statistical methodology (if present) are included.	**x**			
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	**x**			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.		**x**		**Additional sensitivity analyses will be requested and conducted**

Additional Discussion:

During the filing meeting, the team decided that a statistical review might be necessary for the active controlled safety study SFY13370 that collected efficacy data, but a statistical review might be not necessary for efficacy data collected from the long-term open label study LTS11220.
The following will be major potential focus areas in my statistical review:

- Confirmation of key analyses
- Handling of missing data and post-escape data
- Radiographic data analysis
- Labeling, including claims for exploratory endpoints

Comments for Applicant:

1. *For all endpoints proposed for inclusion on the product label, we request additional supportive analyses that include all observed data, including any outcomes collected after escape or discontinuation of study medication.*
2. *You have not provided sensitivity analyses that sufficiently evaluate the potential impact of missing data on the reliability of efficacy results. For co-primary endpoints, please examine the potential effects of missing data and rescue on your results using tipping point sensitivity analyses. These tipping point analyses should include all observed data, including outcomes after patients discontinue study therapy or initiate rescue*

medications, and should vary assumptions about outcomes among the subsets of patients on the sarilumab and placebo arms who withdrew from treatment prior to the planned endpoint. The varying assumptions should include scenarios where dropouts on sarilumab had worse future outcomes than dropouts on placebo. The goal is to identify assumptions under which the conclusions change, i.e., under which there is no longer evidence of a treatment effect. Then, the plausibility of those assumptions can be discussed.

APPEARS THIS WAY ON ORIGINAL

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

YONGMAN KIM
12/29/2015

GREGORY P LEVIN
12/29/2015