CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

761054Orig1s000

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	BLA 761054
Drug Name:	SB2 (Proposed Infliximab Biosimilar)
Indication(s):	All Infliximab Indications
Applicant:	Samsung
Date(s):	Filing Date : 5/20/16
	PDUFA : 1/21/17
Review Priority:	Standard
Biometrics Division:	DBII
Statistical Reviewer:	Pottackal, Ginto
Concurring Reviewers:	Levin, Gregory
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Waheed, Juwaria, Nikolov, Nikolay
Project Manager:	Ford, Christine
Keywords:	BLA review, clinical studies, missing data, biosimilar, infliximab

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Samsung Bioepis has proposed SB2 as a biosimilar to US-Remicade for the treatment of rheumatoid arthritis (RA) and other related indications. The applicant conducted a comparative clinical study (SB2-G31-RA) to evaluate the efficacy, safety, pharmacokinetics and immunogenicity of SB2 compared to EU-Remicade in patients with rheumatoid arthritis. The applicant claims that the results from this trial show similar efficacy between the products.

The primary endpoint of the study was the proportion of patients who remained in the study and achieved an American College of Rheumatology 20% (ACR20) response at Week 30.
The adjusted treatment difference in ACR20 response rate between the SB2 and EU-Remicade treatment groups was −2.95% and the 90% confidence interval of the adjusted treatment difference was (-9.60, 3.70) which was contained within the similarity margin of [−12%, +12%] recommended by FDA. The ACR20 response probabilities over time comparing the two treatments up to Week 30 also supported similarity.

Up to Week 30, 78 (13.4%) patients had withdrawn from the study: 44 patients (15.4%) from the SB2 treatment group and 34 patients (11.6%) from the EU-Remicade treatment group. We conducted tipping point analyses to explore the sensitivity of results to violations in assumptions about the missing data. Confidence intervals for the differences between SB2 and EU-Remicade failed to rule out concerning losses in efficacy only under the assumption that patients who dropped out on SB2 had much worse outcomes than dropouts on EU-Remicade. Given the similar proportions of patients and distributions of reasons for early withdrawal on the two treatment arms, in addition to the similar baseline characteristics between dropouts on the two arms, an assumption of such large differences between the outcomes in dropouts on the two treatments seems implausible. That is, the finding of similar efficacy is highly credible notwithstanding the number of dropouts.

To reliably evaluate whether there are clinically meaningful differences between two products, a comparative clinical study must have assay sensitivity, or the ability to detect meaningful differences between the products, if such differences exist. Historical evidence of sensitivity to drug effects and appropriate trial conduct may be used to support the presence of assay sensitivity and a conclusion that the treatments are similarly effective rather than similarly ineffective. Based on an evaluation of five historical, randomized, placebo-controlled clinical trials of infliximab, we concluded that (1) the design of the historical trials were largely similar to that of comparative clinical Study SB2-G31-RA; and (2) there were relatively large and consistent treatment effects across the five historical studies. There were some issues identified with study conduct, including a relatively high rate of study withdrawal and potential eligibility criteria violations at one clinical site, but the assay sensitivity of this study was sufficient to allow the favorable assessment of the similarity of the SB2 to the US-Remicade.

2 INTRODUCTION

2.1 Background

The Biologics Price Competition and Innovation Act of 2009 (BPCI Act) created an abbreviated licensure pathway for the approval of biosimilar products. Section 351(k) of the PHS Act (42 U.S.C. 262(k)), added by the BPCI Act, outlined the application requirements for a proposed biosimilar product.

In Section 351(*i*) of the PHS Act, a *biosimilar* is defined as follows: "the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components" and "there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity, and potency of the product." In the guidance document *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product* (FDA, April 2015), FDA recommends that applicants use a *stepwise approach* to demonstrate biosimilarity. The stepwise approach will typically include comparative analytical, pharmacokinetic (PK), and clinical studies. FDA intends to consider the *totality-of-the-evidence* when reviewing the applicant's demonstration of biosimilarity.

Rheumatoid Arthritis (RA) is an autoimmune disease characterized by inflammation in the synovium of joints, malaise, morning stiffness, and fatigue. If not treated, RA may lead to significant disabilities including bone erosion and joint deformity, over 10-20 years. Increased levels of tumor necrosis factor alpha (TNF-α) have been detected in RA patients, indicating that it may have a role in inducing inflammatory response.

Infliximab is a monoclonal antibody that inhibits the activity of TNF- α. FDA has approved infliximab for the treatment of RA, Crohn's disease, ulcerative colitis, psoriasis, psoriatic arthritis, and ankylosing spondylitis. Samsung Bioepis has submitted a BLA for SB2, a proposed biosimilar biological product to EU-Remicade. The applicant provided reports on direct physico-chemical and biological comparisons between SB2, EU-Remicade and US-Remicade. US-Remicade was used as the comparator product in the analytical and PK similarity studies whereas EU-Remicade was used as the comparator in the comparative clinical study. The similarity between EU- and US-Remicade was demonstrated using analytical and PK bridging studies.

The applicant has submitted results from several nonclinical, analytical, and clinical studies to support the claim of no clinically meaningful differences between SB2 and US-Remicade. This review primarily considers the safety and efficacy evaluation of SB2 in the comparative clinical study in RA.

2.2 History of Product Development

The primary focus of the clinical development program was to demonstrate similar profiles of PK, efficacy and safety of SB2 compared to the reference product. The development program includes three-way physico-chemical comparisons between SB2, EU-Remicade, and US-Remicade. It contains a single-dose PK study in healthy subjects and a comparative clinical efficacy/safety study in RA patients. Study SB2-G11-NHV was a randomized, single-blind, three-arm, parallel group, single-dose study to compare the PK, safety/tolerability and immunogenicity of three formulations of infliximab (SB2, US Remicade and EU-Remicade) in healthy subjects. Study SB2-G31-RA was a randomized, double-blind, parallel group, multicenter study to evaluate the efficacy, safety/tolerability and immunogenicity of SB2 compared to EU-Remicade in subjects with moderate to severe RA despite methotrexate (MTX) therapy for up to 54 weeks.

The clinical development program was designed by taking into consideration the advice provided by FDA. In February 2012, a pre-IND meeting was held to discuss the chemical, pharmaceutical, and biological development, the non-clinical development, and the clinical development of SB2. A clarification was requested in March 2012. In December 2012, a BPD meeting was held for the clinical development of SB2 followed by a clarification request in March 2013. In March 2014, a BPD meeting was held for the chemical, pharmaceutical, and biological development and the clinical development of SB2. In the Type 4 meeting in December 2015, FDA stated disagreement with the applicant's plan for the primary analysis in Study SB2-G31-RA to be carried out in a per-protocol population using a 95% confidence interval and a similarity margin of ±15%. FDA recommended that the primary analysis be carried out in all randomized patients and stated that it expects the overall type I error rate to be controlled at 5%, i.e., a 90% confidence interval for the difference in ACR20 responses can be compared to the margin. Furthermore, FDA recommended a similarity margin with a lower bound no greater in magnitude than -12%. The applicant agreed with this recommendation and carried out additional analyses to calculate 90% CIs for the difference in ACR20 in the full analysis and per-protocol sets. As the double blind period of the study had already completed, the results from the revised analysis were not included in the clinical study report. However, these results were reported in the submission in the Integrated Summary of Efficacy and in the Summary of Clinical Efficacy report.

2.3 Specific Studies Reviewed

The statistical review will focus on the efficacy results from Study SB2-G31-RA, which was a randomized, double-blind, parallel group, multicenter comparative clinical study to evaluate the efficacy, safety, pharmacokinetics and immunogenicity of SB2 compared to EU-Remicade in

patients with rheumatoid arthritis. There were two stages in this study: a randomized, double-blind period up to week 54 and a transition-extension period up to week 70.

Table 1: List of clinical studies included in development program

Study Phase	Subjects	Location	Study Design	Primary Objective	Primary Endpoints	Study Population	Statistical Analysis
Phase 1: **SB2-G11-NHV**	Healthy subjects	Germany	A randomized, Single-blind, Three-arm, Parallel Group, Single-dose Study	Investigate and compare the PK profiles of SB2, United States of America (US) sourced Remicade and EU sourced Remicade in healthy subjects	AUCinf, AUClast, Cmax	1 (n=53) SB2 2 (n=53) EU sourced Remicade 3 (n=53) US sourced Remicade Total : 159 Subjects	The statistical analysis of the log-transformed primary endpoint
Phase III: **SB2-G31-RA**	RA patients	UK, Czech Republic, Bulgaria, Lithuania, Latvia, Poland, Romania, Bosnia, Ukraine, Korea, India, Mexico, Philippines	A Randomized, Double-blind, Parallel Group, Multicenter Clinical Study	To demonstrate the equivalence of SB2 to EU-Remicade at Week 30, in terms of ACR20 response rate in subjects with moderate to severe rheumatoid arthritis (RA) despite methotrexate (MTX) therapy.	ACR20 response rate at Week 30	Double-blind Period: 584 RA subjects (291 for SB2, 293 for EU Remicade) Transition extension Period: 396 RA subjects (n=201 in SB2/SB2, n= 94 in Remicade/ SB2, n=101 in Remicade/ Remicade)	The adjusted difference, its 95% CI were analyzed by non-parametric method using NParCov SAS macro with baseline C-reactive protein as covariate and stratified by region

Source: Reviewer

2.4 Data Sources

The applicant submitted the data and reports to the CDER electronic data room under the network path, \\CDSESUB1\evsprod\BLA761054\761054.enx. The datasets are available in SAS transport format. The applicant created ADAM datasets for efficacy data analysis. Both ADAM and SDTM datasets are available in the above EDR location.

3 STATISTICAL EVALUATION

The applicant developed a statistical analysis plan (SAP) based on the final version of the Protocol SB2-G31-RA. The SAP provided details of the statistical methods to be used in the analysis of efficacy, safety, pharmacokinetics, and immunogenicity data.

3.1 Data and Analysis Quality

The applicant submitted data files of acceptable quality and it was possible to reproduce the primary analysis dataset, and in particular the primary endpoint results, from the original data source. A final statistical analysis plan (SAP) was submitted and relevant analysis decisions were made prior to unblinding. After detecting some errors in the original data definition file, the applicant resubmitted a modified file. It contains details about different datasets, original variables, and derived variables used for analysis. The applicant submitted SAS codes and SAS macros of primary and secondary analyses.

During the routine inspections conducted by the European Medicines Agency (EMA), (b) (4) [redacted] The FDA Office of Scientific Investigations inspected the sponsor's U.S. agent for this application, Quintiles, and two foreign clinical sites, and did not identify any such issues with study conduct.

3.2 Evaluation of Efficacy

The primary objective of the study SB2-G31-RA was to demonstrate the therapeutic similarity of SB2 with EU-Remicade. The primary efficacy variable for the study was ACR20 response, a composite endpoint defined by the American College of Rheumatology. Other efficacy endpoints included ACR50 response, ACR70 response, individual components of the ACR improvement

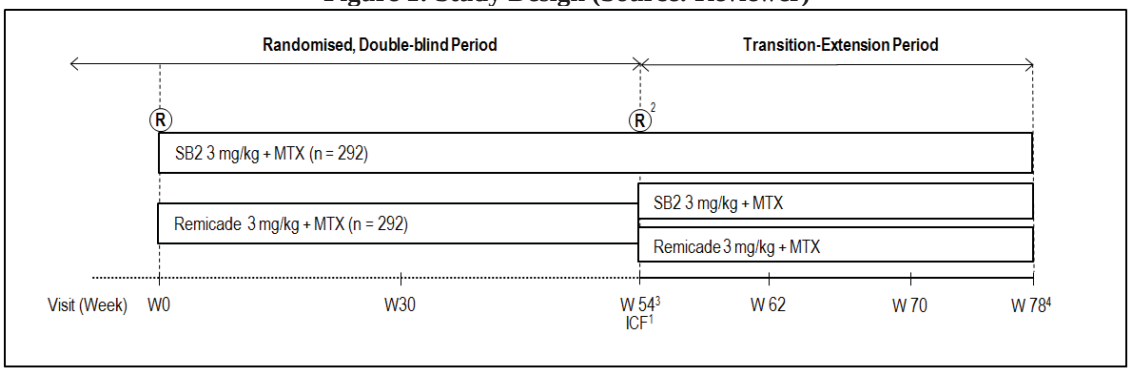
criteria, DAS28, major clinical response and the EULAR response (good response, moderate response or no response).

3.2.1 Study Design

Study SB2-G31-RA was a randomized, double blind, parallel group, multicenter comparative clinical study to evaluate the efficacy, safety, pharmacokinetics, and immunogenicity of SB2 compared to EU-Remicade in subjects with moderate to severe rheumatoid arthritis despite methotrexate therapy. The study consisted of male or female patients aged 18–75 years, who had been diagnosed as having RA according to the revised 1987 ACR criterial for at least 6 months prior to screening. Active disease was defined by the presence of six or more swollen joints, six or more tender joints, and at least two of the following: morning stiffness lasting at least 45 minutes, an erythrocyte sedimentation rate (ESR) greater than 28 mm/h, and a serum C-reactive protein (CRP) concentration greater than 1.0 mg/dL. Patients had been on methotrexate for at least 6 months prior to randomization, with a stable dose of MTX 10-25 mg/week for at least 4 weeks, and they also received ≥ 5-10 mg/week folic acid during the study.

The study contained two distinct periods. i) Randomized double blind period and ii) Transition extension period. A graphical representation of the study design is given in the figure (**Error! Reference source not found.**) below.

Figure 1: Study Design (Source: Reviewer)



3.2.2 Randomized, Double-blind Period

The initial period consisted of 6 weeks of screening and 54 weeks of active treatment stages. 584 subjects with moderate to severe RA were randomized in a 1:1 ratio to receive SB2 3 mg/kg or Remicade 3 mg/kg via a 2 hour (h) intravenous (IV) infusion, at Weeks 0, 2 and 6 and then every 8 weeks until Week 46. From Week 30 the dose level could be increased step-wise by 1.5 mg/kg, up to a maximum of 7.5 mg/kg, every 8 weeks if the subject's RA symptoms were not well controlled by the existing dose. There were approximately 80 investigator sites in Czech

Republic, Lithuania, Poland, Ukraine, Bulgaria, Bosnia and Herzegovina, Romania, the United Kingdom (UK), Latvia, Philippines, and South Korea.

The primary efficacy endpoint was the ACR20 response at Week 30. The ACR20 response was calculated as: at least 20% improvement from baseline in swollen and tender joint counts and at least a 20% improvement from baseline in at least 3 of the following 5 remaining ACR core set measures: subject and physician global assessment using a 100 mm visual analogue scale (VAS), pain assessment using a 100 mm VAS, disability assessment using the health assessment questionnaire disability index (HAQ-DI), and acute phase reactant level (CRP).

The secondary endpoints in the study included ACR20 response at Week 54, ACR 50% response criteria (ACR50) and ACR 70% response criteria (ACR70) at Weeks 30 and Week 54, numeric index of the ACR response (ACR-N) at Week 30 and Week 54, area under the curve (AUC) of ACR-N up to Week 30, disease activity score based on a 28 joint count (DAS28 score) at Week 30 and Week 54,EULAR response at Week 30 and Week 54, AUC of the change in DAS28 from baseline up to Week 30, major clinical responses, and the modified total sharp score at Week 54.

Subjects who discontinued the administration of IP prior to Week 54 were asked to return to the investigator site for the early termination (ET) visit procedures to be performed and to have a follow-up telephone interview, but were not followed up to obtain key efficacy and safety assessments. Subjects who withdrew from the study with missing ACR20 response at Week 30/Week 54 were considered as non-responders at Week 30/Week 54 in analyses in the full analysis set. Major reasons for withdrawal included adverse event, investigator discretion, lack of efficacy, protocol deviation, and withdrawal of consent.

3.2.3 Transition-Extension Period

The transition-extension period of the study ranged from Week 54 to Week 78. In this period, the patients originally randomized to and remaining in the study on the EU-Remicade group were re-randomized in 1:1 ratio to transition to SB2 or continue on EU-Remicade. Subjects originally randomized to the SB2 arm continued the same treatment in this stage. There were 201 subjects in the SB2 arm and 195 subjects in the EU-Remicade arm in the transition period. These 195 subjects in the EU-Remicade arm were randomized to SB2 (94 subjects) or EU-Remicade (101 subjects). Study objectives in this period were to compare the long-term safety, tolerability, immunogenicity and efficacy of SB2 in subjects with RA who transitioned from EU-Remicade treatment to SB2 to subjects who maintained the EU-Remicade treatment.

3.2.4 Data Sets in the study

The following datasets were used in different analyses:

Randomized Set [RAN]: The RAN consisted of all enrolled subjects who received a randomization number at the randomization visit.

Full Analysis Set [FAS]: The FAS consisted of all subjects in the RAN. However, subjects who did not qualify for randomization and were inadvertently randomized into the study were excluded from the FAS, provided these subjects did not receive any IP during that study phase.

Per-protocol Set 1 [PPS1]: The PPS1 consisted of all FAS subjects who completed the Week 30 visit and had an adherence (from Baseline to Week 30) within the range of 80–120% for both the expected number of IP administrations and the expected sum of MTX doses without any major protocol deviations that affected the efficacy assessment. The applicant defined the PPS1 as the primary analysis set. Major protocol deviations that led to exclusion from this set were pre-specified prior to unblinding the treatment codes for analyses. Some of the major protocol deviations include: patient's age out of range (<18 or >75) at screening, RA diagnosis period out of range (< 6 months) at screening, insufficient MTX treatment, joint counts not in the range, etc.

Per-protocol Set 2 [PPS2]: The PPS2 consisted of all FAS subjects who completed the Week 54 visit and had an adherence (from Baseline to Week 54) within the range of 80-120% for both the expected number of IP administrations and the expected sum of MTX doses without any major protocol deviations that affected the efficacy assessment.

Safety Set [SAF]: The SAF consisted of all subjects who received at least 1 dose of double-blind IP during the study period. Subjects were analyzed according to the treatment received.

Pharmacokinetic Population [PK population]: The PK population consisted of all subjects in the SAF who had at least 1 post-dose PK sample collected.

3.3 Statistical Methodologies

3.3.1 Planned Analysis

ACR20 response rate was the primary endpoint of the study. In order to demonstrate the similarity between SB2 and EU-Remicade, the applicant compared ACR20 response rates between the two treatment arms. The null hypothesis of the study was defined as either 1) SB2 is inferior to EU-Remicade or 2) SB2 is superior to EU-Remicade based on a pre-specified similarity margin. According to the statistical analysis plan, the biosimilarity between the two treatments would be concluded if the two-sided 95% confidence interval of the difference in

ACR20 response rate was contained within the similarity margin of [-15%,15%]. The applicant also carried out an analysis using a 90% confidence interval with a similarity margin of [-12%, 12%] based on FDA recommendations at a type 4 meeting on Dec 14, 2015.

A randomization based non-parametric ANCOVA method (Koch, 1998) was used to analyze the primary endpoint by adjusting for effects of region (pooled centers) and baseline CRP value. The primary efficacy analysis for ACR20 response was performed in the per-protocol population (PPS1). No missing data was imputed. In addition to the primary analysis in the PPS1, the applicant performed the same analysis in the full analysis set to explore the robustness of the results.

Similar statistical methods were used to analyze the secondary endpoints such as ACR50 and ACR70 response at Week 30 in the PPS1 and at Week 54 in the PPS2. Continuous ACR-N at Week 30 and Week 54 and the AUC of ACR-N up to Week 30 were analyzed using an analysis of variance (ANOVA) model with treatment group and study center as factors.

Change from Baseline in DAS28 at Week 30 and Week 54, and the AUC of DAS28 up to Week 30 were analyzed using an ANCOVA model, with treatment group and study center as factors, and using DAS28 baseline value as a covariate. Change from baseline value of Modified total Sharp score (mTSS) at Week 54, an endpoint assessing radiographic progression, was analyzed using an ANCOVA model. In addition, the applicant reported 95% confidence intervals for the adjusted difference in rates for binary endpoints and difference in means for continuous endpoints using the full analysis set.

The applicant conducted sensitivity analyses using three different approaches: an analysis of available data without imputation (excluding subjects with missing data at Week 30/Week 54), a non-responder imputation analysis (considering subjects with missing ACR20 response to be non-responders), and a pattern mixture analysis with a multiple imputation approach that assumes missing at random (MAR) missing data except for subjects who withdraw from the study with a primary reason of lack of efficacy.

In addition to the proposed sensitivity analysis methods, FDA recommended that the sponsor conduct additional analyses that more systematically and comprehensively explore the space of plausible missing data assumptions. Specifically, we recommended the inclusion of tipping point analyses that vary assumptions about the missing outcomes on the two treatment arms. As a response, the applicant conducted tipping point analysis and included the results in the Summary of Clinical Efficacy report. The applicant's analyses were based on single imputation, which does not take into account the uncertainty in the imputation process, so we conducted additional supportive tipping point analyses.

3.3.2 Similarity Margin for Study

The determination of an equivalence margin is a critical aspect of the design of the comparative clinical study because it determines the null hypothesis being tested in the primary analysis, i.e., the differences in efficacy that the study will need to rule out at an acceptable significance level. The term equivalence margin is a misnomer because it is not possible to statistically demonstrate that two products are equivalent with respect to a particular endpoint. Instead, we describe the margin as a similarity margin to better reflect the goal of the efficacy evaluation: to determine whether the two products are similar, in that a certain magnitude of difference (the margin) in efficacy can be ruled out.

The applicant initially proposed to conduct the primary efficacy analysis by comparing the 95% confidence interval (CI) of the difference of 2 proportions with the pre-specified equivalence margin of [−15%, 15%]. However, FDA recommended a similarity margin of [-12%, 12%] at a type 4 meeting on Dec 14, 2015. FDA also recommended use of a 90% because it generally expects the type I error probability to be controlled at the overall 5% level in comparative clinical studies. The applicant agreed with this recommendation and performed additional analyses to calculate 90% CIs for the difference in ACR20 in the FAS and PPS. As the double blind period of the study had already completed, the results from the revised analysis were not included in the clinical study report. However, these results were reported in the Integrated Summary of Effectiveness and Summary of Clinical Efficacy report. The lack of a priori agreement between the applicant and FDA on a similarity margin is not of concern in this case because the primary analysis successfully ruled out the ±12% margin recommended by FDA.

Our selection of a ±12% similarity margin was based on discussions with clinicians aimed at weighing the clinical importance of different losses in effect against the feasibility of different study sizes. In a comparative clinical study designed with 90% power to reject absolute differences greater than 12% in magnitude, observed differences larger than approximately 6% will result in failure to establish similarity, as the 90% confidence interval for the estimated difference will not rule out the 12% margin. Therefore, the comparative clinical study will be able to rule out losses in ACR20 response greater than 12% with high (at least 95%) statistical confidence, and will be able to rule out losses greater than around 6% with moderate (at least 50%) statistical confidence. The lower bound of the proposed similarity margin (-12%) also corresponds to the retention of approximately 50% of conservative estimates of treatment effect sizes relative to placebo for infliximab (e.g., see Table 3).

Table 2: Historical Effect of Infliximab on ACR 20 Response in Randomized Clinical Trials of Patients with Active RA Despite Treatment with MTX

Study	Week	MTX + Placebo		MTX + Infliximab		Difference in Response
		N	ACR Response	N	ACR20 Response	
(Maini, 1999)	30	88	20%	86	50%	30%
(Westhovens, 2006)	22	361	24%	360	55%	31%
(Schiff, 2008)	28	110	42%	165	59%	18%
(Zhang, 2006)	18	86	49%	87	76%	27%
(Abe, 2006)	14	47	23%	49	61%	38%
Meta-Analysis (Fixed Effects[1]): Difference (95% CI) 28.4% (23.6%, 33.3%)						
Meta-Analysis (Random Effects[2]): Difference (95% CI) 28.3% (22.6%, 34.1%)						

[1] Based on Mantel-Haenszel weights
[2] Based on DerSimonian-Laird approach

3.4 Patient Disposition, Demographic and Baseline Characteristics

3.4.1 Double Blind Period

Out of 584 randomized subjects, one subject in the SB2 treatment group who did not meet the inclusion/exclusion criteria was excluded from the Full Analysis Set (FAS). A total of 583 (99.8%) subjects were included in the FAS, 478 (81.8%) subjects satisfied the criteria for the PPS1, 410 (70.2%) subjects satisfied the criteria for the PPS2 and 325 (55.7%) subjects were analyzed for PK. Missing ACR responses were treated as non-responders in the FAS and no missing data were imputed in the PPS1 and PPS2.

The following tables summarize the number of patients included in each analysis set and demographic characteristics by treatment arm.

Table 3: Data Sets (Double Blind Period)

	Treatment				***All***	
	SB2		**Remicade**			
	N	**(%)**	**N**	**(%)**	**N**	**(%)**
Double Blind Phase						
Full Analysis Set Population	290	99.66	293	100.00	583	99.83
PPS1 Population	231	79.38	247	84.30	478	81.85
PPS2 Population	202	69.42	208	70.99	410	70.21
Safety Population	290	99.66	293	100.00	583	99.83
PKS Population	165	56.70	160	54.61	325	55.65

Source: Reviewer

From the baseline patient characteristics given in the table below, the two treatment arms were generally comparable and had similar patient profiles. There were 291 patients in the SB2 arm and 293 patients in EU-Remicade arm. Patients in the EU-Remicade group were slightly older than in the SB2 group. The majority of the study population was older than 65 years (85%). Around 87% of patients were whites, 80% were female and the mean age was 52 years. Duration of methotrexate used in the EU-Remicade group was slightly higher than in the SB2 group (53.05 vs. 48.44 months). There were no large imbalances in demographic and disease characteristics between the two study groups.

Table 4: Baseline Demographic Characteristics (Double Blind Period)

	SB2		**Remicade**		**Total**	
N	N=291		N=293		N=584	
Age	51.6	(11.92)	52.63	(11.74)	52.12	(11.83)
Age group						
< 65 years	251	(86.25)	248	(84.64)	499	(85.45)
>= 65 years	40	13.75	45	15.36	85	14.55
Sex						
Male	59	(20.27)	57	(19.45)	116	(19.86)
Female	232	(79.73)	236	(80.55)	468	(80.14)
Race						
White	252	(86.6)	254	(86.69)	506	(86.64)
Other	2	(0.69)	0	(0)	2	(0.34)
Asian	37	(12.71)	39	(13.31)	76	(13.01)
Ethnicity						
Mixed ethnicity	1	(0.34)	0	(0)	1	(0.17)
Indian (Indian subcontinent)	1	(0.34)	1	(0.34)	2	(0.34)
Hispanic or Latino	5	(1.72)	3	(1.02)	8	(1.37)
Other	284	(97.59)	289	(98.63)	573	(98.12)
Weight (kg)	72.27	(15.81)	71.92	(16.51)	72.10	(16.15)
Height (cm)	164.58	(9.28)	164.79	(8.57)	164.69	(8.92)
BMI (kg/m^2)	26.62	(5.25)	26.49	(5.97)	26.56	(5.62)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Table 5: Baseline Disease Characteristics

		SB2	Remicade	Total
Baseline Disease	**N**	290	293	583
HAQ-DI (0-3)	**Mean**	1.5	1.5	1.5
	SD	0.6	0.6	0.6
Physician Global Assessment	**Mean**	61.7	61.8	61.8
	SD	15.6	15.8	15.7
Subject global Assessment	**Mean**	62.9	62.7	62.8
	SD	17.5	18.7	18.1
Swollen joint count (0-66)	**Mean**	14.6	14.9	14.8
	SD	7.8	7.7	7.8
Subject Pain Assessment	**Mean**	61.2	63.3	62.3
	SD	18.6	20.0	19.3
Tender Joint Count	**Mean**	23.6	23.9	23.8
	SD	12.3	12.2	12.2
Duration of RA (years)	**Mean**	6.6	6.3	6.4
	SD	6.0	5.9	5.9
Duration of methotrexate used (months)	**Mean**	48.4	53.1	50.7
	SD	45.6	49.5	47.6
Weekly dose of MTX (mg) at baseline	**Mean**	14.7	14.7	14.7
	SD	4.1	4.2	4.2
Erythrocyte Sedimentation Rate - Baselin	**Mean**	46.7	44.5	45.6
	SD	22.3	19.2	20.9
CRP- Baseline (mg/L)	**Mean**	13.7	12.5	13.1
	SD	19.2	18.8	19.0

Source: Reviewer

3.4.2 Transition Extension Period

The demographic characteristics in the transition extension period were similar across the different study groups.

Table 6: Demographic Characteristics (Transition Extension Period)

N	SB2 ⇒ SB2 N=201		EU-Remicade ⇒ SB2 N=94		EU-Remicade ⇒ Remicade N=101		Total N=396	
Age	51.8	12.13	52.9	10.9	51.4	11.2	52	11.6
Age group								
< 65 years	171	85.1	80	85.11	90	89.11	341	86.11
>= 65 years	30	14.9	14	14.89	11	10.89	55	13.89
Sex								
Male	43	21.3	17	18.09	22	21.78	82	20.71
Female	158	78.6	77	81.91	79	78.22	314	79.29
Race								
White	183	91.0	87	92.55	88	87.13	358	90.40
Other	1	0.5	.	.	.	.	1	0.25
Asian	17	8.4	7	7.45	13	12.8	37	9.34
Weight (kg)	72.72	14.6	72.2	14.9	73.1	17.3	72.7	(15.4)
Height (cm)	165.18	9.0	165.6	8.0	165.4	7.5	165.4	(7.5)
BMI (kg/m^2)	26.64	5.0	26.3	5.1	26.8	6.4	26.6	(5.4)

Source: Reviewer

Cell contents are mean (standard deviation) or frequency (percent)

Table 7: Baseline Disease Characteristics (Transition Extension Period)

		EU-Remicade ⇒EU-Remicade	EU-Remicade⇒SB2	SB2⇒SB2
HAQ-DI (0-3)	N	101	94	201
	Mean	1.0	1.0	1.0
	Std	0.7	0.6	0.7
Physician Global Assessment	Mean	25.0	24.5	25.1
	Std	17.1	18.1	18.0
Subject global Assessment	Mean	35.8	35.5	34.8
	Std	21.9	22.6	23.3
Swollen joint count (0-66)	Mean	4.0	2.7	3.5
	Std	6.1	4.4	5.2
Subject Pain Assessment	Mean	36.0	35.9	35.6
	Std	22.6	23.4	23.8
Tender Joint Count	Mean	8.2	6.1	7.2
	Std	10.5	7.0	9.2
Duration of RA (years)	Mean	6.7	6.3	6.3
	Std	6.1	5.4	6.2
Duration of methotrexate used (mons)	Mean	52.1	49.7	51.1
	Std	50.6	45.4	46.8
Weekly dose of MTX (mg) at baseline	Mean	15.2	14.3	14.7
	Std	4.0	3.9	4.1
Erythrocyte Sedimentation Rate - Baseline	Mean	45.3	45.7	43.0
	Std	19.7	23.0	17.5
CRP- Baseline (mg/L)	Mean	13.7	13.8	12.0
	Std	18.8	21.9	19.1

Source: Reviewer

The disposition of subjects in the study was generally balanced between two study arms. There were 291 subjects in the SB2 and 293 in EU-Remicade group. Out of 584 total randomized subjects, 506 (86.6%) subjects completed week 30 and 452 (77.4%) subjects completed week 54. Up to week 30, 78 (13.4%) patients had withdrawn from the study: 44 (15.1%) patients from

SB2 and 34 (11.6%) patients from EU-Remicade. Withdrawal from the study was therefore slightly higher in SB2 than EU-Remicade at week 30 of the study. However, by week 54, such difference in dropout was no longer apparent: 59 (20.2%) patients on SB2 and 64 (21.8%) patients on EU-Remicade dropped out (Figure 3). Adverse events and withdrawal of consent were the major reasons for withdrawal from the study. Withdrawal from the study due to adverse events was slightly more common in the SB2 group (9.3% vs 7.2%).

Table 8: Reasons for Withdrawal through Week 30

	SB2		EU- Remicade		All	
	N	**(%)**	**N**	**(%)**	**N**	**(%)**
Randomized	291		293		584	
Completed Week 30	247	84.9	259	88.4	506	86.6
Withdrew before Week 30	44	15.1	34	11.6	78	13.4
Reason for Withdrawal						
Adverse event	21	7.2	10	3.4	31	5.3
Investigator discretion	1	0.3	3	1.0	4	0.7
Lack of efficacy	5	1.7	5	1.7	10	1.7
Protocol deviation	.	.	3	1.0	3	0.5
Subject lost to follow-up	.	.	1	0.3	1	0.2
Withdrew consent	17	5.8	12	4.1	29	5

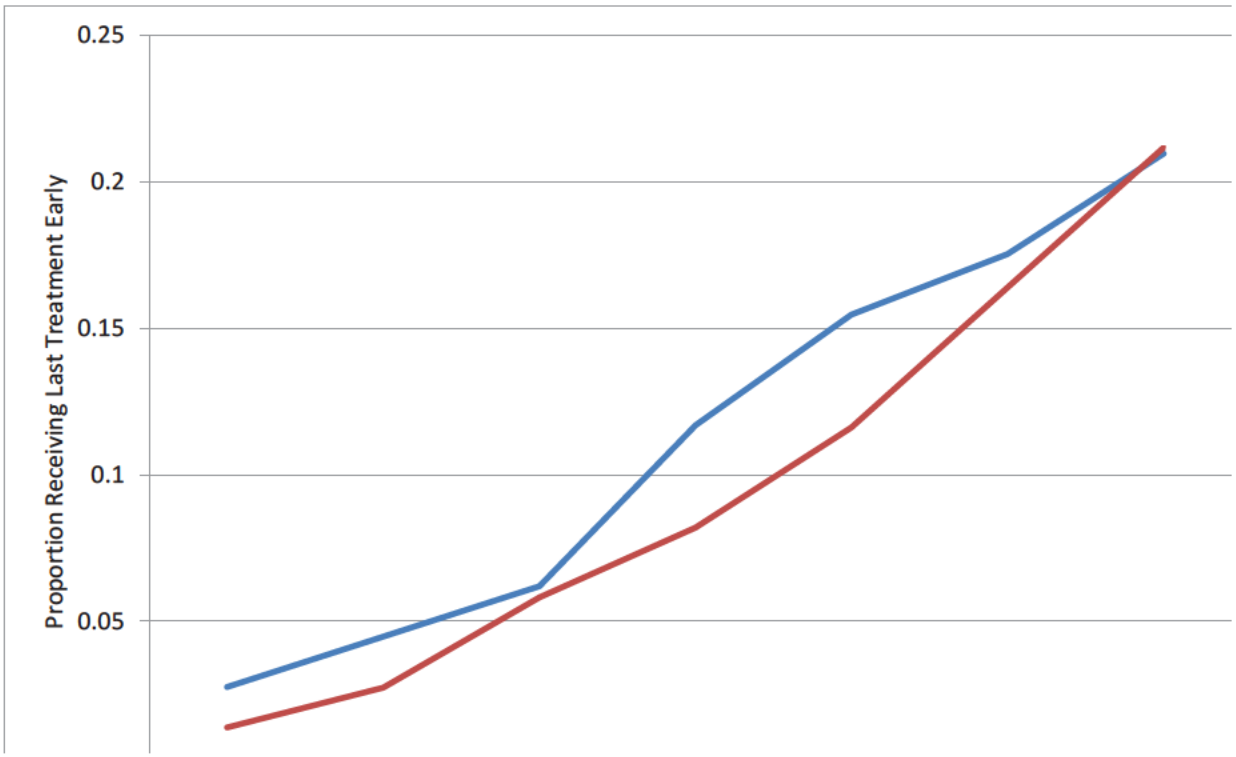
Source: Reviewer

Table 9: Reasons for Withdrawal through Week 54

	SB2		Remicade		All	
	N	**(%)**	**N**	**(%)**	**N**	**(%)**
Randomized	291		293		584	
Completed Week 54	228	78.4	225	76.8	453	77.6
Withdrew before Week 54	59	20.2	64	21.8	123	13.4
Reason for Withdrawal						
Adverse event	27	9.3	21	7.2	48	8.2
Investigator discretion	4	1.4	4	1.4	8	1.4
Lack of efficacy	5	1.7	6	2.0	11	1.9
Protocol deviation	.	.	1	0.3	1	0.2
Subject lost to follow-up	.	.	5	1.7	5	0.9
Pregnancy	.	.	1	0.3	1	0.2
Withdrew consent	23	7.9	26	8.9	49	8.4
Eastern Ukraine sites	4	1.4	4	1.4	8	1.4

Source: Reviewer

Figure 2: Patient Withdrawal over Time (Source: Reviewer)



3.5 Results and Conclusions

3.5.1 Primary Efficacy Analysis

The results from the primary efficacy analysis using the per-protocol set are given in Table 10. The proportion of patients who obtained an ACR20 response at Week 30 was found to be similar between the groups, with a slightly higher rate in the EU-Remicade group (66.0%) compared to the SB2 group (64.1%); the estimated absolute difference was -1.88% (90% CI: -8.91,+ 5.16; 95% CI: -10.26, +6.51). Both the 90% and 95% confidence intervals were well contained within the FDA-recommended similarity margin of [-12%, 12%]

Table 10: Primary analysis of ACR20 response rate at Week 30 (Per-protocol Set)

Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
SB2 (N=231)	148/231	(64.1%)	**−1.88%**	(-8.91, 5.16)	(-10.26, 6.51)
EU-Remicade (N=247)	163/247	(66.0%)			

Source: Reviewer

The analysis of ACR20 response in the full analysis set is given in Table 11 below. The estimated absolute difference was -2.95% (90% CI: -9.60, 3.70; 95% CI: -10.87, 4.97). Both the 90% and 95% confidence intervals were well contained within the FDA-recommended similarity margin of [-12%, 12%]. The lower CI bound of -9.60% also corresponds to the preservation of approximately 60% of conservative estimates of the effect of infliximab from historical trials.

Table 11: Analysis of ACR20 response rate at Week 30 (Full Analysis Set)

Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
SB2 (N=290)	161/290	(55.52%)	**-2.95%**	(-9.60, 3.70)	(-10.87, 4.97)
EU-Remicade (N=293)	173/293	(59.04%)			

Source: Reviewer

More than half of the non-responders were patients who completed the study and did not satisfy the ACR20 response criteria. The majority of the remaining non-responders were patients who withdrew from the study prior to Week 30. There were no large differences between the treatment arms in the distributions of reasons for non-responses. (Table 12)

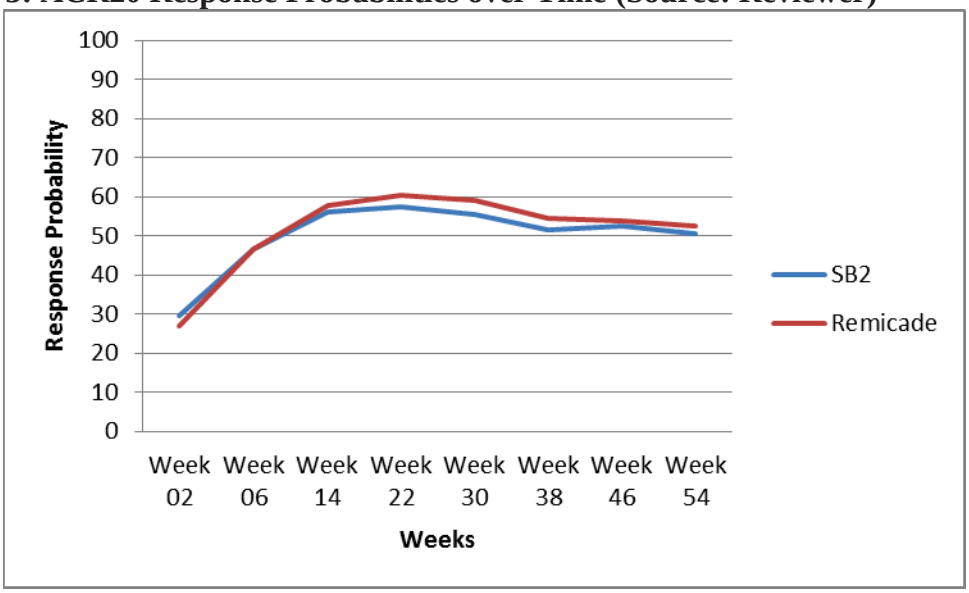
Table 12: Proportions of Non-Responders, and Distributions of Reasons for Non-Response, with Respect to Composite ACR20-Based Primary Endpoint at Week 30

	SB2		***Remicade***	
	N	**(%)**	**N**	**(%)**
Non-responder	129	44.48	120	40.96
ACR20 criteria not met	87	30.00	89	30.38
Withdraw from study	42	14.48	31	10.58
Adverse event	20	6.90	9	3.07
Investigator discretion	1	0.34	2	0.68
Lack of efficacy	5	1.72	5	1.71
Protocol deviation	0	0.00	3	1.02
Subject lost to follow-up	0	0.00	1	0.34
Withdrew consent	16	5.52	11	3.75

Source: Reviewer

The proportions of patients remaining in the study and achieving ACR20 responses over time during the study period were similar between the treatment arms (Figure 3).

Figure 3: ACR20 Response Probabilities over Time (Source: Reviewer)



Mean changes from baseline in the components of the ACR composite endpoint were also similar between the arms in all randomized patients who completed the study (Table 8), as well as in the per-protocol population (results not shown).

Table 13: Mean Changes from Baseline in the ACR Components at Week 30

	SB2		*Remicade*		*Difference*	*90% CI*
	N	**Mean**	**N**	**Mean**		
Swollen Joint count	253	-8.02	265	-7.96	0.0573	(-0.75, 0.87)
Tender Joint Count	253	-15.20	265	-14.33	0.8733	(-0.88, 2.63)
HAQ Score	253	-0.45	265	-0.53	-0.0748	(-0.16, 0.01)
Patient Pain	253	-21.90	264	-25.93	-4.0268	(-7.75, -0.30)
Patient Global	253	-23.80	265	-25.18	-1.3904	(-5.02, 2.24)
Physician Global	253	-32.71	265	-32.82	-0.1076	(-3.22, 3.00)
ESR	253	-15.39	267	-15.35	0.0353	(-3.05, 3.12)
CRP	252	-3.65	268	-5.08	-1.4276	(-4.41, 1.56)

Source: Reviewer

3.5.2 Secondary Efficacy Analysis

The comparative analyses of secondary endpoints also showed similar efficacy between the two treatment groups. Secondary endpoints in the study included ACR20 response at Week 54, ACR50 and ACR 70 at Weeks 30 and Week 54, ACR-N at Week 30 and Week 54, area under the curve of ACR-N up to Week 30, and disease activity score based on 28 joint counts (DAS 28 score) at Week 30 and Week 54, EULAR response at Week 30 and Week 54, AUC of the change

in DAS28 from baseline up to Week 30, major clinical responses, and modified total sharp score at Week 54.

3.5.2.1 Analysis of ACR20 response rate at Week 54

ACR20 response rate at Week 54 was found to be similar between the treatment groups. For patients in the PPS2 set, the adjusted difference rate was -3.07% with a 90% CI of (-10.56, 4.43) and for patients in the full analysis set, it was -1.15% with a 90% CI of (-7.88, 5.57).

Table 14 : Analysis of ACR20 response rate at Week 54

Dataset	*Treatment*	*n/N*	*%*	*Adjusted Difference Rate*	*90% CI*	*95% CI*
Per-protocol Set 2	SB2 (N=202)	132/202	(65.35%)	**-3.07%**	(-10.56, 4.43)	(-11.99, 5.86)
	EU-Remicade (N=208)	144/208	(69.23%)			
Full Analysis Set	SB2 (N=290)	147/290	(50.69%)	**- 1.15%**	(-7.88, 5.57)	(-9.16, 6.86)
	EU-Remicade (N=293)	154/293	(52.56%)			

Source: Reviewer

3.5.2.2 Additional Secondary Endpoints

In addition to the similar results obtained from the analysis of binary ACR20 response, the analysis of different continuous endpoints also showed similarity between the two groups. As continuous endpoints may be more sensitive to detect differences in treatment effects, such results are reassuring. For example, the analysis of the continuous endpoint ACR-N at Week 30 indicates that the treatment effects were similar between the two groups. From Table 15, the difference between the two treatment mean changes was -0.87 with a 90% confidence interval (-5.16, 3.40).

Table 15 : ACR-N at Week 30

TRT	MEAN	Difference Between Means	95% Confidence Limits	90% Confidence Limit
EU-Remicade	37.81	-0.87	(-5.98, 4.22)	(-5.16, 3.40)
SB2	36.63			

Source: Reviewer

Table 16 presents results from the analysis of radiographic progression via the change from baseline value of modified total sharp score (mTSS) at Week 54. In contrast to other endpoints measuring disease signs and symptoms, the mTSS is intended as a surrogate measure of disease progression. The result shows that the average score in both the groups are similar (Difference: -0.0011; CI: (-0.4798,0.4775)).

Table 16 : Change from baseline value of Modified total sharp score (mTSS) at Week 54

TRT	MEAN	Difference Between Means	95% Confidence Limits	90% Confidence Limit
EU-Remicade	0.4709	-0.0011	(-0.5718, 0.5696)	(-0.4798,0.4775)
SB2	0.4698			

Source: Reviewer

Results from other secondary analyses are given in the appendix and also show similar efficacy between the two treatment groups.

3.5.3 Transition Extension Period

At Week 54, subjects receiving EU-Remicade from the randomized, double-blind period of the SB2-G31-RA study were randomized again in a 1:1 ratio to either continue on EU-Remicade (Remicade/Remicade) or be transitioned to SB2 (Remicade/SB2) up to Week 70. ACR20 responses rates across various time points show comparable results between the different study arms.

Table 17: ACR 20 Responses Over Time in Transition Extension Period

ACR Response	*Time Point*	*SB2*		*EU-Remicade*			
				SB2		EU-Remicade	
		n/N	(%)	n/N	(%)	n/N	(%)
ACR20	Week 54	132/201	65.67	67/94	71.28	70/101	69.31
	Week 62	129/192	67.19	68/94	72.34	67/100	67.00
	Week 70	118/180	65.56	61/87	70.11	67/96	69.79
	Week 78	123/187	65.78	54/88	61.36	64/96	66.67

Source: Reviewer

3.5.4 Assay Sensitivity and the Constancy Assumption

In order to reliably evaluate whether there are clinically meaningful differences between two products, a comparative clinical study must have assay sensitivity, or the ability to detect meaningful differences between the products, if such differences exist. In addition, to reliably

evaluate whether the experimental treatment retains a certain proportion of the effect of the reference product versus placebo, the constancy assumption must be reasonable. This is the assumption that estimates of the effect of the reference product from historical, placebo-controlled trials are unbiased for the setting of the comparative clinical study. The absence of a placebo arm in an active-controlled study makes it difficult to determine whether evidence of similarity between the experimental and control arms implies that the two products were similarly effective or similarly ineffective. As discussed in the ICH E10 guidelines and in the literature, historical evidence of sensitivity to drug effects and appropriate trial conduct may be used to support the presence of assay sensitivity and a conclusion that the treatments are similarly effective.

Table 18 describes key characteristics of five historical randomized, double-blind, parallel-group, placebo-controlled clinical trials of infliximab in patients with active RA despite treatment with methotrexate, alongside key characteristics of SB2-G31-RA. Important aspects of the design of the historical studies, including key inclusion/exclusion criteria, permitted concomitant medications, and baseline disease severity, were largely similar if not identical across the six studies. One notable difference was the timing of the ACR20 assessment, which ranged from Week 14 to Week 30. However, the ATTRACT study demonstrated large treatment effects as early as Week 6 , and there was no apparent trend in effect size as a function of the timing of endpoint assessment across the historical studies. Estimated treatment effects with respect to ACR20 for the five historical trials were displayed earlier in Table 2. The estimated effects ranged from 18% to 38% on the absolute difference scale, with an overall estimated effect size of 28%. Thus, the information in Table 18 and Table 2 indicates that (1) the design of the five historical placebo-controlled clinical trials were largely similar to that of comparative clinical Study SB2-G31-RA; and (2) there were relatively large and consistent treatment effects across the five historical studies.

This evidence of historical sensitivity to effects of infliximab in similarly designed clinical trials provides some support for a conclusion that SB2-G31-RA had assay sensitivity. It is also important that a study designed to evaluate similarity has quality conduct, because conduct issues such as violations in eligibility criteria, poor adherence, cross-over between arms, or missing data tend to bias results toward the alternative hypothesis of equivalence. In Study SB2-G31-RA, 13.4% of patients discontinued treatment and the study prior to Week 30. This proportion is slightly greater than the historical discontinuation rates, which ranged from 5% to 11%). This is potentially concerning because adherence at a level lower than that which is best achievable in real clinical practice will tend to bias comparisons between treatments toward equivalence and therefore decrease the sensitivity of the comparative study. Decreased adherence on the active control may also result in decreased efficacy and therefore violations in the constancy assumption. In addition, because patients who discontinued treatment were not retained for safety and efficacy assessments through the double-blind period, this led to substantial missing

data in important analyses. (b) (4) during an EMA inspection. However, the FDA Office of Scientific Investigations inspected the sponsor's U.S. agent for this application, Quintiles, and two foreign clinical sites, and did not identify any such issues with study conduct.

We also examined whether the within-group responses in the comparative clinical study were similar to those observed in previous placebo-controlled trials. The 59% ACR20 response rate on EU-Remicade in Study SB2-G31-RA is in line with the historical rates, which ranged from 50% to 76%.

In summary, there are some concerns about study conduct, including inspection issues identified at one clinical site and the high rates of treatment discontinuation and missing data in Study SB2-G31-RA, an issue that will be discussed in greater detail in 3.5.5. However, the design, conduct, and within-group responses rates of Study SB2-G31-RA were largely similar to those characteristics in five historical clinical trials that demonstrated relatively large and consistent treatment effects of infliximab over placebo. Therefore, the totality of available information largely supports the sufficiency of the assay sensitivity of Study SB2-G31-RA, in addition to the constancy assumption.

Table 18: Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials[1] of Infliximab in RA and Study SB2-G31-RA

	Maini, (1999)	Westhovens, (2006)	Schiff, (2008)	Zhang, (2006)	Abe, (2006)	**Study SB2-G31-RA**
Selected Inclusion/ Exclusion Criteria	≥ 6 SJ, ≥6 TJ, 2 of: morning stiffness ≥ 45 min, ESR >28 mm/h, CRP >2 mg/dL	≥ 6 SJ, ≥6 TJ	Disease duration ≥ 1 year, ≥ 10 SJ, ≥ 12 TJ, CRP ≥1 mg/dL	≥3 SJ, ≥8 TJ, 2 of: morning stiffness ≥ 45 min, ESR >28 mm/h, CRP >1.5 mg/dL	≥6 SJ, ≥6 TJ, 2 of: morning stiffness ≥ 45 min, ESR >28 mm/h, CRP >2 mg/dL	≥6 SJ, ≥6 TJ, ESR >28 mm/h, CRP >1.0 mg/dL
Anti-TNF experience allowed?	No	No	No	Yes	No	No
Concomitant DMARDs	Stable MTX	Stable MTX + additional DMARDs allowed	Stable MTX	Stable MTX + additional DMARDs allowed	Stable MTX (Low Dose)	Stable MTX
Region / Country	NA, EU	NA, EU, AU, SA	NA, EU, AU, AF, SA	China	Japan	EU, AS, NA
Baseline Characteristics of Study Population[2]	SJ: 19; TJ: 32; Disease Duration: 8 yrs; HAQ: 1.8	SJ: 15; TJ: 22 Disease Duration: 8 yrs; HAQ: 1.5	SJ: 20; TJ: 2; Disease Duration: 7 yrs; HAQ: 1.7	Disease Duration: 7 yrs	SJ: 15; TJ: 19; Disease Duration: 9 yrs	SJ: 14.8; TJ: 24; Disease Duration: 6.4 yrs; HAQ: 1.5
Time of ACR20 Evaluation	Week 30	Week 22	Week 28	Week 18	Week 14	Week 30
ACR20 Response on Infliximab	50%	55%	59%	76%	61%	59%
Withdrawal on Infliximab	11%	7%	8%	10%	5%	13.4%

Source: Reviewer

Abbreviations: SJ=swollen joint count; TJ=tender joint count; DMARD=disease-modifying anti-rheumatic drug; HAQ = Health Assessment Questionnaire; MTX = Methotrexate; NA=North America; EU=Europe; AS=Asia

[1] Based on best attempts to identify/estimate characteristics from literature review

[2] Means or medians, depending on what was reported in publication

3.5.5 Potential Effect of Missing Data

This section addresses the effect of missing data on the reliability of the comparative efficacy results in the study. As we noted in Table 8, up to Week 30, 78 (13.4%) patients had withdrawn from the study: 44 patients (15.4%) from the SB2 treatment group and 34 patients (11.6%) from the EU-Remicade treatment group. These patients were excluded in the primary analysis in the per-protocol set, such that results depend on the unverifiable assumption that the non-randomized subsets of protocol adherers on the two arms were comparable. Furthermore, in the key supportive analysis in the full analysis set, patients who dropped out were considered non-responders, such that the primary endpoint was a composite measure of treatment success defined by adherence to the treatment through Week 30 and achieving an ACR20 response at Week 30. Comparing treatments with respect to this composite measure of treatment success may confound differences between treatments in efficacy with differences in tolerability. The composite measure could fail to identify clinically meaningful differences in efficacy, for example, if the proposed biosimilar was better tolerated than the reference product but had lesser efficacy in the subset of patients who adhere. Therefore, it is important to evaluate differences in the components of the composite primary endpoint. This includes an evaluation of ACR20 at Week 30 in all randomized patients regardless of adherence (an evaluation of the de facto or intention-to-treat estimand), in addition to de facto evaluations of the components of ACR20. However, such evaluations are subject to some missing data (because patients who discontinued treatment were not followed up for assessment) and rely on strong and unverifiable assumptions, such as the assumption that outcomes in patients who withdrew early are missing at random. Therefore, we requested from the applicant, and conducted our own, tipping point analyses to explore the sensitivity of results to violations in assumptions about the missing data (i.e., to various missing-not-at-random assumptions). Moreover, the 2010 National Research Council (NRC) Report *The Prevention and Treatment of Missing Data in Clinical Trials* recommends that "examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting."

Table 19 displays results from our tipping point analyses: estimated de facto differences between SB2 and EU-Remicade in the ACR20 response at Week 30, with varying assumptions about the differences on each treatment arm between outcomes in patients who withdrew from the study early and outcomes in patients who completed the study. In order for the 90% CI to fail to rule out a 12% absolute loss in the probability of ACR20 response, the response among SB2 dropouts would need to be around 70 percentage points lower than the response in SB2 completers, while the response among EU-Remicade dropouts would need to be worse by about 35 percentage points than the response among EU-Remicade completers. As a point of reference, the response probabilities among completers on SB2 and EU-Remicade were 64% and 66%, respectively. Given the similar proportions of patients and distributions of reasons for early withdrawal on the two treatment arms, in addition to the similar baseline characteristics between dropouts on the

two arms, an assumption of such large differences between the outcomes in dropouts on the two arms seems implausible. Therefore, these tipping point sensitivity analyses largely support the findings of the key efficacy analyses in Study SB2-G31-RA. The applicant reported similar conclusions based on its tipping point analyses (results not shown).

Table 19 : Tipping Point Analysis

		Shift for EU-Remicade[1]				
		-0.7	**-0.525**	**-0.35**	**-0.175**	**0**
Shift for SB2[2]	**-0.7**	-0.04 (-0.1, 0.03)	-0.06 (-0.11, 0)	-0.08 (-0.13, -0.02)	-0.1 (-0.15, -0.04)	-0.12 (-0.17, -0.06)
	-0.525	-0.01 (-0.07, 0.05)	-0.03 (-0.09, 0.03)	-0.05 (-0.11, 0.01)	-0.07 (-0.13, -0.01)	-0.09 (-0.15, -0.03)
	-0.35	0.02 (-0.04, 0.08)	0 (-0.06, 0.05)	-0.02 (-0.08, 0.03)	-0.04 (-0.1, 0.01)	-0.06 (-0.12, -0.01)
	-0.175	0.04 (-0.01, 0.1)	0.02 (-0.03, 0.08)	0 (-0.05, 0.06)	-0.02 (-0.07, 0.04)	-0.04 (-0.09, 0.02)
	0	0.07 (0.01, 0.13)	0.05 (-0.01, 0.11)	0.03 (-0.02, 0.09)	0.01 (-0.04, 0.06)	-0.01 (-0.06, 0.04)

Source: Reviewer

Cell contents are estimated difference (90% confidence interval). Shaded cells represent assumptions under which the confidence interval fails to rule out the ±12% margin.

[1] Assumed difference in Week 30 ACR20 response between completers and dropouts on EU-Remicade. Response in EU-Remicade completers was 0.66.

[2] Assumed difference in Week 30 ACR20 response between completers and dropouts on SB2 Response in SB2 completers was 0.64.

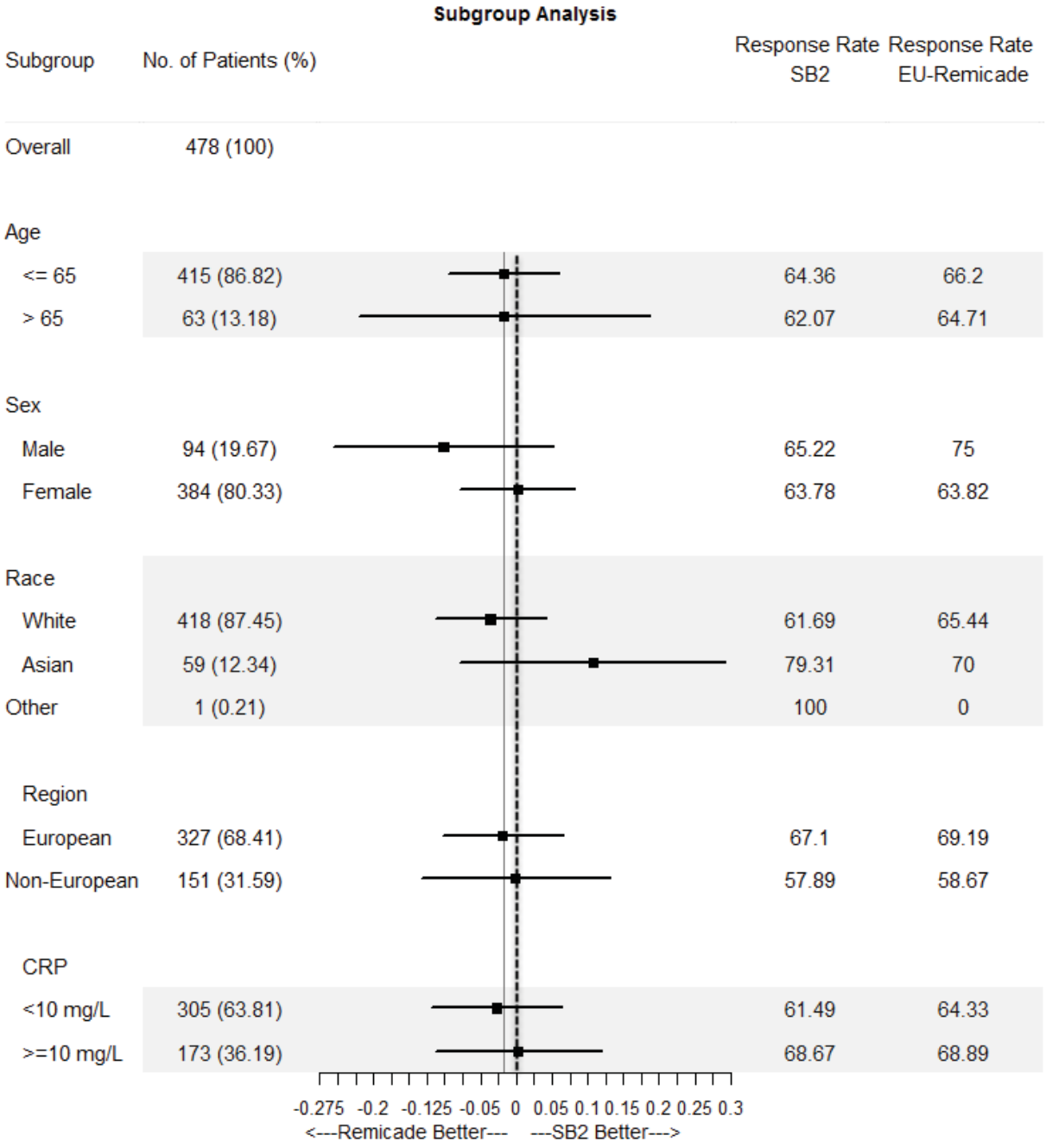
3.6 Evaluation of Safety

Dr. Juwaria Waheed, the Medical Reviewer, conducted the complete safety evaluation. The details of the safety evaluation can be found on Dr. Waheed's report.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The subgroup analyses compared efficacy results across treatment arms within different subgroups defined by sex, region, age, race and CRP values. Figure 4 given below shows estimated differences stratified by subgroups were largely centered around similarity, and there were no striking trends across the two studies. The number of patients other than White or Asian in the study was too small to calculate sufficiently reliable estimated differences to report. There were no U.S. sites in the study so subgroup analyses in the United States are not possible.

Figure 4: Estimated Differences Between SB2 and EU-Remicade in the Probability of Remaining in the Study and Achieving an ACR20 Response at Week 30, Stratified by Selected Subgroups, in Study SB2-G31-RA. Gray Vertical Line Represents Estimated Difference in Overall Population, and Dashed Vertical Line Represents No Difference (Source: Reviewer)



5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

During this statistical review, we identified the following important issues:

- Margin selection and evidence of similarity

The determination of a similarity margin is a critical aspect of the design of a comparative clinical study because it determines the null hypothesis being tested in the primary analysis, i.e., the differences in efficacy that need to be ruled out at an acceptable significance level. The applicant prespecified a primary analysis comparing a 95% CI for the difference in Week 30 ACR20 responses to a similarity margin of ±15% and later performed an additional analysis after database lock to compare a 90% CI to a similarity margin to ±12% in response to feedback from FDA. The lack of a priori agreement between the applicant and FDA on a similarity margin is not of concern in this case because the primary analysis successfully ruled out the ±12% margin recommended by FDA. We selected a margin of ±12% based on meta-analyses of historical effects of infliximab and discussions with clinicians aimed at weighing the clinical importance of different losses in effect against the feasibility of different study sizes. Results from the primary analysis in the per-protocol set (90% CI: -8.91%, 5.16%) and a supportive analysis in the full analysis set (90% CI: -9.60, 3.70;) were well contained within the FDA-recommended similarity margin of [-12%, 12%]. In addition, there were similar improvements from baseline in the components of the composite primary endpoint, as well as additional important secondary endpoints, on the two treatment arms. Therefore, the totality of the evidence from the comparative clinical studies supports a demonstration of no clinically meaningful differences between SB2 and US-Remicade.

- Potential effect of missing data on the reliability of efficacy results

This issue was discussed in detail in 3.5.5. Up to Week 30, 78 (13.4%) patients had withdrawn from the study: 44 patients (15.1%) from the SB2 treatment group and 34 patients (11.6%) from the EU-Remicade treatment group. This led to substantial missing data in important analyses, such as the evaluations of ACR20 and DAS28 at Week 30 in all randomized patients regardless of adherence. Because such evaluations rely on strong and unverifiable assumptions, such as the assumption that outcomes in patients who withdraw early are missing at random, we conducted tipping point analyses to explore the sensitivity of results to violations in this assumption. Confidence intervals for the differences between SB2 and EU-Remicade failed to rule out concerning losses in efficacy only under the assumption that patients who dropped out on SB2 had much worse outcomes than dropouts on EU-Remicade. Given the similar proportions of patients and distributions of reasons for early withdrawal on the two treatment arms, in addition to the similar baseline characteristics

between dropouts on the two arms, an assumption of such large differences between the outcomes in dropouts on the two treatments seems implausible. Therefore, these tipping point sensitivity analyses largely support the findings of the key efficacy analyses in Study SB2-G31-RA.

- Assay sensitivity and the constancy assumption

This issue was discussed in detail in 3.5.4. It is critical that a comparative clinical study has assay sensitivity, or the ability to detect meaningful differences between products, if such differences exist. In addition, the constancy assumption should be reasonable. This is the assumption that estimates of the reference product effect from historical, placebo-controlled trials are unbiased for the setting of the comparative study. Our evaluation of the literature indicated historical sensitivity to effects of infliximab over placebo in five clinical trials with similar designs to that of comparative clinical Study SB2-G31-RA. Within-group responses in the study were also similar to those of historical trials. It is also important that a study designed to evaluate similarity has appropriate conduct because conduct issues tend to bias results toward the alternative hypothesis of equivalence. Despite some concerns about inspection issues at a clinical site and the high rates of treatment discontinuation and missing data, the totality of available information largely supports the sufficiency of the assay sensitivity of Study SB2-G31-RA, in addition to the constancy assumption.

5.2 Collective Evidence

The collective evidence from the comparative clinical study in rheumatoid arthritis supports a conclusion of no clinically meaningful differences between SB2 and US-Remicade. The adjusted treatment difference in ACR20 response rates between the SB2 and EU-Remicade treatment groups in the analysis in the full analysis set was −2.95% and the 90% CI of the adjusted treatment difference was (-9.60, 3.70), which was contained within the similarity margin of [−12%, +12%] recommended by FDA. ACR20, ACR50, and ACR70 responses over time, mean changes from baseline in the components of the ACR composite endpoint and the disease activity score (DAS28), and other secondary efficacy endpoint results, were also similar between the treatment arms. There was substantial missing data in important analyses, but tipping point analyses largely support the findings of key efficacy results in Study SB2-G31-RA. In addition, the totality of available information largely supports the assay sensitivity of Study SB2-G31-RA, in addition to the constancy assumption.

6 APPENDIX: Additional Tables and Figures

Table 20 : Analysis of ACR50 response rate at Week 30

Study data	Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
Per-protocol Set 1	SB2 (N=231)	82/231	(35.5%)	**-2.13%**	(-9.31, 5.06)	(-10.69, 6.43)
	EU-Remicade (N=247)	94/247	(38.06%)			
Full Analysis Set	SB2 (N=290)	89/290	(30.69%)	**- 2.53%**	(-8.86, 3.79)	(-10.07, 5.00)
	EU-Remicade (N=293)	99/293	(33.79%)			

Source: Reviewer

Table 21 : Analysis of ACR70 response rate at Week 30

Study data	Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
Per-protocol Set 1	SB2 (N=231)	42/231	(18.18%)	**-.25%**	(-6.13, 5.62)	(-7.26, 6.75)
	EU-Remicade (N=247)	47/247	(19.03%)			
Full Analysis Set	SB2 (N=290)	45/290	(15.52%)	**- 1.08%**	(-6.10, 3.94)	(-7.06, 4.91)
	EU-Remicade (N=293)	50/293	(17.06%)			

Source: Reviewer

Table 22 : Analysis of ACR50 response rate at Week 54

Study data	Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
Per-protocol Set 2	SB2 (N=202)	84/202	(41.58%)	**3.43%**	(-4.26, 11.12%)	(-5.74, 12.60)
	EU-Remicade (N=208)	81/208	(38.94%)			
Full Analysis Set	SB2 (N=290)	93/290	(32.06%)	**3.07%**	(-3.08, 9.22%)	(-4.26, 10.40)
	EU-Remicade (N=293)	87/293	(29.69%)			

Source: Reviewer

Table 23 : Analysis of ACR70 response rate at Week 54

Study data	Treatment	n/N	%	Adjusted Difference Rate	90% CI	95% CI
Per-protocol Set 2	SB2 (N=202)	45/202	(15.52%)	**-1.07%**	(-7.82, 5.69)	(-9.12, 6.98%)
	EU-Remicade (N=208)	50/208	(17.06%)			
Full Analysis Set	SB2 (N=290)	53/290	(18.28%)	**1.10%**	(-4.08, 6.28)	(-5.08, 7.28%)
	EU-Remicade (N=293)	52/293	(17.75%)			

Source: Reviewer

Table 24 : ACR-N at Week 30

TRT	MEAN	Difference Between Means	95% Confidence Limits	90% Confidence Limit
EU-Remicade	37.81	-0.8793	(-5.98, 4.22)	(-5.16, 3.40)
SB2	36.63			

Source: Reviewer

Table 25 : ACR-N at Week 54

TRT	MEAN	Difference Between Means	95% Confidence Limit	90% Confidence Limit
EU-Remicade	39.77	-0.5674	(-6.13, 5.00)	(-5.24, 4.10)
SB2	38.82			

Source: Reviewer

Figure 5: ACR-N Response (Source: Reviewer)

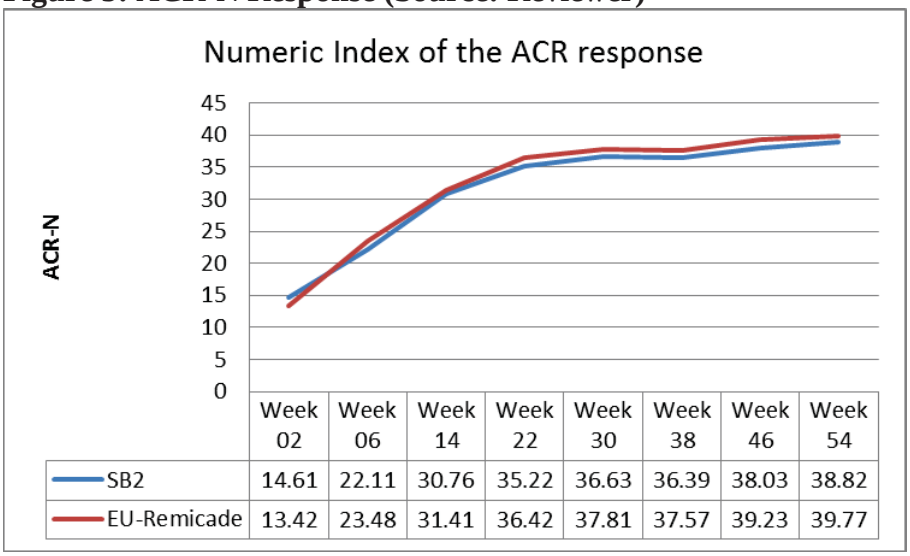


Table 26 : ACR-AUC up to Week 30

TRT	MEAN	Difference Between Means	95% Confidence Limit	90% Confidence Limit
EU-Remicade	6237.14	-105.6737	(-862.37, 651.02)	(-740.36 529.01)
SB2	6131.46			

Source: Reviewer

Figure 6 : ACR-AUC up to Week 30 (Source: Reviewer)

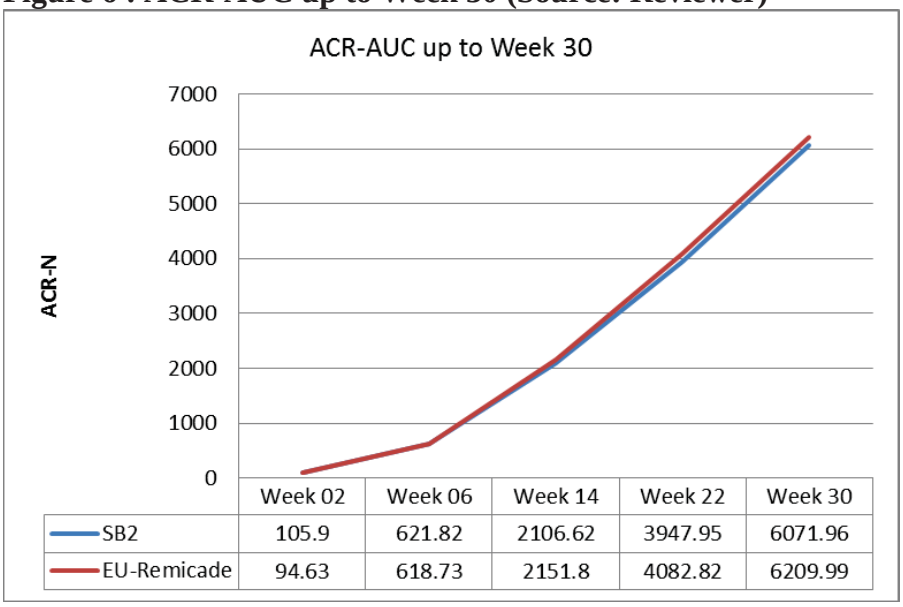
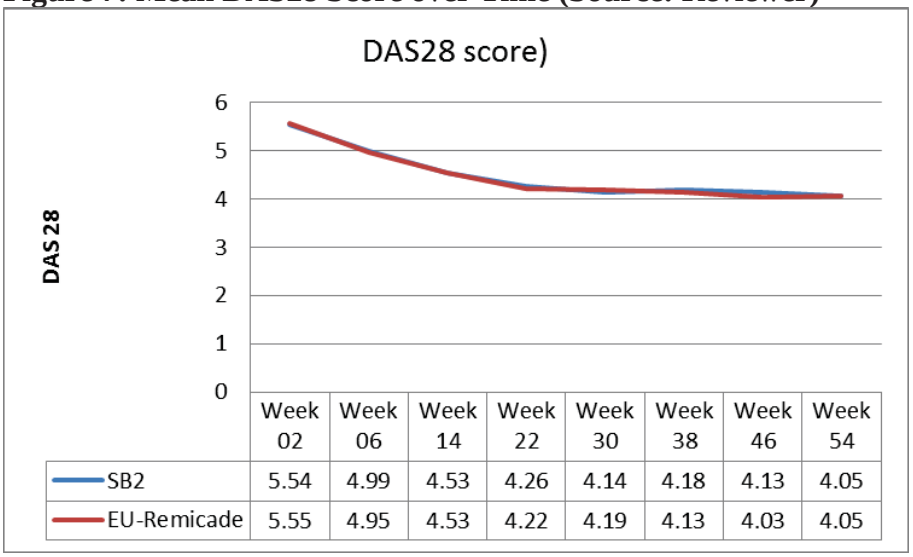


Table 27: Change from baseline in DAS28 at Week 30

TRT	MEAN	Difference Between Means	95% Confidence Limit	90% Confidence Limit
EU-Remicade	-2.3861	-0.0110	(-0.2128, 0.1907)	(-0.2517, 0.2295)
SB2	-2.3972			

Source: Reviewer

Figure 7: Mean DAS28 Score over Time (Source: Reviewer)



7 Bibliography

- Abe, T. T. (2006). A multicenter, double-blind, randomized, placebo controlled trial of infliximab combined with low dose methotrexate in Japanese patients with rheumatoid arthritis. *The Journal of rheumatology*, 37-44.
- FDA. (2016). *Non-Inferiority Clinical Trials to Establish Effectiveness- Guidance for Industry.*
- FDA. (April 2015). *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product : Guidance for Industry.*
- Koch, G. G. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. . *Statistics in medicine*, 1863-1892.
- Maini, R. S. (1999). Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet*, 1932-1939.
- Schiff, M. K. (2008). Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. *Annals of the rheumatic diseases*, 1096-1103.
- Westhovens, R. Y. (2006). The safety of infliximab, combined with background treatments, among patients with rheumatoid arthritis and various comorbidities: a large, randomized, placebo-controlled trial. *Arthritis & Rheumatism*, 1075-1086.
- Zhang, F. C. (2006). Infliximab versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a preliminary study from China. *APLAR Journal of Rheumatology*, 127-130.

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

/s/

GINTO J POTTACKAL
12/15/2016

GREGORY P LEVIN
12/15/2016

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	761054
SERIAL NO.:	0000
DATE RECEIVED BY THE CENTER:	March 21, 2016
DRUG NAME:	SB2 (proposed biosimilar to Remicade, Samsung Bioepis)
DOSAGE FORM:	Lyophilized powder,100 mg/vial
INDICATIONS:	Crohn's Disease (CD), Pediatric Crohn's Disease (Pediatric CD), Ulcerative Colitis (UC), Pediatric Ulcerative Colitis (Pediatric UC), Rheumatoid Arthritis (RA) in combination with methotrexate, Ankylosing Spondylitis (AS), Psoriatic Arthritis (PsA) and Plaque Psoriasis (PsO).
APPLICANT:	Samsung Bioepis Inc.
REVIEW FINISHED:	December 12, 2016
NAME OF STATISTICAL REVIEWER:	Yu-Ting Weng
NAME OF PROJECT MANAGER:	Christine Ford

_____

Reviewer: Yu-Ting Weng, Mathematical Statistician, CDER/OTS/OB/DB VI
Secondary reviewer: Meiyu Shen, Ph.D., Team Leader, CDER/OTS/OB/DB VI

Concur: _____
Yi Tsong, Ph.D., Division Director, DBVI

Distribution: CDER/OTS/OB/DB VI/Yi Tsong
CDER/OTS/OB/DB VI/Meiyu Shen
CDER/OTS/OB/Lillian Patrician
CDER/OBP/Xianghong (Emily) Jing
CDER/OBP/Christopher Downey
CDER/OMPT/CDER/OND/ODEII/DPARP/Christine Ford
CDER/TBBS

TABLE OF CONTENTS

List of Tables

1 EXECUTIVE SUMMARY AND RECOMMENDATION

The CMC statistics reviewer in the Office of Biostatistics analyzed the comparative results of two critical quality attributes (QAs): TNF-α neutralization assay and TNF-α binding assay, which were recommended for equivalence testing analysis by the Office of Biotechnology Products. Tier 1 statistical equivalence testing was conducted using equivalence margins of $\pm 1.5\sigma_R$, where σ_R represents US-licensed reference product variability or the comparator variability. 10 batches of SB2 (test product) combined 6 batches of Drug Product (DP) and 4 batches of Drug Substance (DS) and 46 batches of US-licensed Remicade (reference product), and 40 batches of EU-approved Remicade were used for equivalence testing of TNF-α neutralization assay (potency). The results are summarized in Table 1.

Table 1. Results of equivalence testing for TNF-α neutralization assay (potency)

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(10, 46)	-3.76	(-7.10, -0.44)	(-9.33, 9.33)	Yes
SB2 vs. EU	(10, 40)	-3.35	(-6.92, 0.22)	(-10.36, 10.36)	Yes
EU vs. US	(40, 46)	-0.41	(-2.79,1.96)	(-9.33, 9.33)	Yes

*The 90% confidence interval (CI) is adjusted by the sample size imbalance.

10 batches combined 6 batches of DP and 4 batches of DS of SB2, 41 batches of US-licensed Remicade, and 37 batches of EU-approved Remicade were included in the TNF-α binding assay dataset for the statistical equivalence testing. The results are shown in Table 2.

Table 2. Results of equivalence testing for TNF-α binding assay

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(10, 41)	-2.11	(-4.49, 0.26)	(-5.90, 5.90)	Yes
SB2 vs. EU	(10, 37)	-2.40	(-5.05, 0.25)	(-7.21, 7.21)	Yes
EU vs. US	(37, 41)	0.29	(-1.38,1.96)	(-5.90, 5.90)	Yes

*The 90% confidence interval (CI) is adjusted by the sample size imbalance.

As shown in Tables 1 and 2, the results from the statistical equivalence testing of TNF-α neutralization assay (potency) and TNF-α binding assay demonstrate that the proposed biosimilar SB2 is highly similar to US-licensed Remicade. In addition, the results support the analytical bridge between US-licensed Remicade and EU-approved Remicade.

2 INTRODUCTION

On March 21, 2016, the applicant (Samsung Bioepis) submitted to the US Food and Drug Administration (FDA) a 351(k) BLA which included an analytical similarity assessment of comparing SB2 and US-licensed Remicade.

On May 13, 2016, the Agency requested the sponsor to provide more data for all Tier 1 QAs.

Question 1. The applicant's analytical similarity exercise included five independent DP lots. As the Agency noted in the meeting minutes for the BPD Type 2 and Type 4 meetings held July 20, 2015 and December 14, 2015, respectively, data from only five lots may not be sufficient for the analytical similarity assessment. The Agency notes that five intended-commercial DS lots have been produced that are not included in the analytical similarity assessment. It is unclear whether DP lots have been produced from these additional DS lots. To support the analytical similarity assessment, provide data for all Tier 1 (equivalence testing) analytical tests for these five DS lots or their subsequently produced DP lots. If feasible to obtain, DP data will provide the strongest evidence to support analytical similarity to the US-licensed reference product. DS data may be acceptable for attributes that do not change significantly between DS lots and their resulting DP lots.

Question 2. For Tier 1 QAs (TNF-α neutralization assay and TNF-α binding assay), please provide the testing results from each block (each block has one relative potency) as the relative potency is determined as an average (geometric mean) from 3 to 4 blocks of data. For example, for batch A, the individual relative potency values from the 4 blocks are 96%, 101%, 102%, 98%, then you calculate the relative potency for this batch as (96% × 101% × 102% × 98%)1/4. Those individual block values, 96%, 101%, 102%, 98%, are the data points we are requesting.

On August 5, 2016, the applicant provided the following data:

- All Tier 1 QAs' testing results from each block for 4 intended-commercial DS, 1 intended-commercial DP, and 5 independent DP SB2 lots.
- All Tier 1 QAs' testing results from each block for all US-licensed Remicade and EU-approved Remicade lots.

The applicant characterized multiple batches of US-licensed Remicade and EU-approved Remicade using a comprehensive set of analytical methods during the SB2 development. In addition, the applicant recalculated the 90% Confidence Intervals for all Tier 1 QAs based on the Agency's recommended sample size imbalanced adjusted approach.

The Agency carefully evaluated data for the TNF-α neutralization assay and TNF-α binding assay provided in the initial BLA submission. Samsung Bioepis' statistical equivalence testing (Tier 1 approach) is provided in Section 4, and our independent statistical equivalence testing analyses are present in Section 5.

3 DATA ANALYZED

Samsung Bioepis submitted the analytical data on August 5, 2016. The TNF-α neutralization assay data of 46 US-licensed Remicade lots, 40 EU-approved Remicade lots, and 10 SB2 lots are summarized in Table 3. The TNF-α binding assay data of 41 US-licensed Remicade lots, 37 EU-approved Remicade lots, and 10 SB2 lots are also summarized in Table 3.

Table 3. Number of batches from each product

Product	Number of batches	
	TNF-α neutralization assay (potency)	TNF-α binding assay
US-licensed Remicade	46	41
SB2	10	10
EU-approved Remicade	40	37

4 APPLICANT'S STATISTICAL EQUIVALENCE TESTING

In this submission, Samsung Bioepis conducted Tier 1 statistical equivalence testing with the margin defined as $1.5\hat{\sigma}_R$ for TNF-α neutralization assay (potency) and TNF-α binding assay. To demonstrate statistical equivalence for TNF-α neutralization bioassay (potency) and TNF-α binding assay in this context, the entire two-sided CI must fall within$(-1.5\hat{\sigma}_R, 1.5\hat{\sigma}_R)$. Samsung Bioepis applied the Agency's recommended sample size imbalanced adjusted CI approach to calculate the two-sided CI. In addition, Satterthwaite approximation was applied for obtaining the degree of freedom (DF) of the sample size imbalanced adjusted CI because there is no assumption of equal variance between the test and reference products. However, the DF using in Satterthwaite method is incorrect and the correct version is provided in the following section. After the communication, Samsung Bioepis recalculated the 90% CIs for all Tier 1 QAs using the sample size imbalanced adjusted approach with the correct DF in the amendment on August 5, 2016.

5 FDA STATISTICAL ANALYSES

To evaluate analytical similarity, the Agency recommended Samsung Bioepis to apply a tiered approach in the Agency's responses to IND meetings with Samsung Bioepis. That is, product QAs amendable to statistical evaluation are assigned to three tiers based on their criticality. The quality attributes with potential highest risk in product quality, efficiency, safety and PK/PD are generally assigned to Tier 1, in which analytical similarity is assessed by statistical equivalence test. QAs with lower impact are generally assigned to Tier 2 and their analytical similarity is evaluated by Quality Range approach. That is, a high percentage of the biosimilar data should be covered by $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is the sample mean, $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots, and the multiplier X typically ranges from 2 to 4. The QAs with the lowest risk are generally assigned to Tier 3 and their analytical similarity is evaluated by side-by-side comparison using graphic display.

This review focuses on the equivalence test in Tier 1.

5.1 Statistical method

Let μ_T and μ_R be respectively the population mean of the QA for the test product and the population mean of the QA for the reference product. Let σ_R be the standard deviation of the QA

of interest for the reference product. In order to conclude the equivalence in the QA of interest between the test product and the reference product, we aim to reject the null hypothesis of the following null and alternative hypotheses:

$$H_0 : \mu_T - \mu_R \le \theta_1 \text{ or } \mu_T - \mu_R \ge \theta_2$$
$$H_1 : \theta_1 < \mu_T - \mu_R < \theta_2$$

where $\theta_1 = -1.5\sigma_R$, $\theta_2 = 1.5\sigma_R$, θ_1 and θ_2 are equivalence margins.

We reject H_0 if 90% confidence interval for the mean difference in the QA of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. In other words, we conclude that the equivalence in the QA of interest between the test product and the reference product if 90% confidence interval for the mean difference in the QA of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. This specific equivalence margin was set as 1.5 times the standard deviation of the quality attribute for the reference product to ensure an adequate power for the case in which a small but sufficient number of lots are available for testing. For example, the probability of rejecting H_0 in the above two one-sided tests procedure with the equivalence margin being $\pm(-1.5\sigma_R, 1.5\sigma_R)$ is 87% if the true mean difference is $0.125\sigma_R$ for a sample size of 10 test product lots and 10 reference product lots. First, we estimate σ_R by the sample variability of the reference product (or by the sample variability of EU-approved Remicade in the comparison between SB2 and EU-approved Remicade), and then θ_1 and θ_2 are treated as a constant, but not a random variable in the statistical analysis.

Let X_{Tj} be the observed value of the QA of interest for Batch *j* of the test product (the proposed biosimilar product) and X_{Rj} be the observed value of the QA of interest for Batch *j* of the reference product. Since the two products are manufactured by two manufacturers, two products are independent. $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij} \Big/ n_i$, and $S_i^2 = \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i\right) \Big/ (n_i - 1)$, where n_i is the number of lots in the i^{th} product, $i = T, R$.

Under the unequal variance of the test product and the reference product, the (1-2α)*100% CI of the mean difference in the QA of interest can be calculated as:

$$\left(\overline{X}_T - \overline{X}_R - t_\alpha(\nu)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R}}, \overline{X}_T - \overline{X}_R + t_\alpha(\nu)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R}} \right). \quad (1)$$

where $t_\alpha(\nu)$ is the 1-α quantile and ν is the degrees of freedom calculated by Satterthwaite's approximation.

If $n_R > 1.5n_T$, the (1-2α)*100% sample size imbalanced adjusted CI of the mean difference in the QA of interest can be calculated as:

$$\left(\bar{X}_T - \bar{X}_R - t_\alpha(\nu^*)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R^*}},\ \bar{X}_T - \bar{X}_R + t_\alpha(\nu^*)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R^*}} \right). \quad (2)$$

where $n_R^* = \min(n_R, 1.5n_T)$ and $\nu^* = \dfrac{\left(\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R^*}\right)^2}{\frac{1}{n_T - 1}\left(\frac{S_T^2}{n_T}\right)^2 + \frac{1}{n_R - 1}\left(\frac{S_R^2}{n_R^*}\right)^2}$.

If $n_T > 1.5n_R$, we can apply a similar approach as above with $n_T^* = \min(1.5 \times n_R, n_T)$ for the CI calculation. In the following analyses, we use α=0.05.

5.2 FDA statistical equivalence testing for TNF-α neutralization assay

The TNF-α neutralization assay data points of SB2, US-licensed Remicade, and EU-approved Remicade are displayed in Figure 1. There appears a small mean difference among the three products. The variability of SB2 is smallest among three products.

10 batches of SB2, 46 batches of US-licensed Remicade, and 40 batches of EU-approved Remicade are included for the statistical equivalence testing for the TNF-α neutralization assay. Descriptive statistics for the TNF-α neutralization assay data are listed in Table 4.

Figure 1. Scatter plot of TNF-α neutralization assay for US-licensed Remicade, SB2, and EU-approved Remicade

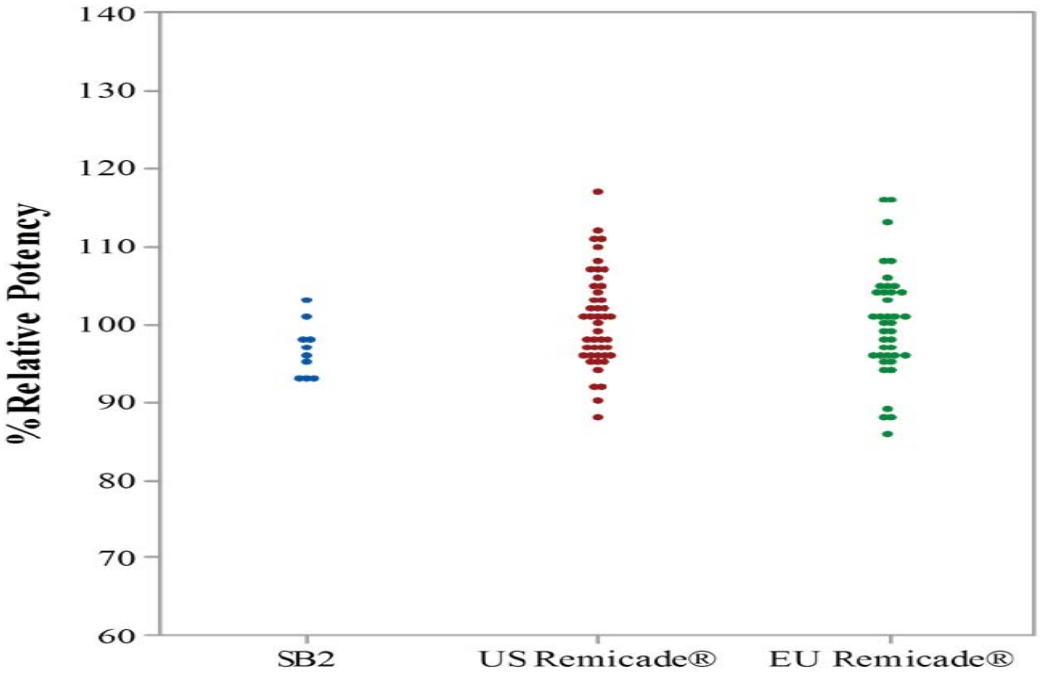


Table 4. Descriptive statistics for the TNF-α neutralization assay data

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Remicade	46	100.74	6.22	88	117.51
SB2	10	96.98	3.67	92.63	103.54
EU-approved Remicade	40	100.33	6.91	86.26	116.44

Because there is no assumption of equal variance between the test and reference products, Satterthwaite approximation is applied for obtaining the degree of freedom of the 90% sample size imbalanced adjusted CI for the mean difference between US-licensed Remicade and SB2. From Table 5, the result shows that the TNF-α neutralization assay of SB2 is equivalent to the TNF-α neutralization assay of US-licensed Remicade. Similarly, the TNF-α neutralization assay of SB2 is equivalent to the TNF-α neutralization assay of EU-approved Remicade, and the TNF-α neutralization assay of EU-approved Remicade is equivalent to the TNF-α neutralization assay of US-licensed Remicade.

Table 5. Equivalence testing results for the TNF-α neutralization assay

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(10, 46)	-3.76	(-7.10, -0.44)	(-9.33, 9.33)	Yes
SB2 vs. EU	(10, 40)	-3.35	(-6.92, 0.22)	(-10.36, 10.36)	Yes
EU vs. US	(40, 46)	-0.41	(-2.79,1.96)	(-9.33, 9.33)	Yes

*The 90% confidence interval is adjusted by the sample size imbalance.

5.3 FDA statistical equivalence testing for TNF-α binding assay

The TNF-α binding assay data points of SB2, US-licensed Remicade, and EU-approved Remicade are displayed in Figure 2. There appears a small mean difference among the three products. The variability of SB2 is smallest among three products.

10 batches of SB2, 41 batches of US-licensed Remicade, and 37 batches of EU-approved Remicade are included in the TNF-α binding assay dataset for the statistical equivalence testing. Descriptive statistics for the TNF-α binding assay data of SB2, US-licensed Remicade, and EU-approved Remicade are listed in Table 6.

From Table 7, the result shows that the equivalence of TNF-α binding assay between SB2 and US-licensed Remicade is supported. The equivalence of TNF-α binding assay between SB2 and

EU-approved Remicade is supported. The equivalence of TNF-α binding assay between US-licensed Remicade and EU-approved Remicade is supported.

Figure 2. Scatter plot of TNF-α binding assay for US-licensed Remicade, SB2, and EU-approved Remicade

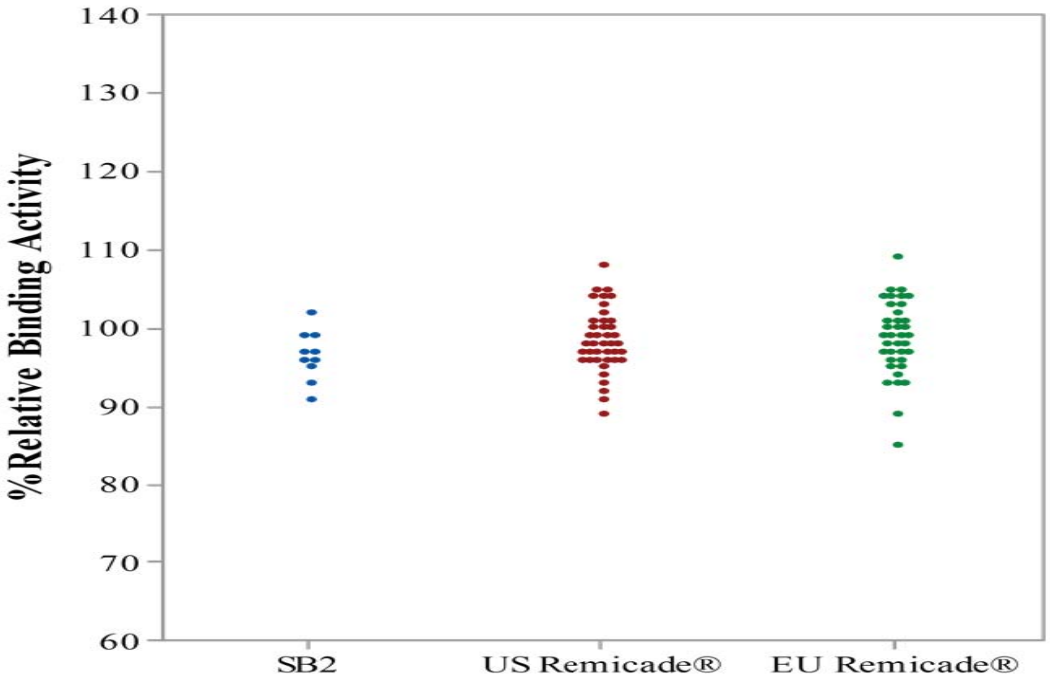


Table 6. Descriptive statistics for the TNF-α binding assay data

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Remicade	41	98.64	3.94	89.49	107.77
SB2	10	96.53	3.05	91.14	101.94
EU-approved Remicade	37	98.93	4.80	84.87	108.89

Table 7. Equivalence testing results for the TNF-α binding assay

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(10, 41)	-2.11	(-4.49, 0.26)	(-5.90, 5.90)	Yes
SB2 vs. EU	(10, 37)	-2.40	(-5.05, 0.25)	(-7.21, 7.21)	Yes
EU vs. US	(37, 41)	0.29	(-1.38,1.96)	(-5.90, 5.90)	Yes

*The 90% confidence interval is adjusted by the sample size imbalance.

5.4 Sensitivity analysis

For some batches, the number of within-batch replicates is different due to the failure of the sample parallelism test and the fraction of batches with parallelism failure is summarized in Table 8.

Table 8. Fraction of lots with parallelism failure for each product

Product	Quality Attribute	Fraction of batches with parallelism failure
SB2	TNF- α Neutralization	3/10
	TNF- α Binding	0/10
US-licensed Remicade	TNF- α Neutralization	10/46
	TNF- α Binding	7/41
EU-approved Remicade	TNF- α Neutralization	11/40
	TNF- α Binding	6/37

Then, the descriptive statistics and 90% CI for both Tier 1 QAs are recalculated after we take out batches with the failure of the sample parallelism test.

5.4.1 TNF-α neutralization assay

Seven batches of SB2, 36 batches of US-licensed Remicade, and 29 batches of EU-approved Remicade are included for the statistical equivalence testing for the TNF-α neutralization assay. Descriptive statistics for the TNF-α neutralization assay data are listed in Table 9. There appears a small mean difference among the three products. The variability of SB2 is smallest among three products.

Table 9. Descriptive statistics for the TNF-α neutralization assay data

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Remicade	36	100.76	6.64	88	117.51
SB2	7	95.97	3.39	92.63	101.85
EU-approved Remicade	29	99.41	6.69	86.26	115.83

The 90% sample size imbalanced adjusted CI for the mean difference between US-licensed Remicade and SB2 is recalculated in Table 10. The result shows that the TNF-α neutralization assay of SB2 is equivalent to the TNF-α neutralization assay of US-licensed Remicade. Similarly, the TNF-α neutralization assay of SB2 is equivalent to the TNF-α neutralization assay of EU-approved Remicade, and the TNF-α neutralization assay of EU-approved Remicade is equivalent to the TNF-α neutralization assay of US-licensed Remicade.

Table 10. Equivalence testing results for the TNF-α neutralization assay

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(7, 36)	-4.79	(-8.81, -0.77)	(-9.96, 9.96)	Yes
SB2 vs. EU	(7, 29)	-3.44	(-7.50, 0.60)	(-10.03, 10.03)	Yes
EU vs. US	(29, 36)	-1.35	(-4.12, 1.44)	(-9.96, 9.96)	Yes

*The 90% confidence interval is adjusted by the sample size imbalance.

5.4.2 TNF-α binding assay

10 batches of SB2, 34 batches of US-licensed Remicade, and 31 batches of EU-approved Remicade are included in the TNF-α binding assay dataset for the statistical equivalence testing. Descriptive statistics for the TNF-α binding assay data of SB2, US-licensed Remicade, and EU-approved Remicade are listed in Table 11. There appears a small mean difference among the three products. The variability of SB2 is smallest among three products.

From Table 12, the result shows that the equivalence of TNF-α binding assay between SB2 and US-licensed Remicade is supported. The equivalence of TNF-α binding assay between SB2 and EU-approved Remicade is supported. The equivalence of TNF-α binding assay between US-licensed Remicade and EU-approved Remicade is supported.

Table 11. Descriptive statistics for the TNF-α binding assay data

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Remicade	34	98.94	4.12	89.49	107.77
SB2	10	96.53	3.05	91.14	101.94
EU-approved Remicade	31	98.48	4.83	84.87	108.89

Table 12. Equivalence testing results for the TNF-α binding assay

Comparison	# of lots	Mean difference, %	90% CI for mean difference, %	Equivalence margin, %	Equivalent
SB2 vs. US	(10, 34)	-2.41	(-4.84, 0.02)	(-6.18, 6.18)	Yes
SB2 vs. EU	(10, 31)	-1.95	(-4.62, 0.71)	(-7.25, 7.25)	Yes
EU vs. US	(31, 34)	-0.46	(-2.33, 1.41)	(-6.18, 6.18)	Yes

*The 90% confidence interval is adjusted by the sample size imbalance.

6 CONCLUSION AND RECOMMENDATION

The results from the statistical equivalence testing of the TNF-α neutralization and the TNF-α binding assay support a demonstration that the proposed biosimilar SB2 is highly similar to US-licensed Remicade. The statistical analyses of the TNF-α neutralization and the TNF-α binding

assay in the three pair-wise comparisons (SB2, US-licensed Remicade, and EU-approved Remicade) also support the scientific bridge to justify the relevance of the data obtained from clinical studies that compared EU-approved Remicade and the SB2 product to support a demonstration of biosimilarity to US-licensed Remicade.

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

/s/

YU-TING WENG
12/12/2016

MEIYU SHEN
12/13/2016

YI TSONG
12/13/2016