CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

761058Orig1s000

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 761-058
Drug Name:	BI 695501 (Proposed Humira Biosimilar)
Indication(s):	Rheumatoid Arthritis (RA), Juvenile Idiopathic Arthritis (JIA) (4 years of age and older), Psoriatic Arthritis (PsA), Ankylosing Spondylitis (AS), Adult Crohn's Disease (CD), Ulcerative Colitis (UC), Plaque Psoriasis (Ps)
Applicant:	Boehringer Ingelheim
Date(s):	Received: 10/27/2016 BsUFA: 8/27/2016
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	James Travis, Ph.D.
Concurring Reviewers:	Gregory Levin, Ph.D.
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Stefanie Freeman, M.D. (Clinical Reviewer) Nikolay Nikolov, M.D. (Clinical Team Leader)
Project Manager:	Sadaf Nabavian, Pharm. D.

Keywords: BLA, clinical studies, biosimilar, missing data

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

This review considers BI 695501, a proposed biosimilar to US-licensed Humira (adalimumab). BI 695501 was evaluated in a single multi-dose comparative clinical study in patients with rheumatoid arthritis. It is my conclusion that BI 695501 met the agreed-upon efficacy requirements for a conclusion of similarity to US-licensed Humira. However, the upper bound of the confidence interval of the difference between treatment effects was larger than previously seen for other biosimilar products and so it will be especially important to ensure that there are no increases in the rates of safety events.

The applicant conducted a single comparative clinical study in patients with moderate to severely active RA for least six months. This was defined as at least six swollen joints (66 joint count) and at least six tender joints (68 joint count) at screening, and either an erythrocyte sedimentation rate (ESR) of at least 28 mm/hour or a C-reactive protein (CRP) level of at least 1.0 mg/dL at screening. Patients must also have been receiving and tolerating oral or parenteral methotrexate therapy at a dose of 15 to 25 mg per week. Patients were allowed to have exposure to no more than one biologic agent. This exposure must have been at least four months prior to screening for this study.

The applicant had two primary endpoints for this study. These were the percentage of patients who met the ACR20 responder criteria at Week 12 and 24. For all previously approved biosimilars with a comparative clinical study in rheumatoid arthritis, the primary analysis evaluated whether the 90% confidence of the difference in the percentage of ACR20 responders was contained in the similarity interval (−12%, 12%) in order for a conclusion of similarity to be made. For this product however, the Agency accepted an asymmetric similarity interval of (−12%, 15%).

I agree with the applicant's margin, but I do not agree with the justification they provided. An important purpose of the upper bound of the similarity interval is to be a surrogate for the safety of the biosimilar. Since the applicant's justification only looked at the dose response for efficacy and not the safety it was in my opinion lacking. However, as Dr. Freeman reports that she did not find any clinically meaningful differences between US-licensed Humira and BI 695501 in either the safety or immunogenicity profiles this would not impact the approvability of this application.

INTRODUCTION

1.1 Overview

The applicant has submitted a Biologics License Application (BLA) under section 351(k) of the Public Health Service (PHS) act to support marketing of BI 695501 as a biosimilar to US-licensed Humira (adalimumab).

The applicant conducted a single comparative clinical study, 1297.2 (VOLTAIRE-RA), to evaluate the similarity of BI 695501 to US-licensed Humira in treating patients with rheumatoid arthritis. This study is summarized in Table 1.

Table 1: Overview of Clinical Studies

Study	Population	Design	Treatment Arms	Number of Patients	Dates
1297.2 (VOLTAIRE-RA) ClinicalTrials.gov identifier: NCT02137226	Rheumatoid Arthritis	48-week, Randomized, double-blind, Parallel Group	BI 695501	324	Feb 4, 2015 (First patient)
			US-Licensed Humira	321	Apr 29, 2016 (Cut-off date for Week 32 Analysis)

The development program for BI 695501 was conducted under IND 110,467, which was originally submitted in November, 2010. I will now summarize the interactions between the Agency and the applicant where the design of the comparative clinical study was discussed.

A Pre-IND meeting was held on April 28, 2011. At this meeting the overall development program for BI 695501 was discussed and feedback was given by the Agency regarding its current thinking on the design of comparative clinical studies.

The next meeting between the Agency and the applicant was a BPD type 3 meeting which was held on July 1, 2013. In the meeting package for this meeting the applicant submitted a draft of the protocol for study 1297.2. At this meeting there was discussion of two points that impacted the design of this comparative clinical study. First, the applicant asked whether it would be acceptable to compare BI 695501 to pooled data from EU- and US-sourced Humira. The Agency stated that the applicant would need to provide bridging PK data showing that EU-approved and US-licensed Humira meet the pre-specified acceptance criteria for analytical and PK similarity. No feedback was provided regarding the sponsor's proposed similarity margin at this time.

In February 2014, the applicant submitted a revised version of the draft protocol (version 2.0) and requested feedback on the proposed study design. The Agency provided a response to this request in May 2014, (b) (4)

(b) (4)

(See Table 2 for a summary of the historical effect of adalimumab). A margin of no more than ±14% was recommended.

The Agency also stated that revisions to the protocol were needed to address missing data in the efficacy analyses. In the historical studies patients who discontinued study treatment early or who were lost to follow-up were considered non-responders, and so this approach was suggested in order to maintain similarity to the historical studies that were used to establish the effect of adalimumab over placebo.

Table 2: Meta-Analysis of Historical Treatment Effect of Adalimumab with Respect to the Absolute Difference in ACR20 Response at Week 12

Study	MTX + Placebo N (% ACR Response)	MTX + Adalimumab N (% ACR Response)	Difference in % Response
(Keystone et al., 2004)	200 (25%)	207 (57%)	34%
(Weinblatt et al., 2003)	62 (23%)	67 (66%)	43%
(Kim et al., 2007)	63 (25%)	65 (57%)	32%
(Chen et al., 2009)	12 (33%)	35 (54%)	21%
Meta-Analysis (fixed effects): Difference (95% CI)			34.0% (27.1%, 40.8%)
Meta-Analysis (random effects): Difference (95% CI)			34.1% (27.3%, 41.0%)
Heterogeneity p-value			0.54

Source: FDA comments provided on May 27, 2014

The next meeting was held on March 10, 2015. This meeting was a BPD Type 2 meeting. At this meeting the applicant was informed that the Agency had a new recommendation on the margin selection for the comparative clinical study. The Agency proposed a new margin of ±12%. However, the Agency also noted that a 90% rather than 95% confidence interval for the difference in ACR20 response rates could be used for the evaluation of similarity. In a post-meeting comment the Agency stated that an asymmetric margin (for example, -12% to +15%) could be considered if an adequate justification was provided for the chosen margin. The Agency's recommended similarity margin was based on considerations aimed at weighing the clinical importance of various differences in effect against the feasibility of different study sizes.

On August 19, 2015 the applicant requested a BPD Type 2 meeting. This meeting was denied, but a written response was issued to response to the applicant's questions. The Sponsor at that time had proposed to compare the ACR20 responses across both Week 12 and Week 24 rather than at the individual time points. The Agency recommended selecting the Week 12 ACR20 response rate as the primary endpoint. The sponsor had also proposed (b) (4)

The Agency stated that this would not be acceptable. The sponsor was also informed that the

primary analysis population should include all randomized patients, and should not exclude patients who failed to provide post-baseline efficacy assessments.

At this time the applicant also selected a similarity margin of -12% to +15% for the study and provided a justification that I will summarize. Table 3 shows the results of the ARMADA trial (Weinblatt et al., 2003) which they said shows that the maximum level of efficacy is achieved with a regimen of 40 mg every other week and that increasing the dose to 80 mg every other week does not provide additional benefit. The applicant also referred to the findings published in the literature (Pouw et al., 2015) which they said showed that maximum efficacy was achieved with adalimumab concentrations of 5-8 µg/mL, which they said is achieved with a 40 mg multiple dose. The Agency stated in the response sent in November 2015 that they agreed with the proposed asymmetric margin.

Table 3: Patients with ACR20, ACR50, and ACR70 responses at Week 24 [Extracted from (Weinblatt, et al., 2003)]

Response Criteria Met	Placebo (n=62)	Adalimumab dosage (every other week)		
		20 mg (n=69)	40 mg (n=67)	80 mg (n=73)
ACR20	9 (14.5)	33 (47.8)	45 (67.2)	48 (65.8)
P†		<0.001	<0.001	<0.001
ACR50	5 (8.1.)	22 (31.9)	37 (55.2)	31 (42.5)
P‡		0.0003	<0.001	<0.001
ACR70	3 (4.8)	7 (10.1)	18 (26.9)	14 (19.2)
P‡		NS	<0.001	0.020

Source: August 19, 2015 briefing package
* Values are the number (%) of patients who met the American College of Rheumatology criteria for 20%, 50%, and 70% improvement (ACR20, ACR50, and ACR70, respectively) at week 24. Patients who did not complete the 24-week study were defined as non-responders, NS = not significant.
† Adalimumab versus placebo, by Dunnett's test; statistical significance was set at $P \leq 0.05$.
‡ Adalimamab versus placebo, by unadjusted *t*-test; statistical significance was set at $P \leq 0.05$.
Source: August 19, 2015 Type 2 BPD Meeting Package.

The final communication between the applicant and the Agency concerning the design and analysis of the clinical comparative study occurred in December 2015. This was a response to a clarification question that the applicant asked following the November 2017 response.

In summary, the history of the similarity margin is shown in Table 4. Though the final version of the margin was not added to the protocol until three months prior to the Week 32 cutoff (see Table 1), it was suggested by the Agency the previous year, in March 2015 and proposed by the applicant in a submission to the Agency on August 19, 2015. The first independent Data Monitoring Committee (IDMC) meeting was held on August 28, 2015, and a blinded assessment of the sample size assumptions was also performed at this time. The conclusion of the IDMC was that no adjustments to the sample size were needed.

Table 4: History of the Similarity Margin

Protocol Version	Protocol Date	Endpoint	Time point	Confidence Level	Margin
1.0	Aug 2013	[redacted] (b) (4)			
2.0	Feb 2014				
3.0	Nov 2014				
4.0	Feb 2015				
5.0	Apr 2015	ACR20	Week 12	90%	(-12%, 12%)
		ACR20	Week 24	95%	(-15%, 15%)
6.0	Jan 2016	ACR20	Week 12	90%	(-12%, 15%)
		ACR20	Week 24	95%	(-15%, 15%)

Source: Reviewer

1.2 Data Sources

The data from the clinical study was provided electronically in the SAS transport format and can be found at the following location in the CDER electronic document room: \\cdsesub1\evsprod\BLA761058\0000\m5

2 STATISTICAL EVALUATION

2.1 Data and Analysis Quality

The submitted data were of acceptable quality and were adequately documented. I was able to reproduce results of all the primary and secondary analyses.

2.2 Study Design and Endpoints

Study 1297.2 was a randomized, double-blind, parallel arm, multiple dose, comparative clinical study. This study was conducted to evaluate the similarity of BI 695501 to US-licensed Humira in the treatment of patients with rheumatoid arthritis. Since only US-licensed Humira was used for this study, for brevity it will be referred to as Humira for the remainder of the discussion of this study.

The population for this study consisted of patients who were between 18 and 80 years of age, had a diagnosis of moderately to severely active RA for at least 6 months, defined as at least six swollen and six tender joints at screening, and had either an erythrocyte sedimentation rate (ESR) of more than 28 mm/hour or a C-reactive protein (CRP) level greater than 1.0 mg/DL at screening.

Patients must have been receiving a stable dose of oral or parenteral methotrexate for at least 12 weeks immediately prior to Day 1. The dose must have been between 15 to 25 mg per week. Prior exposure to at most one non-adalimumab biologic agent was allowed, provided the last exposure was at least 4 months prior to screening. Patients were randomized in a 1:1 ratio to BI 695501 or Humira via injection every two weeks. Dose increases were not permitted. Randomization was stratified according to region and to prior exposure to a biologic agent.

The pre-specified primary endpoints for this study were the proportion of patients who met the criteria for an ACR20 response at Week 12 and 24. A patient was considered an ACR20 responder if he or she had a 20% reduction from baseline in the number of swollen and tender joints, as well as a 20% reduction in at least 3 of the 5 following assessments:

1. Patient's assessment of pain (100 mm Visual Analog Scale – VAS).
2. Patient's global assessment of disease activity (VAS).
3. Physician's global assessment of disease activity (VAS).
4. Patient's assessment of physical function, as measured by the Health Assessment Questionnaire – Disability Index (HAQ-DI).
5. Acute phase reactant (CRP).

The change from baseline in Disease Activity Score 28 – Erythrocyte Sedimentation Rate (DAS28-ESR) at Weeks 12 and 24 were the only secondary endpoints for this study. The applicant also explored a number of other exploratory efficacy endpoints. These were as follows:

- The proportion of patients meeting ACR20 response criteria at Week 48.
- The proportion of patients meeting ACR50 response criteria at Weeks 12, 24, and 48.
- The proportion of patients meeting ACR70 response criteria at Weeks 12, 24, and 48.
- Individual parameters of the ACR improvement criteria at Week 12 and 24.
- The change from baseline in DAS28(CRP) at Weeks 12, 24, and 48.
- The change from baseline in DAS28(ESR) at Week 48.
- The proportion of patients who meet the ACR/European League against Rheumatism (EULAR) definition of remission at Weeks 12, 24, and 48.
- The proportion of patients with EULAR response (good response, moderate response, or no response) at Weeks 12, 24, and 48.
- Change from baseline in SF-36 v2 at Weeks 12, 24, 48.

2.3 Statistical Methodologies

2.3.1 Planned Analyses

Both of the primary efficacy endpoints were analyzed using a logistic regression model. The models included the patient's DAS28(ESR) score at baseline as a covariate, and whether the patient had any prior exposure to any biologic anti-rheumatic agents.

The applicant used a combination of two approaches to impute missing data for its primary analysis for patients with non-calculable ACR20 responses. Patients who discontinued for reasons that the applicant considered "informative" were imputed as non-responders (NRI). Informative reasons included withdrawal due to adverse event, lack of efficacy, physician's decision and lost to follow-up. All other patients were imputed using a multiple imputation (MI) approach. This was performed in two steps; first intermittent missing observations for each of the individual components were imputed using a Markov chain Monte Carlo (MCMC) method (Schafer, 1997) which assumed missing at random and multivariate normality. Second, missing component scores after discontinuation were imputed using a regression approach (Rubin, 1987). The model for each component in this approach included treatment, prior exposure to a biological agent, region, and previous observed data for the particular component. The imputed

ACR20 scores were then calculated using the individual imputed components. Given that patients who discontinued for "informative" reasons were considered non-responders, this primary analysis implicitly targeted an estimand that combines the treatment effect of the drug on the signs and symptoms of rheumatoid arthritis with the treatment effect on tolerability and overall compliance in the study.

The applicant used a novel approach to generate the confidence intervals for the primary analysis. This approach, described in (Reeve, 2016), uses the parameter estimates from the logistic regression model to generate confidence intervals for the absolute risk difference that have been adjusted using the baseline covariates, which are the patient's DAS28(ESR) score at baseline and whether the patient had prior exposure to any biologic anti-rheumatic agents. This method was applied to each imputed dataset. The confidence intervals for each imputed dataset were combined using Rubin's rules (Rubin, 1987) to produce the final confidence intervals.

The applicant also planned a number of sensitivity analyses to explore the effect of missing data, analysis population choice and analysis method on the study conclusion. These analyses are summarized as follows:

- *Missing Data Analysis #1:* The primary analysis was repeated, but the last observation was carried forward (LOCF) for any patient who had missing scores due to "non-informative" reasons.
- *Missing Data Analysis #2:* The primary analysis was repeated. Observed data obtained after treatment discontinuation was used for this analysis rather than imputing the missing data for patients who discontinued for non-informative reasons. Patients who discontinued due to informative reasons were still considered non-responders.
- *Analysis Population:* The primary analysis was repeated using the applicant's per protocol set.
- *Analysis method:* The data was reanalyzed without using any adjustment for covariates.

I do not believe that these analyses sufficiently explored the missing data assumptions and so I requested additional analyses which are described in Section 2.3.2.

The history of the similarity margins for this study is discussed in Section 1.1. For this study the applicant selected two different similarity margins for the two ACR20 endpoints. For the Week 12 endpoint the 90% confidence interval of the difference in ACR20 response rates between patients receiving BI 695501 and Humira must have been greater than -12% and less than 15%. This was the primary analysis aligning with recommendations from the Agency. For the Week 24 endpoint the 95% confidence interval of the difference must have been between -15% and 15%.

2.3.2 Additional Analyses

There were several additional analyses that the applicant performed after submission of the application at my request. The first was a tipping point analysis where patients who discontinued due to the informative reasons used in the primary analysis were considered non-responders. Observed data was used where possible for patients who discontinued due to non-informative reasons. This tipping point analysis targeted the same estimand used in the primary analysis, but systematically varied the assumptions about the ACR20 response among the subjects with

missing data (rather than relying on the missing-at-random assumption underlying the multiple imputation approach used in the applicant's primary analysis). The tipping point analysis is intended to explore the sensitivity of the study conclusion to possible differences in the assumed response rate.

In addition to the estimand targeted in the primary analysis, another estimand of interest to me was the difference in the rate of ACR20 response where treatment discontinuation for an informative reason (due to adverse events, physician's decision, lack of efficacy or reasons indicated as other) is no longer consider indicative of non-response. The intent for this estimand is similar to missing data analysis #2 described above. It differed in that we wanted to know what happened to the clinical outcomes for patients who discontinued for informative reasons, rather than just considering them non-responders.

This estimand was explored in two analyses. For the first analysis, the data was reanalyzed using the primary analysis approach without the non-responder imputation. Second, an additional tipping point analysis was performed where observations for all patients with missing data were multiply imputed based on systematically varying assumptions for all patients with incomplete ACR20 scores. Again the non-responder imputation was not applied in this analysis.

Finally, I will present the results of analyses of the individual components that are used in the ACR scores and both the ESR and CRP versions of the DAS28. These analyses used the same multiple imputation approach described for the primary analysis, without the non-responder imputation. The scores were analyzed using analysis of covariance models, with treatment group and prior exposure to another biologic agent as factors and baseline score as a covariate. The DAS28 scores were imputed directly and not component-wise.

2.4 Patient Disposition, Demographic and Baseline Characteristics

The patient demographic characteristics at baseline for this study are shown in Table 5.

Table 5: Demographic Characteristics at Baseline (Safety Population)

Characteristic (Unit)	BI 695501 (N=324)	Humira (N=321)
Age (years)		
Mean (SD)	53.7 (12.04)	53.6 (11.32)
Gender (n [%])		
Male	57 (17.6)	52 (16.2)
Female	267 (82.4)	269 (83.8)
Race (n [%])		
American Indian or Alaska Native	0	0
Asian	8 (2.5)	6 (1.9)
Black or African American	6 (1.9)	7 (2.2)
Native Hawaiian of Other Pacific Islander	0	0
White	309 (95.4)	304 (94.7)
Other	1 (0.3)	4 (1.2)
Ethnicity (n [%])		
Asia	6 (1.9)	6 (1.9)
Europe	231 (71.3)	228 (71.0)
Latin America	25 (7.7)	26 (8.1)
USA	62 (19.1)	61 (19.0)
Weight (kg)		
Mean (SD)	73.1 (16.87)	75.1 (170.07)
BMI (kg/m^2)		
Mean (SD)	27.04 (5.422)	27.86 (6.309)

Source: Applicant's Study Report Table 12

The baseline disease status for the patients in the safety population is shown in Table 6.

Table 6: Baseline Disease Status (Safety Population)

Characteristic (Unit)	BI 695501 (N=324)	Humira (N=321)
Duration of RA (years) Mean (SD)	7.36 (7.27)	7.06 (6.765)
ESR (mm/hour) Mean (SD)	45.5 (19.16)	43.2 (17.99)
DAS28(ESR) Mean (SD)	6.59 (0.812)	6.56 (0.815)
DAS28(CRP) Mean (SD)	5.68 (0.843)	5.68 (0.873)
Prior Exposure to a Biological Agent (n [%])		
Yes	85 (26.2)	86 (26.8)
No	239 (73.8)	235 (73.2)
Previous DMARD Therapies Mean (SD)	2.2 (1.37)	2.4 (1.51)
Methotrexate Dosage (mg/week) Mean (SD)	16.30 (3.619)	16.79 (3.911)
Patients with positive ADA n (%)	11 (3.4)	21 (6.5)
Patients with positive nAb n (%)	9 (2.8)	16 (5.0)
RF and anti CCP antibodies n (%)		
Total	324 (100)	321 (100)
Both negative	0	1 (0.3)
RF positive or indeterminate or anti-CCP positive	324 (100)	320 (99.7)

Source: Applicant's Study Report Table 13

The patient disposition in the study is shown in Table 7. The amount of discontinuation from the study prior to both Week 12 and Week 24 was low (2% and 3% at Week 12 and 6% and 5% at Week 24 for the BI 695501 and Humira arms, respectively). There were no large differences between the two treatment arms.

Table 7: Patient Disposition

	Week 12		Week 24	
	BI 695501 n (%)	Humira n (%)	BI 695501 n (%)	Humira n (%)
Randomized	324	321	324	321
Discontinued from treatment	10 (3%)	15 (5%)	26 (8%)	27 (8%)
Adverse event	1	6	4	10
Primary lack of efficacy	0	0	2	1
Secondary lack of efficacy	0	0	1	0
Withdrawal by patient	5	3	11	6
Physician decision	1	1	4	2
Lost to follow-up	2	2	3	2
Other	1	3	1	6
Discontinued from study	5 (2%)	11 (3%)	20 (6%)	16 (5%)
Adverse event	0	3	3	3
Primary lack of efficacy	0	0	1	0
Withdrawal by patient	3	2	10	4
Physician decision	1	1	1	2
Lost to follow-up	1	2	3	2
Other	0	3	2	5

Source: Reviewer

The enrollment by location is shown in Table 8.

Table 8: Study Enrollment by Location

Location	Enrollment
EU	290
United States	123
Ukraine	110
Chile	51
Serbia	32
Russian Federation	27
Korea, Republic of	5
Thailand	5
Malaysia	2

Source: Reviewer

2.5 Evaluation of Efficacy

The results of my reproduction of the applicant's primary analysis are shown in Table 9. The 90% confidence intervals for the difference in ACR20 response between BI 695501 and Humira were within the agreed upon bounds for both Week 12 and Week 24. The applicant's primary analysis excluded 6 patients (3 per arm) who did not provide any post-baseline efficacy assessments. These patients were included in my analyses, but did not change the overall conclusion.

Table 9: Primary Efficacy: Estimates and CIs for Differences in ACR20 Response Rate at Week 12 and 24 (NRI and MI)

	Treatment	N	ACR20 Response Proportion (%)	Difference in Proportion (BI 695501 – Humira, %)	
				Estimate	90% CI
Week 12	BI 695501	324	66.4	5.9	(−0.8, 12.7)
	Humira	321	60.5		
Week 24	BI 695501	324	68.4	4.4	(−2.3, 11.1)
	Humira	321	64.0		

Source: Reviewer's reproduction of Study Report Table 14.
Similarity Requirements: 90% confidence interval within (-12%, 15%).

First, I will present the sensitivity analyses that the applicant conducted. I excluded the first sensitivity analysis as single imputation methods such as LOCF underestimate the variability of the outcome and so I do not believe this is a useful analysis to present. In Table 10 I show the results of what I have labeled missing data analysis #2. In this analysis, where the clinical outcomes after treatment discontinuation were included for patients who discontinued for reasons other than those considered informative by the applicant, estimates of the ACR20 response were slightly higher than for the primary analysis. The overall conclusion is however the same as the 90% confidence interval of the treatment difference still falls within the pre-specified interval.

Table 10: Off-treatment data collection: Estimates and CIs for Differences in ACR20 Response Rate at Week 12 and 24 (NRI and MI)

	Treatment	N	Proportion (%)	Difference in Proportion (BI 695501 – Humira, %)	
				Estimate	90% CI
Week 12	BI 695501	324	67.6	5.0	(−1.7, 11.8)
	Humira	321	62.6		
Week 24	BI 695501	324	71.0	2.2	(−4.4, 8.9)
	Humira	321	68.8		

Source: Applicant's Study Report Table 16.
Similarity Requirements: 90% confidence interval within (-12%, 15%).

Table 11 shows the results of the applicant's per-protocol analysis. Again, we see slightly higher estimates of the ACR20 response rate, but no change in the overall conclusion.

Table 11: Per Protocol Analysis: Estimates and CIs for Differences in ACR20 Response Rate at Week 12 and 24 (NRI and MI)

				Difference in proportion (BI 695501 – Humira, %)	
	Treatment	N	Proportion (%)	Estimate	90% CI
Week 12	BI 695501	324	67.1	4.2	(−2.9, 11.2)
	Humira	321	62.9		
Week 24	BI 695501	324	70.3	1.6	(−5.3, 8.4)
	Humira	321	64.5		

Source: Applicant's Study Report Table 15.
Similarity Expectations: 90% confidence interval within (-12%, 15%).

Presented in Table 12 are the results of my reproduction of the applicant's unadjusted analysis. The applicant excluded the same six patients from this analysis who were excluded from the primary analysis. My analysis includes these patients. For this analysis we see a small change in the ACR20 response rates. However, the confidence intervals are considerably wider than seen in the primary analysis which is what we would expect if there is a correlation between the variables used in the model and the clinical outcome. See Section 3.2 for an exploration of the effects of the baseline disease severity on the ACR20 response rates.

Table 12: Unadjusted analysis: Estimates and CIs for differences in ACR20 Response Rate at Week 12 and 24 (NRI and MI)

				Difference in proportion (BI 695501 – Humira, %)	
	Treatment	N	Proportion (%)	Estimate	90% CI
Week 12	BI 695501	324	66.4	5.9	(−1.9, 13.9)
	Humira	321	60.4		
Week 24	BI 695501	324	68.3	4.4	(−3.2, 11.9)
	Humira	321	64.0		

Source: Applicant's Study Report Table 14.2.1.10.
Similarity Requirements: 90% confidence interval within (-12%, 15%).

The next analysis I will present is the results of the applicant's tipping point analysis. The results are shown in Table 13 for Week 12 and Table 14 for Week 24. These analyses thoroughly explored the possible rates of response for the missing patients who discontinued due to non-informative reasons and failed to find any points where the conclusion was changed. Patients who discontinued for informative reasons were considered non-responders for this analysis.

Table 13: Tipping Point Analysis: ACR20 Estimates and 90% Confidence intervals at Week 12

Assumed ACR20 response rate for BI 695501 (shift from adjusted response rate)	Assumed ACR20 response rate for Humira (shift from adjusted response rate)				
	0%	22.9% (−40%)	42.9% (−20%)	62.9% (no shift)	100%
0%	5.0 (−2.0, 11.5)	4.3 (−2.6, 11.1)	3.7 (−3.2, 10.6)	3.0 (−3.8, 9.8)	2.1 (−4.9, 8.8)
27.5% (−40%)	5.3 (−1.5, 12.1)	4.6 (−2.2, 11.4)	4.0 (−2.8, 10.8)	3.4 (−3.4, 10.2)	2.5 (−4.3, 9.3)
47.5% (−20%)	5.6 (−1.1, 12.4)	4.9 (−1.9, 11.8)	4.4 (−2.5, 11.2)	3.7 (−3.1, 10.5)	2.8 (−4.0, 9.6)
67.5% (no shift)	6.0 (−0.7, 12.7)	5.3 (−1.5, 12.1)	4.7 (−2.1, 11.6)	4.1 (−2.7, 10.9)	3.2 (−3.6, 10.0)
100%	7.1 (0.2, 13.5)	6.4 (−0.3, 13.1)	5.8 (−0.9, 12.6)	5.2 (−1.5, 11.9)	4.3 (−2.6, 10.8)

Source: Table 3 Submission number 0010 (14)

Table 14: Tipping Point Analysis: ACR20 Estimates and 90% Confidence intervals at Week 24

Assumed ACR20 response rate for BI 695501 (shift from adjusted response rate)	Assumed ACR20 response rate for Humira (shift from adjusted response rate)				
	0%	29.3% (−40%)	49.3% (−20%)	69.3% (no shift)	100%
0%	0.7 (−6.3, 7.2)	−0.2 (−7.0, 6.6)	−0.9 (−7.7, 6.0)	−1.5 (−8.3, 5.3)	−2.8 (−9.8, 3.8)
31.7% (−40%)	2.3 (−4.5, 9.0)	1.4 (−5.5, 8.3)	0.7 (−6.1, 7.6)	0.1 (−6.7, 6.9)	−1.2 (−8.0, 5.7)
51.7% (−20%)	3.2 (−3.5, 9.8)	2.3 (−4.4, 9.1)	1.6 (−5.0, 8.3)	1.0 (−5.7, 7.7)	−0.3 (−7.0, 6.4)
71.7% (no shift)	4.6 (−2.0, 11.1)	3.7 (−2.9, 10.3)	3.0 (−3.5, 9.6)	2.4 (−4.2, 9.0)	1.1 (−5.5, 7.7)
100%	6.3 (−0.3, 12.3)	5.4 (−0.9, 11.8)	4.7 (−1.6, 11.1)	4.1 (−2.3, 10.5)	2.8 (−3.8, 8.9)

Source: Reviewer

Next, I will present the results of the analyses described in section 2.3.2, where an alternative estimand was discussed. In particular, these analyses evaluate the difference in ACR20 response probability regardless of treatment discontinuation. The primary analysis was conducted for this estimand and the results are shown in Table 15 and the results of the tipping point analyses are shown in Table 16 for Week 12 and Table 17 for Week 24. For both weeks the ACR20 response rates are higher than seen in the original analysis, though the differences between the treatment groups are lower.

For the tipping point analyses, there are no points at which the conclusion changes for Week 12 (Table 16). For Week 24 (Table 17), however, the conclusions changes for the top right and bottom left entries in the table. Both these analyses correspond to the situations where 0% of the patients with missing Week 24 outcomes are assumed to be responders for one treatment arm, and 100% of the patients with missing outcomes are assumed to be responders for the other arm. These are obviously the most extreme cases possible, and seem implausible given the similar proportions of patients and reasons for dropout on the two treatment arms. Since the conclusion only changes in these extreme situations we can be reassured about the strength of the original findings.

Table 15: Sensitivity Analysis: Analysis of the ACR20 Response Rate at Week 12 and 24 Regardless of Treatment Discontinuation (MI)

				Difference in Proportion (BI 695501 – Humira, %)	
	Treatment	N	Proportion (%)	Estimate	90% CI
Week 12	BI 695501	324	68.2	4.5	(−2.3, 11.3)
	Humira	321	63.7		
Week 24	BI 695501	324	73.7	3.1	(−3.5, 9.7)
	Humira	321	70.6		

Source: Table 5 Submission number 0010 (14)

Table 16: Tipping Point Analysis regardless of Treatment Discontinuation: ACR20 Estimates and 90% Confidence intervals at Week 12

Assumed ACR20 response rate for BI 695501 (shift from adjusted response rate)	Assumed ACR20 response rate for Humira (shift from adjusted response rate)				
	0%	24.3% (−40%)	44.3% (−20%)	64.3% (No shift)	100%
0%	5.0 (−2.0, 11.5)	3.9 (−2.9, 10.7)	3.0 (−3.9, 10.0)	2.0 (−4.9, 8.9)	−0.1 (−7.1, 6.6)
28.1% (−40%)	5.7 (−1.1, 11.5)	4.6 (−2.2, 11.5)	3.8 (−3.1, 10.7)	2.8 (−4.2, 9.7)	0.7 (−6.2, 7.5)
48.1% (−20%)	6.3 (−0.4l, 13.0)	5.2 (−1.6, 12.0)	4.4 (−2.5, 11.3)	3.4 (−3.5, 10.2)	1.3 (−5.5, 8.1)
68.1% (no shift)	6.9 (0.2, 13.6)	5.8 (−0.9, 12.5)	5.0 (−1.8, 11.8)	3.9 (−2.9, 10.7)	1.9 (−4.9, 8.6)
100%	8.0 (1.2, 14.4)	7.0 (0.4, 13.6)	6.1 (−0.6, 12.8)	5.1 (−1.6, 11.8)	3.0 (−3.8, 9.4)

Source: Table 6 Submission number 0010 (14)

Table 17: Tipping Point Analysis regardless of Treatment Discontinuation: ACR20 Estimates and 90% Confidence intervals at Week 24

Assumed ACR20 response rate for BI 695501 (shift from adjusted response rate)	Assumed ACR20 response rate for Humira (shift from adjusted response rate)				
	0%	31.7% (−40%)	51.7% (−20%)	71.7% (no shift)	100%
0%	0.7 (−6.3, 7.2)	−1.2 (−8.1, 5.7)	−2.7 (−9.7, 4.2)	−4.1 (−11.1, 2.8)	−5.9 (−13.0, 0.8)
34.6% (−40%)	3.5 (−3.1, 10.1)	1.6 (−5.3, 8.5)	0.1 (−6.8, 6.9)	−1.3 (−8.1, 5.5)	−3.0 (−9.8, 3.7)
54.6% (−20%)	5.3 (−1.3, 11.9)	3.4 (−3.5, 10.2)	1.9 (−4.9, 8.6)	0.5 (−6.3, 7.2)	−1.3 (−8.0, 5.5)
74.6% (no shift)	7.2 (0.8, 13.6)	5.3 (−1.3, 11.9)	3.8 (−2.7, 10.3)	2.4 (−4.0, 8.8)	0.7 (−5.8, 7.2)
100%	9.9 (3.7, 15.6)	8.0 (1.9, 14.1)	6.5 (0.4, 12.6)	5.1 (−1.0, 11.2)	3.4 (−2.9, 9.1)

Source: Reviewer
Cells highlighted in grey indicate analyses where the similarity margin is no longer ruled out.

Presented next are the results of my analyses of the ACR50 and ACR70 response rates for Week 12 and 24 (Table 18). Again the response rates are relatively similar between groups. However, no margins were specified for similarity for these endpoints.

Table 18: Estimate and CIs for Differences in ACR50 and ACR70 Response Rate at Week 12 and 24 (NRI and MI)

Parameter	Week	BI 695501 Response (%) N=324	Humira Response (%) N=321	Difference in Proportion (BI 695501 – Humira, %) Estimate	90% CI
ACR50	12	28.9%	30.6%	−1.7	(−8.3, 4.8)
	24	36.5%	36.3%	0.3	(−6.7, 7.3)
ACR70	12	10%	10.9%	−0.9	(−5.3, 3.5)
	24	13.5%	17.7%	−4.3	(−9.2, 0.6)

Source: Reviewer.

My analyses of the individual ACR20 components are shown in Table 19. All presented analyses found relatively similar mean changes from baseline, with confidence intervals for differences much smaller than the changes from baseline. No margins were pre-specified for these comparisons.

Table 19: Mean Change from Baseline in the ACR20 Components

Component	Week	BI 695501 (N=324)	Humira (N=321)	Difference (90% CI)
Tender Joint Count (68)	12	−14.4	−13.8	−0.59 (−1.78, 0.61)
	24	−15.7	−15.3	−0.42 (−1.66, 0.82)
Swollen Joint Count (66)	12	−10.7	−10.3	−0.39 (−1.16, 0.37)
	24	−12.1	−11.5	−0.54 (−1.34, 0.25)
HAQ Score	12	−0.4	−0.4	0.01 (−0.05, 0.07)
	24	−0.5	−0.5	0.03 (−0.04, 0.09)
Patient Global	12	−24.6	−23.4	−1.2 (−3.86, 1.47)
	24	−27.1	−26.7	−0.46 (−3.32, 2.41)
Physician Global	12	−32.4	−33.5	1.04 (−1.22, 3.29)
	24	−36.2	−38.2	1.92 (−0.41, 4.26)
Patient Pain	12	−25.3	−25.5	0.23 (−2.59, 3.06)
	24	−27.7	−27.2	−0.45 (−3.5, 2.59)
CRP (mg/L)	12	−6.8	−5.9	−0.96 (−3.3, 1.38)
	24	−6.7	−6.5	−0.16 (−1.67, 1.35)

Source: Reviewer

Table 20 shows the results of my analyses of the change in both types of the Disease Activity Score 28 (DAS28) score and the erythrocyte sedimentation rate (ESR). Again, we see relatively similar mean changes from baseline, with confidence intervals for differences much smaller than the level of change seen.

Table 20: Mean Change from Baseline in the DAS28 Scores and ESR

Component	Week	BI 695501 (N=324)	Humira (N=321)	Difference (90% CI)
ESR (mm/h)	12	−16.8	−17.8	1.07 (−0.92, 3.07)
	24	−18.2	−17.7	−0.5 (−2.93, 1.94)
DAS28(ESR)	12	−2.1	−2	−0.09 (−0.24, 0.06)
	24	−2.4	−2.5	0.04 (−0.12, 0.2)
DAS28(CRP)	12	−1.9	−1.8	−0.13 (−0.27, 0.01)
	24	−2.1	−2.2	0.03 (−0.12, 0.18)

Source: Reviewer

In reviewing this comparative clinical study it is important to understand the similarity of the trial design to the historical studies used in setting the similarity margin (see Table 2). Significant differences in the design could impact the validity of the selected margin. In Table 21 I compared the key features of the study design for Study 1297.2 with the historical studies used to determine the similarity margin for the study.

Overall, the design of the study was comparable to the historical studies. The main difference between this study and the previous studies was that this study included patients who had prior exposure to non-adalimumab anti-TNF inhibitors. There were also several small differences in the inclusion criteria, including a reduction in the required number of tender joints and inclusion of a requirement for the erythrocyte sedimentation rate. These changes do not appear to have had a noticeable effect on the enrolled population.

Study 1297.2 was also conducted in a much wider range of geographic locations than previous studies, which are summarized in Table 8.

Table 21: Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials of Adalimumab in RA and Comparative Clinical Study 1297.2 (Voltaire-RA)

	Study				
	(Keystone, et al., 2004)	(Weinblatt, et al., 2003)	(Kim, et al., 2007)	(Chen, et al., 2009)	1297.2 (VOLTAIRE-RA)
Selected inclusion/ exclusion criteria	≥9 TJC; ≥6 SJC; CRP >1 mg/dL; RF+; ≥1 join erosion	≥9 TJC; ≥6 SJC	≥9 TJC; ≥6 SJC	≥9 TJC; ≥6 SJC	≥6 TJC ≥6 SJC ESR > 28 mm/hr or CRP > 1.0 mg/dL
Biologic DMARD experience allowed?	No	No	No	No	Yes Allowed exposure to a single non-adalimumab biologic agent, ~27% had prior biologic exposure
Concomitant DMARDS	Stable MTX, cortico-steroids, NSAIDS	Stable MTX, corticosteroids, NSAIDS	Stable MTX	Stable MTX	Stable MTX, corticosteroids, NSAIDS
Region/Country	US & Canada	US & Canada	Korea	Taiwan	US, EU, Chile, Korea, Malaysia, Russia, Ukraine, Serbia & Thailand

Baseline Characteristics of Study Population[2]	TJC: 27; SJC: 19; Disease Duration: 11 yrs; HAQ-DI: 1.5	TJC: 28; SJC: 17; Disease Duration: 12 yrs; HAQ-DI: 1.6	TJC: 19; SJC: 12; Disease Duration: 6 yrs; KHAQ-DI: 1.4	TJC: 33; SJC: 22; Disease Duration: 6 yrs; HAQ-DI: 1.7	TJC: 25 SJC: 17 Disease Duration: 7 yrs HAQ-DI: 1.4
Time of ACR20 Evaluation	Week 24	Week 24	Week 24	Week 12	Week 12 & 24
ACR20 Response on Humira	63%	67%	62%	54%	Week 12: 60.5% Week 24: 64%
Withdrawal on Humira	22% by Week 52	7% by Week 16 (34% escaped to ADA)	9%	N.A.	15% treatment discontinuation, 6% study discontinuation by Week 24

Source: Reviewer and Amjevita Review by Dr. Yongman Kim
Abbreviations: SJC=swollen joint count; TJC=tender joint count; DMARD=disease-modifying anti-rheumatic drug; bDMARD = Biologic DMARD; EU=Europe; US=United States
[1] Based on best attempts to identify/estimate characteristics from literature review
[2] Means or medians, depending on what was reported in publication

2.6 Evaluation of Safety

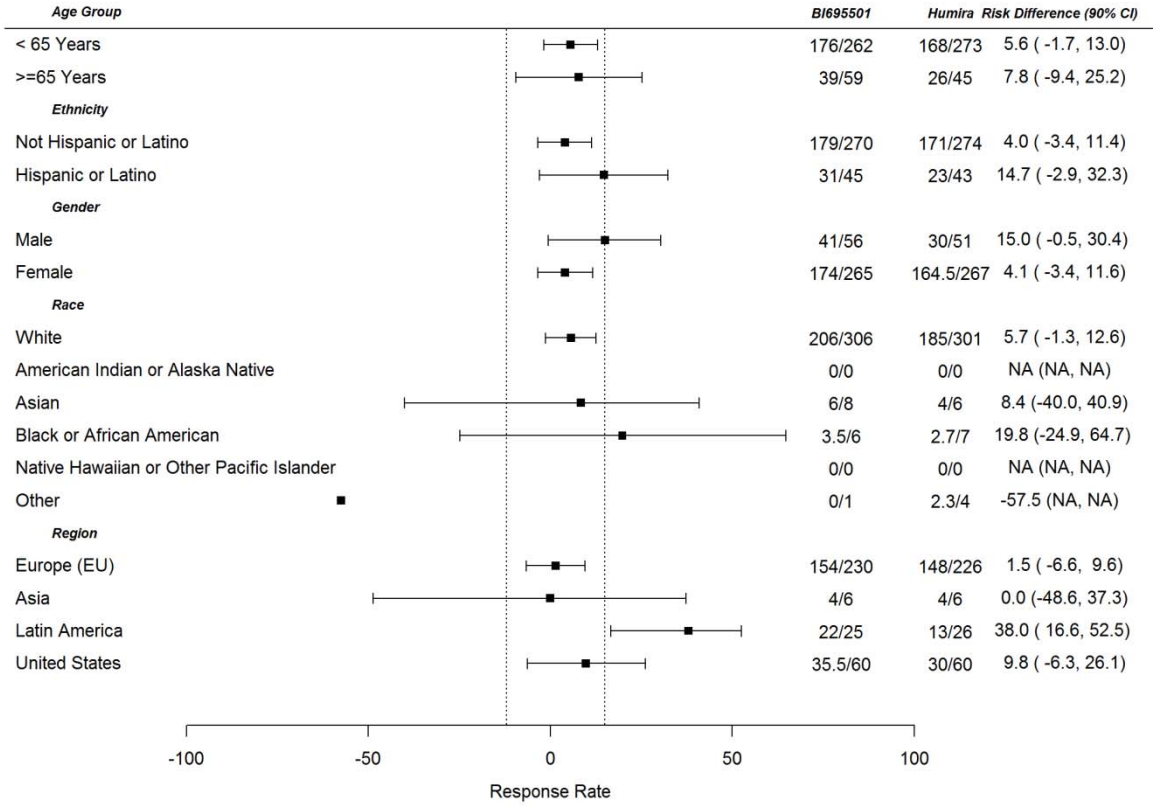
For an evaluation of the safety of this product the reader should refer to the review by Dr. Stefanie Freeman.

3 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

3.1 Gender, Race, Age, and Geographic Region

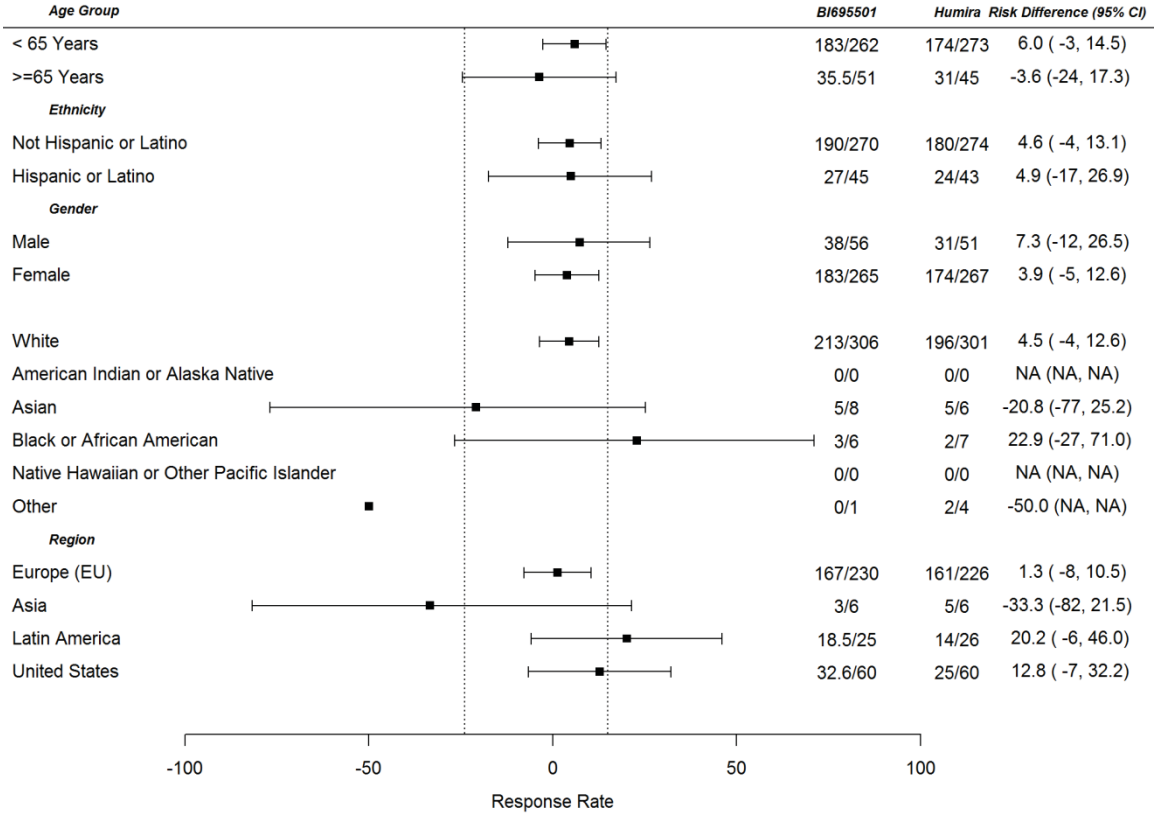
The results of the analyses of the primary endpoint by demographic subgroup are shown in Figure 1 for Week 12 and Figure 2 for Week 24. The vertical lines correspond to the similarity margins of −12% and +15%. There do not appear to be any findings that would raise concerns in these analyses.

Figure 1: Forest Plot of Week 12 ACR20 Response Rates and 90% CIs by Demographic Subgroup



Source: Applicant's Study Report Figure 14.2.4.1

Figure 2: Forest Plot of Week 24 ACR20 Response Rates and 90% CIs by Demographic Subgroup

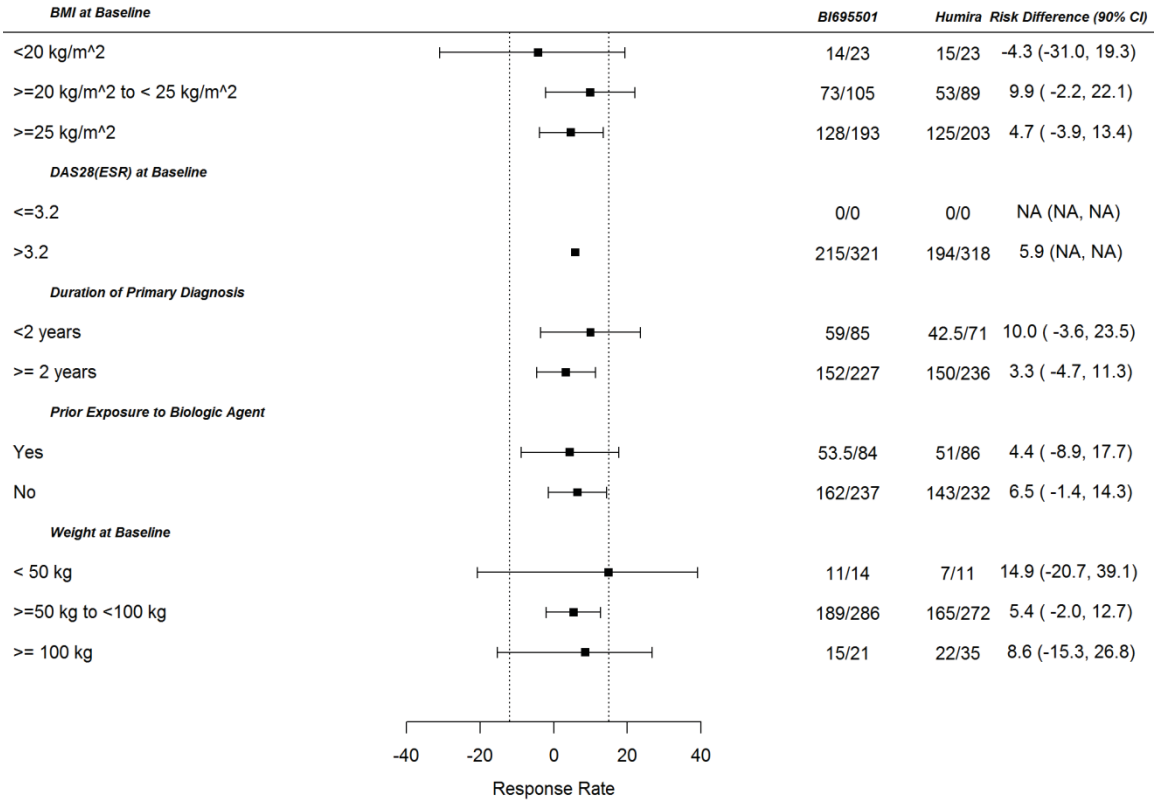


Source: Applicant's Study Report Figure 14.2.4.2

3.2 Other Special/Subgroup Populations

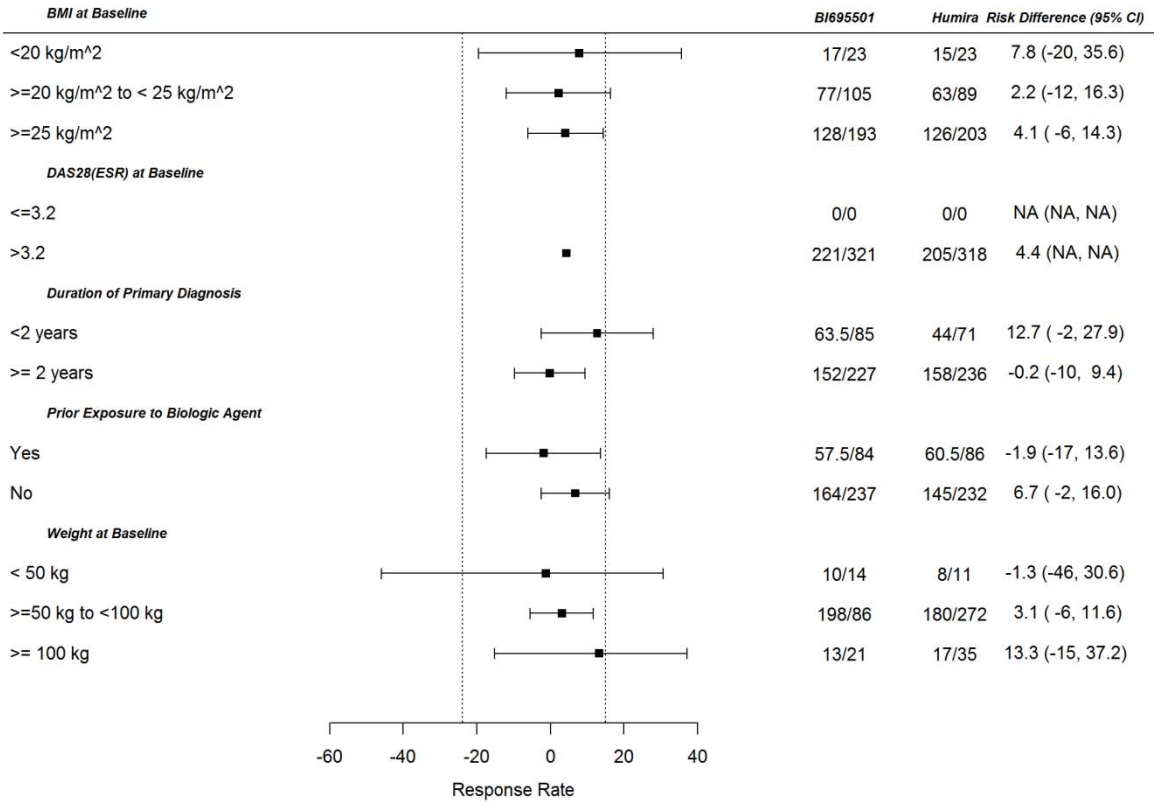
The applicant also analyzed the ACR20 response rates using several baseline disease status metrics. The results of these analyses are shown in Figure 3 for Week 12 and Figure 4 for Week 24. Again there did not appear to be any findings that would raise concern.

Figure 3: Forest Plot of Week 12 ACR20 Response Rates and 90% CIs by Baseline Disease Status Subgroup



Source: Applicant's Study Report Figure 14.2.4.1

Figure 4: Forest Plot of Week 24 ACR20 Response Rates and 90% CIs by Baseline Disease Status Subgroup



Source: Applicant's Study Report Figure 14.2.4.2

4 SUMMARY AND CONCLUSIONS

4.1 Statistical Issues

The first main statistical issue was the selection of the similarity margin for this study. This issue can be further divided into two parts,

1. The timing of the selection.
2. The justification provided.

This discussion is especially important, because the 90% confidence interval for week 12 fell outside of the margin of ±12% set near April 2015 on the request of the Agency (see Table 4 in Section 1.1). As discussed in section 1.1, the applicant sought agreement for a change in the similarity margin approximately 9 days before the first IDMC meeting and finalization of a blinded sample size re-evaluation. I do not believe that this change was not due to any findings the applicant may have had at this time as the applicant does not appear to have had any access to comparative interim results.

Next, I shall discuss what I perceive are the issues with the justification provided by the applicant for the final margin of -12% to +15%. As discussed in Section 1.1, the Agency agreed with this margin during IND development. However, the justification the applicant provided was focused on the fact that they expected the selected dose to be on the plateau of the dose-

response curve, and so they expected that additional exposure would not provide any additional efficacy. I believe that this justification is incomplete and misses an important aspect that we need to consider.

An important purpose of the upper bound of the similarity margin is to act as a surrogate for safety of the proposed biosimilar compared to the reference product. That is, if patients consistently experienced additional efficacy over the reference product, then an increase in adverse drug reactions might also be expected in the studied indication or another approved indication. If the applicant had instead explored the dose-response relationship for safety and immunogenicity of the reference product and provided evidence that additional exposure did not result in additional safety issues then the justification would have been more convincing.

In summary, I believe a justification based on the safety and efficacy dose-response relationship is acceptable in cases like this where we would not expect an increased dose to provide any additional benefits or harms. Any observed increases in efficacy would therefore most likely be due to chance, especially since in this case we don't see an increase in benefit across all the considered endpoints. Although in this case the applicant did not provide any evidence regarding the dose-response relationship for the safety or immunogenicity of the reference product, since, as reported by the medical officer, Dr. Stefanie Freeman, we did not see any clinically meaningful differences between US-licensed Humira and BI 695501 in either the safety or immunogenicity profiles it is my opinion that the approvability of this product should not be impacted.

The second statistical issue was the validity of the methodology chosen for the primary analysis. As described in Section 2.3.1 this was a novel approach that was created specifically for this application. The design of this approach appears acceptable. This is further supported by the results of an unadjusted analysis (see Table 12) which still met the criteria for similarity.

4.2 Conclusions and Recommendations

It is my conclusion that there were no meaningful differences between BI 695501 and US-licensed Humira in the single comparative clinical efficacy study performed by the applicant and so it is my recommendation that BI 695501 be approved as a biosimilar to US-licensed Humira.

In addition to meeting the pre-specified similarity margin for the ACR20 response rate there were no major differences in any of the other analyses that I or the applicant performed. These included a number of sensitivity analyses to explore the effect of missing data on the ACR20 analysis, analyses of the ACR50 and ACR70 response rates, analyses of the changes in the individual components of the ACR20 response criteria, and analyses of the Disease Activity Score 28 joint (DAS28) endpoints.

4.3 Labeling Recommendations

I have no recommendations for the labeling of this product.

5 BIBLIOGRAPHY

Chen, D. Y., Chou, S. J., Hsieh, T. Y., Chen, Y. H., Chen, H. H., Hsieh, C. W., & Lan, J. L. (2009). Randomized, double-blind, placebo-controlled, comparative study of human anti-TNF antibody adalimumab in combination with methotrexate and methotrexate alone in Taiwanese patients with active rheumatoid arthritis. *J Formos Med Assoc, 108*(4), 310-319. doi:10.1016/S0929-6646(09)60071-1

Keystone, E. C., Kavanaugh, A. F., Sharp, J. T., Tannenbaum, H., Hua, Y., Teoh, L. S., . . . Chartash, E. K. (2004). Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum, 50*(5), 1400-1411. doi:10.1002/art.20217

Kim, H.-Y., Lee, S.-K., Song, Y. W., Yoo, D.-H., Koh, E.-M., Yoo, B., & Luo, A. (2007). A randomized, double-blind, placebo-controlled, phase III study of the human anti-tumor necrosis factor antibody adalimumab administered as subcutaneous injections in Korean rheumatoid arthritis patients treated with methotrexate. *APLAR Journal of Rheumatology, 10*(1), 9-16. doi:10.1111/j.1479-8077.2007.00248.x

Pouw, M. F., Krieckaert, C. L., Nurmohamed, M. T., van der Kleij, D., Aarden, L., Rispens, T., & Wolbink, G. (2015). Key findings towards optimising adalimumab treatment: the concentration–effect curve. *Annals of the Rheumatic Diseases, 74*, 513-518.

Reeve, R. (2016). Confidence interval of difference of proportions in logistic regression in presence of covariates. *Statistical Methods in Medical Research*. doi:10.1177/0962280216631583

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman and Hall.

Weinblatt, M. E., Keystone, E. C., Furst, D. E., Moreland, L. W., Weisman, M. H., Birbara, C. A., . . . Chartash, E. K. (2003). Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum, 48*(1), 35-45. doi:10.1002/art.10697

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

/s/

JAMES E TRAVIS
07/13/2017

GREGORY P LEVIN
07/13/2017

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	761058
SERIAL NO.:	0000
DATE RECEIVED BY THE CENTER:	October 27, 2016
DRUG NAME:	BI 695501 (proposed biosimilar to US-licensed Humira® (adalimumab))
DOSAGE FORM:	Solution for Injection 40 mg/0.8ml
INDICATIONS:	Based on the registered adult and pediatric indications of US-licensed Humira: Rheumatoid Arthritis, Juvenile Idiopathic Arthritis, Ankylosing spondylitis, Psoriatic Arthritis, Psoriasis, Adult Crohn's disease and ulcerative colitis
APPLICANT:	Boehringer Ingelheim
REVIEW FINISHED:	June 20, 2017
NAME OF STATISTICAL REVIEWER:	Li Xing
NAME OF SECONDARY REVIEWER	Meiyu Shen
NAME OF PROJECT MANAGER:	Sadaf Nabavian

_____
Meiyu Shen, Ph.D., Lead Mathematical Statistician

Concur: _____
Yi Tsong, PhD, Division Director, DBVI

Distribution: CDER/OTS/OB/DB VI/Meiyu Shen
CDER/OTS/OB/DB VI/Yi Tsong
CDER/OTS/OB/Lillian Patrician
CDER/OBP/ Richard Ledwidge
CDER/OBP/ Howard A Anderson
CDER/TBBS

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY AND RECOMMENDATION

The CMC statistics reviewer in the Office of Biostatistics analyzed the comparative results of 2 critical quality attributes: sTNF-α neutralization assay and TNF-α binding assay SPR (RLCA), which were recommended for equivalence testing analysis by the Office of Biotechnology Products. Tier 1 statistical equivalence testing was conducted using equivalence margins of ±1.5 σ_R, where σ_R represents US-licensed reference product variability or the comparator variability.

Thirteen lots of BI 695501, 55 lots of US-licensed Humira, and 86 lots of EU-approved Humira were used for equivalence testing of sTNF-α neutralization assay. The results are summarized in Table 1.

Table 1 Results of equivalence testing for sTNF-neutralization assay

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
BI 695501 vs. US	(13, 55)*	-0.478	(-5.98, 5.02)	(-14.17, 14.17)	Yes
BI 695501 vs. EU	(13, 86)*	2.841	(-3.21, 8.89)	(-17.52, 17.52)	Yes
EU vs. US	(86, 55)*	-3.319	(-6.32, -0.32)	(-14.17, 14.17)	Yes

*The 90% confidence interval for the mean difference between the test and comparator is adjusted by the sample size imbalance.

Thirteen lots of BI 695501, 43 lots of US-licensed Humira, and 53 lots of EU-approved Humira are included in the TNF-α binding assay SPR (RLCA) dataset for the statistical equivalence testing. Note that SPR is the abbreiviate of the Surface Plasmon Resonance and RLCA is the abbreviate of the response level correlation assay. The results are shown in Table 2.

Table 2 Results of equivalence testing for TNF-α binding assay SPR (RLCA)

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
BI 695501 vs. US	(13, 43)*	-0.145	(-0.44, 0.15)	(-0.89, 0.89)	Yes
BI 695501 vs. EU	(13, 53)*	-0.293	(-0.57, -0.02)	(-0.80, 0.80)	Yes
EU vs. US	(53, 43)	0.148	(-0.05, 0.34)	(-0.89, 0.89)	Yes

*The 90% confidence interval for the mean difference between the test and comparator is adjusted by the sample size imbalance.

As shown in Tables 1 and 2, the results from the statistical equivalence testing of sTNF-α neutralization assay and TNF-α binding assay SPR (RLCA) support a demonstration that the

proposed biosimilar BI 695501 is highly similar to US-licensed Humira and also support the analytical bridge between US-licensed Humira and EU-approved Humira.

2 INTRODUCTION

On October 27, 2016, the applicant (Boehringer Ingelheim) submitted to the US Food and Drug Administration (FDA) a 351(k) BLA, which included an analytical similarity assessment of comparing BI 695501 and US-licensed Humira.

The applicant characterized multiple lots of US-licensed Humira and EU-approved Humira using a comprehensive set of analytical methods during the BI 695501 development.

The Agency carefully evaluated data for the sTNF-α neutralization assay and TNF-α binding assay SPR (RLCA) provided in this BLA. Our comments regarding Boehringer Ingelheim's 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

Boehringer Ingelheim submitted the analytical data on October 27, 2016. Note that in Table 4, the sTNF-α neutralization assay data of 55 US-licensed Humira lots, 86 EU-approved Humira lots, 13 BI 695501 lots were submitted by Boehringer Ingelheim.

Boehringer Ingelheim also provided and analyzed the TNF-α binding assay SPR (RLCA) for 53 lot values of EU-approved Humira, 13 lot values of BI 695501, and 43 lot values of US-licensed Humira.

Table 4 Number of lots from each product

Product	Number of lots	
	sTNF-α neutralization assay	TNF-α binding assay SPR (RLCA)
US-licensed Humira	55	43
BI 695501	13	13
EU-approved Humira	86	53

4 APPLICANT'S STATISTICAL EQUIVALENCE TESTING

In this submission, Boehringer Ingelheim conducted Tier 1 statistical equivalence testing with the margin defined as $1.5\hat{\sigma}_R$ for sTNF-α neutralization assay and TNF-α binding assay SPR (RLCA). Boehringer Ingelheim followed the analysis guidelines provided by FDA.

5 FDA STATISTICAL ANALYSES

To evaluate analytical similarity, the Agency recommended that Boehringer Ingelheim apply a tiered approach in the Agency's responses to IND meetings with Boehringer Ingelheim. That is, product quality attributes 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. Quality attributes 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 (Mean – X*SD, Mean + X*SD) defined by the reference product. Here, the multiplier X typically ranges from 2 to 4. The quality attributes 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 means of the quality attribute for the test product and the population mean of the quality attribute for the US-licensed Herceptin product. Let σ_R be the standard deviation of the quality attribute of interest for the US-licensed Herceptin. In order to conclude the equivalence in the quality attribute of interest between the test product and the US-licensed Herceptin product, we aim to reject the null hypothesis of the following null and alternative hypotheses:

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

$$H_1 : \theta_1 < \mu_T - \mu_R < \theta_2$$

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

We reject H_0 if the 90% confidence interval for the mean difference in the quality attribute of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. In other words, we conclude the equivalence in the quality attribute of interest between the test product and the US-licensed Herceptin product if the 90% confidence interval for the mean difference in the quality attribute 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 US-licensed Herceptin 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 biosimilar lots and 10 US-licensed Herceptin lots. First we estimate σ_R by the sample variability of the US-licensed Herceptin product (or by the sample variability of EU-approved Herceptin in the comparison between MYL-1401O and EU-approved Herceptin) and then in the statistical analysis, θ_1 and θ_2 are treated as a constant, not a random variable.

Let X_{Tj} be the observed value of the quality attribute of interest for Batch *j* of the test product (the proposed biosimilar product) and X_{Rj} be the observed value of the quality attribute of interest for Batch *j* of the US-licensed Herceptin product. Since the two products are manufactured by two manufacturers, two groups are independent. $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij} \Big/ n_i$, and $S_i^2 = \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i\right) \Big/ \left(n_i - 1\right)$, 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 US-licensed Herceptin product, the (1-2α)*100% confidence interval of the mean difference in the quality attribute of interest can be calculated as:

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

Here $t_\alpha(\nu)$ 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% confidence interval of the mean difference in the quality attribute of interest can be calculated as:

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

Here $n_R^* = \min(n_R, 1.5n_T)$ 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 the number of biosimilar lots, n_T, is 50% more than the number of reference lots, n_R, we can apply a similar approach as above with $n_T^* = \min(1.5 \times n_R, n_T)$ for the confidence interval calculation. In the following analyses, we use α=0.05.

5.2 FDA statistical equivalence testing for sTNF-α neutralization assay

The sTNF-α neutralization assay data points of BI 695501, US-licensed Humira, and EU-approved Humira are displayed in Figure 1. There is no obvious appearance of mean difference among the 3 products.

Thirteen lots of BI 695501, 55 lots of US-licensed Humira, and 86 lots of EU-approved Humira are included for the statistical equivalence testing for the sTNF-α neutralization assay. Descriptive statistics for the sTNF-α neutralization assay data are listed in Table 5.

Figure 1 Scatter plot of sTNF-α neutralization assay for US-licensed Humira, BI 695501, and EU-approved Humira

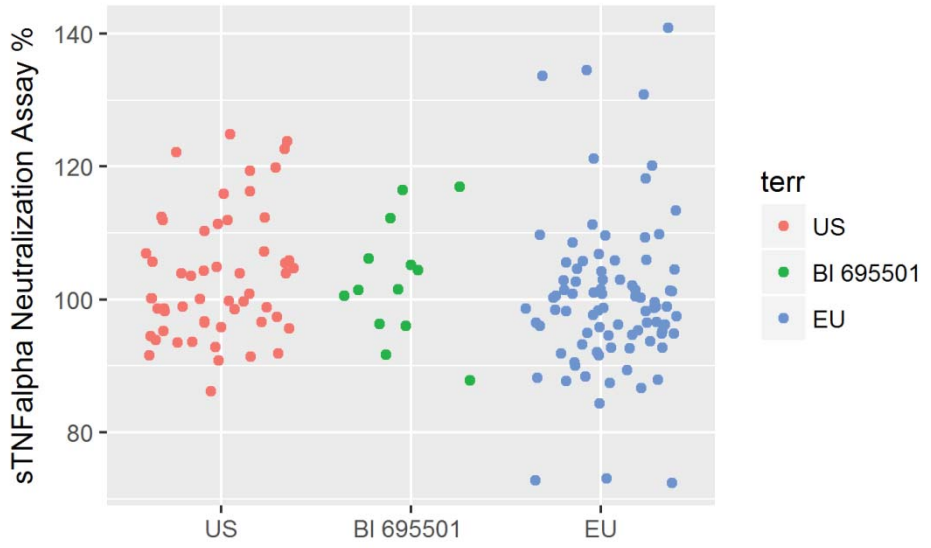


Table 5 Descriptive statistics for the sTNF-α neutralization assay data

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Humira	55	103.29	9.44	86.10	124.80
BI 695501	13	102.81	8.83	87.80	116.90
EU-approved Humira	86	99.97	11.68	72.40	140.80

Since we don't assume equal variance of test and reference products, we use Satterthwaite approximation for obtaining 90% confidence interval for the mean difference between US-licensed Humira and BI 695501. From Table 6, it is seen that the sTNF-α neutralization assay of BI 695501 is equivalent to the sTNF-α neutralization assay of US-licensed Humira. Similarly, the sTNF-α neutralization assay of BI 695501 is equivalent to the sTNF-α neutralization assay of EU-approved Humira, and the sTNF-α neutralization assay of EU-approved Humira is equivalent to the sTNF-α neutralization assay of US-licensed Humira.

Table 6 Equivalence testing results for the sTNF-α neutralization assay

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
BI 695501 vs. US	(13, 55)*	-0.478	(-5.98, 5.02)	(-14.17, 14.17)	Yes
BI 695501 vs. EU	(13, 86)*	2.841	(-3.21, 8.89)	(-17.52, 17.52)	Yes
EU vs. US	(86, 55)*	-3.319	(-6.32, -0.32)	(-14.17, 14.17)	Yes

*The 90% confidence interval for the mean difference between the test and comparator is adjusted by the sample size imbalance.

5.3 FDA statistical equivalence testing for TNF-α binding assay SPR (RLCA)

The TNF-α binding assay SPR (RLCA) data points of BI 695501, US-licensed Humira, and EU-approved Humira are displayed in Figure 2.

Thirteen lots of BI 695501, 43 lots of US-licensed Humira, and 53 lots of EU-approved Humira are included in the TNF-α binding assay SPR (RLCA) dataset for the statistical equivalence testing. Descriptive statistics for the TNF-α binding assay SPR (RLCA) data of BI 695501, US-licensed Humira, and EU-approved Humira are listed in Table 7.

From Table 8, it is seen that the equivalence of TNF-α binding assay SPR (RLCA) between BI 695501 and US-licensed Humira is supported. The equivalence of TNF-α binding assay SPR (RLCA) between BI 695501 and EU-approved Humira is supported. The equivalence of TNF-α binding assay SPR (RLCA) between US-licensed Humira and EU-approved Humira is supported.

Figure 2 Scatter plot of TNF-α binding assay SPR (RLCA) for US-licensed Humira, BI 695501, and EU-approved Humira

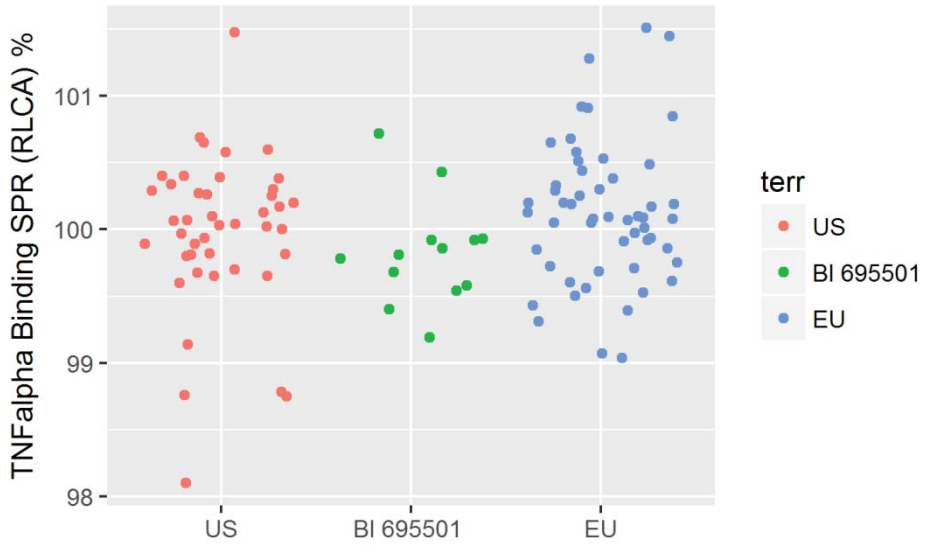


Table 7 Descriptive statistics for the TNF-α binding assay SPR (RLCA) data

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Humira	43	99.97	0.59	98.10	101.48
BI 695501	13	99.83	0.40	99.19	100.72
EU-approved Humira	53	100.12	0.54	99.03	101.51

Table 8 Equivalence testing results for the TNF-α binding assay SPR (RLCA)

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
BI 695501 vs. US	(13, 43)*	-0.145	(-0.44, 0.15)	(-0.89, 0.89)	Yes
BI 695501 vs. EU	(13, 53)*	-0.293	(-0.57, -0.02)	(-0.80, 0.80)	Yes
EU vs. US	(53, 43)	0.148	(-0.05, 0.34)	(-0.89, 0.89)	Yes

*The 90% confidence interval for the mean difference between the test and comparator is adjusted by the sample size imbalance.

6 CONCLUSION AND RECOMMENDATION

The results from the statistical equivalence testing of the sTNF-α neutralization assay and the TNF-α binding assay SPR (RLCA) support a demonstration that the proposed biosimilar BI 695501 is highly similar to US-licensed Humira. The statistical analyses of the sTNF-α neutralization assay and the TNF-α binding assay SPR (RLCA) in the three pair-wise comparisons (BI 695501, US-licensed Humira, and EU-approved Humira) also support the scientific bridge to justify the relevance of the data obtained from clinical studies that compared EU-approved Humira and the BI 695501 product to support a demonstration of biosimilarity to US-licensed Humira.

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

/s/

LI XING
06/23/2017

MEIYU SHEN
06/23/2017

YI TSONG
06/30/2017