

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**761024Orig1s000**

**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 761-024

**Drug Name:** ABP 501

**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:** Amgen

**Date(s):** Submitted: November 25, 2015  
BsUFA: September 25, 2016

**Review Priority:** Standard

**Biometrics Division:** Division of Biometrics II

**Statistical Reviewer:** Yongman Kim, Ph.D.

**Concurring Reviewer:** Gregory Levin, Ph.D. (Statistics Team Leader)

**Statistical Supervisor:** Thomas Permutt, Ph.D. (Division Director)

**Medical Division:** Division of Pulmonary, Allergy, and Rheumatology Products

**Clinical Team:** Keith Hull, 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

<b>LIST OF TABLES</b> .....	<b>3</b>
<b>LIST OF FIGURES</b> .....	<b>3</b>
<b>1 EXECUTIVE SUMMARY</b> .....	<b>4</b>
<b>2 INTRODUCTION</b> .....	<b>5</b>
2.1 OVERVIEW .....	5
2.2 HISTORY OF PRODUCT DEVELOPMENT .....	5
2.3 SPECIFIC STUDIES REVIEWED .....	6
2.4 DATA SOURCES .....	7
<b>3 STATISTICAL EVALUATION</b> .....	<b>7</b>
3.1 DATA AND ANALYSIS QUALITY .....	7
3.2 STUDY DESIGN .....	7
3.3 STATISTICAL METHODOLOGIES .....	8
3.3.1 <i>Planned Analyses</i> .....	8
3.3.2 <i>Additional Reviewer Analyses</i> .....	9
3.3.3 <i>Similarity Margin for Study 20120262</i> .....	10
3.4 EVALUATION OF EFFICACY .....	11
3.4.1 <i>Patient Disposition, Demographic, and Baseline Characteristics</i> .....	11
3.4.2 <i>Key results in Study 20120262</i> .....	13
3.4.3 <i>Assay Sensitivity and the Constancy Assumption</i> .....	16
3.4.4 <i>Potential Effect of Missing Data</i> .....	18
3.5 EVALUATION OF SAFETY .....	21
<b>4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS</b> .....	<b>21</b>
<b>5 SUMMARY AND CONCLUSIONS</b> .....	<b>22</b>
5.1 STATISTICAL ISSUES .....	22
5.2 COLLECTIVE EVIDENCE .....	24
<b>APPENDIX</b> .....	<b>25</b>
<b>REFERENCES</b> .....	<b>29</b>

## LIST OF TABLES

Table 1. Overview of Key Clinical Study.....	7
Table 2. Historical Effect of Adalimumab on ACR20 Response in Randomized Clinical .....	11
Table 3. Baseline Characteristics in RA Patients in Study 20120262 .....	12
Table 4. Patient Dropout, by Reason for Withdrawal, in Study 20120262 .....	13
Table 5. Protocol-Specified Primary Analysis: Proportions of Responders with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262 .....	14
Table 6. FDA-Suggested Primary Analysis: Proportions of Responders, and Distributions of Reasons for Non-Response, with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262 .....	15
Table 7. Mean Changes from Baseline in the ACR Components and DAS28-CRP at Week.....	15
Table 8. Per-Protocol Analysis: Proportions of Responders with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262 .....	16
Table 9. Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials <sup>1</sup> of Adalimumab in RA and Comparative Clinical Study.....	17
Table 10. Tipping Point Analysis in Study 20120262: Inference on the Difference Between ABP 501 and US-licensed Humira in the Probability of Week 24 ACR20 Response under Varying Assumptions About the Differences on Each Treatment Arm Between Responses in Patients who Withdrew from the Study Early and Responses in Patients who Completed the Study .....	19
Table 11. Tipping Point Analysis in Study 20120262: Inference on the Difference Between.....	20

## LIST OF FIGURES

Figure 1. Patient Withdrawal over Time in Study 20120262 (Source: Reviewer).....	13
Figure 2. ACR20/50/70 Response <sup>1</sup> Probabilities over Time in Study 20120262.....	14
Figure 3. Estimated Differences Between ABP 501 and US-licensed Humira in the Probability of Remaining in the Study and Achieving an ACR20 Response at Week 24, Stratified by Selected Subgroups, in Study 20120262. Solid Vertical Line Represents No Difference. (Source: Reviewer) .....	22
Figure 4. Mean Disease Activity Score (DAS28-CRP) among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer) .....	25
Figure 5. Mean Swollen Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer).....	25
Figure 6. Mean Tender Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer).....	26
Figure 7. Mean Health Assessment Questionnaire (HAQ) Physical Ability Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer).....	26
Figure 8. Mean Patient Pain Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer).....	27
Figure 9. Mean Patient Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer) .....	27
Figure 10. Mean Physician Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer) .....	28
Figure 11. Empirical Distribution Function for Change from Baseline in Disease.....	28

## 1 EXECUTIVE SUMMARY

This review considers the therapeutic protein product ABP 501 as a potential biosimilar to US-licensed Humira (adalimumab). We focus on Study 20120262, a 24-week, randomized, double-blind, parallel-group clinical trial that compared the efficacy and safety of ABP 501 and US-licensed Humira in 526 patients with active rheumatoid arthritis (RA) who had an inadequate response to methotrexate.

In Study 20120262, the primary endpoint was the proportion of patients who remained in the study and achieved an American College of Rheumatology 20% (ACR20) response at Week 24. Approximately 71.2% of patients randomized to ABP 501 and 72.1% of patients randomized to US-licensed Humira were ACR20 responders, for an estimated absolute difference between treatments of -0.4% (90% confidence interval [CI]: -6.8%, +6.1%). The 90% CI successfully ruled out the similarity margin of  $\pm 12\%$  that the Agency has determined reasonable. ACR20, ACR50, and ACR70 responses over time, in addition to mean changes from baseline in the components of the ACR composite endpoint, and the disease activity score (DAS28-CRP), were also similar between the treatment arms.

Patients who discontinued treatment early were also withdrawn from the clinical studies. Approximately 6% of randomized patients failed to complete the 24-week double-blind treatment, which was relatively low when compared to typical RA trials. But the dropout led to missing data in important analyses, such as the evaluations of ACR20 and DAS28-CRP at Week 24 in all randomized patients regardless of adherence. Therefore, we assessed tipping point analyses to explore the sensitivity of results to violations in assumptions about the missing data. Confidence intervals for the differences between ABP 501 and US-licensed Humira successfully ruled out concerning losses in efficacy under the plausible range of assumptions about outcomes among patients who dropped out on ABP 501 and on US-licensed Humira. 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 should 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 four published historical, randomized, placebo-controlled clinical trials of adalimumab, we concluded that (1) the design of the historical trials were largely similar to that of the comparative clinical Study 20120262; and (2) there were relatively large and consistent treatment effects across the four historical studies. We did not identify any issues with the quality of study conduct, with the exception of the differing rates of study withdrawal between the two arms (8% for ABP 501 vs. 4% for US-licensed Humira), likely by random chance. The totality of available information supports the assay sensitivity of Study 20120262.

## 2 INTRODUCTION

### 2.1 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 ABP 501 as a biosimilar to US-licensed Humira (adalimumab). Section 351(i) of the PHS Act defines biosimilarity to mean “that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components” and that “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.” As noted in the FDA guidance for industry *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product* [1], protein products are typically more complex than small molecule drugs and analytical methods may not be able to identify all relevant structural differences between the proposed biosimilar and the reference product. Because even minor differences in structure (e.g., higher order structure such as protein folding) may significantly affect safety, purity, or potency, comparative data from clinical studies designed to rule out important differences in safety and efficacy will often need to be part of the evaluation of biosimilarity.

Adalimumab is a monoclonal antibody that inhibits the activity of tumor necrosis factor (TNF), an inflammatory cytokine thought to play a role in many disease processes. Adalimumab was first approved in the United States in 2002 and is currently indicated for the treatment of adult and pediatric Crohn's disease (CD), ulcerative colitis, rheumatoid arthritis (RA) in combination with methotrexate, juvenile idiopathic arthritis (JIA) in patients 2 years of age and older, ankylosing spondylitis (AS), psoriatic arthritis (PsA), plaque psoriasis, hidradenitis suppurativa (HS), and uveitis (UV) . The approved dose for treatment of RA, AS, and PsA is 40 mg/kg every other week. The approved dose for JIA is 20 mg every other week for patients with weight ranging from 15 kg to 30 kg and 40 mg every other week for patients with weight greater than 30 kg. The approved dose for CD, ulcerative colitis, and hidradenitis suppurativa is 160 mg at Day 1 and 80 mg at Day 15, followed by 40 mg every other week. The approved dose for plaque psoriasis is 80 mg at Day 1, followed by 40 mg every other week.

The applicant has submitted results from several nonclinical, analytical, and clinical studies to support the biosimilarity of ABP 501 to US-licensed Humira. The proposed indications for ABP 501 sought by Amgen are: 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), and Plaque Psoriasis (Ps). This review primarily considers the efficacy evaluation of ABP 501 in clinical Study 20120262.

### 2.2 History of Product Development

The clinical development program for ABP 501 was introduced to the Division of Pulmonary, Allergy, and Rheumatology Products under IND 111,714. Following are descriptions of several

interactions with the applicant during product development, which are potentially relevant to this review.

At a Pre-IND Type B meeting in August 2011, FDA recommended that the applicant use a more sensitive endpoint (i.e., continuous variable) such as Hybrid ACR, DAS28, or ACRn for the comparative clinical study. FDA also recommended the use of a 2-sided comparative efficacy analysis for the comparative clinical study. At a Biosimilar Biological Product Development (BPD) Type 2 meeting in May 2013, FDA recommended that if the applicant proceeds with an equivalence trial design as proposed, the applicant should either utilize an endpoint such as ACR20 for which there are data available to justify an equivalence margin or provide a scientific justification for the proposed equivalence margin for DAS28. FDA also recommended that the applicant evaluate several different time points early in treatment, e.g., weeks 1, 2, 3, 4, 6, 8, etc., as secondary endpoints. At a BPD Type 2 meeting in January 2015, FDA stated that the use of last observation carried forward (LOCF) to impute missing ACR20 data at Week 24 is not acceptable. LOCF relies on the strong and unverifiable assumption that patient outcomes prior to withdrawal would have remained constant through Week 24. In addition, as a single-imputation approach, LOCF does not appropriately take into account the uncertainty in the imputation process. FDA also acknowledged that RA Study 20120262 enrollment was complete, the database was locked in January 2015 and the study had been unblinded; hence it was impracticable to make changes to the protocol or statistical analysis plan (SAP) at the time of the meeting. FDA requested that the applicant provide data from historical randomized clinical trials of adalimumab to justify the adequacy of the proposed similarity margin of (0.738, 1/0.738) for the ratio of ACR20 responses. FDA recommended that the similarity margin based on the absolute difference scale for the proposed comparative clinical study in rheumatoid arthritis be no greater in magnitude than  $\pm 12\%$ . The proposed margin of  $\pm 12\%$  was based on considerations aimed at weighing the clinical importance of various differences in effect against the feasibility of different study sizes. FDA also recommended that a margin based on the absolute difference scale be used, as it is considered more important than other metrics, such as risk ratio, from a clinical perspective for an evaluation of benefit-risk. As an Information Request after filing, FDA requested that the applicant examine the potential effects of missing data on the applicant's results using tipping point sensitivity analyses for the primary endpoint.

### **2.3 Specific Studies Reviewed**

The applicant has submitted results from two completed comparative clinical studies. Study 20120262 was a randomized, double-blind, parallel-group clinical trial to compare the efficacy of ABP 501 and US-licensed Humira in 526 patients with active RA who had an inadequate response to methotrexate (MTX). Study 20120263 was a randomized, double-blind, active comparator-controlled clinical trial to compare the immunogenicity, safety and efficacy of ABP 501 and US-licensed Humira in 350 patients with moderate to severe plaque psoriasis. Our evaluation of the similarity of ABP 501 and US-licensed Humira centers on Study 20120262, the randomized, double-blind comparative study in RA patients, the comparative clinical study in which a comparison of efficacy and safety was the primary objective. Readers

are referred to the statistical review of Dr. Kathleen Fritsch for a summary of results from Study 20120263. Table 1 provides a summary of the comparative clinical study that is the focus of this review.

Table 1. Overview of Key Clinical Study

Study	Population	Design	Treatment Arms	Number of Patients	Dates*
20120262	RA	24-week, R, DB, PG	ABP 501 US-licensed Humira	264 262	10/2013- 11/2014

Source: Reviewer

\*Dates correspond to the start and the end of the study.

Abbreviations: RA = rheumatoid arthritis; R = randomized; DB = double-blind; PG = parallel group

## 2.4 Data Sources

Data were submitted by the applicant to the CDER electronic data room in SAS transport format. Protocols, correspondence, data listings, program code, and study reports were accessed under the network path <\\cdsesub1\evsprod\bla761024\761024.enx>.

## 3 STATISTICAL EVALUATION

### 3.1 Data and Analysis Quality

The submitted datasets were of acceptable quality and were adequately documented. We were able to reproduce the results of all important primary and secondary analyses.

### 3.2 Study Design

Study 20120262 was a 24-week, randomized, double-blind, parallel-group clinical trial to compare the safety and efficacy of ABP 501 and US-licensed Humira in 526 patients with active rheumatoid arthritis despite treatment with methotrexate. The study consisted of patients of ages 18 to 80 years who had been diagnosed with RA, as determined by meeting 2010 American College of Rheumatology (ACR) or European League Against Rheumatism (EULAR) classification criteria for at least 3 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 one of the following: 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 12 consecutive weeks, with a stable dose (7.5 to 25 mg/week) for at least 8 weeks, and they also received folic acid during the study. Patients previously treated with two or more biological therapies for RA or who had received disease-modifying antirheumatic drugs (DMARDs) other than methotrexate (e.g., leflunomide, cyclosporine, azathioprine, or cyclophosphamide) in the past 4 weeks were excluded. Subjects were randomized 1:1 to ABP 501 or US-licensed Humira administered via subcutaneous (SC) injection at a dose of 40

mg every 2 weeks until week 22. No dose reductions or changes were allowed. Randomization was stratified by region (Eastern Europe versus Western Europe versus North & Latin America) and prior biologic use for RA (with prior biologic use capped at 40% of the study population).

Withdrawal from the treatment was equivalent to withdrawal from the study because patients who stopped taking the therapy early were not followed up for safety and efficacy assessment for the remainder of the 24-week treatment period. Possible protocol-specified reasons for withdrawal included adverse event, loss to follow-up, significant protocol violation, and withdrawal of consent from the study. If possible, an early withdrawal visit was conducted no later than 2 weeks after the last dose of study medication. The many potential reasons for stopping treatment, combined with the fact that the applicant did not continue to collect information on patients who stopped therapy early, led to missing data in intention-to-treat safety and efficacy analyses (see 5.1 for further discussion).

The pre-specified primary efficacy endpoint was the proportion of patients achieving an ACR20 response at Week 24. An ACR20 response was defined as at least 20% improvement from baseline in both the tender and swollen joint counts, in addition to at least 20% improvement in at least three of the following: patient assessment of pain on a visual analog scale (VAS), patient global assessment of disease status (VAS), physician global assessment of disease status (VAS), Health Assessment Questionnaire Disability Index (HAQ-DI), and serum C-reactive Protein (CRP) concentration. Secondary efficacy endpoints included the components used to define ACR20 response, the Disease Activity Score in 28 joints with CRP (DAS28-CRP), ACR50 response, and ACR70 response. Most were evaluated at Weeks 2, 4, 8, 12, 18, and 24.

### **3.3 Statistical Methodologies**

#### **3.3.1 Planned Analyses**

In Study 20120262, a sample size of 500 patients was planned to rule out a similarity margin of (0.738, 1/0.738) in terms of risk ratio at the 5% overall significance level with 90% power under the alternative hypothesis of no difference, assuming a response rate of 63% in both groups and 15% dropout by week 24. The primary analysis was based on a log-binomial regression model adjusting for region and prior biologic use in which the null hypothesis would be rejected if the 90% confidence interval (CI) for the ratio in ACR20 response proportions was contained within the similarity margin. The last observation carried forward (LOCF) approach was used to impute missing data for patients who discontinued treatment early (and therefore the study, as well), or had missing or incomplete data for the evaluation of ACR20 at Week 24.

The applicant also carried out a supportive analysis that FDA suggested during regulatory interactions, in which the difference in ACR20 response proportions was recommended as the main metric with a similarity margin of  $\pm 12\%$ , and patients who withdrew early were treated as non-responders (see 3.3.3 for additional discussion). The analysis was based on a binomial regression model with identity-link function adjusting for region and prior biologic use.

Analyses of ACR20, ACR50, and ACR70 responses were also based on the log-binomial regression model adjusting for region and prior biologic use. Mean changes from baseline in DAS28-CRP were evaluated by a mixed-effects model repeated measures (MMRM) with region and prior biologic use, baseline scores, visit week, treatment, and treatment-by-visit interaction.

All analyses were carried out in both the all-randomized population and the per-protocol population. The per-protocol population was defined as patients who completed the treatment period and did not have a protocol violation that would affect evaluation of the primary objective of the study. The following were considered major protocol deviations: mis-stratification at randomization, missing baseline and/or week 24 ACR measures, noncompliance of inclusion/exclusion criteria, inappropriate joint count and/or ESR/CRP, and receipt of certain protocol-prohibited medications.

### **3.3.2 Additional Reviewer Analyses**

We conducted several additional analyses to support those carried out by the applicant. The applicant's planned primary analysis was specified in 2011 and was based on comparing a 90% confidence interval for the ratio in Week 24 ACR20 responses to a similarity margin of (0.738, 1/0.738). FDA recommendations for these studies were under discussion and had not been established at that time. In 2011, FDA agreed to the applicant's proposal. Further discussion of this protocol occurred in 2013 and 2015. In 2015, FDA's thinking on similarity studies had evolved and recommendations regarding the use of the absolute risk difference scale and a 12% margin were made. The applicant did not incorporate these recommendations into the protocol since the recommendations were received after database lock. At the time of this review, we do not agree with the similarity scale and margin and the LOCF missing data handling approach. In RA, FDA prefers the absolute difference scale because it is the most clinically relevant scale for a benefit risk evaluation and directly reflects the public health impact. In addition, the absolute difference in ACR20 is used for phase 3 trials of new drugs and biologics in RA, so it is well understood and accepted by clinicians. The LOCF method for missing data is generally not appropriate since it relies on strong and unverifiable assumptions. Therefore, we (and the applicant) undertook an additional supportive analysis using a similarity margin of  $\pm 12\%$  for the risk difference instead of risk ratio and treating dropouts as non-responders.

The applicant performed limited secondary analyses. Therefore, we carried out several additional supportive analyses that we considered important. We compared mean changes from baseline in important continuous secondary efficacy endpoints using linear regression models adjusting for the baseline value of the endpoint and the stratification factors. These endpoints included the ACR components and DAS28-CRP. Such continuous endpoints may be more sensitive to small but important differences between treatments in efficacy than the primary binary ACR response endpoint. In addition, we gave importance to endpoints that directly measure how patients function or feel in daily life, such as the tender and swollen joint counts and HAQ-DI score in

RA. Although the primary ACR20 endpoint is largely composed of such direct measures, it is also based on the changes in CRP, which is a surrogate endpoint.

We also compared the utility of the two treatments by presenting empirical distribution function plots for these continuous endpoints in which patients who discontinued the assigned treatment were assigned the worst outcomes.

We carried out all key analyses in all randomized patients to evaluate mean differences between treatment groups at key time points in all randomized patients regardless of adherence to the treatment or to the protocol (i.e., the intention-to-treat or de facto estimand). We also carried out analyses in the per-protocol population to evaluate mean differences between treatment groups at key time points in the subset of patients who tolerate and adhere. Draft FDA Guidance [2] and ICH guidelines [3] indicate that the evaluation of both estimands is important in the context of a study designed to establish similarity between treatments. The de facto evaluation is critical because, unlike the per-protocol evaluation, it preserves the integrity of randomization and therefore guarantees reliable inference regarding possible differences in effects of the treatment strategies (if there are no missing data). However, in the presence of true differences between treatments, the per-protocol difference may be larger and easier to detect than the de facto difference because of the restriction to the subsets of patients who adhere.

Because patients were not followed after treatment discontinuation, there were missing outcome data at Week 24 in the comparative clinical study. Therefore, evaluations of de facto estimands based on data with LOCF imputation rely on untestable assumptions about the unobserved missing values at the follow-up time of interest (e.g., 24 weeks). This assumption may not be plausible given the known efficacy of adalimumab and the fact that early symptomatic improvement on treatment within a patient who does not tolerate or adhere to the treatment regimen might go away within a few weeks of treatment discontinuation. In addition, the subsets of patients who withdrew from the study on the two treatment arms may have been inherently different with respect to important, unmeasured prognostic characteristics, thus leading to different future (unobserved) outcomes. Furthermore, FDA suggested an additional approach treating dropouts as non-responders, but this analysis also has a limitation (see 3.4.4).

Therefore, we carried out additional analyses to explore the sensitivity of results to violations in the assumptions about the missing data. We also requested the applicant to conduct tipping point analyses to determine how much worse outcomes in patients who discontinued early on ABP 501 (relative to ABP 501 completers) would have to be than outcomes in dropouts on US-licensed Humira (relative to US-licensed Humira completers) such that there would be a concerning difference in efficacy. This allows for a follow-up discussion of the plausibility of those assumptions under which the conclusions change.

### **3.3.3 Similarity Margin for Study 20120262**

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 pre-specified a similarity margin of (0.738, 1/0.738) with respect to the risk ratio. The applicant provided justification for the margin based on historical data from a randomized clinical trial of adalimumab (Keystone[4]) and the goal of preserving at least 50% of the effect size of the reference product. We do not agree with the applicant's selection of historical studies, as three important studies [5-7] are not included in the meta-analysis, and we do not agree with the proposed (0.738, 1/0.738) margin. Furthermore, we consider the risk difference metric as more important. We believe that a margin of  $\pm 12\%$  for the risk difference is more appropriate.

Our selection of a  $\pm 12\%$  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 roughly 50% of conservative estimates of treatment effect sizes relative to placebo for adalimumab (Table 2).

Table 2. Historical Effect of Adalimumab on ACR20 Response in Randomized Clinical Trials of Patients with Active RA Despite Treatment with Methotrexate (MTX)

Study	Week	MTX + Placebo		MTX + Adalimumab		Difference in % Response
		N	ACR Response	N	ACR Response	
Keystone [4]	24	200	30%	207	63%	34%
Weinblatt [5]	24	62	15%	67	67%	53%
Kim [6]	24	63	37%	65	62%	25%
Chen [7]	12	12	33%	35	54%	21%
Meta-Analysis (fixed effects <sup>1</sup> ): Difference (95% CI)						35.0% (28.2%, 41.9%)
Meta-Analysis (random effects <sup>2</sup> ): Difference (95% CI)						35.4% (22.5%, 48.2%)
Heterogeneity p-value						0.04

Source: Reviewer

<sup>1</sup> Based on Mantel-Haenszel weights

<sup>2</sup> Based on DerSimonian-Laird weights

### 3.4 Evaluation of Efficacy

#### 3.4.1 Patient Disposition, Demographic, and Baseline Characteristics

Baseline characteristics for Study 20120262 are presented in Table 3. There were no large imbalances in the distributions of baseline characteristics across the treatment arms. In the study, there were 526 subjects enrolled at 92 sites in 12 countries worldwide. Ninety-five percent of patients were White, 81% were female, and the mean age was 56 years. The average swollen and tender joint counts were 14 and 24, respectively, and the average disease activity score (DAS28-CRP; scale: 0 - 10) was 5.7.

As described previously, the design of the clinical study was such that subjects who stopped treatment early were also withdrawn from the study. There were many pre-specified reasons for withdrawal, such as adverse event, lack of efficacy, and protocol deviation. As a result, there were patient dropouts. The proportions of patients withdrawing over time in Study 20120262 are displayed by treatment group in Figure 1. Approximately 6% of all randomized patients failed to complete the 24-week double-blind treatment period and the dropout rate of ABP 501 arm (8%) was higher than the rate of US-licensed Humira arm (4%) in Study 20120262 (Table 4). The distributions of reasons for dropout were largely similar between ABP 501 and US-licensed Humira in the study. There was slightly higher dropout due to adverse events on ABP 501 (2%) than US-licensed Humira (1%) in the study, but such small differences would not be unusual by random chance if there was no true difference between treatments.

Table 3. Baseline Characteristics in RA Patients in Study 20120262

	ABP 501	US-licensed Humira	Overall
N	264	262	526
Female	214 (81%)	212 (81%)	426 (81%)
Age (years)	55.4 (11.9)	56.3 (11.5)	55.9 (11.7)
Age Group (years)			
< 35	15 (6%)	12 (5%)	27 (5%)
35-50	64 (24%)	58 (22%)	122 (23%)
50-65	126 (48%)	127 (48%)	253 (48%)
≥ 65	59 (22%)	65 (25%)	124 (24%)
Race			
White	251 (95%)	249 (95%)	500 (95%)
Black	9 (3%)	12 (4%)	21 (3%)
Asian	3 (1%)	0 (0%)	3 (1%)
Other	1 (1%)	1 (1%)	2 (1%)
Weight (kg)	74.9 (15.3)	76.9 (17.0)	75.9 (16.2)
Height (cm)	164.1 (8.8)	165.8 (9.3)	164.9 (9.1)
BMI (kg/m <sup>2</sup> )	27.8 (5.3)	27.9 (5.6)	27.9 (5.4)
Region			
Eastern Europe	169 (64%)	168 (64%)	337 (64%)
Western Europe	22 (8%)	20 (8%)	42 (8%)
North and Latin America	73 (28%)	74 (28%)	147 (28%)
Swollen Joint Count	14.7 (9.1)	14.1 (8.0)	14.4 (8.5)
Tender Joint Count	24.3 (14.4)	23.9 (13.5)	24.1 (13.9)

HAQ-DI Score	1.5 (0.6)	1.5 (0.6)	1.5 (0.6)
Patient Pain Score	58.3 (21.8)	60.6 (22.4)	59.5 (22.1)
Patient Global Assessment	6.5 (1.9)	6.6 (1.9)	6.5 (1.9)
Physician Global Assessment	6.8 (1.3)	6.7 (1.6)	6.8 (1.5)
CRP (mg/dL)	13.9 (20.7)	14.7 (19.4)	14.3 (20.0)
DAS28-CRP	5.7 (0.9)	5.7 (0.9)	5.7 (0.9)

Source: Reviewer

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

Figure 1. Patient Withdrawal over Time in Study 20120262 (Source: Reviewer)

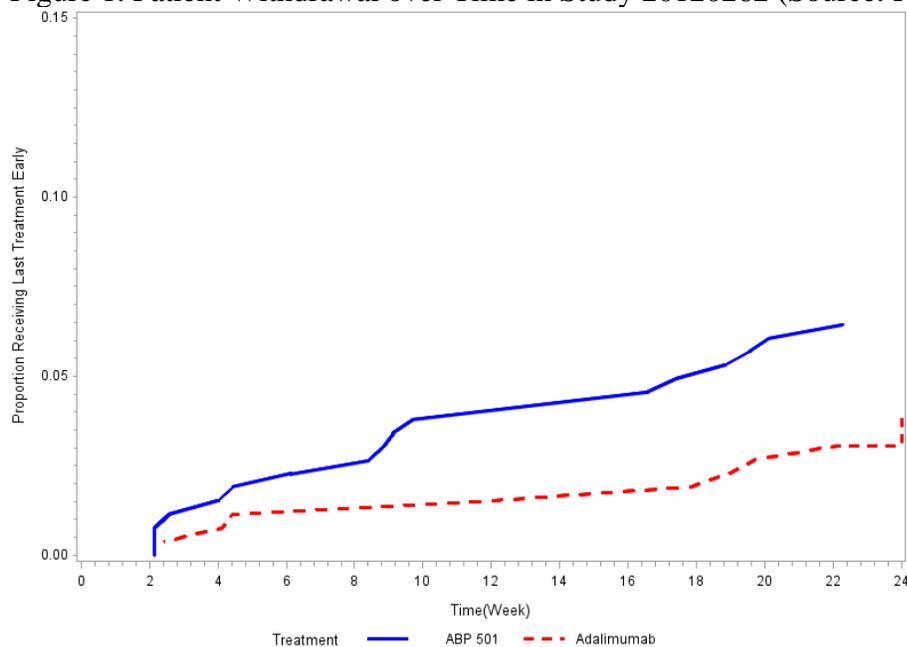


Table 4. Patient Dropout, by Reason for Withdrawal, in Study 20120262

	ABP 501	US-licensed Humira	Overall
N	264	262	526
Completed	243 (92%)	251 (96%)	494 (94%)
Withdrew from Study	21 (8%)	11 (4%)	32 (6%)
Adverse Event	6 (2%)	2 (1%)	8 (2%)
Patient consent withdrawn	11 (4%)	6 (2%)	17 (3%)
Patient lost to follow-up	2 (1%)	2 (1%)	4 (1%)
Significant protocol violation	1 (1%)	0 (0%)	1 (0%)
Other	1 (1%)	1 (1%)	2 (1%)

Source: Reviewer

### 3.4.2 Key results in Study 20120262

Table 5 displays results from the primary efficacy analysis in Study 20120262. Approximately 74.6% of patients randomized to ABP 501 and 72.4% of patients randomized to US-licensed

Humira achieved an ACR20 response at Week 24, for an estimated risk ratio between treatments of 1.04 (90% CI: 0.95, 1.13). The 90% CI ruled out the margin of (0.738, 1/0.738) proposed by the applicant.

Table 6 displays results from the FDA-suggested primary efficacy analysis. Approximately 71.2% of patients randomized to ABP 501 and 72.1% of patients randomized to US-licensed Humira remained in the study and achieved an ACR20 response at Week 24, for an estimated absolute difference between treatments of -0.4% (90% CI: -6.8%, +6.1%). The 90% CI ruled out the margin of  $\pm 12\%$  that the Agency has determined reasonable. The lower CI bound of -6.8% also corresponds to the preservation of approximately 75% of conservative estimates of the effect of adalimumab from historical trials (Table 2). Approximately 70% of the non-responders were patients who completed the study and did not satisfy the ACR20 response criteria. Most of the remaining non-responders were patients who withdrew from the study prior to Week 24. There were no large differences between the treatment arms in the distributions of reasons for non-response (Table 6).

In a supportive analysis of ACR20 response in the subset of patients who completed the study and adhered to the protocol (per-protocol population), 76.5% and 76.4% responded on ABP 501 and US-licensed Humira, respectively, for an estimated difference of 0.4% (90% CI: -6.0%, +6.9%) meeting the similarity margin of  $\pm 12\%$  (Table 8).

The proportions of patients remaining in the study and achieving ACR20 responses at Weeks 2, 4, 8, 12, 18, and 24, in addition to ACR50 and ACR70 response probabilities over time, were similar between the treatment arms (Figure 2). Mean changes from baseline in the components of the ACR composite endpoint and the disease activity score (DAS28-CRP) were also similar between the arms in all randomized patients who completed the study (Table 7). In particular, the 95% CI of (-0.20, 0.21) and the 90% CI of (-0.18, 0.17) for the estimated mean difference in Week 24 DAS28-CRP change ruled out the margin of  $\pm 0.6$  proposed by the applicant. See 3.4.4 for additional discussion on the potential effect of missing data on these comparisons. On both treatment arms, improvements in these continuous secondary endpoints were evident as early as Week 12, and trends over time were similar (see Appendix: Figures 5 - 10). Empirical distribution functions with worst possible values assigned for dropouts were also comparable between the treatment arms for key continuous efficacy endpoints (e.g., see DAS28-CRP comparison in Figure 11).

Table 5. Protocol-Specified Primary Analysis: Proportions of Responders with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262

	ABP 501 (N=264)	US-licensed Humira (N=262)
Responder <sup>1</sup>	194/260 (74.6%)	189/261 (72.4%)
Ratio: 1.039 (90% CI: 0.954, 1.133) <sup>2</sup>		

Source: Applicant

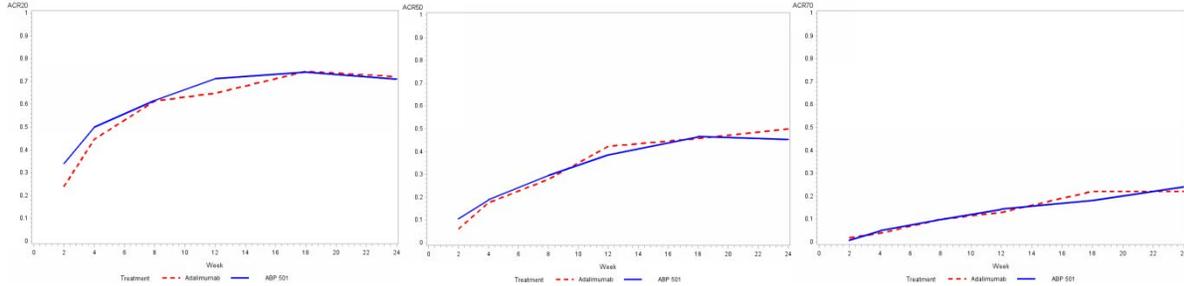
Abbreviations: CI = confidence interval

<sup>1</sup> Defined by meeting ACR20 response criteria after applying LOCF method for missing ACR20 data at Week 24;

Patients who did not have post-baseline ACR measures were excluded from the analysis.

<sup>2</sup> Ratio between ABP 501 and US-licensed Humira and CI based on a generalized linear model adjusted for geographic region and prior biologic use for RA as covariates in the model

Figure 2. ACR20/50/70 Response<sup>1</sup> Probabilities over Time in Study 20120262 (Source: Reviewer)



<sup>1</sup> Defined by remaining in the study and meeting ACR20 response criteria at Weeks 2, 4, 8, 12, 18, and 24

Table 6. FDA-Suggested Primary Analysis: Proportions of Responders, and Distributions of Reasons for Non-Response, with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262

	ABP 501 (N=264)	US-licensed Humira (N=262)
Responder <sup>1</sup>	188 (71.2%)	189 (72.1%)
Difference: -0.4% (90% CI: -6.8%, 6.1%) <sup>2</sup>		
Non-Responder	76 (28.8%)	73 (27.9%)
ACR20 Criteria Not Met	55 (20.8%)	62 (23.7%)
Withdrew from Study	21 (8.0%)	11 (4.2%)
Adverse Event	6 (2.3%)	2 (0.8%)
Patient consent withdrawn	11 (4.2%)	6 (2.3%)
Patient lost to follow-up	2 (0.8%)	2 (0.8%)
Significant protocol violation	1 (0.4%)	0 (0.0%)
Other	1 (0.4%)	1 (0.4%)

Source: Reviewer

Cell contents are frequency (percent of column total)

Abbreviations: CI = confidence interval

<sup>1</sup> Defined by remaining in the study through Week 24, and meeting ACR20 response criteria at Week 24

<sup>2</sup> Difference between ABP 501 and US-licensed Humira and CI based on a generalized linear model adjusted for geographic region and prior biologic use for RA as covariates in the model

Table 7. Mean Changes from Baseline in the ACR Components and DAS28-CRP at Week 24 in Study 20120262 Completers

	ABP 501 (N=264)		US-licensed Humira (N=262)		Difference (95% CI) <sup>2</sup>
	N <sup>1</sup>	Mean	N <sup>1</sup>	Mean	
Swollen Joint Count	246	-10.5	253	-10.3	-0.2 (-1.1, 0.7)
Tender Joint Count	246	-15.4	253	-14.8	-0.7 (-2.2, 0.9)
HAQ Score	246	-0.44	253	-0.47	0.03 (-0.06, 0.12)
Patient Pain	246	-31.7	253	-30.9	-0.8 (-4.6, 3.1)
Patient Global	246	-3.00	253	-2.96	-0.04 (-0.41, 0.33)
Physician Global	246	-4.37	253	-4.27	-0.10 (-0.40, 0.21)

CRP	243	-5.97	251	-6.03	0.05 (-1.67, 1.78)
DAS28-CRP	243	-2.25	251	-2.26	0.01 (-0.20, 0.21)

Source: Reviewer

Abbreviations: CI = confidence interval

<sup>1</sup> Number of patients with complete data included in analysis

<sup>2</sup> Mean difference between ABP 501 and US-licensed Humira and CI based on a linear regression model adjusted for baseline value, geographic region and prior biologic use for RA as covariates in the model

Table 8. Per-Protocol Analysis: Proportions of Responders with Respect to Composite ACR20-Based Primary Endpoint at Week 24 in Study 20120262

	ABP 501 (N=230)	US-licensed Humira (N=233)
Responder <sup>1</sup>	176/260 (76.5%)	178/233 (76.4%)
Difference: 0.4% (90% CI: -6.0%, +6.9%) <sup>2</sup>		

Source: Applicant

Abbreviations: CI = confidence interval

<sup>1</sup> Defined by meeting ACR20 response criteria at Week 24

<sup>2</sup> Difference between ABP 501 and US-licensed Humira and CI based on a generalized linear model adjusted for geographic region and prior biologic use for RA as covariates in the model

### 3.4.3 Assay Sensitivity and the Constancy Assumption

In order to reliably evaluate whether there are clinically meaningful differences between two products, a comparative clinical study should 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 [8] and in the literature [9], 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 9 describes key characteristics of four historical randomized, double-blind, parallel-group, placebo-controlled clinical trials of adalimumab in patients with active RA despite treatment with methotrexate, alongside key characteristics of Study 20120262. 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 five studies. One notable difference was the allowance of anti-TNF experience. The historical placebo-controlled trials did not allow anti-TNF experience while the comparative clinical trial allowed it (although the proportion was relatively small at 28%). This difference might reflect

the change in medical convention of using anti-TNF therapy more frequently in the current clinical setting. Estimated treatment effects with respect to ACR20 for the four historical trials were displayed earlier in Table 2. The estimated effects ranged from 21% to 43% on the absolute difference scale, with an overall estimated effect size of 34%. Thus, the information in Tables 2 and 9 indicates that (1) the designs of the four historical placebo-controlled clinical trials were largely similar to that of comparative clinical Study 20120262; and (2) there were relatively large and consistent treatment effects across the four historical studies.

This evidence of historical sensitivity to effects of adalimumab in similarly designed clinical trials provides some support for a conclusion that Study 20120262 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 20120262, there were only 10 (1.9%) patients with failed eligibility criteria and only 2 patients received the wrong treatment prior to Week 24. Also, approximately 6% of patients discontinued treatment prior to Week 24 - this proportion is lower than the historical discontinuation rates, which ranged from 7% to 22% (Table 9). With this high level of adherence, any potential concern about bias toward equivalence due to low adherence is mitigated. Since the discontinuation rate on the active control was only 4%, potential concerns about decreased efficacy relative to historical studies and violations in the constancy assumption are also mitigated. However, because patients who discontinued treatment were not retained for safety and efficacy assessments through the double-blind period, it is still worthwhile to assess the potential impact of missing data due to dropout on the similarity assessment.

We also examined whether the within-group responses in the comparative clinical study were similar to those observed in previous placebo-controlled trials. The 72% ACR20 response rate on US-licensed Humira in Study 20120262 is slightly higher than historical rates, which ranged from 54% to 67%.

In summary, we did not identify any issues with study conduct. We will discuss the potential impact of missing data on the similarity assessment in detail in 3.4.4. The design, conduct, and within-group responses rates of Study 20120262 were largely similar to those characteristics in four historical clinical trials that demonstrated relatively large and consistent treatment effects of adalimumab over placebo. Therefore, the totality of available information supports the assay sensitivity of Study 20120262, in addition to the constancy assumption.

Table 9. Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials<sup>1</sup> of Adalimumab in RA and Comparative Clinical Study 20120262

Study				
Keystone [4]	Weinblatt [5]	Kim [6]	Chen [7]	Study 20120262

Selected inc/exc criteria	≥9 TJC; ≥6 SJC; CRP >1 mg/dL; RF+ or ≥1 joint erosion	≥9 TJC; ≥6 SJC	≥9 TJC; ≥6 SJC	≥9 TJC; ≥6 SJC	≥6 TJC; ≥6 SJC; ESR >28 mm/hr or CRP >1 mg/dL; RF+ or ACCP+
Anti-TNF experience allowed?	No	No	No	No	Yes (28%)
Concomitant DMARDS	Stable MTX, corticosteroids, NSAIDS	Stable MTX, corticosteroids, NSAIDS	Stable MTX	Stable MTX	Stable MTX
Region/Country	US & Canada	US & Canada	Korea	Taiwan	EU, NA, & LA
Baseline Characteristics of Study Population <sup>2</sup>	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: 24; SJC: 14; Disease Duration: 9 yrs; HAQ-DI: 1.5
Time of ACR20 Evaluation	Week 24	Week 24	Week 24	Week 12	Week 24
ACR20 Response on Humira	63%	67%	62%	54%	72%
Withdrawal on Humira	22% by Week 52	7% by Week 16 (34% escaped to ADA)	9%	N.A.	6%

Source: Reviewer

Abbreviations: SJC=swollen joint count; TJC=tender joint count; DMARD=disease-modifying anti-rheumatic drug; EU=Europe; NA=North America; LA=Latin America; US=United States

<sup>1</sup> Based on best attempts to identify/estimate characteristics from literature review

<sup>2</sup> Means or medians, depending on what was reported in publication

### 3.4.4 Potential Effect of Missing Data

As described in detail in 3.4.1, there was some early patient withdrawal in Study 20120262. In the FDA-suggested primary analysis, the primary endpoint was a composite measure of treatment success defined by remaining in the study and on treatment through Week 24 and achieving an ACR20 response at Week 24. Therefore, outcomes in patients who withdrew early

were not missing - these patients were non-responders according to the composite endpoint definition. However, 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 24 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 (and other important endpoints such as DAS28-CRP). However, such evaluations are subject to some missing data and rely on the strong and unverifiable assumption that outcomes in patients who withdrew early are missing at random. Therefore, we requested and evaluated the applicant's 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).

Table 10 displays estimated de facto differences between ABP 501 and US-licensed Humira in the ACR20 response at Week 24, 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. As a point of reference, the response probabilities among completers on ABP 501 and US-licensed Humira were 77% and 75%, respectively. As seen in the table, there were no scenarios in which the 90% CI fails to rule out a 12% loss in the ACR20 response. I conducted a similar tipping point analysis on the key secondary endpoint of DAS28-CRP. Table 11 displays estimated de facto differences between ABP 501 and US-licensed Humira in the mean change from baseline in DAS28-CRP at Week 24, 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. As a point of reference, the mean change from baseline in DAS28-CRP at Week 24 among completers on ABP 501 and US-licensed Humira were -2.319 and -2.318, respectively. As seen in the table, under a range of plausible scenarios, the 90% CI rules out  $\pm 0.6$  in the mean change from baseline in DAS28-CRP at Week 24. Therefore, these tipping point sensitivity analyses highly support the findings of the key efficacy analyses in Study 20120262.

Table 10. Tipping Point Analysis in Study 20120262: Inference on the Difference Between ABP 501 and US-licensed Humira in the Probability of Week 24 ACR20 Response under Varying Assumptions About the Differences on Each Treatment Arm Between Responses in Patients who Withdrew from the Study Early and Responses in Patients who Completed the Study

Shift for ABP 501 <sup>1</sup>	Shift for US-licensed Humira <sup>1</sup>				
		-0.700	-0.525	-0.350	-0.175

<b>-0.700</b>	0.002 (-0.063, 0.067)	-0.005 (-0.072, 0.061)	-0.014 (-0.080, 0.052)	-0.019 (-0.085, 0.047)	-0.023 (-0.088, 0.042)
<b>-0.525</b>	0.013 (-0.052, 0.078)	0.005 (-0.060, 0.071)	-0.003 (-0.068, 0.063)	-0.008 (-0.073, 0.057)	-0.012 (-0.077, 0.053)
<b>-0.350</b>	0.024 (-0.041, 0.088)	0.016 (-0.049, 0.081)	0.008 (-0.057, 0.072)	0.003 (-0.062, 0.067)	-0.001 (-0.066, 0.063)
<b>-0.175</b>	0.035 (-0.030, 0.100)	0.027 (-0.038, 0.093)	0.019 (-0.046, 0.084)	0.014 (-0.051, 0.079)	0.010 (-0.055, 0.075)
<b>0.000</b>	0.048 (-0.016, 0.111)	0.040 (-0.025, 0.105)	0.032 (-0.033, 0.096)	0.026 (-0.038, 0.090)	0.023 (-0.041, 0.086)

<sup>1</sup> Assumed difference in Week 24 ACR20 response between completers and dropouts. Responses in ABP 501/US-licensed Humira completers were 0.77/0.75.

Source: Applicant (Response to IR post BLA submission)

Cell contents are estimated difference (90% confidence interval).

Table 11. Tipping Point Analysis in Study 20120262: Inference on the Difference Between ABP 501 and US-licensed Humira in the Mean Change from Baseline in DAS28-CRP at Week 24 under Varying Assumptions About the Differences on Each Treatment Arm Between Mean Changes in Patients who Withdrew from the Study Early and Mean Changes in Patients who Completed the Study

<b>Shift for ABP 501<sup>2</sup></b>	<b>Shift for US-licensed Humira<sup>1</sup></b>				
	<b>+4</b>	<b>+3</b>	<b>+2</b>	<b>+1</b>	<b>0</b>
<b>+4</b>	0.15 (-0.12, 0.42)	0.19 (-0.07, 0.45)	0.23 (-0.02, 0.49)	0.27 (0.02, 0.53)	0.32 (0.06, 0.57)
<b>+3</b>	0.07 (-0.19, 0.33)	0.11 (-0.14, 0.36)	0.15 (-0.09, 0.40)	0.20 (-0.04, 0.43)	0.24 (0.00, 0.47)
<b>+2</b>	-0.01 (-0.26, 0.24)	0.03 (-0.21, 0.27)	0.07 (-0.16, 0.30)	0.12 (-0.11, 0.34)	0.16 (-0.07, 0.38)
<b>+1</b>	-0.09 (-0.33, 0.15)	-0.05 (-0.28, 0.18)	-0.01 (-0.23, 0.22)	0.04 (-0.18, 0.26)	0.08 (-0.14, 0.30)
<b>0</b>	-0.17 (-0.41, 0.07)	-0.13 (-0.36, 0.1)	-0.09 (-0.31, 0.14)	-0.04 (-0.26, 0.17)	0.00 (-0.22, 0.21)

Source: Reviewer

Cell contents are estimated difference (90% confidence interval).

<sup>1</sup> Assumed difference in Week 24 mean DAS28-CRP change between completers and dropouts on US-licensed Humira. Mean change in US-licensed Humira completers was -2.318.

<sup>2</sup> Assumed difference in Week 24 mean DAS28-CRP change between completers and dropouts on ABP 501. Mean change in ABP 501 completers was -2.319.

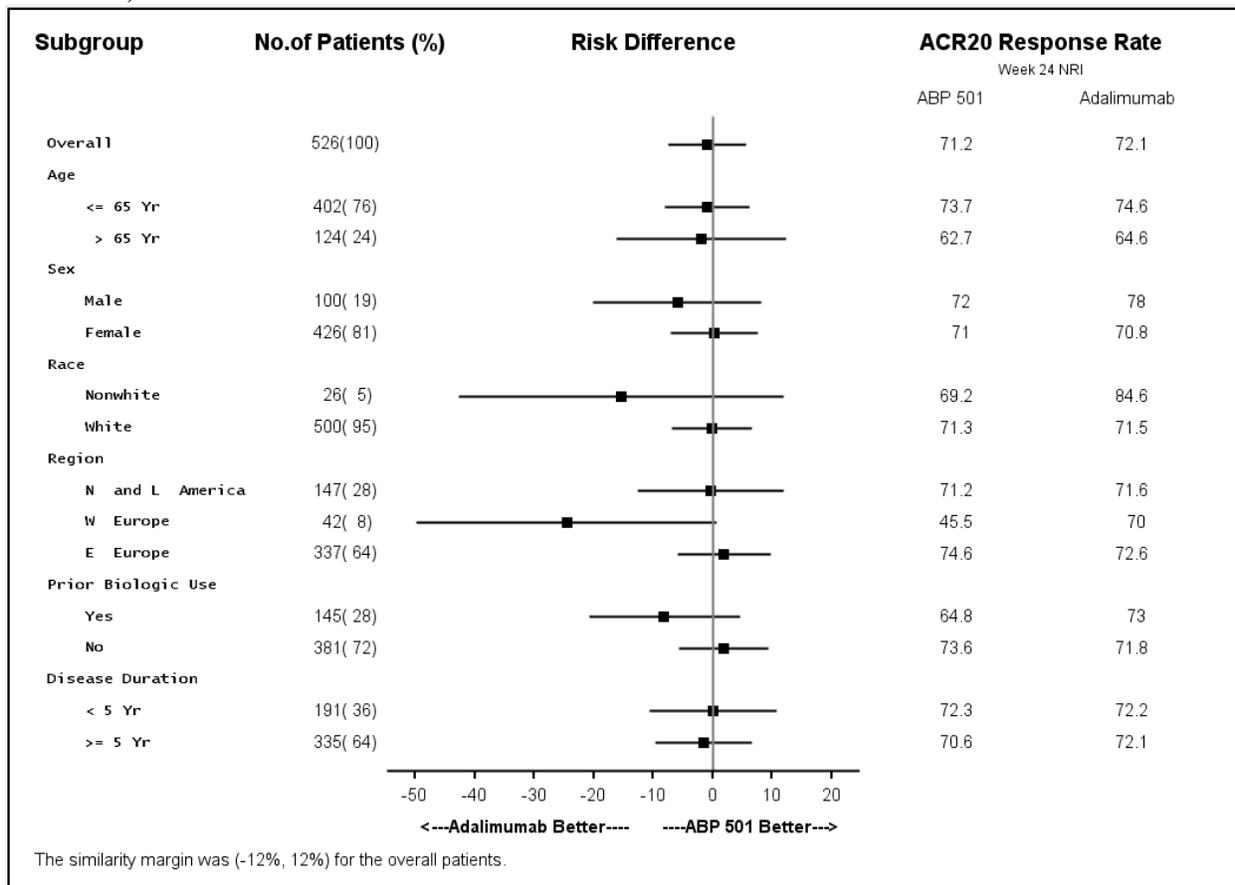
### **3.5 Evaluation of Safety**

Dr. Keith Hull, the Medical Reviewer, conducted the complete safety evaluation, and the reader is referred to Keith Hull's review for more detailed information on safety.

## **4 Findings in Special/Subgroup Populations**

Figure 3 presents the results of subgroup analyses by sex, race (White versus non-White), age ( $\leq 65$ ,  $> 65$ ), and geographic region (North & Latin American versus Western European versus Eastern European) in Study 20120262. As would be expected, there was considerable heterogeneity in the estimated differences in response probabilities comparing ABP 501 and US-licensed Humira across the many subgroups (some very small in size). However, estimated differences were largely centered around similarity. The numbers of non-White patients and the number of Western European patients were very small, leading to very wide confidence intervals around the estimated differences.

Figure 3. Estimated Differences Between ABP 501 and US-licensed Humira in the Probability of Remaining in the Study and Achieving an ACR20 Response at Week 24, Stratified by Selected Subgroups, in Study 20120262. Solid 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 pre-specified a similarity margin of (0.738, 1/0.738) with respect to the risk ratio.

The applicant provided justification for the margin based on historical data from a randomized clinical trial of adalimumab (Keystone [4]) and the goal of preserving at least 50% of the effect size of the reference product. We do not agree with the applicant's selection of historical studies, as three important studies [5-7] are not included in the meta-analysis, and we do not agree with the proposed (0.738, 1/0.738) margin. Furthermore, we consider the risk difference metric as more important. We believe that a margin of  $\pm 12\%$  for the risk difference is more appropriate.

We selected a margin of  $\pm 12\%$  based on meta-analyses of historical effects of adalimumab and discussions with clinicians aimed at weighing the clinical importance of different losses in effect against the feasibility of different study sizes. Despite the lack of agreement on an appropriate similarity margin, results from the primary analysis of Study 20120262 (90% CI: -6.8%, +6.1%) successfully ruled out the  $\pm 12\%$  margin we consider to be reasonable. 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 study supports a demonstration of no clinically meaningful differences between ABP 501 and US-licensed Humira.

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

This issue was discussed in detail in 3.3.2 and 3.4.4. In Study 20120262, 6% of patients failed to complete the 24-week double-blind period. Although this relatively low dropout rate did not lead to substantial missing data, we assessed the potential impact of the missing data in important analyses, such as the evaluations of ACR20 and DAS28-CRP at Week 24 in all randomized patients regardless of adherence. Because the applicant's primary analysis based on LOCF relies on strong and unverifiable assumptions about the missing data, we requested and evaluated tipping point analyses from the applicant to explore the sensitivity of results to violations in the assumptions. Confidence intervals for the differences between ABP 501 and US-licensed Humira continued to rule out concerning losses in efficacy under a reasonably wide range of assumptions about the missing data, including assumptions that patients who dropped out on ABP 501 had considerably worse outcomes than dropouts on US-licensed Humira. Therefore, these tipping point sensitivity analyses highly support the findings of the key efficacy analyses in Study 20120262.

The missing data in important analyses of endpoints at specific follow-up times was largely due to the design of the study, in particular, the fact that patients who discontinued treatment early were also withdrawn from the study. Future comparative clinical studies in RA should clearly differentiate treatment discontinuation from study withdrawal, and ideally the only reason for study withdrawal should be a patient's withdrawal of consent for additional follow-up. This will help prevent missing data and improve the reliability of key results.

- Assay sensitivity and the constancy assumption

This issue was discussed in detail in 3.4.3. 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 adalimumab over placebo in four clinical trials with similar designs to that of comparative clinical Study 20120262. Within-group responses in Study 20120262 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 the rates of treatment discontinuation and missing data, the totality of available information supports the assay sensitivity of Study 20120262, in addition to the constancy assumption.

## **5.2 Collective Evidence**

The collective evidence from the comparative clinical study in rheumatoid arthritis supports a demonstration of no clinically meaningful differences between ABP 501 and US-licensed Humira. In Study 20120262 in RA, 71.2% of ABP 501 patients and 72.1% of US-licensed Humira patients were ACR20 responders, for an estimated absolute difference between treatments of -0.4% (90% CI: -6.8%, +6.1%). The confidence interval successfully ruled out the similarity margin of  $\pm 12\%$  that the Agency has determined reasonable. ACR20, ACR50, and ACR70 responses over time, in addition to mean changes from baseline in the components of the ACR composite endpoint, and the disease activity score (DAS28-CRP) were also similar between the treatment arms. There was missing data in important analyses, but tipping point analyses highly support the findings of key efficacy results. In addition, the totality of available information supports the assay sensitivity of Study 20120262, in addition to the constancy assumption.

## APPENDIX

Figure 4. Mean Disease Activity Score (DAS28-CRP) among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

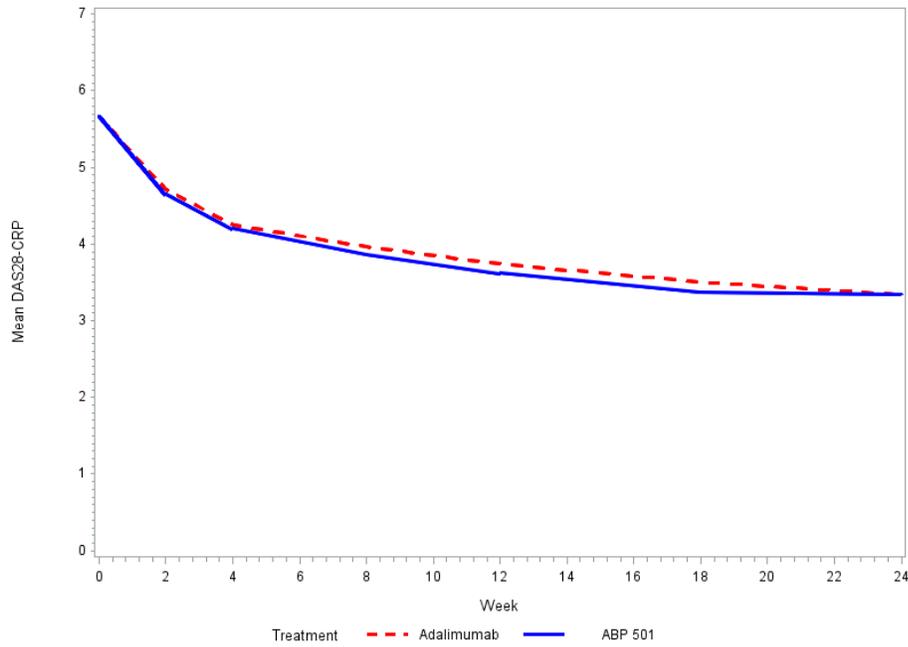


Figure 5. Mean Swollen Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

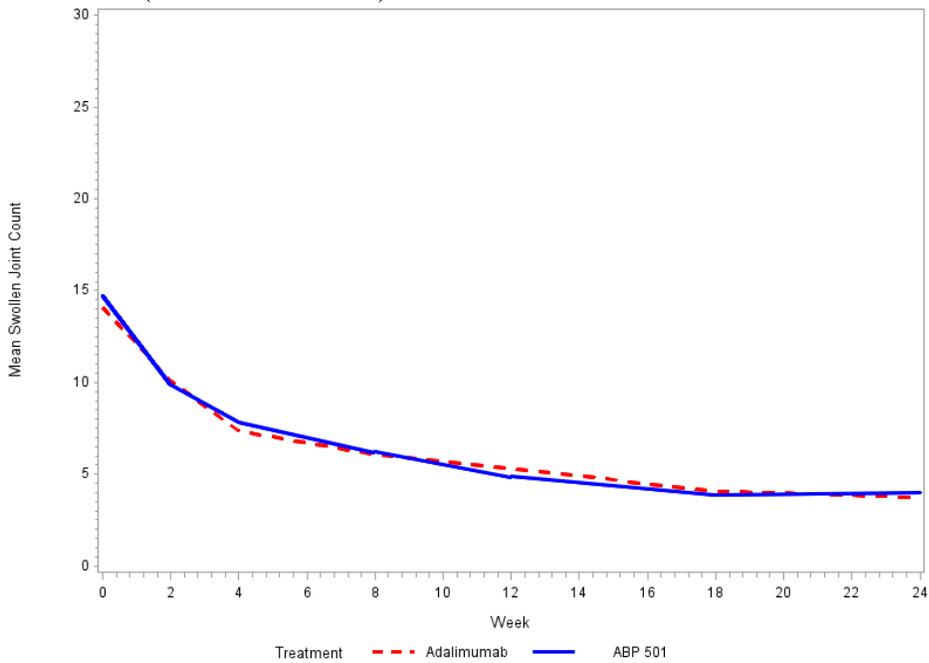


Figure 6. Mean Tender Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

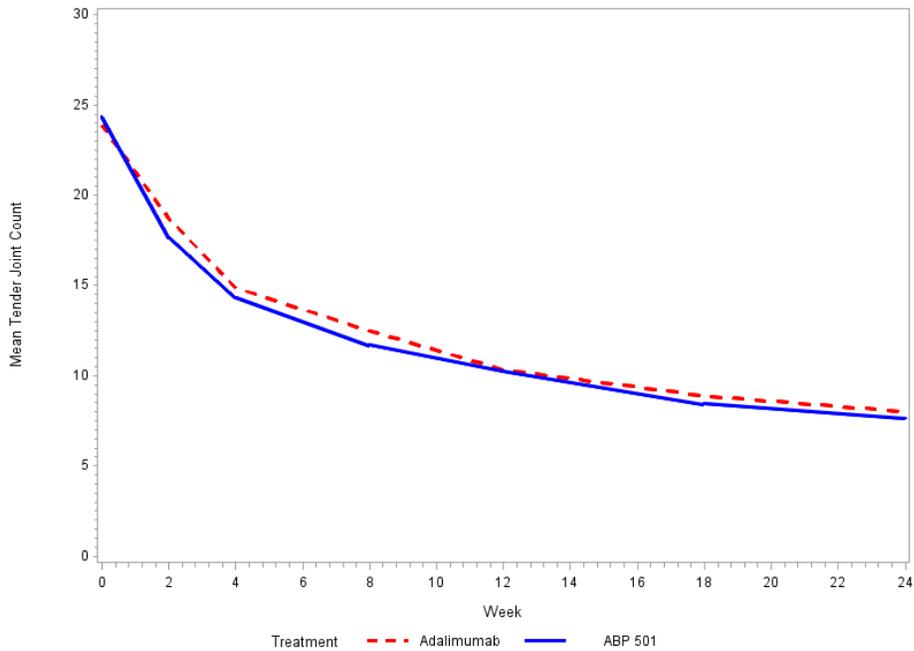


Figure 7. Mean Health Assessment Questionnaire (HAQ) Physical Ability Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

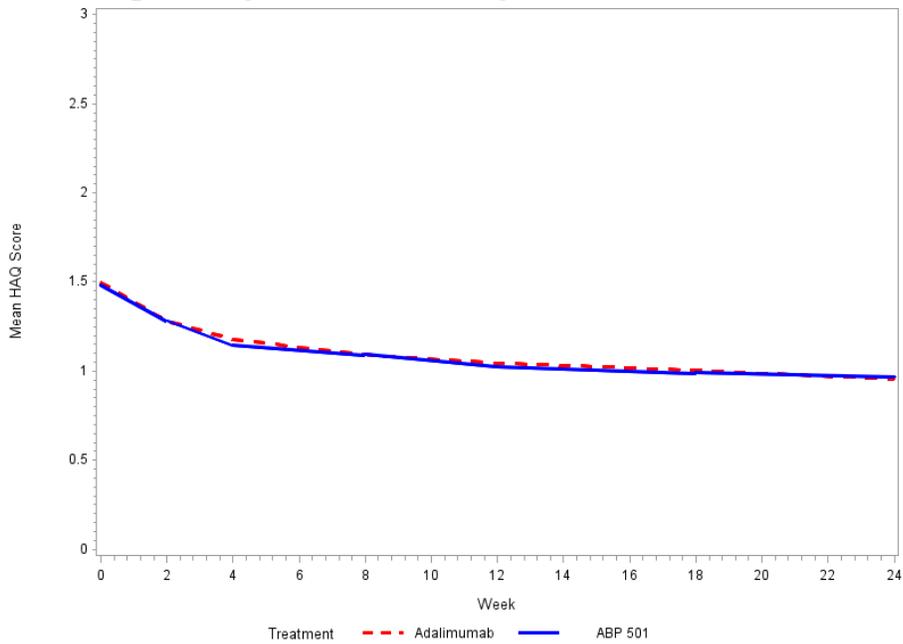


Figure 8. Mean Patient Pain Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

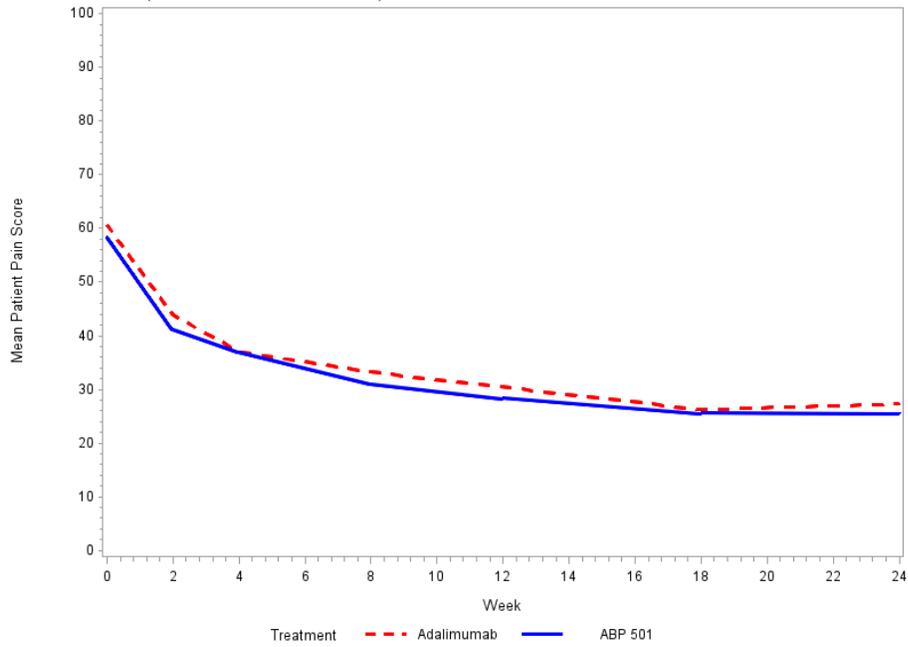


Figure 9. Mean Patient Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

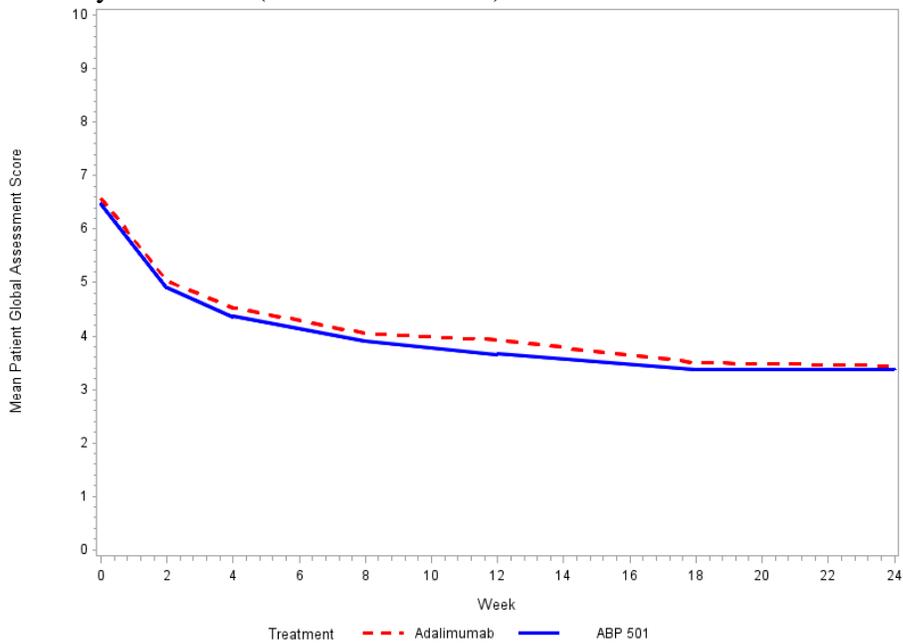


Figure 10. Mean Physician Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

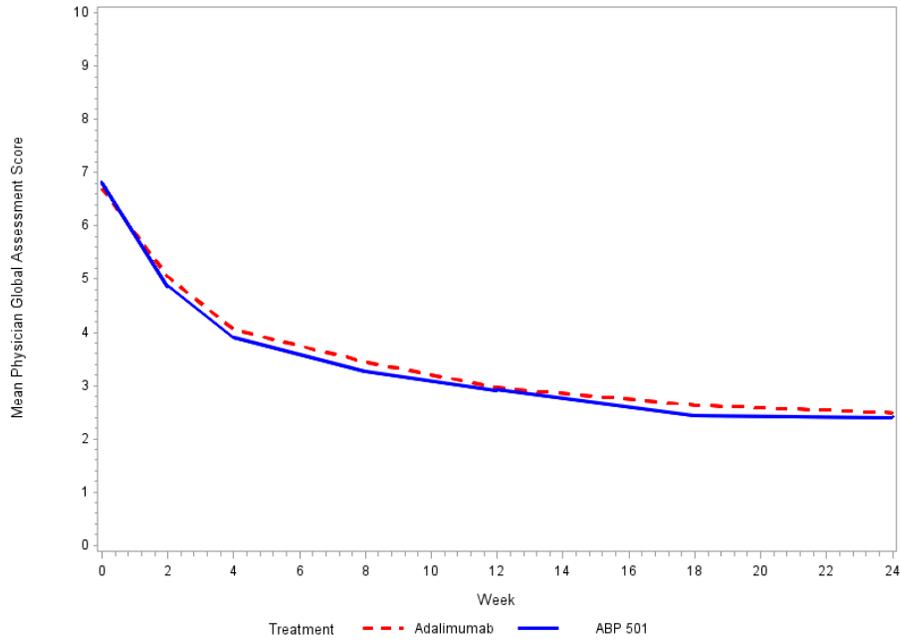
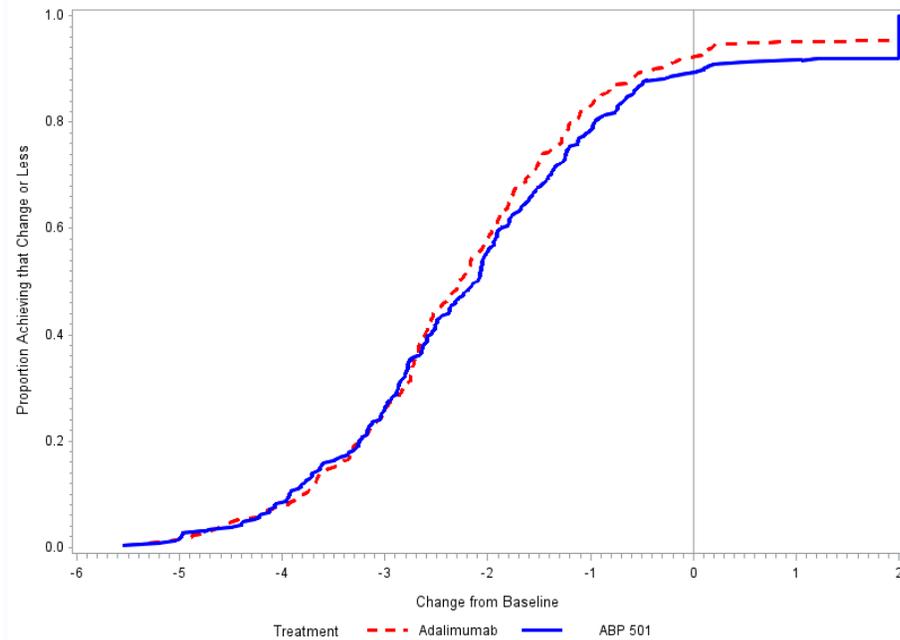


Figure 11. Empirical Distribution Function for Change from Baseline in Disease Activity Score (DAS28-CRP) at Week 24 in Study 20120262 (Source: Reviewer)



## References

- [1] Food and Drug Administration. Guidance for industry: Scientific considerations in demonstrating biosimilarity to a reference product, 2015.
- [2] Food and Drug Administration. Draft guidance for industry: Non-inferiority clinical trials, 2010.
- [3] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. E9: Statistical principles for clinical trials, 1998.
- [4] 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 & Rheumatism*, 50(5), 1400-1411.
- [5] 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 & Rheumatism*, 48(1), 35-45.
- [6] 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.
- [7] 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. *Journal of the Formosan Medical Association*, 108(4), 310-319.
- [8] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. E10: Choice of control group and related issues in clinical trials, 2000.
- [9] Mark Rothmann, Brian Wiens, and Ivan Chan. *Design and Analysis of Non-Inferiority Trials*. Boca Raton, FL: Chapman and Hall/CRC.

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

/s/  
-----

YONGMAN KIM  
09/15/2016

GREGORY P LEVIN  
09/15/2016



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/Serial Number:** 761024 / 0

**Drug Name:** ABP 501

**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:** Amgen

**Dates:** Submitted: 11/25/2015  
BSUFA: 9/23/2016

**Review Priority:** Standard review

**Biometrics Division:** Division of Biometrics III

**Statistics Reviewer:** Kathleen Fritsch, Ph.D.

**Concurring Reviewer:** Mohamed Alosh, Ph.D.

**Medical Division:** Division of Dermatology and Dental Products  
Division of Pulmonary, Allergy, and Rheumatology Products

**Clinical Team:** Denise Cook, M.D. / Gordana Diglisic, M.D.  
Keith Hull, M.D. / Nikolay Nikolov, M.D.

**Project Manager:** Sadaf Nabavian

**Keywords:** biosimilar

## Table of Contents

<b>1</b>	<b>EXECUTIVE SUMMARY .....</b>	<b>3</b>
<b>2</b>	<b>INTRODUCTION.....</b>	<b>4</b>
2.1	Overview .....	4
2.2	Data Sources.....	5
<b>3</b>	<b>STATISTICAL EVALUATION .....</b>	<b>5</b>
3.1	Data and Analysis Quality .....	5
3.2	Evaluation of Efficacy .....	6
3.2.1	Study Design and Statistical Analysis .....	6
3.2.2	Subject Disposition .....	8
3.2.3	Baseline Characteristics .....	10
3.2.4	Primary Efficacy Endpoint .....	12
3.2.5	Secondary Endpoints .....	15
3.2.6	Interpretation of Comparative Clinical Studies .....	18
3.3	Evaluation of Safety.....	20
3.3.1	Extent of Exposure .....	20
3.3.2	Adverse Events .....	21
3.3.3	Immunogenicity .....	23
<b>4</b>	<b>FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>24</b>
4.1	Gender, Race, Age, and Geographic Region .....	24
4.2	Other Special/Subgroup Populations.....	24
<b>5</b>	<b>SUMMARY AND CONCLUSIONS .....</b>	<b>25</b>
5.1	Statistical Issues and Collective Evidence .....	25
5.2	Conclusions and Recommendations.....	26
	<b>REFERENCES .....</b>	<b>26</b>
	<b>SIGNATURES/DISTRIBUTION LIST .....</b>	<b>27</b>

## 1 Executive Summary

ABP 501 is a proposed biosimilar to US-licensed Humira (adalimumab). As part of the development program, the applicant conducted a comparative clinical study of ABP 501 versus European Union (EU)-approved Humira in subjects with moderate to severe psoriasis (Study 263). Study 263 was a randomized, double-blind comparative clinical study of ABP 501 and EU-approved Humira in subjects age 18 to 75 years old with moderate to severe plaque psoriasis. The study enrolled 350 subjects, 175 randomized to the ABP 501 arm and 175 randomized to the EU-approved Humira arm, of which 347 received at least one dose of study product. Subjects were enrolled in Europe, Canada, and Australia. The primary endpoint was the percent improvement in PASI (Psoriasis Area Severity Index) from Week 1 to Week 16. The pre-specified similarity margin for the confidence interval for the difference in means was  $\pm 15$ . At Week 16, subjects who achieved at least PASI 50 response (at least 50% improvement from baseline) continued into the second treatment period. All subjects originally randomized to ABP 501 continued treatment with ABP 501 through Week 48. Subjects originally randomized to EU-approved Humira were re-randomized 1:1 to either continue treatment with EU-approved Humira or transition to ABP 501 through Week 48. Subjects were evaluated in the second treatment period for efficacy, safety, and immunogenicity outcomes.

The mean percent improvement in PASI at Week 16 was similar on the ABP 501 and EU-approved Humira arms and the confidence interval for the difference was within the pre-specified margin of  $\pm 15$ . In the applicant's full analysis set (FAS), defined as all subjects randomized and dispensed medication who had at least one post-baseline efficacy assessment, the mean percent improvement in PASI values on the ABP 501 and EU-approved Humira arms were 80.9 vs 83.1. Results on the per protocol population and an analysis population that includes all subjects randomized and dispensed medication whether or not they had post-baseline efficacy assessments were similar and also fell within the pre-specified margin. See Table 1. The results of the secondary endpoints of PASI 75, clear or almost clear on the static Physician's Global Assessment, and reduction from baseline in body surface area were consistent with the primary endpoint.

**Table 1 – Percent Improvement in PASI at Week 16**

	ABP 501	EU-approved Humira	Difference <sup>d</sup>	90% Conf. Int.
Full Analysis Set <sup>a</sup> (LOCF)	N=172 80.9	N=173 83.1	-2.2	(-6.6, 2.2)
Sensitivity Analysis <sup>b</sup> (LOCF)	N=174 80.0	N=173 83.1	-3.1	(-7.5, 1.4)
Per protocol <sup>c</sup> (Observed)	N=155 82.6	N=152 85.3	-2.6	(-6.2, 0.9)

<sup>a</sup> Randomized, dispensed medication, and at least one post-baseline efficacy assessment

<sup>b</sup> Randomized, dispensed medication

<sup>c</sup> Completed the treatment period without protocol violations that affected the evaluation of the primary objective

<sup>d</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Because Study 263 was conducted completely outside the US, the applicant did not discuss the proposed similarity margin with the FDA prior to conducting the study. The applicant did not provide a rationale for their choice of similarity margin in the protocol or study report. Therefore, this reviewer evaluated the applicant's proposed margin using information from the published literature on the percent improvement in PASI from published placebo-controlled studies of Humira and other TNF- $\alpha$  inhibitors. Based on this evaluation, the assumptions of consistency and assay sensitivity appear reasonable for Study 263, and the confidence interval for the primary endpoint of percent improvement in PASI is sufficiently narrow to conclude that the study met the criteria for demonstrating similarity.

Adverse event rates were similar on both the ABP 501 and EU-approved Humira arms. During the initial treatment period, 10% of ABP 501 subjects and 14% of EU-approved Humira subjects developed neutralizing antibodies. Among the subjects who continued into the second treatment period, 20% of subjects on EU-approved Humira/EU-approved Humira arm, 25% on the EU-approved Humira/ABP 501 arm, and 14% on the ABP 501/ABP 501 arm developed neutralizing antibodies during the study.

Thus we conclude that the results on the ABP 501 and EU-approved Humira arms are similar and that Study 263 supports a demonstration of no clinically meaningful differences between ABP 501 and US-licensed Humira.

## **2 Introduction**

### **2.1 Overview**

ABP 501 is being developed as a proposed biosimilar to US-licensed Humira (adalimumab) under Section 351(k) of the Public Health Service (PHS) Act. Section 351(i) of the PHS Act defines biosimilarity to mean “that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components” and that “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.” As part of their development program, the applicant has conducted two comparative clinical studies of ABP 501 and a 3-way pharmacokinetic similarity study. Study 262 evaluated ABP 501 and US-licensed Humira in subjects with rheumatoid arthritis. Study 263 evaluated ABP 501 and EU-approved Humira in subjects with plaque psoriasis. Study 217 was a 3-way pharmacokinetic similarity study (ABP 501 vs. US-licensed Humira vs. EU-approved Humira) in healthy volunteers. This review will evaluate Study 263. The design details for Study 263 are summarized in Table 2.

Study 263 was conducted outside the US and the protocol was not submitted to the FDA prior to conducting the study. Although the details of Study 263 were not discussed with FDA, other components of the development program were discussed at Biosimilar Biological Product Development meetings held on August 24, 2011 and June 10, 2015.

**Table 2 – Characteristics of Study 263**

Study Number	20120263 (Study 263)
Study Design	Part 1: ABP 501 vs. EU-approved Humira (Week 1 to Week 16) Part 2: Subjects with PASI 50 continue in study. ABP 501 subjects continue treatment with ABP 501 through Week 52. EU-approved Humira subjects are randomized 1:1 to transition to ABP 501 or continue EU-approved Humira through Week 52.
Inclusion criteria	Subjects age 18-75 years with stable moderate to severe plaque psoriasis for at least 6 months Body Surface Area $\geq$ 10%, PASI $\geq$ 12, and static Physician's Global Assessment (sPGA) $\geq$ 3.
Treatment regimen	80 mg at Week 1, 40 mg at Week 2 and every other week thereafter.
Primary endpoint	Percent reduction in PASI at Week 16
Secondary endpoints	PASI 75, sPGA response (0 or 1), change in BSA
Treatment arms and Sample Size	ABP 501 - 175 EU-approved Humira - 175
Study location	Australia, Canada, France, Germany, Hungary, Poland

## 2.2 Data Sources

This reviewer evaluated the applicant's clinical study report for Study 263, clinical summaries, and proposed labeling. The submission was in eCTD format and was entirely electronic. Both SDTM and analysis datasets were submitted. The analysis datasets for Study 263 used in this review are archived at <\\cdsesub1\evsprod\bla761024\0001\m5\datasets\20120263\analysis\adam\datasets>.

## 3 Statistical Evaluation

### 3.1 Data and Analysis Quality

The databases for Study 263 required minimal data management prior to performing the analyses, and no requests for information regarding the datasets for Study 263 were made to the applicant.

### 3.2 Evaluation of Efficacy

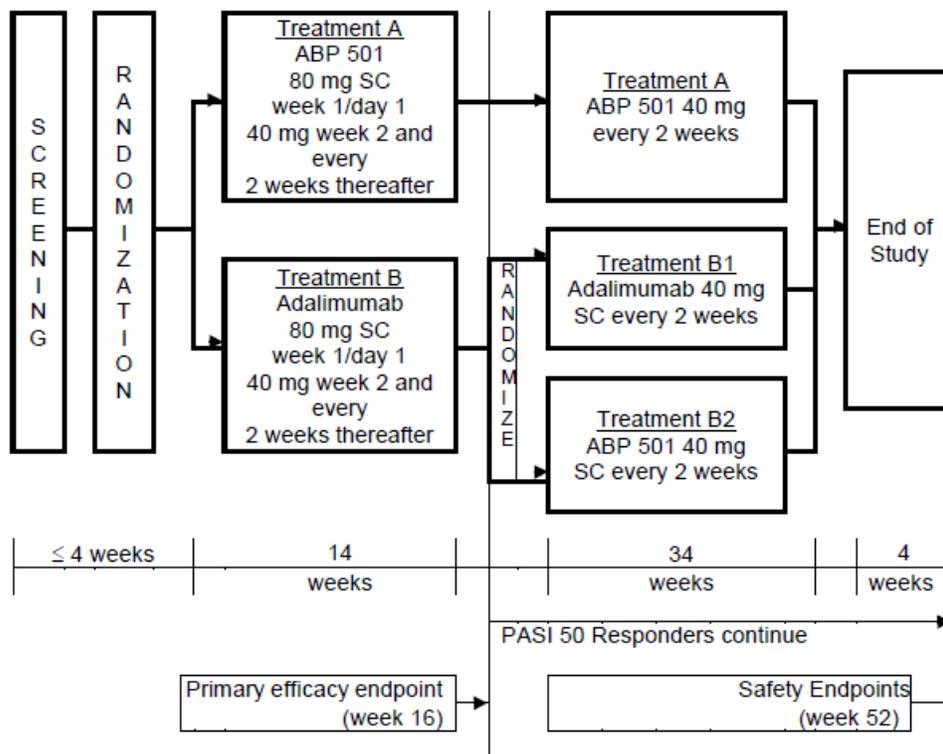
#### 3.2.1 Study Design and Statistical Analysis

Study 263 was a randomized, double-blind comparative clinical study of ABP 501 and EU-approved Humira in subjects with moderate to severe plaque psoriasis. The study included data (including immunogenicity) on subjects transitioning from EU-approved Humira to ABP 501. The study enrolled subjects age 18 to 75 with stable moderate to severe plaque psoriasis for at least 6 months, involving at least 10% body surface area (BSA), PASI  $\geq$  12, and static Physician's Global Assessment (sPGA)  $\geq$  3 (moderate,

severe, or very severe). Subjects were to be candidates for systemic therapy or phototherapy and were to have previously failed, had inadequate response, intolerance to, or contraindication to at least one conventional anti-psoriatic systemic therapy.

The study enrolled 350 subjects, 175 randomized to the ABP 501 arm and 175 randomized to the EU-approved Humira arm, of which 347 received at least one dose of study product. Subjects were enrolled at 49 centers in 6 countries (Australia, Canada, France, Germany, Hungary, and Poland). Randomization was stratified by geographic region (Eastern Europe, Western Europe, Other) and prior biologic use for psoriasis (yes/no). Subjects received subcutaneous injection of 80 mg at Week 1, 40 mg at Week 2 and 40 mg every 2 weeks thereafter. The primary timepoint for efficacy assessment was Week 16 (15 weeks after treatment was initiated at Week 1). At Week 16, subjects who achieved at least PASI 50 response (at least 50% improvement from baseline) continued into the second treatment period. Subjects originally randomized to ABP 501 continued treatment with ABP 501 through Week 48. Subjects originally randomized to EU-approved Humira were re-randomized 1:1 to either continue treatment with EU-approved Humira or undergo a single transition to ABP 501 through Week 48. Subjects were followed through Week 52. See Figure 1.

**Figure 1 – Design of Study 263**



Source: pg. 17 of <https://cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf>.

Subjects were evaluated for efficacy at screening and Weeks 1, 4, 8, 12, 16, 32, and 50. Efficacy was assessed using the PASI scale, BSA, and sPGA. The PASI score is derived from assessments for erythema, plaque elevation, and scaling over four body regions

(head, trunk, upper limbs, and lower limbs). PASI scores can range from 0 to 72. The sPGA scale was a 6-point scale with 0 = clear, 1 = almost clear, 2 = mild, 3 = moderate, 4 = severe, and 5 = very severe. The protocol states that the sPGA scale is used to measure the severity of disease in terms of induration, scaling, and erythema, but does not otherwise list any morphological descriptions for the categories of the sPGA scale.

The primary endpoint was the percent improvement in PASI from Week 1 to Week 16. The secondary endpoints were PASI 75 (at least 75% reduction from baseline in the PASI score), sPGA response (0 or 1; clear or almost clear), and change in BSA. Secondary endpoints were assessed at Weeks 16, 32, and 50.

The protocol specified that the percent improvement in PASI at Week 16 would be analyzed with a 95% confidence interval (CI) for the difference in means using estimates from an ANCOVA model adjusted for baseline PASI score and the stratification factors (geographic region and prior biologic use for psoriasis). The pre-specified similarity margin was  $\pm 15$ . Study 263 was conducted outside the US and the applicant did not discuss the study design with FDA prior to conducting the study. Accordingly, FDA did not provide any comments on the endpoints, margin, or analysis methods at the design stage. Although the protocol for Study 263 specified 95% confidence intervals for the primary endpoint, FDA also analyzed the data using 90% confidence intervals, as the FDA has generally recommended 90% confidence intervals (corresponding to a Type I error rate of 5%) for comparative clinical studies in biosimilarity development programs. Note also that FDA had advised the applicant to use a 90% confidence interval in their comparative clinical study in rheumatoid arthritis subjects (Study 262).

The primary analysis population was the full analysis set (FAS), defined in the protocol as all subjects initially randomized in the study. However, in their analyses, the applicant included in the FAS only subjects who had been randomized, dispensed medication, and who had at least one post-baseline efficacy assessment. In the study, 350 subjects were randomized, 347 received at least one dose of investigational product, and 345 had at least one post-baseline assessment. Analyses on the per protocol population were supportive. The per protocol population included subjects who completed the specified treatment period without protocol violations that affected the evaluation of the primary objective. For the second part of the study, the re-randomized analysis set included all subjects who were re-randomized at Week 16.

For the primary endpoint of percent improvement in PASI, missing data in the FAS were imputed using last observation carried forward (LOCF). An observed case analysis and the per protocol analysis were supportive. The protocol also stated that a sensitivity analysis would be conducted in which a number of covariates (age group, race, sex, disease duration, neutralizing antibody status, concomitant topical steroid use, and prior use of systemic or phototherapies) were included in the ANCOVA model and then assessed using backward selection. Percent improvement in PASI would also be analyzed using a repeated measures analysis using data from visits through Week 16.

Analyses for the secondary endpoints were considered descriptive. Confidence intervals for the difference in PASI 75 response and sPGA response were computed using estimates from a generalized linear model with the stratification factors (geographic region and prior biologic use for psoriasis) and baseline PASI score or baseline sPGA score, respectively, as covariates. Missing data for the FAS was handled with LOCF, non-responder imputation, or observed cases. Change in BSA was analyzed with an ANCOVA model with the stratification factors and baseline BSA as covariates.

### 3.2.2 Subject Disposition

Study 263 randomized 350 subjects, 175 each to the ABP 501 and EU-approved Humira arms. Three subjects were not dispensed treatment medication (1 on the ABP 501 arm and 2 on the EU-approved Humira arm). Two subjects had no post-baseline efficacy assessments (both on the ABP 501 arm). Approximately 5% of subjects on each arm discontinued treatment during the initial treatment period. The most common reasons for treatment discontinuation were adverse events and consent withdrawn. See Table 3. Most subjects continued into the second treatment period (152 (87%) of ABP 501 subjects and 156 (89%) of EU-approved Humira subjects), where subjects on the EU-approved Humira arm were randomized to continue EU-approved Humira or undergo a single transition to ABP 501 and subjects on the ABP 501 arm continued ABP 501. Approximately 90% of the subjects who entered Treatment Period 2 completed the study. See Table 4.

**Table 3 - Disposition of Subjects in Treatment Period 1**

	ABP 501	EU-approved Humira
Subjects Randomized	175	175
Subjects Treated	174 (99%)	173 (99%)
Discontinued treatment by Week 16	8 (5%)	10 (6%)
Adverse event	4 (2%)	5 (3%)
Consent withdrawn	3 (2%)	2 (1%)
Lost to follow-up	--	1 (<1%)
Protocol violation	1 (<1%)	2 (1%)
Completed efficacy assessments at Week 16 <sup>a</sup>	165 (94%)	167 (95%)
Did not complete efficacy assessments at Week 16	10 (6%)	8 (5%)

<sup>a</sup> Day 92- 119

Source: pg 37-38 of [\\cdsub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis.

**Table 4 – Disposition of Subjects in Treatment Period 2**

	Treatment in Period 1		
	ABP 501 N=175	EU-approved Humira N=175	
Completed through Week 16	164 (94%)	162 (93%)	
Re-randomized at Week 16	152 (87%)	156 (89%)	
Not re-randomized at Week 16	23 (13%)	29 (11%)	
<PASI 50 at Week 16	11 (6%)	6 (3%)	
Missed Week 16 visit or discontinued study	12 (7%)	23 (13%)	
	Treatment in Period 2		
	ABP 501 N=152	EU-Hum N=79	ABP 501 N=77
Completed Treatment Period 2	138 (89%)	71 (90%)	69 (90%)
Discontinued Treatment Period 2	17 (11%)	8 (10%)	8 (10%)
Consent withdrawn	8 (5%)	3 (4%)	3 (4%)
Other	8 (5%)	4 (5%)	2 (3%)
<i>Adverse event</i>	5 (3%)	1 (1%)	1 (1%)
<i>Lack of efficacy</i>	2 (1%)	3 (4%)	1 (1%)
<i>Non-compliance</i>	1 (<1%)	--	--
Lost to follow-up	1 (<1%)	1 (1%)	2 (3%)
Physician decision	--	--	1 (1%)

Source: pg 40 of [\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

Approximately 11% of subjects were excluded from the per protocol population. The most common reasons for being excluded from the per protocol population were not completing treatment through Week 16 and being mis-stratified at randomization. The rates were similar on the two arms. See Table 5.

**Table 5 – Primary Reason for Per Protocol Population Exclusion**

	ABP 501 N=175	EU-approved Humira N=175
Subjects excluded from Per Protocol Population	18 (10%)	22 (13%)
Did not complete treatment through Week 16	7 (4%)	8 (5%)
Did not have previous failure to psoriatic systemic therapy	--	1 (<1%)
Incorrect treatment received	1 (<1%)	2 (1%)
Mis-stratification at randomization	7 (4%)	6 (3%)
Prior use of 2 or more biologic therapies	--	4 (2%)
Prohibited medications during study	3 (2%)	1 (<1%)

Source: reviewer analysis.

### 3.2.3 Baseline Characteristics

The baseline demographics were generally balanced across the treatment groups in Study 263. The mean age was about 45 years, with about 6% of subjects age 65 and older. The majority of subjects were male (65%) and white (93%). The mean weight at baseline was 89 kg. Approximately 40% of subjects were enrolled in Eastern Europe, 25% in Western Europe, and 35% in Australia or Canada. See Table 6.

**Table 6 – Baseline Demographics (Randomized Subjects)**

	ABP 501 N=175	EU-approved Humira N=175
<i>Age (years)</i>		
Mean	45.1	44.0
Range	18-74	18-73
18 to 64 years	164 (94%)	163 (93%)
65 + years	11 (6%)	12 (7%)
<i>Gender</i>		
Female	63 (36%)	59 (34%)
Male	112 (64%)	116 (66%)
<i>Race</i>		
White	167 (95%)	157 (90%)
Black	--	2 (1%)
Asian	5 (3%)	8 (5%)
Other	1 (<1%)	5 (3%)
Unknown	2 (1%)	3 (2%)
<i>Geographic Region</i>		
Eastern Europe	71 (41%)	70 (40%)
Western Europe	43 (25%)	43 (25%)
Other	61 (35%)	62 (35%)
<i>Weight (kg)</i>	N=174	N=173
Mean (SD)	88.9 (23.6)	89.3 (19.4)
Range	48.0-200.6	52.9-166.1

Source: pg 45 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-ffic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

To be enrolled in the study, subjects were to have stable moderate to severe plaque psoriasis for at least 6 months involving at least 10% body surface area (BSA), PASI  $\geq$  12, and static Physician's Global Assessment (sPGA)  $\geq$  3 (moderate, severe, or very severe). At baseline subjects had a mean PASI score of 20 and a mean BSA of 27%. Approximately 60% of subjects had an sPGA score of moderate. About 18% had prior use of a biologic for psoriasis. See Table 7.

**Table 7 – Baseline Disease Characteristics (Subjects Randomized and Dispensed Medication)**

	ABP 501 N=174	EU-approved Humira N=173
<i>PASI</i>		
Mean (SD)	19.7 (8.1)	20.5 (7.9)
Range	12.0 - 61.8	12.0 - 52.2
<i>BSA</i>		
Mean (SD)	25.3 (15.0)	28.5 (16.8)
Range	10 - 82	10 - 90
<i>sPGA</i>		
Moderate	106 (61%)	102 (59%)
Severe	61 (35%)	61 (35%)
Very Severe	7 (4%)	10 (6%)
<i>Prior biologic use for psoriasis</i>	N=175	N=175
Yes	33 (19%)	30 (17%)
No	142 (81%)	145 (83%)

Source: pg 46 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

### 3.2.4 Primary Efficacy Endpoint

The primary efficacy endpoint was the percent change in PASI from Week 1 to Week 16. The protocol specified that the percent improvement in PASI at Week 16 would be analyzed with a 95% confidence interval (CI) for the difference in means using estimates from an ANCOVA model adjusted for baseline PASI score and the stratification factors (geographic region and prior biologic use for psoriasis). The pre-specified similarity margin was  $\pm 15$ . The applicant also presented 90% confidence intervals. The primary analysis population was the full analysis set (FAS), defined in the protocol as all subjects initially randomized in the study. However, in their analyses, the applicant included in the FAS only subjects who had been randomized, dispensed treatment medication, and who had at least one post-baseline efficacy assessment. Missing data was handled with LOCF. Study 263 met the pre-specified similarity criterion for the primary endpoint of percent improvement in PASI at Week 16. For the applicant's primary analysis in the FAS population, both the 95% and 90% confidence intervals for the difference in mean percent improvement in PASI was within the pre-specified margin of  $\pm 15$ . See Table 8.

**Table 8 – Percent Reduction in PASI at Week 16 (FAS/LOCF)**

	ABP 501 N=172	EU-approved Humira N=173
Baseline (Week 1) PASI <sup>a</sup>	19.7 (8.1)	20.5 (7.9)
Week 16 PASI <sup>a</sup>	3.7 (5.1)	3.3 (5.8)
Percent Improvement <sup>a</sup>	80.9 (24.2)	83.1 (25.2)
Difference <sup>b</sup>		-2.2
95% CI		(-7.4, 3.0)
90% CI		(-6.6, 2.2)

<sup>a</sup> Mean (SD)

<sup>b</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Source: pg 52 of <\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-ffic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf> and reviewer analysis

The applicant conducted sensitivity analyses for the primary endpoint using the per protocol population and observed cases. The results of these analyses are similar to the analysis in the FAS population. See Table 9. FDA conducted additional sensitivity analyses for the handling of missing data. Although the applicant’s FAS population was defined in the protocol as all randomized subjects, the applicant’s analysis excluded two subjects who were dispensed medication but had no post-baseline efficacy assessments. Both subjects were on the ABP 501 arm and received both the Week 1 and Week 2 doses. Therefore, this reviewer conducted an additional sensitivity analysis including all subjects who were randomized and dispensed medication, using baseline observation carried forward for the subjects with no post-baseline assessments. The results of the sensitivity analysis are similar to the results of the applicant’s primary analysis, but with a slightly larger estimated treatment difference of -3.1 and 90% confidence interval of (-7.5, 1.4).

This reviewer also conducted sensitivity analyses using alternate imputations for missing data for the percent improvement in PASI endpoint, where subjects with missing data on one arm are imputed assuming no improvement from baseline (0%) and subjects with missing data on the other arm are imputed assuming full improvement (100%). These results are also presented in Table 9. While these two imputations shift the estimated treatment difference to -6.3 and +2.3, the 90% confidence bounds for both sensitivity analyses remain within the bounds of -11 to +7 and thus the confidence bounds remain within the pre-specified margins of ±15 even under relatively extreme imputation assumptions. Thus the results of the sensitivity analyses for handling missing data are consistent with the primary analysis.

**Table 9 - Sensitivity Analyses for the Percent Improvement in PASI at Week 16**

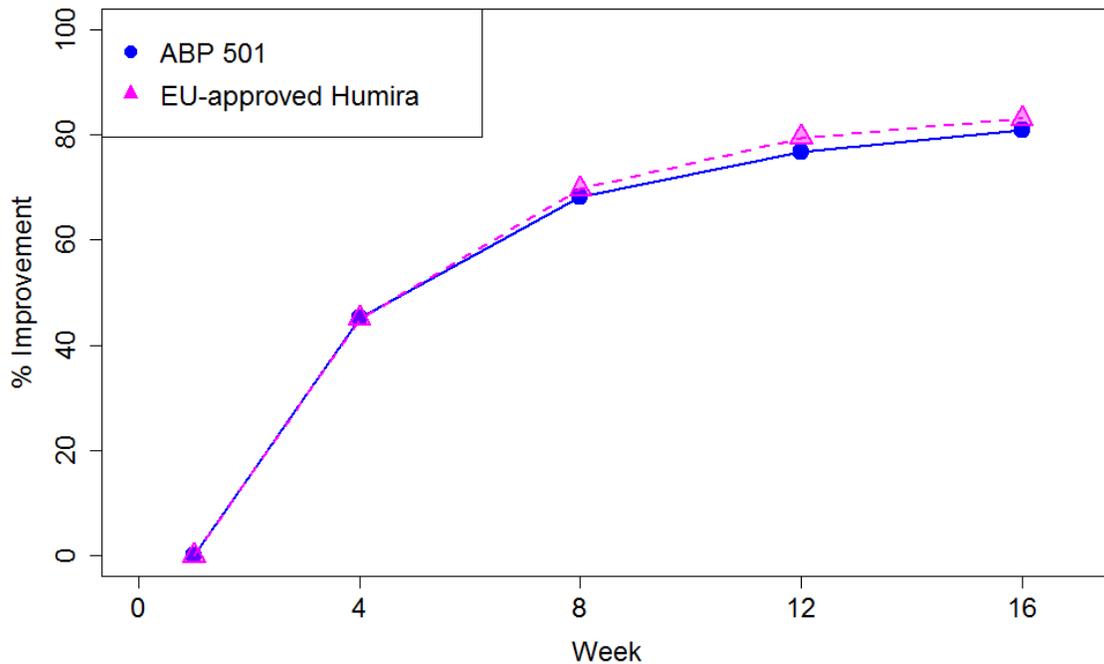
	ABP 501	EU-approved Humira	Difference <sup>a</sup>	90% Conf. Int.
<b>Applicant’s sensitivity analyses</b>				
Per protocol	N=155 82.6	N=152 85.3	-2.6	(-6.2, 0.9)
Observed Cases	N=165 82.6	N=167 84.1	-1.5	(-5.5, 2.6)
<b>Reviewer’s sensitivity analyses</b>				
LOCF (including subjects with no post-baseline assessments)	N=174 80.0	N=173 83.1	-3.1	(-7.5, 1.4)
ABP 501 missing as 0%/EU-approved Humira missing as 100%	78.3	84.6	-6.3	(-10.9, -1.8)
ABP 501 missing as 100%/EU-approved Humira missing as 0%	83.5	81.1	2.3	(-2.0, 6.7)

<sup>a</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Source: pg 274, 277 of [\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](https://cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf) and reviewer analysis

During the initial treatment period, PASI assessments were conducted at baseline (Week 1) and Weeks 4, 8, 12, and 16. The percent reduction in PASI over time for ABP 501 and EU-approved Humira were similar at each study visit. See Figure 2.

**Figure 2 – Percent Improvement in PASI during Treatment Period 1 (FAS, LOCF)**



Source: reviewer analysis

Subjects with at least PASI 50 at Week 16 were to continue into the second treatment period, where subjects originally randomized to ABP 501 continued on ABP 501 and subjects originally randomized to EU-approved Humira were randomized 1:1 to remain on EU-approved Humira or transition to ABP 501. During the second treatment period, the percent improvement in PASI remained relatively constant among the re-randomized subjects from Week 16 to Week 50. See Table 10.

**Table 10 - Percent Improvement in PASI after Re-randomization (Observed Cases)**

	ABP 501 / ABP 501		EU-Hum / EU-Hum		EU-Hum / ABP 501	
	N	Mean	N	Mean	N	Mean
Week 16	152	86.6	79	88.0	77	88.2
Week 32	143	87.6	72	88.2	71	87.0
Week 50	134	87.2	70	88.1	69	85.8

Source: pg 280-281 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

### 3.2.5 Secondary Endpoints

The secondary endpoints were PASI 75, sPGA response (clear or almost clear), and reduction in BSA. The applicant also assessed PASI 50 and PASI 90, though these analyses were not pre-specified in the protocol. The protocol stated that the secondary endpoints would be analyzed with descriptive statistics, including 95% confidence intervals for the treatment difference. The protocol did not specify margins for interpreting the confidence intervals. The response rates for PASI 75 and sPGA at Week 16 were each approximately 7-8% lower on the ABP 501 arm than on the EU-approved Humira arm. Similarly, the reduction from baseline in BSA was slightly lower on the ABP 501 arm than the EU-approved Humira arm. The 90% confidence intervals for the PASI 75 and BSA reduction endpoints do not include 0, but in both cases the 95% confidence intervals do. Both the 90% and 95% confidence intervals for sPGA response include 0. See Table 11.

**Table 11 - Secondary Endpoints at Week 16 (FAS/LOCF)**

	ABP 501 N=172	EU-approved Humira N=173	Difference <sup>a</sup>	90% Conf. Int.	95% Conf. Int.
PASI 75	74.4%	82.7%	-7.7%	(-15.2, -0.3)	(-16.6, 1.2)
sPGA (clear/almost clear)	58.7%	65.3%	-7.4%	(-15.6, 0.9)	(-17.2, 2.5)
Reduction in BSA					
Baseline (Week 1)	25.3	28.5			
Week 16	7.4	6.4			
Reduction	18.0	22.1	-1.9	(-3.8, -0.1)	(-4.1, 0.2)

<sup>a</sup> Model estimate adjusted for prior biologic use, region, and baseline score

Source: pg 354, 368, and 388 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

The PASI and sPGA scales are correlated as both scales measure the same underlying signs of erythema, scaling, and plaque elevation. Thus, it is not unexpected that endpoints based on these scales and BSA assessments would generally trend in the same direction. In addition, we would expect some variation in the magnitude of effect for different analyses when multiple analyses are conducted in a study. Because the 90% confidence interval for PASI 75 excluded 0 (although the 95% confidence interval included 0) and the fact that PASI 75 has been used as a primary endpoint in many clinical trials for psoriasis, this reviewer further evaluated the distribution of PASI scores and related endpoints (PASI 50, PASI 90, and absolute reduction in PASI). PASI 50 and PASI 90 response rates are presented in Table 12 along with the PASI 75 response rates at Week 16. Table 12 also presents the absolute reduction in PASI score from baseline to Week 16. When PASI 50 and PASI 90 are considered, the estimated treatment differences are smaller (-2.7% and +0.3%) than for PASI 75 (-7.7%). In addition the estimated treatment difference for the absolute reduction in PASI was less than 1 unit, with a narrow confidence interval that contains 0.

**Table 12 –Supportive Endpoints based on PASI Score at Week 16 (FAS/LOCF)**

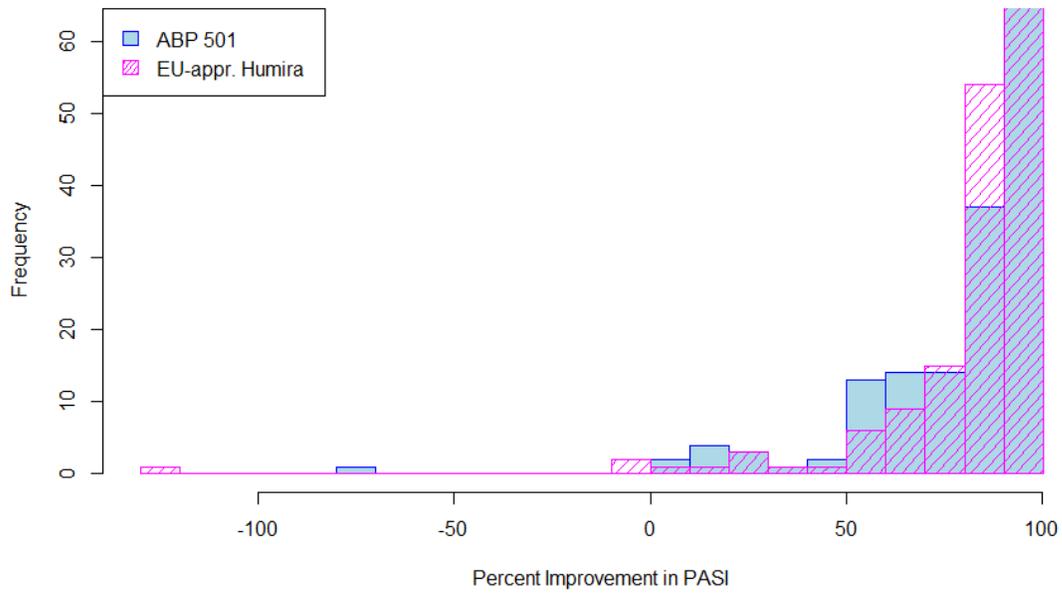
	ABP 501 N=172	EU-approved Humira N=173	Difference <sup>a</sup>	90% Conf. Int.	95% Conf. Int.
PASI 50	92.4%	94.2%	-2.7%	(-7.0, 1.6)	(-7.8, 2.4)
PASI 75	74.4%	82.7%	-7.7%	(-15.2, -0.3)	(-16.6, 1.2)
PASI 90	47.1%	47.4%	0.3%	(-8.4, 9.0)	(-10.0, 10.7)
Reduction in PASI					
Baseline (Wk 1)	19.8 (8.1)	20.5 (7.9)			
Week 16	3.7 (5.1)	3.3 (5.8)			
Reduction	16.0 (8.1)	17.2 (9.2)	-0.58	(-1.5, 0.4)	(-1.7, 0.5)

<sup>a</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Source: pg 343 and 439 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](https://cdsesub1.evsprod/bla761024/0001/m5/53-clin-stud-rep/535-rep-effic-safety-stud/plaque-psoriasis/5351-stud-rep-contr/20120263/02-csr-20120263-rpt-body.pdf) and reviewer analysis

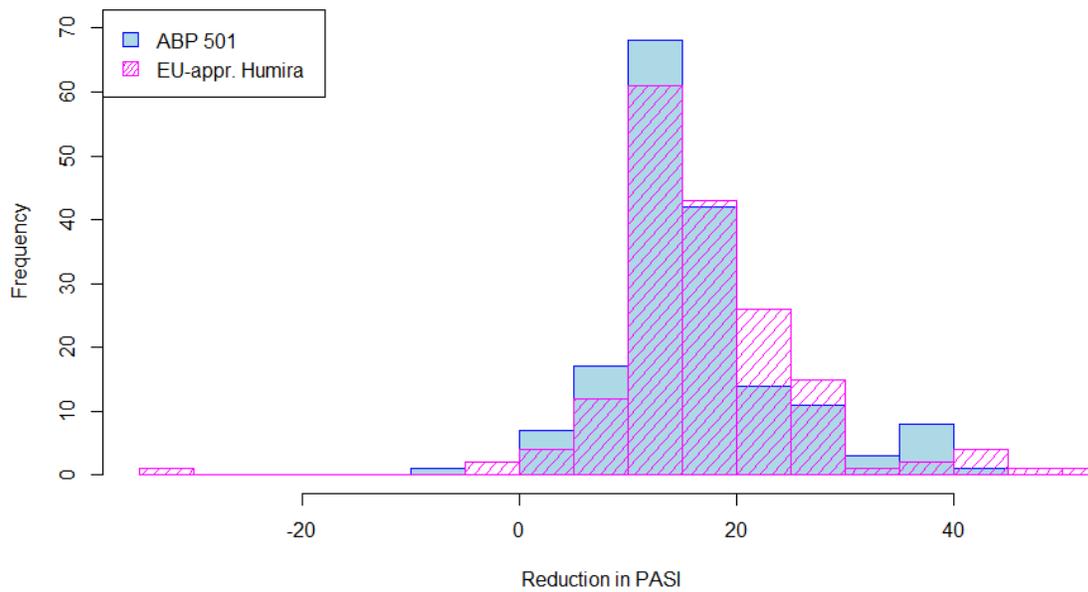
The overlaid histograms for the percent improvement and absolute reduction in PASI for ABP 501 and EU-approved Humira are presented in Figure 3 and Figure 4. The distribution of percent improvement in PASI is highly skewed with a few outliers, while the distribution for the absolute reduction in PASI is more symmetric. The slight difference in observed means for the two samples can be seen as a slight shift in location in each pair of histograms. The differences between the two samples appear to be magnified when dichotomizing the percent improvement in PASI using 75% improvement as the cutoff point, as opposed to other potential cutoff points. Thus, when considering the full distributions, the supportive PASI endpoints are consistent with the primary analysis of the mean percent improvement in PASI, and support the conclusion of the primary endpoint of no clinically meaningful differences between the treatments.

**Figure 3 – Histogram of Percent Improvement in PASI at Week 16 (FAS/LOCF)**



Source: reviewer analysis.

**Figure 4 – Histogram of Absolute Reduction in PASI at Week 16 (FAS/LOCF)**



Source: reviewer analysis.

### 3.2.6 Interpretation of Comparative Clinical Studies

Study 263 was a comparative clinical study of ABP 501 and EU-approved Humira; it did not include a placebo arm. Thus we need to evaluate whether the study has adequate assay sensitivity (the ability to detect meaningful differences if they were to exist) and have confidence that the pre-specified margin is appropriate. Three placebo-controlled trials of Humira have been published (Gordon (2006), Saurat (2008), and Menter (2008)). Each of these studies had PASI 75 as the primary endpoint, but all three also presented the percent improvement in PASI results at either Week 12 or Week 16. Note that for Study 263, baseline was defined as Week 1, while in the published studies baseline was defined as Week 0. Therefore for comparative purposes, the primary timepoint in Study 263 will be referred to as Week 15 in this section. The key design criteria and results for the published Humira studies are presented in Table 13. The Gordon study had less restrictive inclusion criteria ( $BSA \geq 5$ , no requirement on PASI), but the Saurat and Menter studies had similar inclusion criteria to Study 263 ( $BSA \geq 10$ ,  $PASI \geq 10$  or 12, and  $sPGA \geq \text{Moderate}$ ). The percent improvement in PASI scores from Study 263 on the EU-approved Humira arm (83) was generally consistent with the percent improvement in PASI scores from the published Humira studies at Weeks 12-16 (70-81). Because the means for the percent improvement in PASI on the placebo arm (14-22) were generally much smaller than the means for the Humira arm, the assay sensitivity assumption appears reasonable for Study 263.

**Table 13 – Study Characteristics and Results of Published Humira Studies**

	Gordon (2006)	Saurat (2008)	Menter (2008)	Study 263
Selected inclusion criteria	$BSA \geq 5$	$BSA \geq 10$ $PASI \geq 10$ $sPGA \geq \text{Mod}$	$BSA \geq 10$ $PASI \geq 12$ $sPGA \geq \text{Mod}$	$BSA \geq 10$ $PASI \geq 12$ $sPGA \geq \text{Mod}$
Region/Country	US, Canada	Europe, Canada	US, Canada	Europe, Canada, Australia
Baseline PASI Mean ( <i>Humira</i> )	PASI = 16.7	PASI = 20.2	PASI = 19.0	PASI = 20.5
% Imp. in PASI <i>Humira</i> <i>Placebo</i>	(Week 12) 70 14	(Week 16) 81 22	(Week 12) 76 15	(Week 15 <sup>a</sup> ) 83 --
PASI 75 <i>Humira</i> <i>Placebo</i>	(Week 12) 53% (n=50) 4% (n=52)	(Week 16) 80% (n=108) 19% (n=53)	(Week 16) 71% (n=814) 7% (n=398)	(Week 15 <sup>a</sup> ) 83% (n= 173) --

<sup>a</sup> 15 weeks after the baseline visit

Study 263 had a pre-specified similarity margin of  $\pm 15$  for the primary endpoint of percent improvement in PASI. The applicant did not provide a rationale in their protocol

for the size of the proposed margin, and the margin was not discussed with FDA prior to conducting the study. While ideally the similarity margin would be selected based on a consensus of what magnitude of difference for the endpoint is not clinically meaningful, in practice sample sizes may be constrained by feasibility concerns. This reviewer took two approaches to assess the applicant’s margin. The first approach computed the percent preservation of effect, to ensure that the test product would maintain at least some benefit relative to placebo. However, the goal of a comparative clinical study is to support the demonstration of no clinically meaningful differences. Therefore this reviewer also evaluated what margins would lead to an adequately powered study for a given sample size.

Although the Gordon, Saurat, and Menter studies included mean values for the percent improvement in PASI at either Week 12 or 16, none of the studies included standard deviations, which are needed to construct confidence intervals. Thus, alternate sources are needed to find reasonable estimates of the standard deviation for this endpoint. Two publications for studies of other TNF- $\alpha$  inhibitors (Enbrel and Remicade) presented standard deviations for the percent improvement in PASI endpoint (Table 14). Based on these publications, standard deviation estimates in the range of 20 to 30, may be a reasonable approximations for the purpose of constructing confidence intervals to aid in the evaluation the applicant’s proposed margin.

**Table 14 - Published Estimates of the Standard Deviation for the Percent Improvement in PASI Endpoint in Trials of Other TNF- $\alpha$  Inhibitors**

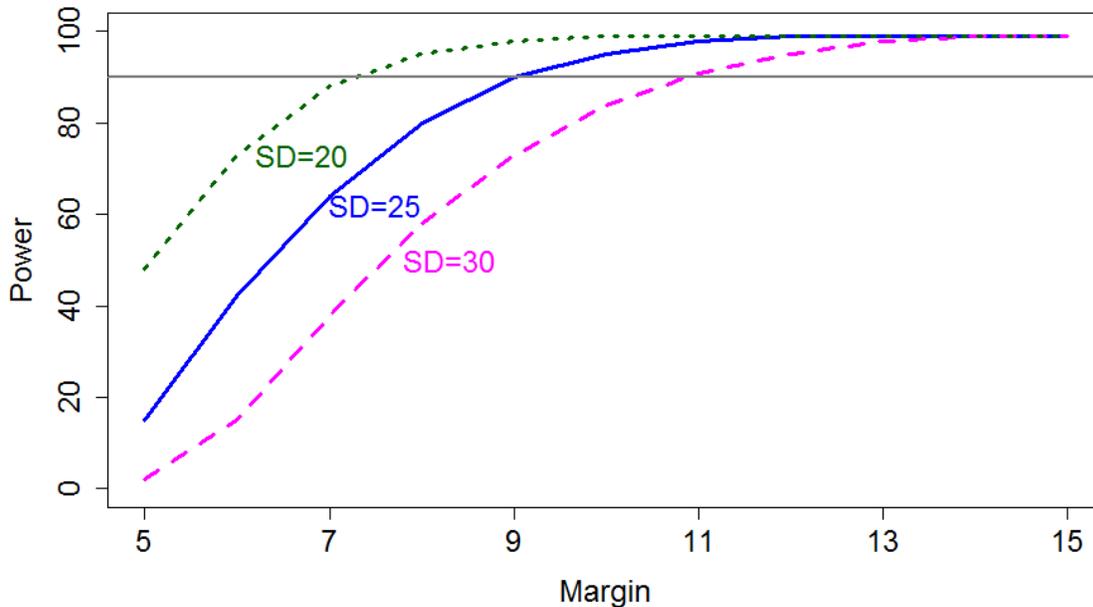
Study	Product	Week	N	Mean	Standard Deviation
Leonardi (2003)	Enbrel	12	164	64.2	30.7
Reich (2005)	Remicade	10	301	85.5	21.4

This reviewer calculated the percent preservation of the margin relative to the point estimate and an approximate lower 95% confidence bound for the treatment effect for the percent improvement in PASI. These calculations use the point estimate for percent improvement in PASI (61) and sample sizes ( $n_1 = 814$ ,  $n_2 = 398$ ) from the largest of the three Humira studies (Menter) and a standard deviation estimate in the upper end of the range observed in the Leonardi and Reich studies ( $SD=30$ ). An approximate 95% confidence interval for the treatment effect for percent improvement in PASI for Humira would be  $61 \pm 3.6 = (57.4, 64.6)$ . Thus a lower bound margin of -15 maintains at least 75% of the expected treatment effect using the point estimate of 61 and at least 74% of the expected treatment effect using the lower 95% confidence bound of 57.4.

Although lower bound margin of -15 maintains a substantial portion of the expected treatment effect, because the estimated treatment effect relative to placebo is large, even retaining a substantial portion of the treatment effect relative to placebo could lead to clinically meaningful differences between treatments. Thus, the relationship between the study power and various margins for a given sample size is also of interest. Using the sample size originally proposed in the protocol of 340 subjects and the assumption that

the two treatments have the same effect, we can get a sense of what margins would lead to a design with adequate power. Figure 5 displays the relationship between study power and margin, assuming the true treatment difference is 0, total sample size of 340 subjects (170 per arm), symmetric margins, 90% confidence level, and standard deviations of 20, 25, and 30. Using the more conservative standard deviation estimate of 30, we see that a study of the proposed design and sample size would be powered at 90% for margins with magnitude of about  $\pm 11$  or greater. We note that in Study 263, the 90% confidence interval for the percent improvement in PASI was (-6.6, 2.2), and the endpoint would have met the similarity criteria for margins with magnitude  $\pm 7$  or greater. Thus the confidence interval for the primary endpoint of percent improvement in PASI is sufficiently narrow to conclude that the study met the criteria for demonstrating similarity.

**Figure 5 – Study Power versus Margin Magnitude (Assuming True Treatment Difference = 0, N=340 and Symmetric Margins)**



Source: reviewer analysis

### 3.3 Evaluation of Safety

#### 3.3.1 Extent of Exposure

The extent of exposure to study drug was similar for subjects randomized to ABP 501 and EU-approved Humira in the first treatment period, with approximately 90 days of drug exposure on each arm and approximately 87% of subjects receiving all 8 planned doses in the first treatment period. The mean total dose in the first treatment period was similar on both arms. All subjects received at least 2 doses. See Table 15. Exposure was also similar across the arms during the second treatment period. See Table 16.

**Table 15 – Extent of Drug Exposure in Treatment Period 1**

	ABP 501 N=174	EU-approved Humira N=173
Exposure Days		
Mean (SD)	89.5 (12.5)	89.9 (9.2)
Range	6-99	36-99
Total Dose Received (mg)		
Mean (SD)	349.9 (36.9)	350.8 (28.4)
Range	120 - 360	200-360
Number of Doses Administered		
1	--	--
2	3 (2%)	--
3	--	--
4	1 (<1%)	4 (2%)
5	1 (<1%)	--
6	3 (2%)	4 (2%)
7	13 (8%)	16 (9%)
8	153 (88%)	149 <sup>a</sup> (86%)

<sup>a</sup> One subject received the initial 80 mg dose as two 40 mg doses two days apart for a total of 9 injections  
Source: pg 457 of <\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf> and reviewer analysis

**Table 16 – Extent of Drug Exposure in Treatment Period 2**

	ABP 501/ ABP 501 N=152	EU-appr. Hum./ EU-appr. Hum. N=79	EU-appr. Hum./ ABP 501 N=77
Exposure Days			
Mean (SD)	211.9 (43.8)	208.8 (51.1)	211.2 (45.5)
Range	13 - 233	1 - 232	15 – 232
Total Dose Received (mg)			
Mean (SD)	634.1 (124.6)	627.3 (146.4)	626.5 (131.0)
Range	80-720	40-720	80-680

Source: pg 459 of <\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf> and reviewer analysis

### 3.3.2 Adverse Events

Similar rates of adverse events, serious adverse events, and study discontinuations due to adverse events occurred on the ABP 501 and EU-approved Humira arms. No deaths occurred during the study. See Table 17.

**Table 17 – Summary of Adverse Events (Safety Population)**

	ABP 501 N=174	EU-approved Humira N=173	
Treatment Period 1			
Any Adverse Events	117 (67%)	110 (64%)	
Serious Adverse Events	6 (3%)	5 (3%)	
Discontinued Study due to AE	7 (4%)	5 (3%)	
Treatment Period 2	ABP 501/ ABP 501 N=152	EU-appr. Hum./ EU-appr. Hum. N=79	EU-appr. Hum./ ABP 501 N=77
Any Adverse Events	108 (71%)	52 (66%)	54 (70%)
Serious Adverse Events	4 (3%)	4 (5%)	4 (5%)
Discontinued Study due to AE	4 (3%)	1 (1%)	2 (3%)

Source: pg 69-70 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

Adverse events of special interest were infections, malignancies, hypersensitivity, demyelinating diseases, hematological reactions, heart failure, lupus-like syndrome, liver enzyme elevations, and injection site reactions. No cases of demyelinating disease, heart failure, or lupus-like syndromes were reported during the study. Rates of observed adverse events of special interest were similar on the ABP 501 and EU-approved Humira arms. See Table 18.

**Table 18 – Adverse Events of Special Interest (Safety Population)**

	ABP 501 N=174	EU-approved Humira N=173	
Treatment Period 1			
Infections	59 (34%)	58 (34%)	
Hypersensitivity	8 (5%)	7 (4%)	
Injection site reactions	3 (2%)	9 (5%)	
Liver enzyme elevations	4 (2%)	2 (1%)	
Hematological reactions	--	3 (2%)	
Malignancies	1 (<1%)	1 (<1%)	
Treatment Period 2	ABP 501/ ABP 501 N=152	EU-appr. Hum./ EU-appr. Hum. N=79	EU-appr. Hum./ ABP 501 N=77
Infections	67 (44%)	29 (37%)	37 (48%)
Hypersensitivity	8 (5%)	2 (3%)	3 (4%)
Injection site reactions	2 (1%)	3 (4%)	--
Liver enzyme elevations	9 (6%)	2 (3%)	2 (3%)
Hematological reactions	--	1 (1%)	1 (1%)
Malignancies	1 (<1%)	--	--

Source: pg 88-90 of [\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

### 3.3.3 Immunogenicity

During the initial treatment period, 17/174 (10%) ABP 501 subjects and 24/173 (14%) EU-approved Humira subjects developed neutralizing antibodies. Eleven of the ABP 501 subjects and 18 of the EU-approved Humira subjects with neutralizing antibodies continued into the second treatment period. Among the subjects who received EU-approved Humira in the first treatment period and were re-randomized in the second treatment period, 16/79 (20%) of subjects remaining on EU-approved Humira developed neutralizing antibodies during the study (9 in the first treatment period and 7 in the second treatment period) compared with 19/77 (25%) of subjects who transitioned to ABP 501 (9 in the first treatment period and 10 in the second treatment period). Among the subjects who remained in the study and received ABP 501 during both treatment periods, 21/152 (14%) developed neutralizing antibodies during the study (11 in the first treatment period and 10 during the second treatment period). See Table 19.

**Table 19 – Neutralizing Antibodies (NAb)**

Treatment in Period 1	ABP 501 N=174		EU-approved Humira N=173		
Treatment in Period 2	Not Re- randomized N=22	ABP 501 N=152	Not Re- randomized N=17	EU-appr. Humira N=79	ABP 501 N=77
First Positive Result for NAb in Treatment Period 1	6	11	6	9	9
<i>Total</i>	<i>17</i>		<i>24</i>		
First Positive Result for NAb in Treatment Period 2	7 <sup>a</sup>	10	1 <sup>a</sup>	7	10
Any Positive Result for NAb during Study	13	21	7	16	19

<sup>a</sup> For subjects not re-randomized, first positive result may have occurred during post-treatment follow-up  
Source: pg 1413-1415 of <\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf> and reviewer analysis

## 4 Findings in Special/Subgroup Populations

### 4.1 Gender, Race, Age, and Geographic Region

The mean percent improvement in PASI values at Week 16 were generally consistent across gender. The study enrolled too few non-white subjects and subjects over the age of 65 to have meaningful comparisons for these subgroups. Results were also generally consistent across geographic regions. Geographic region (Eastern Europe, Western Europe, and Other) was a stratification factor in the initial randomization. See Table 20.

**Table 20 – Percent Improvement in PASI at Week 16 by Gender, Race, Age Group, and Geographic Region (FAS)**

	ABP 501 N=172	EU-approved Humira N=173	Difference <sup>a</sup>	90% Conf. Int.
Gender				
Female	N=63 77.7 (31.9)	N=58 76.8 (36.4)	0.9	(-9.5, 11.23)
Male	N=109 82.8 (18.4)	N=115 86.2 (16.2)	-3.5	(-7.3, 0.3)
Race				
White	N=164 80.7 (24.7)	N=157 84.4 (23.8)	-3.7	(-8.1, 0.8)
Non-White	N=6 86.2 (12.2)	N=13 72.1 (29.7)	10.2	(-16.2, 36.5)
Age				
<65 years	N=161 81.2 (24.3)	N=161 83.1 (25.7)	-1.9	(-6.4, 2.7)
≥ 65 years	N=11 76.4 (24.7)	N=12 83.3 (17.4)	-4.5	(-20.7, 11.8)
Geographic Region				
Eastern Europe	N= 71 84.4 (19.8)	N=70 88.4 (15.7)	-4.1	(-9.1, 0.9)
Western Europe	N=41 75.0 (32.3)	N=43 78.4 (22.5)	-3.0	(-13.1, 7.1)
Other	N=60 80.8 (22.2)	N=60 80.1 (33.8)	0.1	(-8.6, 8.8)

<sup>a</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Source: pg 299-317 of [\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](https://cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf) and reviewer analysis

#### **4.2 Other Special/Subgroup Populations**

In addition to geographic region, the randomization was also stratified by prior use of biologics for psoriasis (yes/no). A relatively small proportion of subjects (18%) had prior biologic use. In general, the results were consistent across prior biologic use. See Table 21.

**Table 21 – Percent Improvement in PASI at Week 16 by Prior Biologic Use**

	ABP 501 N=172	EU-approved Humira N=173	Difference <sup>a</sup>	90% Conf. Int.
Prior Biologic Use				
Yes	N=32 79.5 (32.3)	N=30 76.0 (43.3)	3.3	(-12.8, 19.4)
No	N=140 81.2 (22.1)	N=143 84.5 (19.3)	-3.3	(-7.4, 0.7)

<sup>a</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Source: pg 287 of [\\cdsesub1\evsprod\bla761024\0001\m5\53-clin-stud-rep\535-rep-effic-safety-stud\plaque-psoriasis\5351-stud-rep-contr\20120263\02-csr-20120263-rpt-body.pdf](#) and reviewer analysis

## 5 Summary and Conclusions

### 5.1 Statistical Issues and Collective Evidence

The mean percent improvement in PASI at Week 16 was similar on the ABP 501 and EU-approved Humira arms and the confidence interval for the difference was within the pre-specified margin of  $\pm 15$ . In the applicant’s full analysis set (FAS), defined as all subjects randomized and dispensed medication who had at least one post-baseline efficacy assessment, the mean percent improvement in PASI values on the ABP 501 and EU-approved Humira arms were 80.9 vs 83.1. Results on the per protocol population and an analysis population that includes all subjects randomized and dispensed medication whether or not they had post-baseline efficacy assessments were similar and also fell within the pre-specified margin.. See Table 22. The results of the secondary endpoints of PASI 75, clear or almost clear on the static Physician’s Global Assessment, and reduction from baseline in body surface area were consistent with the primary endpoint.

**Table 22 – Percent Improvement in PASI at Week 16**

	ABP 501	EU-approved Humira	Difference <sup>d</sup>	90% Conf. Int.
Full Analysis Set <sup>a</sup> (LOCF)	N=172 80.9	N=173 83.1	-2.2	(-6.6, 2.2)
Sensitivity Analysis <sup>b</sup> (LOCF)	N=174 80.0	N=173 83.1	-3.1	(-7.5, 1.4)
Per protocol <sup>c</sup> (Observed)	N=155 82.6	N=152 85.3	-2.6	(-6.2, 0.9)

<sup>a</sup> Randomized, dispensed medication, and at least one post-baseline efficacy assessment

<sup>b</sup> Randomized, dispensed medication

<sup>c</sup> Completed the treatment period without protocol violations that affected the evaluation of the primary objective

<sup>d</sup> Model estimate adjusted for prior biologic use, region, and baseline PASI

Because Study 263 was conducted completely outside the US, the applicant did not discuss the proposed similarity margin with the FDA prior to conducting the study. The applicant did not provide a rationale for their choice of similarity margin in the protocol

or study report. Therefore, this reviewer evaluated the applicant's proposed margin using information from the literature on the percent improvement in PASI from published placebo-controlled studies of Humira and other TNF- $\alpha$  inhibitors. Based on this evaluation, we conclude that assumptions of consistency and assay sensitivity appear reasonable for Study 263, and that the confidence interval for the primary endpoint of percent improvement in PASI is sufficiently narrow to conclude that the study met the criteria for demonstrating similarity.

Adverse event rates were similar on both the ABP 501 and EU-approved Humira arms. During the initial treatment period, 10% of ABP 501 subjects and 14% of EU-approved Humira subjects developed neutralizing antibodies. Among the subjects who continued into the second treatment period, 20% of subjects on EU-approved Humira/EU-approved Humira arm, 25% on the EU-approved Humira/ABP 501 arm, and 14% on the ABP 501/ABP 501 arm developed neutralizing antibodies during the study.

## **5.2 Conclusions and Recommendations**

We conclude that Study 263 met its objective for assessing clinical similarity and that Study 263 supports a demonstration of no clinically meaningful differences between ABP 501 and US-licensed Humira.

## **References**

- Gordon KB et al, J Am Acad Dermatol. 2006 Oct; 55(4): 598-606
- Saurat JH et al, Br J Dermatol. 2008; 158: 558-66
- Menter A et al, J Am Acad Dermatol. 2008; 58(1): 106-15
- Leonardi CL et al, N Engl J of Med. 2003; 349:2014-22
- Reich K et al, Lancet. 2005; 366:1367-74

## **Signatures/Distribution List**

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.  
Date: 9/7/2016

Statistical Team Leader: Mohamed Alosch, Ph.D.

cc:

DDDP/Marcus  
DDDP/Diglisic  
DDDP/Cook  
DDDP/Phillips  
DPARP/Nikolov  
DPARP/Hull  
OBIO/Patrician  
DBII/Kim  
DBII/Levin  
DBIII/Wilson  
DBIII/Alosch  
DBIII/Fritsch

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

/s/  
-----

KATHLEEN S FRITSCH  
09/07/2016

MOHAMED A ALOSH  
09/07/2016



## STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

<b>BLA No.:</b>	761024
<b>SERIAL No.:</b>	0000
<b>DATE RECEIVED BY THE CENTER:</b>	November 24, 2015
<b>DRUG NAME:</b>	ABP 501 (proposed biosimilar to Humira, AbbVie)
<b>DOSAGE FORM:</b>	<ul style="list-style-type: none"><li>• Pre-filled Syringes (PFS): [REDACTED] 50 mg/1.0 mL</li><li>• Prefilled autoinjectors for subcutaneous use: 50 mg/1.0 mL</li></ul>
<b>INDICATIONS:</b>	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)
<b>APPLICANT:</b>	Amgen Inc.
<b>REVIEW FINISHED:</b>	June 1, 2016
<b>NAME OF STATISTICAL REVIEWER:</b>	Meiyu Shen
<b>NAME OF PROJECT MANAGER:</b>	Sadaf Nabavian

\_\_\_\_\_  
Meiyu Shen, PhD, Lead Mathematical Statistician

Concur:

\_\_\_\_\_  
Yi Tsong, Ph.D., Division Director, DBVI

Distribution: CDER/OTS/OB/DB VI/Yi Tsong  
CDER/OTS/OB/Lillian Patrician  
CDER/OBP/Jun Park  
CDER/OBP/Joel Welch  
CDER/TBBS

## TABLE OF CONTENTS

<b>1</b>	<b><i>Executive summary and recommendation</i></b> .....	<b>3</b>
<b>2</b>	<b><i>Introduction</i></b> .....	<b>3</b>
<b>3</b>	<b><i>Data analyzed</i></b> .....	<b>4</b>
<b>4</b>	<b><i>Applicant’s statistical equivalence testing</i></b> .....	<b>4</b>
<b>5</b>	<b><i>FDA statistical analyses</i></b> .....	<b>5</b>
5.1	<b>Statistical method</b> .....	<b>5</b>
5.2	<b>FDA statistical equivalence testing for apoptosis inhibition bioassay</b> .....	<b>6</b>
5.3	<b>FDA statistical equivalence testing for sTNF-<math>\alpha</math> binding</b> .....	<b>8</b>
<b>6</b>	<b><i>Conclusion and recommendation</i></b> .....	<b>9</b>

## 1 EXECUTIVE SUMMARY AND RECOMMENDATION

The CMC statistics reviewer in the Office of Biostatistics analyzed the comparative results of 2 critical quality attributes: Apoptosis inhibition bioassay and sTNF- $\alpha$  binding, 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  $R$  represents US-licensed reference product variability or the comparator variability. 10 batches of ABP 501 and 21 batches of US-licensed Humira, and 17 batches of EU-approved Humira were used for equivalence testing of apoptosis inhibition bioassay (potency). The results are summarized in Table 1.

**Table 1 Results of equivalence testing for apoptosis inhibition bioassay (potency)**

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
ABP 501 vs. US	(10, 21)	-1.43	(-4.50, 1.93)	(-8.57, 8.57)	Yes
ABP 501 vs. EU	(10, 17)	1.12	(-3.37, 5.82)	(-14.04, 14.04)	Yes
EU vs. US	(17, 21)	-2.55	(-6.97, 1.88)	(-8.57, 8.57)	Yes

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

Ten batches of ABP 501, 10 batches of US-licensed Humira, and 10 batches of EU-approved Humira are included in the TNF- $\alpha$  binding dataset for the statistical equivalence testing. The results are shown in Table 2.

**Table 2 Results of equivalence testing for sTNF- $\alpha$  binding**

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
ABP 501 vs. US	(10, 10)	-3.60	(-10.93, 3.73)	(-14.97, 14.97)	Yes
ABP 501 vs. EU	(10, 10)	-3.00	(-9.23, 3.23)	(-10.54, 10.54)	Yes
EU vs. US	(10, 10)	-0.60	(-7.34, 6.14)	(-14.97, 14.97)	Yes

As shown in Tables 1 and 2, the results from the statistical equivalence testing of apoptosis inhibition bioassay (potency) and sTNF- $\alpha$  binding support a demonstration that the proposed biosimilar ABP 501 is highly similar to US-licensed Humira and also support the analytical bridge between US-licensed Humira and EU-approved Humira.

## 2 INTRODUCTION

On November 24, 2015, the applicant (Amgen) submitted to the US Food and Drug Administration (FDA) a 351(k) BLA which included an analytical similarity assessment of comparing ABP 501 and US-licensed Humira.

## Statistical Review of BLA761024

The applicant characterized multiple batches of US-licensed Humira and EU-approved Humira using a comprehensive set of analytical methods during the ABP 501 development.

The Agency carefully evaluated data for the apoptosis inhibition bioassay and sTNF- $\alpha$  binding provided in the initial BLA submission. Our comments regarding Amgen'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

Amgen submitted the analytical data on November 24, 2015. Note that in Table 3, the apoptosis inhibition bioassay data of 21 US-licensed Humira lots, 17 EU-approved Humira lots, 10 ABP 501 lots were submitted by Amgen.

In addition, Amgen provided and analyzed the sTNF-alpha binding for 10 lot values of EU-approved Humira, 10 lot values of ABP 501, and 10 lot values of US-licensed Humira.

**Table 3 Number of batches from each product**

Product	Number of batches	
	apoptosis inhibition bioassay (potency)	sTNF- $\alpha$ binding
US-licensed Humira	21	10
ABP 501	10	10
EU-approved Humira	17	10

### 4 APPLICANT'S STATISTICAL EQUIVALENCE TESTING

In this submission, Amgen conducted Tier 1 statistical equivalence testing with the margin defined as  $1.5\hat{\sigma}_R$  for apoptosis inhibition bioassay (potency) and sTNF- $\alpha$  binding. Amgen performed = the Brown and Forsythe's Test for Homogeneity of Variance to determine if sample variances should be pooled in computing the confidence interval for the difference of means between the test product and reference product. If the p-value exceeds 0.05, then the pooled variance is used to compute the confidence interval. If the p-value is less than 0.05, the confidence interval for unequal variances is employed using the Satterthwaite approximation to determine the degrees of freedom. How to calculate the 90% confidence interval depends on the hypothesis test for equal variance. To demonstrate statistical equivalence for apoptosis inhibition bioassay (potency) and sTNF- $\alpha$  binding in this context, the entire two-sided confidence interval must be contained in the range from  $-1.5\hat{\sigma}_R$  to  $1.5\hat{\sigma}_R$ .

*Reviewer's comments: Applicant's analyses did not adjust the impact of imbalance sample sizes of the test product and the reference product.*

## 5 FDA STATISTICAL ANALYSES

To evaluate analytical similarity, the Agency recommended that Amgen apply a tiered approach in the Agency's responses to IND meetings with Amgen. 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  $(\text{Mean} - X \cdot \text{SD}, \text{Mean} + X \cdot \text{SD})$  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  $\mu_T$  and  $\mu_R$  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 Humira product. Let  $\sigma_R$  be the standard deviation of the quality attribute of interest for the US-licensed Humira. In order to conclude the equivalence in the quality attribute of interest between the test product and the US-licensed Humira product, we aim to reject the null hypothesis of the following null and alternative hypotheses:

$$H_0 : \mu_T - \mu_R \leq \theta_1 \text{ or } \mu_T - \mu_R \geq \theta_2$$

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

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

We reject  $H_0$  if 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 that the equivalence in the quality attribute of interest between the test product and the US-licensed Humira product if 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 Humira 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 Humira lots. First we estimate  $\sigma_R$  by the sample variability of the US-licensed Humira product (or by the sample variability of EU-approved Humira in the comparison between ABP 501 and EU-approved Humira) and then in the statistical analysis,  $\theta_1$  and  $\theta_2$  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 Humira product. Since the two products are manufactured by two manufacturers, two groups are independent.  $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i$ , and

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n_i - 1), \text{ where } n_i \text{ is the number of lots in the } i^{\text{th}} \text{ product, } i = T, R.$$

Under the unequal variance of the test product and the US-licensed Humira product, the  $(1-2\alpha)*100\%$  confidence interval of the mean difference in the quality attribute of interest can be calculated as:

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

Here  $t_\alpha(v)$  is the  $1-\alpha$  quantile and  $v$  is the degrees of freedom calculated by Satterthwaite's approximation.

If  $n_R > 1.5n_T$ , the  $(1-2\alpha)*100\%$  confidence interval of the mean difference in the quality attribute of interest can be calculated as:

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

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

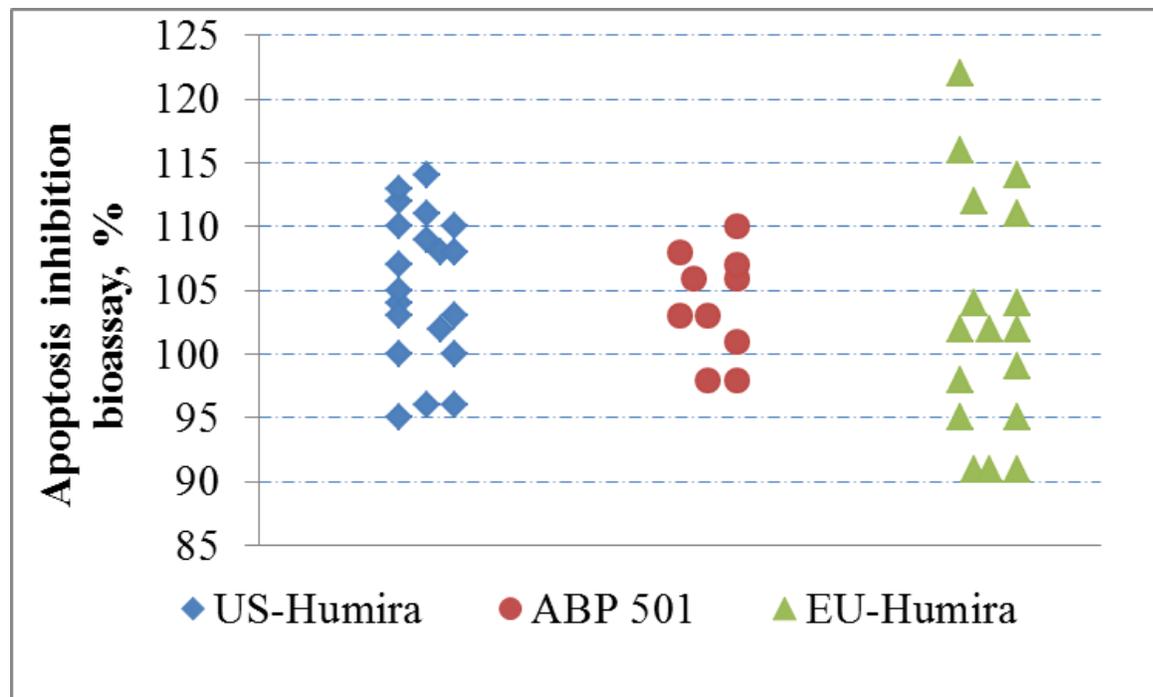
If the number of biosimilar lots,  $n_T$ , is 50% more than the number of reference lots,  $n_R$ , we can apply a similar approach as above with  $n_T^* = \min(1.5 \times n_R, n_T)$  for the confidence interval calculation. In the following analyses, we use  $\alpha=0.05$ .

## 5.2 FDA statistical equivalence testing for apoptosis inhibition bioassay

The apoptosis inhibition bioassay data points of ABP 501, US-licensed Humira, and EU-approved Humira are displayed in Figure 1. There appears a small mean difference among the 3 products. The variability of ABP 501 is smallest among 3 products.

Ten batches of ABP 501, 21 batches of US-licensed Humira, and 17 batches of EU-approved Humira are included for the statistical equivalence testing for the apoptosis inhibition bioassay. Descriptive statistics for the apoptosis inhibition bioassay data are listed in Table 4.

**Figure 1 Scatter plot of Apoptosis inhibition bioassay for US-licensed Humira, ABP 501, and EU-approved Humira**



**Table 4 Descriptive statistics for the apoptosis inhibition bioassay data**

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Humira	21	105.43	5.71	95	114
ABP 501	10	104	4.11	98	110
EU-approved Humira	17	102.88	9.36	91	114

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 ABP 501. From Table 5, it is seen that the apoptosis inhibition bioassay of ABP 501 is equivalent to the apoptosis inhibition bioassay of US-licensed Humira. Similarly, the apoptosis inhibition bioassay of ABP 501 is equivalent to the apoptosis inhibition bioassay of EU-approved Humira, and the apoptosis inhibition bioassay of EU-approved Humira is equivalent to the apoptosis inhibition bioassay of US-licensed Humira.

**Table 5 Equivalence testing results for the apoptosis inhibition bioassay**

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
ABP 501 vs. US	(10, 21)	-1.43	(-4.50, 1.93)	(-8.57, 8.57)	Yes
ABP 501 vs. EU	(10, 17)	1.12	(-3.37, 5.82)	(-14.04, 14.04)	Yes
EU vs. US	(17, 21)	-2.55	(-6.97, 1.88)	(-8.57, 8.57)	Yes

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

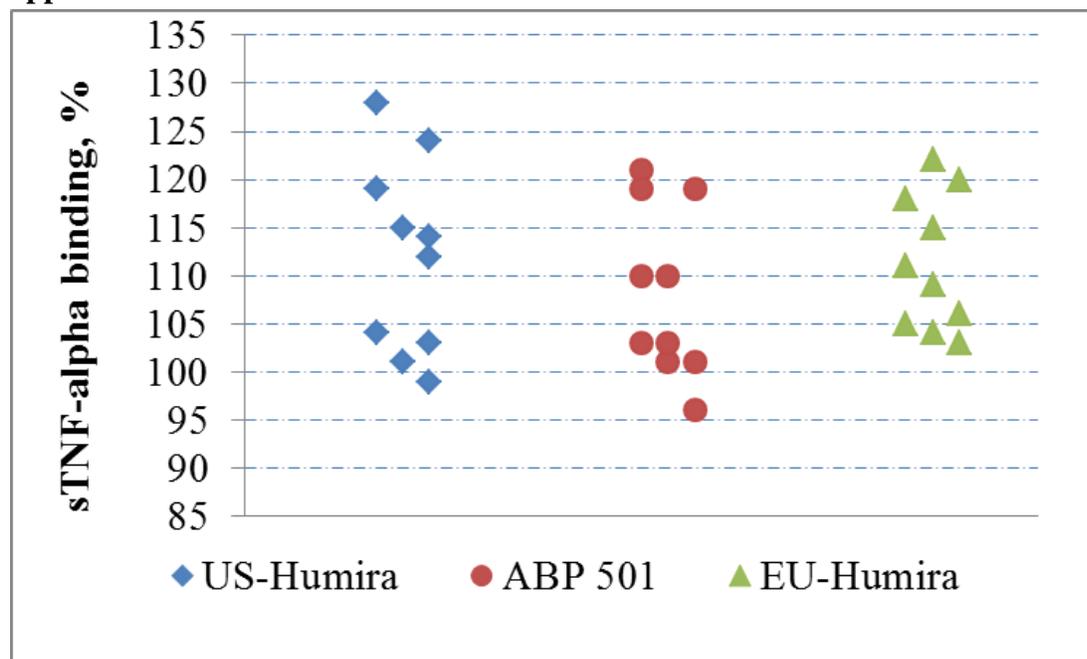
### 5.3 FDA statistical equivalence testing for sTNF- $\alpha$ binding

The sTNF- $\alpha$  binding data points of ABP 501, US-licensed Humira, and EU-approved Humira are displayed in Figure 2. Clearly there is a mean shift between the US-licensed Humira and ABP 501.

Ten batches of ABP 501, 10 batches of US-licensed Humira, and 10 batches of EU-approved Humira are included in the sTNF- $\alpha$  binding dataset for the statistical equivalence testing. Descriptive statistics for the sTNF- $\alpha$  binding data of ABP 501, US-licensed Humira, and EU-approved Humira are listed in Table 6.

From Table 7, it is seen that the equivalence of sTNF- $\alpha$  binding between ABP 501 and US-licensed Humira is supported. The equivalence of sTNF- $\alpha$  binding between ABP 501 and EU-approved Humira is supported. The equivalence of sTNF- $\alpha$  binding between US-licensed Humira and EU-approved Humira is supported.

**Figure 2 Scatter plot of sTNF- $\alpha$  binding for US-licensed Humira, ABP 501, and EU-approved Humira**



**Table 7 Descriptive statistics for the sTNF- $\alpha$  binding data**

Product	Number of batches	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
US-licensed Humira	10	111.9	9.98	99	128
ABP 501	10	108.3	8.88	96	121
EU-approved Humira	10	111.3	7.02	103	122

**Table 8 Equivalence testing results for the sTNF- $\alpha$  binding**

Comparison	# of lots	Mean difference, %	90% confidence interval for mean difference, %	Equivalence margin, %	Equivalent
ABP 501 vs. US	(10, 10)	-3.60	(-10.93, 3.73)	(-14.97, 14.97)	Yes
ABP 501 vs. EU	(10, 10)	-3.00	(-9.23, 3.23)	(-10.54, 10.54)	Yes
EU vs. US	(10, 10)	-0.60	(-7.34, 6.14)	(-14.97, 14.97)	Yes

## 6 CONCLUSION AND RECOMMENDATION

The results from the statistical equivalence testing of the apoptosis inhibition bioassay and the sTNF- $\alpha$  binding support a demonstration that the proposed biosimilar ABP 501 is highly similar to US-licensed Humira. The statistical analyses of the apoptosis inhibition bioassay and the sTNF- $\alpha$  binding in the three pair-wise comparisons (ABP 501, 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 ABP 501 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/  
-----

MEIYU SHEN  
08/17/2016

YI TSONG  
08/17/2016

## STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

**BLA Number:** 761024

**Applicant:** Amgen

**Stamp Date:** 11/24/2015

**Drug Name:** ABP 501

**BLA Type:** 351(k) Biosimilar

**Indication:** Psoriasis

I. On **initial** overview of the NDA/BLA application identify and list any potential Refuse to File issues:

	<b>Content Parameter for RTF</b>	<b>Yes</b>	<b>No</b>	<b>NA</b>	<b>Comments</b>
1	Indexing and reference links within the electronic submission are sufficient to permit navigation through the submission, including access to reports, tables, data, etc.	<b>X</b>			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	<b>X</b>			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated.	<b>X</b>			
4	Data sets in EDR are accessible and conform to applicable guidances (e.g., existence of define.pdf file for data sets).	<b>X</b>			

### IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE?

Yes.

II. Identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

<b>Content Parameter (possible review concerns for 74-day letter)</b>	<b>Yes</b>	<b>No</b>	<b>NA</b>	<b>Comment</b>
Designs utilized are appropriate for the indications requested.	<b>X</b>			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	<b>X</b>			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			<b>X</b>	
Appropriate references for novel statistical methodology (if present) are included.			<b>X</b>	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	<b>X</b>			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	<b>X</b>			

# STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

## 74-DAY LETTER REQUESTS TO THE APPLICANT

None.

### SUBMISSION SUMMARY

The applicant conducted a two-arm comparative clinical study of ABP 501 versus EU-approved adalimumab in subjects with psoriasis (Study 20120263). The study enrolled 350 subjects, of which 347 received at least one dose of investigational product, and 345 received at least one dose and had at least one post-baseline assessment. The study was conducted in Australia, Canada, Germany, France, Hungary, and Poland. The study enrolled subjects age 18-75 with stable moderate to severe plaque psoriasis for at least 6 months, BSA  $\geq 10\%$ , PASI  $\geq 12$ , and sPGA  $\geq 3$ .

The primary endpoint was percent change in PASI from baseline to Week 16. The 95% confidence interval for the difference in percent change in PASI was to be compared with a similarity margin of  $\pm 15\%$ . PASI 75 at Week 16 was the first listed secondary endpoint. The other secondary endpoints were PASI 75 (Weeks 32 and 50), PASI percent improvement (Weeks 32 and 50), sPGA response (clear/almost clear; Weeks 16, 32, and 50), and BSA (Weeks 15, 32, and 50). The protocol was not submitted to the FDA prior to the conduct of the trial.

### Efficacy Endpoints

	ABP 501 N=172	Adalimumab N=173
<b>Primary</b>		
% Improvement in PASI at Week 16	80.9%	83.1%
95% Confidence interval		(-7.4, 3.0)
90% Confidence interval		(-6.6, 2.2)
<b>Secondary</b>		
PASI 75 at Week 16	74.4%	82.7%
95% Confidence interval		(-16.6, 1.2)
90% Confidence interval		(-15.2, -0.3)

**ASSOCIATED IND:** 111714

Reviewing Statistician: Kathleen Fritsch, Ph.D.  
Mathematical Statistician, Biometrics III

Supervisor/Team Leader: Mohamed Alosh, Ph.D.  
Team Leader, Biometrics III

cc:

BLA 761024 / 0

DDDP/Marcus

DDDP/Diglisic

DDDP/Cook

DDDP/Attinello

DPARP/Nabavian

DPARP/Nikolov

DPARP/Hull

OBIO/Patrician

DBII/Levin

DBII/Kim

DBIII/Wilson

DBIII/Alosh

DBIII/Fritsch

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

/s/  
-----

KATHLEEN S FRITSCH  
01/11/2016

MOHAMED A ALOSH  
01/11/2016

## STATISTICAL FILING REVIEW FOR A NEW NDA/BLA

**NDA/BLA Number:** BLA761024

**NDA/BLA Type:** Standard

**Stamp Date:** 11/25/2015

**Applicant:** Amgen

**Drug Name:** ABP 501 (adalimumab biosimilar)

**Indication:** Treatment of rheumatoid arthritis (RA), juvenile idiopathic arthritis in patients 4 years of age and older, psoriatic arthritis, ankylosing spondylitis, adult Crohn's disease, ulcerative colitis, plaque psoriasis (Ps), (b) (4)

**Statistical Team:** Yongman Kim PhD & Gregory Levin PhD

**Clinical Team:** Keith Hull MD & Nikolay Nikolov MD

**Medical Division:** Division of Pulmonary, Allergy, and Rheumatology Products

**Project Manager:** Sadaf Nabavian

### Introduction:

This submission is for an original BLA of ABP 505, a product biosimilar to adalimumab.

The ABP 501 clinical development program included 2 comparative clinical studies, Study 20120262 with RA patients and Study 20120263 with Plaque Psoriasis patients. See the table below for details on the two comparative studies conducted in the clinical development program.

Study ID	No. of Study Centers	Study Start, Enrollment Status, Date	Design Control Type	Study and Control Drugs, Dose, Route and Regimen	Study Objective	No. Subjects by Arm Entered/Complete	Duration of Study <sup>a</sup>	No. M/F Median Age (Range)	Diagnosis Key Inclusion Criteria	Primary Endpoint(s)
Study Reports of Controlled Clinical Studies Pertinent to Claimed Indication in Module 5.3.5.1										
<a href="#">Study 20120262</a>	92 sites Eastern Europe (Poland, Czech Republic, Hungary, Bulgaria, Romania, Russian Federation), Western Europe (Germany, Spain, United Kingdom), North America (United States, Canada), Latin America (Mexico)	October 2013, Complete Enrollment/ Goal November 2014 526 / 500	Randomized, double-blind, active comparator-controlled	ABP 501 vs adalimumab (US), 40 mg SC, Q2W	Efficacy, safety, immunogenicity	ABP 501: 264 entered/ 243 complete Adalimumab: 262 entered/ 251 complete	26 wks	100/426 57.0 yrs (21 to 80)	Men and women ≥ 18 to ≤ 80 yrs of age  Moderate to severe RA for ≥ 3 mos ≥ 6 swollen joints and ≥ 6 tender joints ESR ≥ 28 mm/hr or CRP > 1.0 mg/dL Received MTX ≥ 12 wks and on stable dose ≥ 8 wks	Risk ratio of ACR20 at wk 24

Study ID	No. of Study Centers Locations	Study Start, Enrollment Status, Date Total Enrollment/ Enrollment Goal	Design Control Type	Study and Control Drugs, Dose, Route and Regimen	Study Objective	No. Subjects by Arm Entered/ Complete	Duration of Study <sup>a</sup>	No. M/F Median Age (Range)	Diagnosis Key Inclusion Criteria	Primary Endpoint(s)
Study Reports of Controlled Clinical Studies Pertinent to Claimed Indication in Module 5.3.5.1 (continued)										
<a href="#">Study 20120263</a>	49 sites Eastern Europe (Poland, Hungary) Western Europe (Germany, France), Other (Australia, Canada)	October 2013, Complete March 2015 350 / 340	Randomized, double-blind, active comparator-controlled Subjects qualifying for re-randomization at wk 16: Group A continued treatment with ABP 501; Group B re-randomized to Group B1 or Group B2	ABP 501 vs adalimumab (EU), 80 mg SC, wk 1/day 1, then 40 mg SC Q2W beginning at wk 2	Efficacy, safety, immunogenicity	Group A: ABP 501: 175/164 (wk 16); 152/135 (wk 52) Group B: Adalimumab: 175/162 (wk 16) Group B1: Adalimumab/ ABP 501: 79/71 (wk 52) Group B2: Adalimumab/ ABP 501: 77/69 (wk 52)	52 wks	228/122 43.0 yrs (18 to 74)	Men and women ≥ 18 to ≤ 75 yrs of age Moderate to severe Ps for ≥ 6 mos BSA ≥ 10% involved PASI ≥ 12 sPGA ≥ 3 Subjects achieving ≥ PASI 50 response at wk 16 qualified for re-randomization	PASI % improvement from baseline at wk 16

I will review the RA study and Dr. Kathleen Fritsch will review the Ps study. The key measures of efficacy assessing clinical response in the ABP 501 clinical development program for RA are American College of Rheumatology response criteria, Disease Activity Score for 28 joints, and, physical function as measured by HAQ-DI.

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

- Type B meeting (8/24/2011):
  - FDA recommended that Amgen use a more sensitive endpoint (ie, continuous variable) such as Hybrid ACR, DAS28, or ACRn-N for the pivotal study. FDA recommended the use of a 2-sided comparative efficacy analysis for the pivotal study.
  - At a subsequent meeting held prior to the initiation of the pivotal study in Rheumatoid Arthritis (20120262), Amgen and the Agency agreed with the use of risk ratio of ACR20 at week 24 as a primary endpoint (09 May 2013, Ref ID 3330346). Additionally, DAS28-CRP was incorporated as a secondary endpoint and the data are provided in the CSR. Comparative efficacy was evaluated by a 2-sided equivalence approach.
- BPD Type 2 meeting (5/9/2013):
  - FDA recommended that if Amgen proceeds with an equivalence trial design as proposed, Amgen should either utilize an endpoint such as ACR20 for which there are data available to justify an equivalence margin or provide a scientific justification for the proposed equivalence margin for DAS28. FDA recommended that Amgen evaluate several different time points early in treatment, e.g., weeks 1, 2, 3, 4, 6, 8, etc, as secondary endpoints.
  - Risk Ratio of ACR20 at week 24 was incorporated as the primary endpoint in Study 20120262. Amgen confirms that additional efficacy assessments at early time points as

recommended by the FDA were added to the study protocol 20120262 as secondary endpoints. Assessments were conducted at weeks 2, 4, 8, 12, 18, and 24. Amgen submitted the revised pivotal clinical study Protocol 20120262 (version 2, dated 06 June 2013) along with statistical simulation results to support the use of Risk Ratio of ACR20 as the primary endpoint to IND 111714 on 27 June 2013 (SN 0010).

- BPD Type 2 meeting (1/26/2015):
  - FDA stated that the use of last observation carried forward (LOCF) to impute missing ACR20 data at Week 24 is not acceptable. LOCF relies on the strong and unverifiable assumption that patient outcomes prior to withdrawal would have remained constant through Week 24. In addition, as a single-imputation approach, LOCF does not appropriately take into account the uncertainty in the imputation process. During the Type 2 meeting FDA acknowledged that RA study 20120262 enrollment is complete, the database was locked on 21 January 2015 and the study has been unblinded; hence it was impracticable to make changes to the protocol or SAP at the time of the meeting.
  - The non-responder imputation analysis was pre-specified in the study 20120262 Statistical Analysis Plan (SAP) as a sensitivity analysis. This analysis has been performed and the study disposition summary is included in the Study 20120262 CSR.
  - FDA requested Amgen to provide data from historical randomized clinical trials of adalimumab to justify the adequacy of the proposed similarity margins of (0.738, 1/0.738) for the ratio of ACR20 responses.
  - The rationale for the equivalence margin was based on considerations in the draft US FDA Non-inferiority Clinical Trials Guidance For Industry (2010). The equivalence margin of (0.738, 1/0.738) for the RR of ACR20 responses was chosen based on a published relevant adequate and well-controlled trial (Keystone et al, 2004), and was expected to preserve 50% of the estimated 80% upper confidence bound of the treatment effect of the reference product compared with placebo.
  - Post-Meeting Addendum: Similarity Margin Recommendation FDA recommends that the similarity margin for the proposed comparative clinical study (CCS) in rheumatoid arthritis be no greater in magnitude than  $\pm 12\%$ . The proposed margin of  $\pm 12\%$  is based on considerations aimed at weighing the clinical importance of various differences in effect against the feasibility of different study sizes. FDA also recommend that a margin based on the absolute difference scale be used, as it is considered more important than other metrics, such as risk ratio, from a clinical perspective for an evaluation of benefit-risk.
  - Both the 90% CI and 95% CI for risk difference of ACR20 at week 24 were within the FDA recommended margin of  $\pm 12\%$ . This margin was not pre-specified as the Agency's recommendation was received after the database lock (FDA Type 2 Meeting Minutes, reference ID 3716075). However, this evaluation further confirms clinical equivalence between ABP 501 and adalimumab.

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

**Filing Checklist:**

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

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

**IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE?** \_\_\_Yes\_\_\_

**Potential Review Issues:**

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

**Additional Discussion:**

The following will be major potential focus areas in my statistical review:

- Confirmation of key analyses
- Handling of missing data
- Labeling claims [REDACTED] (b) (4) if included

**Comments for Applicant:**

1. *You have not provided sensitivity analyses that sufficiently evaluate the potential impact of missing data on the reliability of efficacy results in Study 20120262. For the primary endpoint, please examine the potential effects of missing data on your results using tipping point sensitivity analyses. These tipping point analyses should include all observed data, including outcomes after patients discontinue study therapy and should vary assumptions about outcomes among the subsets of patients on the ABP 501 and adalimumab arms who withdrew from the study prior to the planned endpoint. The varying assumptions should include scenarios where dropouts on ABP501 had different future outcomes than dropouts on adalimumab. The goal is to identify assumptions under which the conclusions change, i.e., under which there is no longer evidence of similarity. Then, the plausibility of those assumptions can be discussed.*

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

/s/  
-----

YONGMAN KIM  
01/11/2016

GREGORY P LEVIN  
01/11/2016