



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 ACR<sub>n</sub> 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)

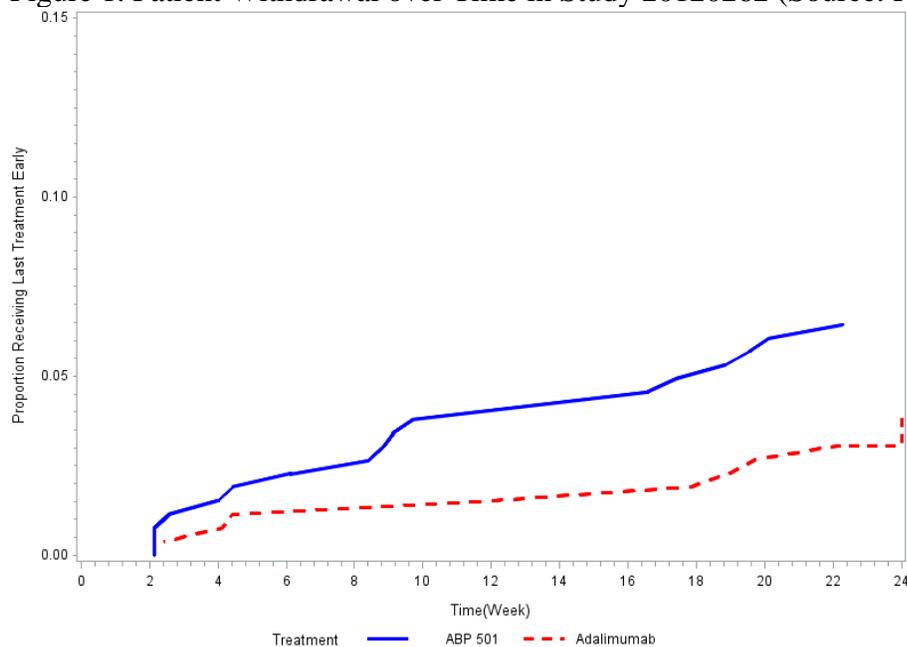


Table 4. Patient Dropout, by Reason for Withdrawal, in Study 20120262

	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

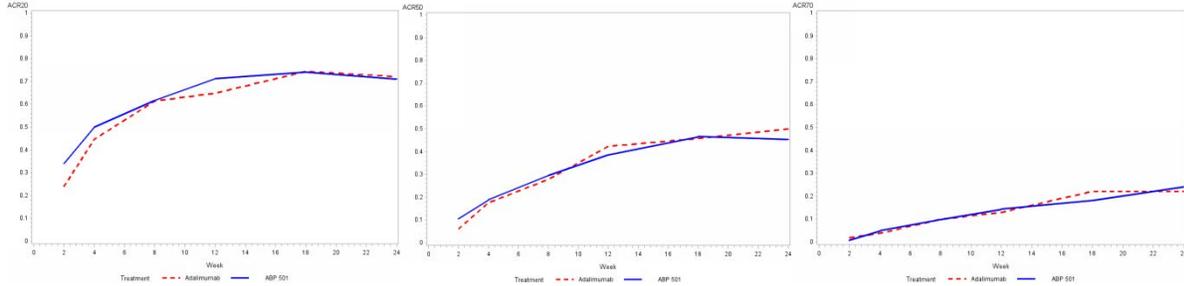
Abbreviations: CI = confidence interval

<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.

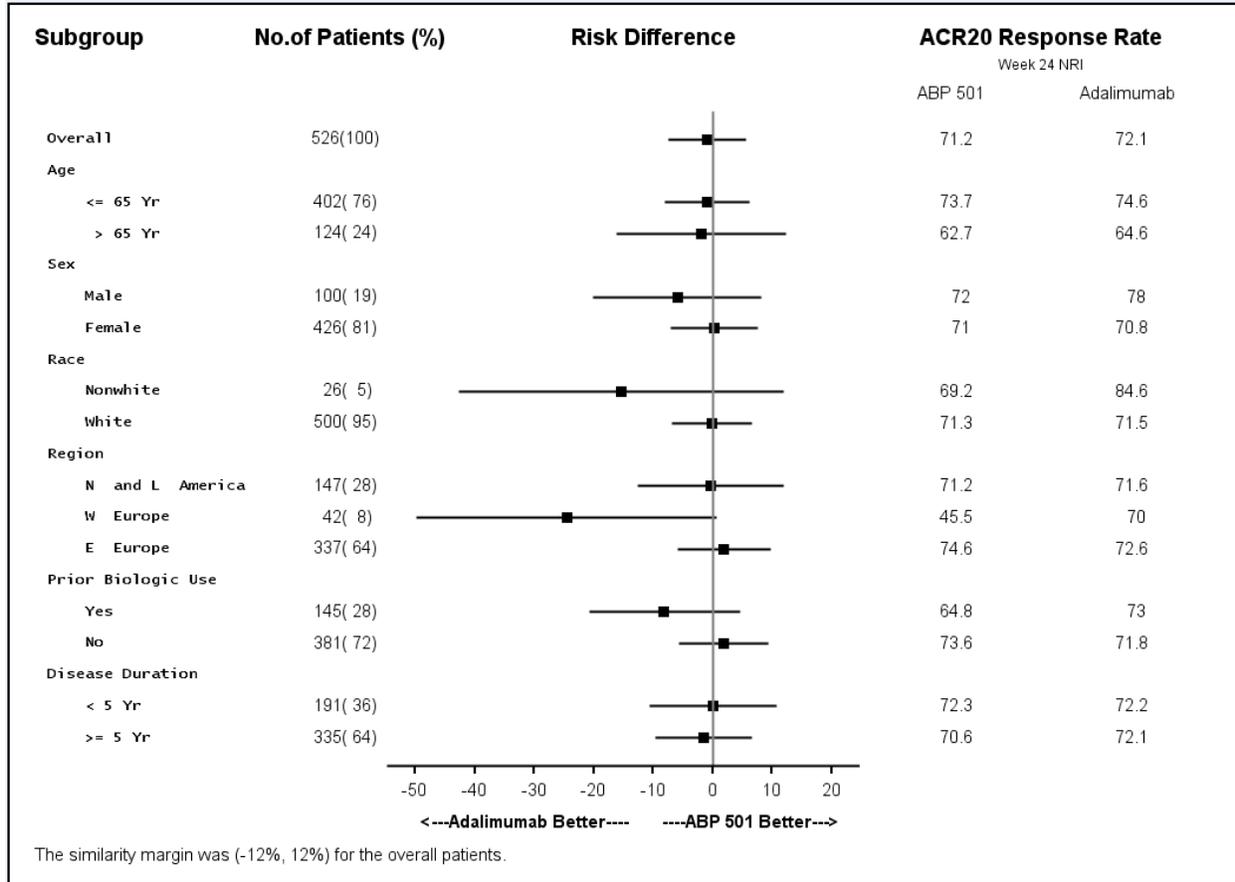
### **3.5 Evaluation of Safety**

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)

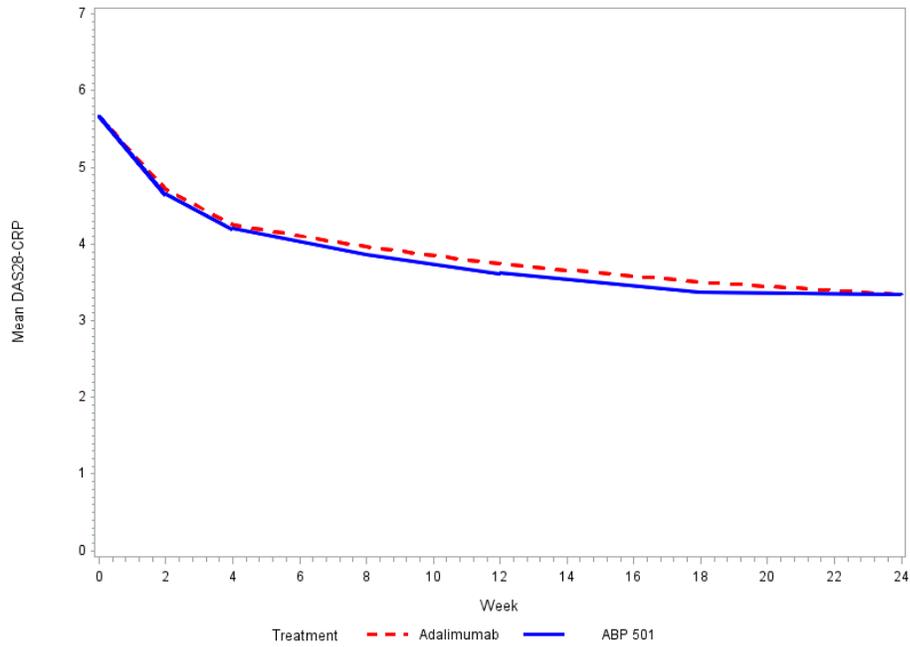


Figure 5. Mean Swollen Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

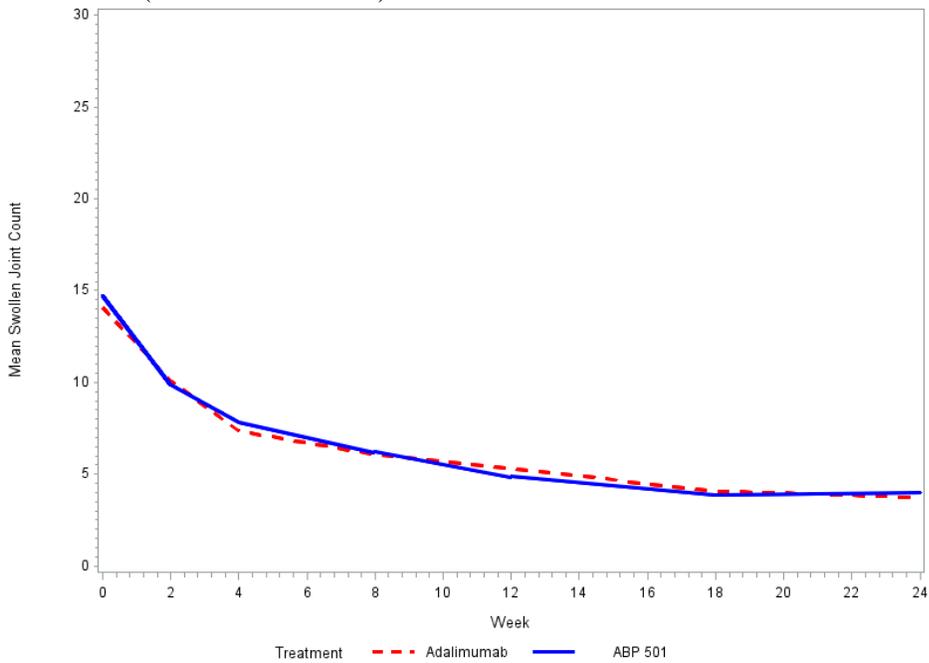


Figure 6. Mean Tender Joint Count among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

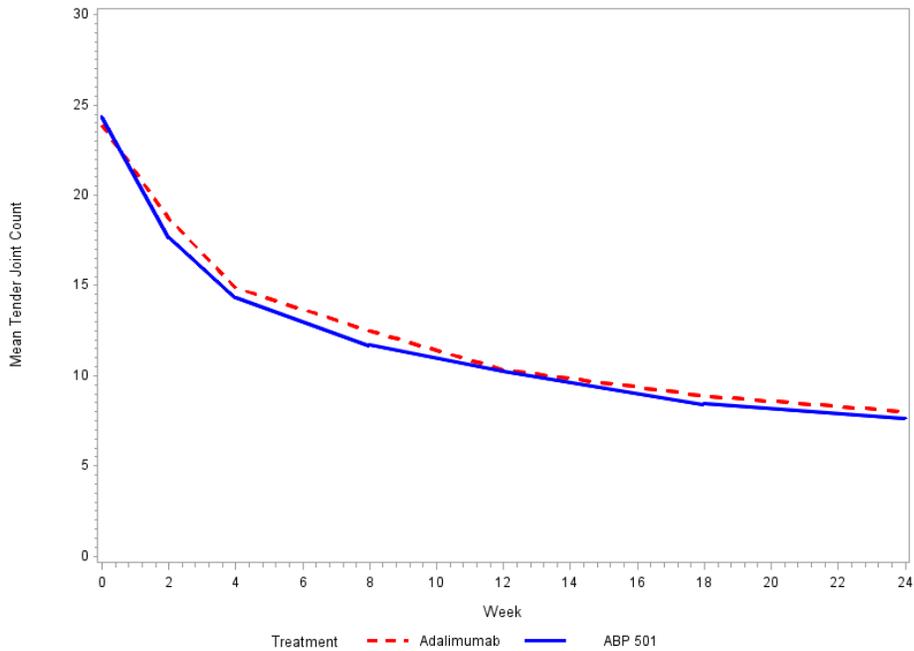


Figure 7. Mean Health Assessment Questionnaire (HAQ) Physical Ability Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

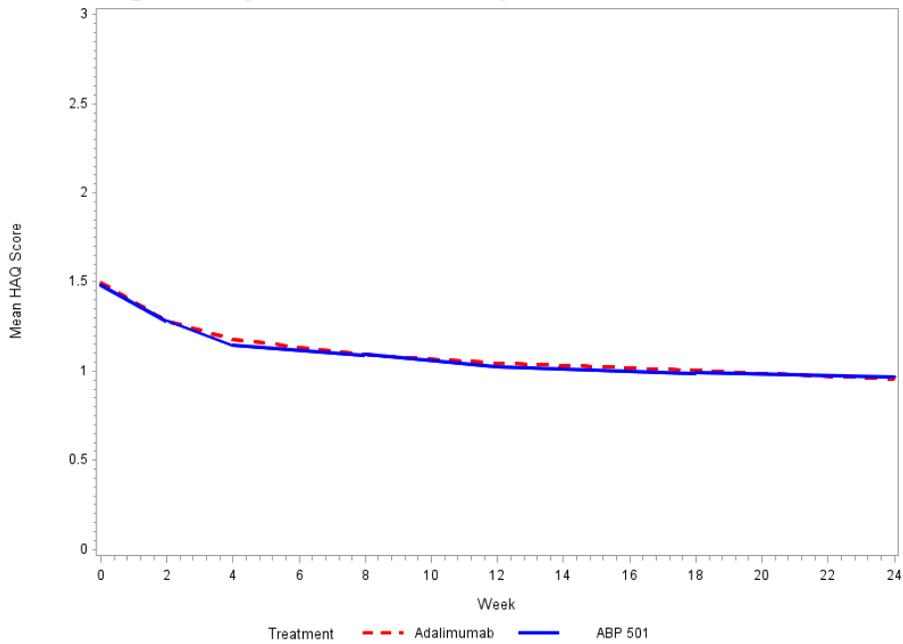


Figure 8. Mean Patient Pain Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

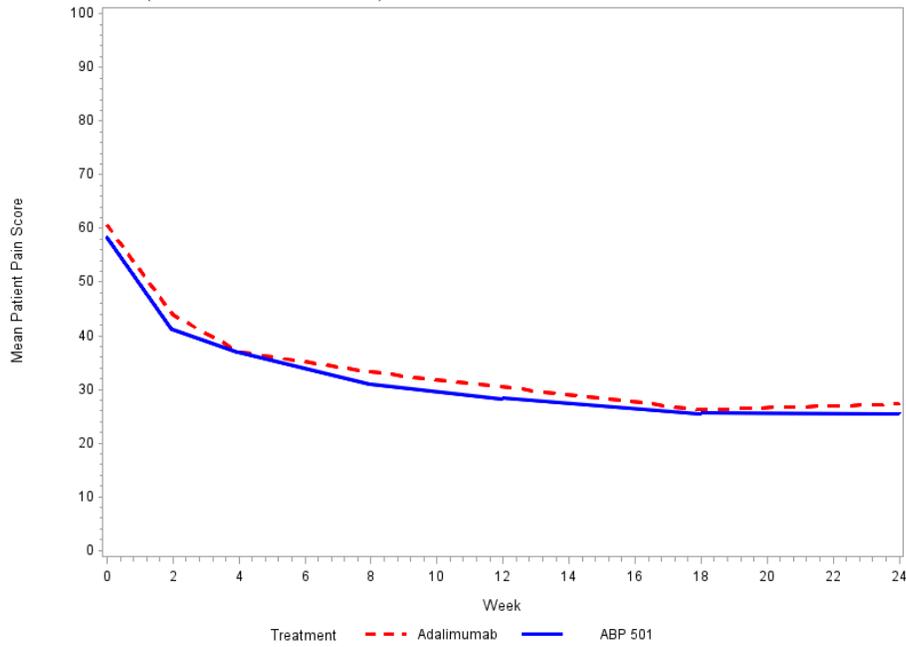


Figure 9. Mean Patient Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

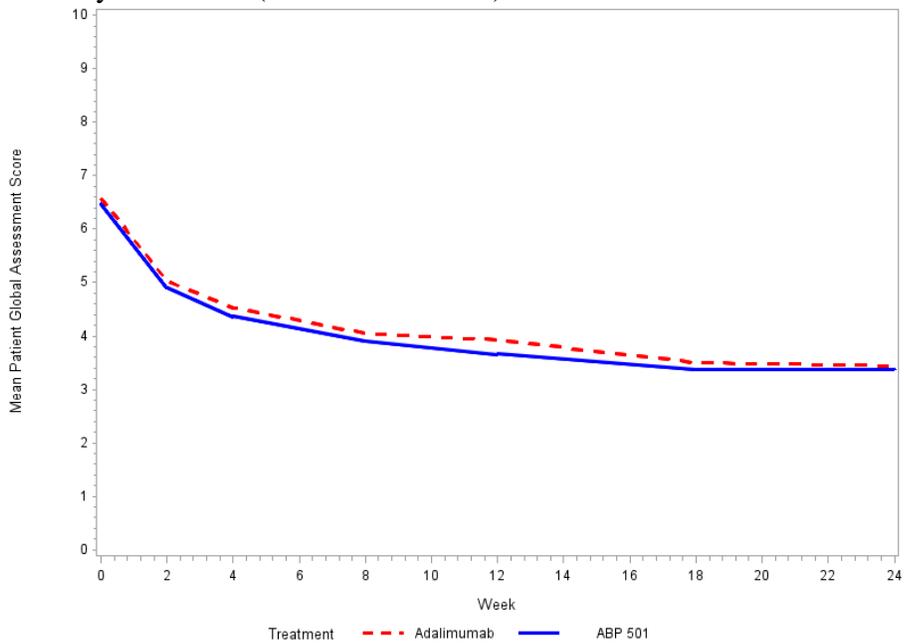


Figure 10. Mean Physician Global Assessment Score among Patients Remaining in Study over Time in Study 20120262 (Source: Reviewer)

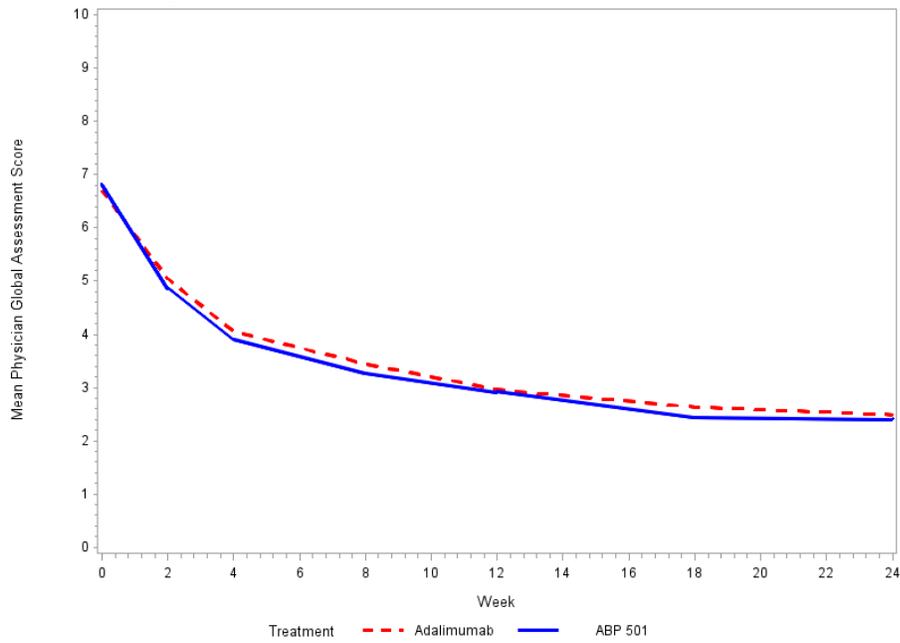
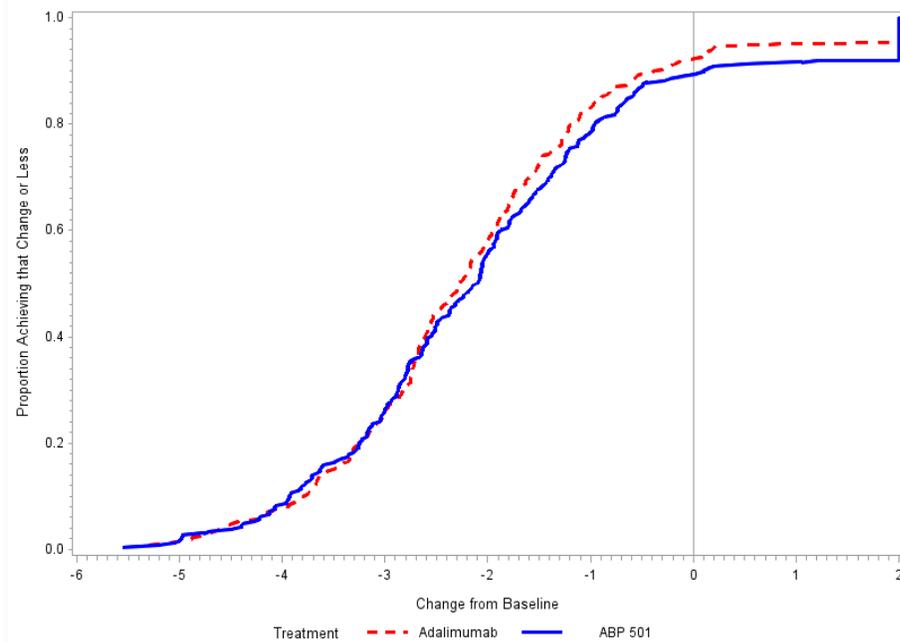


Figure 11. Empirical Distribution Function for Change from Baseline in Disease Activity Score (DAS28-CRP) at Week 24 in Study 20120262 (Source: Reviewer)



## 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