

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

761072Orig1s000

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

BLA / Seq #: BLA 761072 / 0001

Drug Name: PF-06438179 (proposed infliximab biosimilar)

Indication(s): All infliximab indications

Applicant: Pfizer

Date(s): Received: February 14, 2017
BSUFA: December 13, 2017

Review Priority: Standard

Biometrics Division: Division of Biometrics II

Statistical Reviewer: William Koh, PhD, Statistical Reviewer

Concurring Reviewer: Robert Abugov, PhD, Team Leader

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

Clinical Team: Erika Torjusen, MD, Clinical Reviewer
Banu, Karimi-Shah, MD, Clinical Team Leader
Nikolay, Nikolov, MD, Clinical Team Leader

Project Manager: Christine Ford, RPM

Keywords: biosimilar, BSUFA, missing data

Table of Contents

1	EXECUTIVE SUMMARY	5
2	INTRODUCTION	6
2.1	OVERVIEW	6
2.2	HISTORY OF PRODUCT DEVELOPMENT	7
2.3	SPECIFIC STUDIES REVIEWED	8
2.4	DATA SOURCES	8
3	STATISTICAL EVALUATION	8
3.1	DATA AND ANALYSIS QUALITY	8
3.2	EVALUATION OF EFFICACY	9
3.2.1	<i>Study Design and Endpoints</i>	9
3.2.2	<i>Statistical Methodologies</i>	11
3.2.3	<i>Patient Disposition, Study Conduct, Demographic and Baseline Characteristics</i>	15
3.2.4	<i>Results and Conclusions</i>	20
3.3	EVALUATION OF SAFETY	32
3.3.1	<i>Drug Exposure</i>	32
3.3.2	<i>Deaths</i>	32
3.3.3	<i>Adverse Events</i>	32
3.3.4	<i>Adverse Events of Special Interest Determined by Applicant</i>	34
3.4	ASSAY SENSITIVITY AND CONSTANCY ASSUMPTION	34
4	FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	37
4.1	GENDER, RACE, AGE, GEOGRAPHIC REGION, AND KEY BASELINE RA CHARACTERISTICS	37
5	SUMMARY AND CONCLUSIONS	38
5.1	STATISTICAL ISSUES	38
5.2	COLLECTIVE EVIDENCE	39
6	APPENDICES	40
6.1	ADDITIONAL TABLES AND FIGURES	40
6.2	REFERENCES	41
6.3	TIPPING POINT ANALYSIS METHODOLOGY	42

LIST OF TABLES

Table 1 Summary of the Five Historical Randomized Clinical Studies of Infliximab in Patients with Active RA with Inadequate Response to MTX.....	14
Table 2 Disposition of Patients in Study B5371002.....	16
Table 3 Summary of Reasons for Discontinuation from Treatment but were Followed-up in Study B5371002	17
Table 4 Summary of Reasons for Discontinuation from Study in Study B5371002.....	17
Table 5 Patients Excluded from the Per-Protocol Population (up to Week 14) and Reasons	18
Table 6 Baseline Characteristics for all Randomized Patients ^a in Study B5371002.....	19
Table 7 Baseline Disease Characteristics for all Randomized Patients ^a in Study B5371002.....	20
Table 8 Differences in ACR20 Responses for Various Analysis Datasets Based on Exact Binomial Score Test Conducted at Primary Week of Analysis.....	21
Table 9 Change from Baseline for the Individual Components of ACR at Scheduled Visit Weeks Based on all Data Collected Regardless of Study Treatment Discontinuation by Treatment Arm	24
Table 10 Change from Baseline for the Individual Components of ACR at Scheduled Visit Weeks Based on all Data Collected Regardless of Study Treatment Discontinuation by Treatment Arm (continued from Table 9).....	25
Table 11 Regression Analysis of the Change from Baseline in DAS28(CRP) Based on all Patients Regardless of Adherence to Study Treatment and PP Dataset	27
Table 12 Number of Patients who Dose Escalated During the Study after Week 14.....	28
Table 13 Post-hoc Analyses of ACR20 Responses Regardless of Adherence to Study Treatment Based on Whether Patients Dose Escalated at Weeks 14 and 22.....	28
Table 14 Exposure to Study Drug during Treatment Period 1	32
Table 15 Summary of Adverse Events in Treatment Period 1 through 30 Weeks	33
Table 16 Adverse Events by MedDRA System Organ Class and Preferred Term during Treatment Period 1	33
Table 17 Summary of Adverse Events of Special Interest Flagged by the Applicant through TP1.....	34
Table 18 Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials ^a of Infliximab in RA and Comparative Clinical Study B5371002.....	36
Table 19 Number of Patients in each Specific Dataset for Analysis of ACR20 at Week 14	40
Table 20 Probability of ACR20 Response at Other Study Visit Weeks.....	40
Table 21 Listings of Patients who Died during Treatment Period 1 (+ 28 Days)	41
Table 22 Parameters and Notation for Tipping Point Analysis in Presence of Missing Data	42

LIST OF FIGURES

Figure 1 Study Design Schematic for B5371002	10
Figure 2 ACR Responder Curve, Regardless of Adherence to Study Treatment for Various Thresholds at Week 14 Comparing Infliximab EU vs PF-06438179	22
Figure 3 Differences in ACR Response Comparing PF-06438179 vs Infliximab-EU Using all Possible ACR Thresholds Across Different Study Visits up to Week 30, Regardless of Adherence to Study Treatment.....	23
Figure 4 Cumulative Distribution Curves for Patients with Observed DAS28(CRP) at Week 14.....	26
Figure 5 Tipping Point Analyses Results for 90% Asymptotic Lower CL (Heatmap) and Point Estimates (Contour Lines): Inference on the Difference Between PF06438179 and Infliximab-EU in the Probability of Week 14 ACR20 Responses under Varying Assumptions About the Differences on Each Treatment Arm	30
Figure 6 Tipping Point Analyses Results for 90% Asymptotic Upper CL (Heatmap) and Point Estimates (Contour Lines): Inference on the Difference Between PF06438179 and Infliximab-EU in the Probability of Week 14 ACR20 Responses under Varying Assumptions About the Differences on Each Treatment Arm	31
Figure 7 Estimated Differences Between Infliximab-EU and PF-06438179 in the Probability of ACR20 Response at Week 14 and on Randomized Treatment, Stratified by Selected Demographic Subgroups and Key RA Baseline Characteristics.....	37

1 EXECUTIVE SUMMARY

This review evaluates PF-06438179 as a potential biosimilar to EU-licensed infliximab (hereafter referred to as infliximab-EU) for the treatment of rheumatoid arthritis (RA) and other indications. The applicant, Pfizer, conducted a 54-week comparative clinical study (B5371002; NCT02222493) to evaluate the efficacy, safety, pharmacokinetics, and immunogenicity of PF-06438179 compared to infliximab-EU in 649 patients with active RA who had an inadequate response to methotrexate.

In study B5371002, the primary efficacy endpoint was the proportion of patients who achieved an American College of Rheumatology (ACR) 20 response (ACR20) at week 14. At week 14, 63.8% of the patients randomized to infliximab-EU and 61.1% of the patients randomized to PF-06438179 were ACR20 responders who stayed on randomized treatment; the estimated difference on the absolute scale comparing PF-06438179 relative to infliximab-EU was -2.7% (90% exact confidence interval (CI): -9.1%, 3.6%). The limits of the 90% CI were within the prespecified asymmetric similarity margins of -12% and 15% agreed upon with the Agency. At other time points, and for other ACR thresholds, individual components of change from baseline for ACR (swollen joint count out of 66 (SJC66), tender joint count out of 68 (TJC68), patient global assessment, physician global assessment, health assessment questionnaire-disability index (HAQ-DI), patient pain, and C-reactive protein (CRP)) and disease activity score-28 with four components based on CRP (DAS28(CRP)) were similar between the treatment arms.

The finding of similar efficacy was maintained in the face of a broad range of plausible alternative assumptions regarding outcomes among patients with missing data. At week 14, 22 (4%) patients had withdrawn from the study: 12 patients (4%) from the PF-06438179 arm and 10 patients (3%) from the infliximab-EU arm. Tipping point analyses conducted to explore the sensitivity of results to violations in assumptions about the missing data showed that the confidence intervals for differences between PF-06438179 and infliximab-EU arm were robust, ruling out differences of concern in efficacy even under the assumption that outcomes among patients who dropped out on PF-06438179 were much worse or much better than outcomes among patients who dropped out on infliximab-EU. Given the similar proportions of early withdrawal patients, similar distributions of reasons for early withdrawal, similar baseline characteristics between dropouts on the two treatment arms, and implausible differences between the outcomes among dropouts needed to tip results toward non-similarity, the finding of similar efficacy is highly credible.

To reliably determine whether there is biosimilarity between two products, there must be historical constancy for calculation of appropriate similarity margins, i.e., the clinical trials from which the similarity margins were calculated must have been of similar design and address similar patient populations as the proposed comparative clinical study (CCS), and the CCS, as executed, must have assay sensitivity, or the ability to detect meaningful differences between the products if such differences exist. Based on an evaluation of the CCS and the five historical, randomized, placebo-controlled clinical trials used to calculate the similarity margins, we concluded that: (1) there were relatively large and consistent treatment effects across the five historical studies, implying potential assay sensitivity in a properly conducted CCS; and (2) the design and realized outcomes for the CCS (components of ACR20, HAQ-DI, DAS28(CRP), dropout rates for the reference product) were largely similar to those of the historical trials from

which the non-inferiority margin was calculated, implying historical constancy and assay sensitivity. In addition, I did not identify issues of concern with study conduct, other than protocol deviations and minimal dropouts, that could have affected the quality of study and bias the results towards similarity. Therefore, the planning and assay sensitivity of this study were sufficient to allow a credible assessment of similarity between PF-06438179 and infliximab-EU.

2 INTRODUCTION

2.1 Overview

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

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

Rheumatoid arthritis (RA) is an autoimmune disease characterized by inflammation in the synovium of joints, malaise, morning stiffness, and fatigue.

Infliximab is a monoclonal antibody that inhibits the activity of TNF- α . FDA has approved infliximab for the treatment of RA, Crohn's disease, ulcerative colitis, psoriasis, psoriatic arthritis, and ankylosing spondylitis. Currently, there are two other approved biosimilars for infliximab, namely, infliximab-dyyb (Celltrion) and infliximab-abda (Samsung Bioepis).

The applicant has submitted a BLA for PF-06438179, a proposed biosimilar to infliximab produced for the European market, EU-Remicade. The applicant provided reports on direct physicochemical and biological comparisons between PF-06438179, EU-Remicade, and US-Remicade. The similarity between EU- and US-Remicade was demonstrated using analytical and PK bridging studies.

The applicant has submitted results from clinical study B5371002 conducted in adult patients with rheumatoid arthritis to provide supportive claim of no clinically meaningful differences between PF-06438179 and US-licensed infliximab. The proposed indications for PF-06438179 sought by Pfizer are identical to those of the reference product. This review primarily considers the efficacy and safety evaluation of PF-06438179 in clinical studies.

2.2 History of Product Development

The clinical development program for PF-06438179 was introduced to the Division of Pulmonary, Allergy, and Rheumatology Products under IND 114,828. Interactions with the applicant during product development potentially relevant to this review are summarized below.

At the BPD Type 2 meeting on December 18, 2013, FDA expressed concerns regarding the applicant's proposed similarity margin and the applicant's proposal to (b) (4). The applicant had not provided adequate justification of the choice of historical studies, choice of statistical metric (e.g. log relative risk), and choice of confidence limits ((b) (4)) for the estimated effect size. The Agency expressed further concerns regarding significant heterogeneity in the estimated treatment effect from the applicant's meta-analysis and that none of the study-specific estimates were contained within the confidence intervals from both fixed and random effects meta-analysis models. The Agency recommended that adequate justification be included in determination of the proposed equivalence margin. The applicant acknowledged the Agency's concerns (b) (4) and agreed to revise the study protocol to continue follow-up of all patients who discontinued randomized treatment.

At the BPD Type 3 meeting in 2014, the applicant proposed to use absolute differences in ACR20 response rate as the primary endpoint, and included justification for the proposed equivalence margin using that proposed metric. The Agency agreed with the revised metric, and recommended calculation of the non-inferiority margin using a meta-analysis of five historical studies, using a two-sided 95% confidence interval used to estimate uncertainty in the effect of infliximab over placebo, and with 50% preservation of the lower bound of the 95% confidence interval. The Agency recommended that the applicant submit a more comprehensive approach to address missing data consistent with 2010 National Research Council report *The Prevention and Treatment of Missing Data in Clinical Trials*.

The applicant followed-up with a revised proposal for the similarity margin (b) (4) for the proposed clinical study B5371002 based on an information request in 2014 following the Type 3 meeting. The Agency disagreed with the applicant's margin and recommended a $\pm 12\%$ margin based on a balance between study feasibility and power to detect clinically important differences. The Agency reiterated that margins based on the absolute difference scale were preferable to risk-ratios from a clinical perspective because differences are commonly used and well-understood. The Agency also stated a willingness to consider asymmetric biosimilarity margins if the applicant adequately justified them. The Agency noted that additional factors would remain critical to the interpretation of the clinical data, e.g., the ability of the comparative clinical study to rule out clinically meaningful differences, its assay sensitivity, would critically depend on study conduct.

In a subsequent response from the applicant in 2015, the applicant revised the proposed similarity margins, with a lower limit of -12% (ruling out no loss of more than 12% efficacy on the absolute difference scale) and an upper limit of +15% (ruling out no more than 15% better than the reference product on the absolute difference scale). The Agency agreed with the justification of their proposed asymmetric margin.

In the BPD Type 2 meeting in 2016, the Agency disagreed with proposed sensitivity analyses ((b) (4)) to explore the impact of missing data. The Agency instead recommended inclusion of tipping point sensitivity analyses that vary assumptions about the missing outcomes on the two treatment arms. Such analyses would include all observed data (including data collected post treatment discontinuation) and include scenarios where dropouts on PF-06438179 had dissimilar outcomes from dropouts on EU-approved infliximab. During the meeting, the applicant agreed to include supportive efficacy analyses based on data from all randomized patients regardless of adherence to treatment, as well as the tipping point analyses for the primary endpoint.

There were no additional statistical questions raised for study B5371002 during the BPD Type 4 meeting in 2016.

2.3 Specific Studies Reviewed

The applicant submitted results from a comparative clinical trial, B5371002 (NCT02222493), a multicenter, randomized, double-blind, parallel-group study. There were two stages in this study: treatment period 1 up to week 30 and treatment period 2 from week 30 through week 54. The review evaluates the data submitted for treatment period 1.

2.4 Data Sources

Data were submitted by the applicant to CDER in SAS transport format. Protocols, correspondence, data listings, program code, and study reports were accessed under the network path \\cdsesub1\evsprod\BLA761072\761072.enx.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Datasets were submitted in legacy format, i.e., according to the applicant's proprietary format. Additional steps were required to derive specific baseline variables such as prior concomitant medications. In general, the quality of data submitted for efficacy analyses was adequate. Data from treatment period 2 were not submitted for review.

During FDA evaluation of this submission, the applicant noted a data entry error for subject (b) (6) at week 14 (Seq 0022; October 26, 2017). In particular, the joint tenderness status of the third metacarpophalangeal joint of the right hand was incorrectly recorded as present and the joint swelling status was incorrectly recorded as absent. Thus, the patient's ACR20 response was revised from non-responder to responder. I submitted an information request to clarify the process that led to discovery of this error. In response to my information request, the applicant included additional revisions identified by the investigators and provided a summary of the processes that led to these discoveries. After review of the applicant's response (Seq 0025; November 07, 2017), I found that the data revisions were appropriately conducted and did not impact the overall study results. In this review, results reviewed/reported were based on the

original data submission and only included the data error revision for subject (b) (6) at week 14. I address discoveries of additional data entry revisions using sensitivity analyses and provide a summary of the findings in Section 5.1 below.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

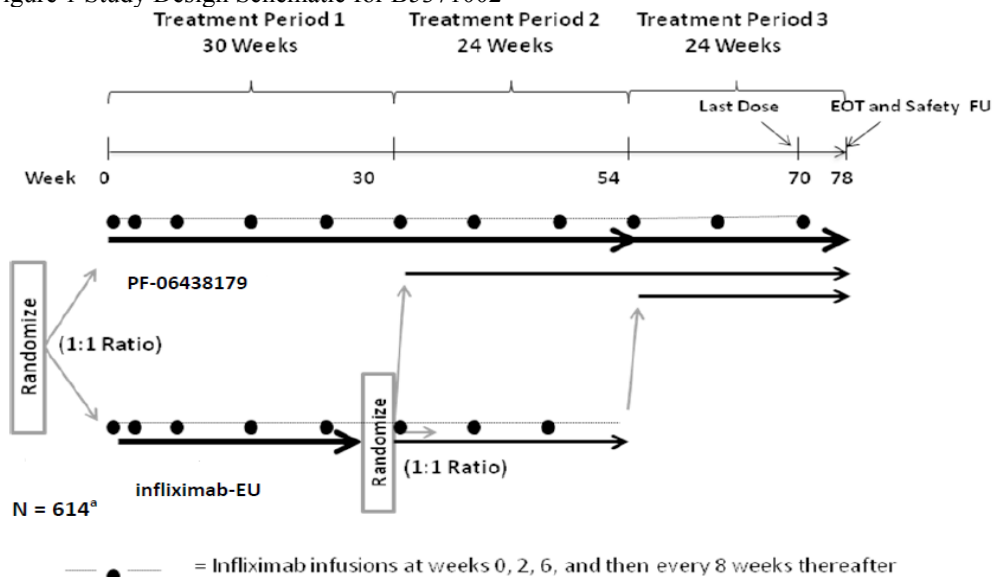
Study B5371002 was a multicenter, randomized, double-blind, parallel-group study designed to determine whether the efficacy and safety of PF-06438179 was similar to infliximab-EU in patients with moderately to severely active rheumatoid arthritis who had an inadequate response to methotrexate therapy.

The study design consisted of three treatment periods (Figure 1). During treatment period 1 (TP1), 649 patients were randomized in a 1:1 ratio at week 0 to infliximab-EU or PF-06438179. TP1 ended with the completion of the week 30 assessment. Treatment period 2 (TP2) started from week 30 and ended on week 53. During TP2, patients initially randomized to infliximab-EU were re-randomized in a 1:1 ratio to infliximab-EU or PF-06438179. Patients initially randomized to PF-06438179 remained on their assigned treatment arm with blinding until completion of the week 54 assessment. During treatment period 3 (TP3), patients on infliximab-EU crossed over to PF-06438179, and stayed on that treatment for an additional 24 weeks until visit week 70. The end-of-study visit was on visit week 78, eight weeks after last dose at visit week 70.

For purposes of enrollment, patients with active RA had at least 6 tender joints (out of 68 assessed) at both screening and baseline, at least 6 swollen joints (out of 66 assessed) at both screening and baseline, and CRP at least 10 mg/L at screening. Those not meeting this entry criteria could be re-tested once within 14 days and randomized if their CRP was at least 10 mg/L. Enrolled patients were also on stable background methotrexate (MTX) between 10 to 25 mg/week, received MTX for at least 12 weeks, and were on a stable MTX dose for at least 4 weeks prior to the first dose of study drug. Other allowable background medications included stable low doses of oral corticosteroids (for at least 4 weeks), a single non-steroidal anti-inflammatory drug (NSAIDs) or Cyclo-oxygenase-2 (Cox-2) inhibitor, with aspirin dose limited to ≤ 325 mg/day, or a stable dose (for at least 2 weeks) of an allowable opioid.

Dosages for TP1 were based on screening body weight, with intravenous infusions of 3 mg/kg administered on weeks 0, 2, and 6. During the maintenance period, from week 14 onward, a one-time dose escalation from 3 mg/kg to 5 mg/kg was made for patients who did not achieve minimum clinical response at week 14, i.e., $\geq 20\%$ improvement from baseline in both SJC66 and TJC68. Patients who had achieved 20% or greater response in either swollen or tender joint counts at week 14 but did not achieve minimal clinical response at week 22 were given a 5 mg/kg of study treatment at week 22. All other patients, who had achieved minimum clinical response at both weeks 14 and 22, remained on the 3 mg/kg dose.

Figure 1 Study Design Schematic for B5371002



[Source: Clinical study report, pp 33]

The originally planned stratified randomization was by current use of sulfasalazine (SSZ: yes/no) and geographic region (North America vs Western Europe/rest of the world). However, because there was a limited number of patients who were using SSZ, the protocol was amended on September 19, 2014 (thereafter known as version 1) to stratify only by geographic region (North America and Western Europe vs rest of the world).

Clinical and laboratory measures, including joint counts, patient's and physician's global assessments of arthritis using a visual analog scale (VAS), patient's assessment of arthritis pain using a VAS, HAQ-DI, and laboratory tests such as CRP, were assessed at weeks 0, 2, 4, 6, 12, 14, 22, 30, 38, 46, 54, 62, and 70. Clinical visit windows were ± 2 days for visit week 2 to visit week 14 and ± 14 days from week 22 onward. For purposes of efficacy evaluation, visit windows for weeks 2 through 14 were ± 7 days and, from week 22 through week 70, were ± 28 days.

Patients could discontinue from study treatment or study at their own request. The original protocol did not state whether patients who discontinued dosing prior to week 30 and had not withdrawn from the study should return for follow-up assessments. Following recommendations from the Agency, this ambiguity was clarified from protocol version 1 onward to specifically state that these patients would return for all scheduled study visits up to week 30. In addition, the protocol distinguished between withdrawal from treatment and withdrawal from study participation.

Of note, the same assessor was used to assess joints at screening, baseline, and week 14. Artificial joints were not assessed. Intra-articular joint injections with corticosteroids or hyaluronate were allowed. Joints were assessed prior to such injections and had their pre-injection status carried forward for the remainder of the study.

For regulatory reporting, the primary analysis was conducted after all patients completed the week 30 assessment or end of TP1. Additional analyses were conducted after the week 54

assessment, including analyses of immunogenicity data. The cut-off date for TP1 as denoted in the clinical study report was June 29, 2016.

The prespecified primary endpoint in this study was the proportion of patients achieving an ACR20 response at week 14. ACR20 is a binary responder type, multi-component endpoint defined by achieving at least 20% improvement from baseline in TJC68 and SJC66, with at least 20% or more improvement from baseline in 3 of the 5 additional measures of disease signs or symptoms: patient global assessment, physician global assessment, patient global assessment of pain, HAQ-DI, and CRP. Patients with dose escalation were considered non-responders.

Other secondary endpoints included ACR50, ACR70, change from baseline in DAS28(CRP)¹, DAS28(CRP) < 2.6, European League Against Rheumatism (EULAR) response, ACR/EULAR remission², and change from baseline in individual components of ACR response.

3.2.2 Statistical Methodologies

3.2.2.1 Planned statistical methodologies

The protocol-defined primary endpoint, ACR20 response while remaining on study treatment at week 14, was analyzed for similarity using an exact binomial test, with the null hypothesis of non-similarity rejected if the 90% confidence bounds for the difference in ACR20 response probabilities at week 14 were within the pre-specified similarity margins. The 90% limits of the confidence interval were obtained based on score statistics using the method of Farrington-Manning [6]. A planned supportive analysis used the unconditional approach based on Santner and Snell 1980.

An additional supportive analysis for the primary endpoint was conducted using a binomial regression model adjusted for treatment group and geographic region. The mean change from baseline for the individual components of ACR across time was provided, along with the 95% CI. A generalized estimating equation model using the binomial family with identity link was fit to evaluate ACR20 in all patients, regardless of adherence to treatment, from week 2 up to week 30, adjusting for treatment arm, categorical visit, treatment by visit interaction, and geographical region. An autoregressive correlation with visit lag of one was assumed for the variance structure.

Secondary efficacy endpoints at all scheduled visits (weeks 2, 4, 6, 12, 22, and 30) were evaluated using descriptive statistics, i.e., unadjusted differences in means with 95% CI. Endpoints evaluated using such analyses included ACR20, ACR50, and ACR70 responses, proportion of patients with EULAR responses (none, moderate, or good), proportion of patients

¹ DAS28(CRP) is a continuous outcome derived by combining tender joints counts out of 28 joints (TJC28), swollen joints out of 28 joints (SJC28), CRP (mg/L), and patient's global assessment of disease activity [GH] by the following equation:

$$\text{DAS28 (CRP)} = 0.56 \times \sqrt{\text{TJC28}} + 0.28 \times \sqrt{\text{SJC28}} + 0.36 \times \log_{10}(\text{CRP} + 1) + 0.014 \times \text{GH} + 0.96.$$

² Either simplified disease activity index or SDAI ≤ 3.3 , sum of tender joint counts, swollen joint counts, patient global assessment, physician global assessment, and CRP or (tender joint counts, swollen joint counts, CRP, and patient global assessment are all ≤ 1).

with DAS28(CRP)<2.6, and change from baseline in continuous endpoints such as ACR components and DAS28(CRP).

A linear mixed-effects repeated measure model was fit to continuous outcomes such as change from baseline in DAS28, adjusting for treatment, categorical visit, treatment and visit interaction, categorical geographical regions, baseline value of the continuous outcome of interest, and interaction of continuous outcome of interest with categorical visit. Within-subject variation across visits was accounted using an unstructured variance-covariance structure, and Kenward-Roger approximation was used to estimate degrees of freedom for pairwise comparisons.

The above statistical analyses were carried out in both intent-to-treat (ITT) and per-protocol (PP) population. The ITT population included all randomized patients. The PP population included all patients who: were randomized and received treatment as planned up to week 14; had an ACR measurement; and did not have major protocol deviations. Major protocol deviations included administration of excluded concomitant medications, interruption of therapy longer than allowed, administration of incorrect treatment, and prohibited modifications of background medication. Patients with incorrect dose escalations on and after week 14 were included in the PP analysis.

The applicant conducted sensitivity analyses using three different approaches: an analysis of available data without imputation (excluding patients with missing data at week 14), a PP analysis of available data (excluding patients with major protocol deviations), and a pattern mixture analysis based on a multiple imputation approach that assumes missingness is at random (MAR). Briefly, multiple imputation-based tipping point analysis of ACR20 at week 14 with delta adjustment was used to investigate the impact of missing data. For ACR20 response at week 14, this model assumed data was missing at random conditional on geographical region, treatment, and the history of ACR20. To evaluate the impact of missing data for the continuous endpoint DAS28(CRP), a pattern mixture multiple imputation analysis was conducted with the last mean carried forward [10].

Subgroup analyses evaluated treatment differences, assuming MAR, according to age (<65 years vs ≥65 years), sex (male/female), race (white, black, and Asian), geographic region (North America and Western Europe, Japan, South Korea, Latin America, and rest of the world), as well as important RA disease characteristics including baseline corticosteroid use (yes/no), duration of prior use of MTX (< 6 months, 6 months to < 1 year, and ≥ 1 year), anti-drug antibody (ADA+/ADA-) status, neutralizing antibody (Nab+/ Nab-) status, baseline use of sulfasalazine (yes/no), and baseline use of anti-malarial treatment (yes/no).

For each subgroup, point estimates and two-sided 95% CIs based on the normal approximation were reported for the difference in ACR20 at week 14 between the two treatment arms. In the protocol, re-definitions of sub-groups for age and prior MTX duration were allowable, to reflect data distribution and balance across different categories. In the applicant's analyses, categories of age and prior MTX duration were revised to: age < 45 years, greater than or equal to 45 years to < 65 years old, ≥ 65 years; prior MTX duration <1 year, or ≥1 year and < 3 years, or ≥ 3 years).

3.2.2.2 Margin selection

The determination of similarity margins is a critical aspect of biosimilarity studies because it precisely defines acceptable differences in efficacy between the reference product and the biosimilar.

The applicant had proposed similarity margins of (b) (4)% absolute on ACR20 response to determine whether PF-06438179 was similar to infliximab-EU. This margin was chosen based on retaining 50% of the lower bound of 95% 2-sided CI on treatment effect from historical studies. The Agency did not agree with the proposed margin and instead recommend that the similarity margin for the proposed comparative clinical study in RA to be no greater in magnitude than $\pm 12\%$. The recommended margin was based on considerations aimed at weighing the clinical importance of various differences in effect against the feasibility of different study sizes.

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% would result in failure to establish similarity, as the 90% confidence interval for the estimated difference would not rule out the 12% margin. Therefore, such a comparative clinical study would be able to rule out losses in ACR20 response greater than 12% with high (at least 95%) statistical confidence, and would 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 corresponded to retention of approximately 50% of conservative estimates of treatment effect size relative to placebo for infliximab (Table 1).

The Agency agreed to consider a proposal for a relaxed upper bound as part of an asymmetric similarity margin, provided the applicant include adequate justification for such an approach as well as the margin chosen. That said, the reliability of the comparative clinical study to rule out clinically meaningful differences depends on the study having assay sensitivity, i.e., the ability to distinguish an effective treatment from a less effective or ineffective treatment.

Assay sensitivity relies on appropriate trial conduct, as poor conduct tends to obscure true differences, increasing the risk of falsely declaring an inferior product to be similar. Therefore, it will be important to evaluate the quality of conduct and the comparability of certain characteristics of the comparative clinical study to those of historical trials that displayed sensitivity to effects of the reference product. Examples of such study characteristics include adherence, concomitant medication use, within-group response rates, and missing data.

The applicant agreed with the FDA to use -12% to rule out any unacceptable loss in efficacy comparing PF-06438179 over infliximab-EU and proposed a relaxed upper margin $+15\%$ to rule out superiority over infliximab. The applicant justified the above choice of a relaxed upper margin based on infliximab saturation of dose response for efficacy based on the ATTRACT Trial. However, the applicant did not discuss whether there could be dose-related safety issues with infliximab. As noted in [2], in a large randomized, placebo-controlled safety study, the relative risk (RR) of developing serious infections (the primary endpoint) by week 22 comparing infliximab 3 mg/kg and placebo was 1.0 (95% CI; 0.3, 3.1), while the RR comparing infliximab

10 mg/kg and placebo was 3.1 (95% CI: 1.2, 7.9). There were 6 (1.7%), 6 (1.7%), and 18 (5.0%) patients with serious infections through week 22 on the placebo, 3 mg/kg, and 10 mg/kg arms, respectively.

Based on the size of the comparative clinical study (325 patients per arm), it was anticipated that there would be reasonable power to detect moderate to large increases in the risk of adverse events (e.g., a 3-fold increase in the risk of serious infections) compared to the reference product. Thus, the Agency agreed that the use of a slightly relaxed upper similarity margin seemed reasonable.

Table 1 Summary of the Five Historical Randomized Clinical Studies of Infliximab in Patients with Active RA with Inadequate Response to MTX

Study	Week	MTX + Placebo		MTX + Infliximab		Difference in % Response
		N	% Response	N	% Response	
Maini et al. [1]	30	88	20%	86	50%	30%
Westhovens et al. [2]	22	361	24%	360	55%	31%
Schiff et al. [3]	28	110	42%	165	59%	18%
Zhang et al. [4]	18	86	49%	87	76%	27%
Abe et al. [5]	14	47	23%	49	61%	38%
Meta-analysis (fixed effects ¹): Difference (95% CI)						28.4% (23.6%, 33.3%)
Meta-analysis (random effects ²): Difference (95% CI)						28.3% (22.6%, 34.1%)
Heterogeneity p-value						0.3

[Source: FDA IR response]

¹: Based on Mantel-Haenszel weights

²: Based on DerSimonian-Laird approach

Abbreviations: MTX=methotrexate; RA=rheumatoid arthritis

3.2.2.3 Reviewer's additional statistical methodologies

Descriptive statistics for all individual components of ACR will be included to describe the efficacy results at all timepoints across the study.

I included cumulative responder curves by computing the probability of ACR response at all thresholds at all study visit weeks to assess for departures from similarity between PF-06438179 and infliximab-EU for probability of ACR response. In addition, post hoc comparisons between PF-06438179 and infliximab-EU for each ACR response threshold were computed based on differences in proportions, with 95% CI based on a normal approximation of the differences in proportions using unpooled standard deviations.

For evaluation of key continuous secondary efficacy endpoints such as DAS28(CRP), the applicant considered the use of mixed models repeated measures for the analysis method. However, the presence of missing data presented additional issues because the proposed statistical model relies on strong and unverifiable assumptions about the missingness mechanism, in addition to assumptions of constant variance and normality across treatment arms. Thus, as supportive analyses, I reported the observed means, standard deviations, and number of patients

available, for the change from baseline in DAS28(CRP) analysis based on all observed data. I further fit the change from baseline in DAS28(CRP) at each scheduled visit using a linear regression with Huber-White sandwich errors adjusting for continuous baseline DAS28(CRP), treatment group, geographical region. Additionally, cumulative distribution curves at week 14 were computed for observed DAS28(CRP) to descriptively assess for potential departures of similarity based on the entire observed distribution regardless of adherence to treatment.

I 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). I 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 [8] and ICH guidelines [9] 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 effect may be larger and easier to detect than the de facto effect because of the restriction to the subsets of patients who adhere. Thus, I further provide a supportive descriptive analysis for the primary endpoint to evaluate the primary endpoint with respect to dose escalation.

The Agency recommended that the applicant conduct additional analyses to more systematically and comprehensively explore the space of plausible missing data assumptions. The tipping point results presented by the applicant evaluated a specific missing not at random assumption depending on the history of the ACR20 and evaluated a limited tipping point parameter space. Thus, I included an additional tipping point analysis that relaxes the dependence on the history of ACR20 collected post baseline. In this tipping point analyses, I computed confidence intervals based on a normal approximation for the difference in ACR20 responses at week 14. Then, I varied the dropouts on both arms independently to completely explore the tipping space. This allows for a follow-up discussion of the plausibility of those assumptions under which the conclusions change.

In the evaluation of safety, the proportion of patients with at least one adverse event (AE) was included for each treatment arm. Risk differences and 95% CI based on normal approximation of the differences in proportions using unpooled standard deviations are presented.

3.2.3 Patient Disposition, Study Conduct, Demographic and Baseline Characteristics

At week 14, approximately 94% of the randomized patients completed randomized treatment. This number of patients decreased to 87% by week 30 (Table 2). The discontinuation rates for both groups appeared similar across the two arms. At week 14, 4% and 7% of the patients from the infliximab-EU and PF-06438179 arm discontinued treatment respectively. The reasons for discontinuation from treatment (Table 3) or discontinuation from study (Table 4) for both treatment groups were numerically similar. Two patients from each of the treatment arm died during Part 1 of the study (See 3.3.2 for more details).

A total of 12% of the patients were excluded in the per-protocol population based on reasons for exclusion collected up to week 14 (Table 5). The reasons for exclusion were balanced across the two arms. The most common reasons for exclusion were not receiving correct doses up to week 6 and not having the week 14 ACR assessment within planned visit schedule window.

Patient demographics and anthropometric variable were similar across the two treatment arms (Table 6). Patients were on average 53 years and older, more frequently female, more frequently white, and neither Hispanic nor Latino. Majority of the randomized patients were enrolled from regions of the world other than North America, Western Europe, Japan, South Korea, or Latin America.

Baseline RA disease characteristics were similar across the two treatment arms (Table 7). The patient's disease duration averaged 6.9 years. Patients had an average of 16 swollen joints, 25 tender joints, and HAQ-DI of 1.6. There were no striking differences between treatments for other baseline disease characteristics or background medication use across the treatment groups. Baseline usage of sulfasalazine was limited to 1% of the total randomized patients. Slightly more patients on the infliximab-EU arm were RF or anti CCP antibody positive.

Table 2 Disposition of Patients in Study B5371002

	Infliximab-EU	PF-06438179	Total
Number randomized	326 (100%)	324 (100%)	650
<i>Number not treated</i>	-	1	1
<i>Number treated ^a</i>	326	323	649
Number who completed randomized treatment in treatment period 1 ^b			
<i>Up to week 14</i>	313 (96%)	300 (93%)	613 (94%)
<i>Up to week 22</i>	296 (91%)	291 (90%)	587 (90%)
<i>Up to week 30</i>	286 (88%)	280 (87%)	566 (87%)
Discontinued study treatment prior to week 30	40 (12%)	43 (13%)	83 (13%)
<i>Discontinued from study</i>	26 (8%)	35 (11%)	61 (11%)
<i>Discontinued from treatment and followed up</i>	14 (4%)	8 (2%)	22 (4%)

[Source: Reviewer]

Counts and percentages were computed relative to row ^a.

^a: One subject was randomized twice and treated as missing on the PF-06438179 arm.

^b: By design, if the last dose at week 22 is taken, the subject is considered to have completed randomized treatment in TP1.

Table 3 Summary of Reasons for Discontinuation from Treatment but were Followed-up in Study B5371002

	Infliximab-EU (N=326) Counts (%) @	PF-06438179 (N=324) Counts (%) @
Number who discontinued treatment prior to week 30	40 (12%)	43 (13%)
<i>Adverse event: Not related to drug</i>	6 (1.8%)	5 (1.5%)
<i>Adverse event: Related to drug</i>	14 (4.3%)	13 (4.0%)
<i>Death</i>	2 (0.6%) ^a	2 (0.6%)
<i>Insufficient clinical response</i>	7 (2.1%)	-
<i>Lost to follow-up</i>	1 (0.3%)	-
<i>No longer willing to participate in study</i>	9 (2.8%)	11 (3.4%)
<i>Non-compliance with study treatment</i>	-	1 (0.3%)
<i>Other</i>	-	4 (1.2%)
<i>Protocol violation</i>	1 (0.3%)	5 (1.5%)
<i>Withdrawn due to pregnancy</i>	-	2 (0.6%)

[Source: Reviewer]

@: Counts and percentages were computed relative to N.

^a: One death occurred on day 235 ((b) (6)) which was within the 30 week visit window cutoff and prior to last subject who completed week 30 assessments ((b) (6)).

Table 4 Summary of Reasons for Discontinuation from Study in Study B5371002

	Infliximab-EU (N=326) Counts (%) @	PF-06438179 (N=324) Counts (%) @
Number who discontinued study prior to week 30	26 (12%)	35 (13%)
<i>Adverse event: Not related to drug</i>	2 (0.6%)	1 (0.3%)
<i>Adverse event: Related to drug</i>	3 (0.9%)	8 (2.5%)
<i>Death</i>	1 (0.3%)	2 (0.6%)
<i>Insufficient clinical response</i>	2 (0.6%)	-
<i>Lost to follow-up</i>	1 (0.3%)	-
<i>No longer willing to participate in study</i>	17 (5.2%)	21 (6.5%)
<i>Protocol violation</i>	-	3 (0.9%)

[Source: Reviewer]

@: Counts and percentages were computed relative to N.

Table 5 Patients Excluded from the Per-Protocol Population (up to Week 14) and Reasons

	Infliximab-EU (N=326) Counts (%)[@]	PF-06438179 (N=324) Counts (%)[@]
Patients excluded from PP for week 14 analysis^a	36 (11.0%)	45 (13.9%)
<i>Incorrectly randomized</i>	-	1 (0.3%)
<i>Inclusion/exclusion not met</i>	2 (0.6%)	2 (0.6%)
<i>GCP non-compliance</i>	2 (0.6%)	2 (0.6%)
<i>Did not receive correct doses up to week 6</i>	16 (4.9%)	15 (4.6%)
<i>Concomitant medications</i>	12 (3.7%)	13 (4.0%)
<i>Missing week 14 ACR measurement^b</i>	12 (3.7%)	21 (6.5%)
Additional patients with protocol deviations not excluded from PP dataset after week 14	38 (11.7%)	39 (12.0%)
<i>Did not receive correct doses prior to week 30</i>	29 (9%)	36 (11%)
<i>Discontinued study treatment^c</i>	11	7
<i>Dose escalated despite not meeting criteria^c</i>	13	22
<i>Missing week 30 ACR measurement^d</i>	9 (3%)	6 (3%)
<i>Concomitant medications after week 14</i>	11 (3%)	6 (2%)

[Source: Reviewer]

[@]: Counts and percentages were computed relative to N.^a: A subject can have more than one protocol deviation.^b: Includes week 14 RA assessment out of protocol defined window (> 14 days); joint assessment for one subject occurred on study day 114^c: Patients could have been listed with more than one of the reasons.^d: Did not come back for week 30 RA assessment; Included one subject from the PF-06438179 arm whose week 30 RA assessment was conducted on day 274

Abbreviations: ACR=American College of Rheumatology; GCP=good clinical practice; PP=per protocol; RA=rheumatoid arthritis

Table 6 Baseline Characteristics for all Randomized Patients^a in Study B5371002

	Infliximab-EU (N=326)	PF-06438179 (N=324)	Total (N=650)
Age (years)	52.8 (12.9)	52.7 (13.3)	52.8 (13.1)
Male	62 (19%)	66 (20%)	128 (20%)
Race			
<i>White</i>	247 (76%)	257 (79%)	504 (78%)
<i>Black</i>	9 (3%)	5 (2%)	14 (2%)
<i>Asian</i>	45 (14%)	46 (14%)	91 (14%)
<i>Other</i>	25 (8%)	15 (5%)	40 (6%)
<i>Unspecified ^a</i>	0 (0%)	1 (0%)	1 (0%)
Geographical Location			
<i>North America/ Western Europe</i>	51 (16%)	50 (15%)	101 (16%)
<i>US patients</i>	44 (14%)	38 (12%)	82 (13%)
<i>Japan</i>	23 (7%)	23 (7%)	46 (7%)
<i>South Korea</i>	5 (2%)	4 (1%)	9 (1%)
<i>Latin America</i>	22 (7%)	22 (7%)	44 (7%)
<i>Rest of the world</i>	225 (69%)	224 (69%)	449 (69%)
<i>Unspecified ^a</i>	0 (0%)	1 (0%)	1 (0%)
Ethnicity			
<i>Hispanic/Latino</i>	32 (10%)	31 (10%)	63 (10%)
<i>Not (Hispanic/Latino)</i>	294 (90%)	292 (90%)	586 (90%)
<i>Unspecified ^a</i>	0 (0%)	1 (0%)	1 (0%)
Weight (kg)	74.2 (20.0)	73.3 (19.8)	73.8 (19.9)
Height (cm)	163.6 (9.5)	163.8 (9.4)	163.7 (9.4)
Body Mass Index (kg/m²)	27.7 (7.0)	27.2 (6.4)	27.4 (6.7)

[Source: Reviewer]

Counts (percentages relative to N) or means (standard deviation) are presented.

^a: One patient from PF-06438179 arm was randomized twice.

Table 7 Baseline Disease Characteristics for all Randomized Patients^a in Study B5371002

		Infliximab-EU (N=326)	PF-06438179 (N=324)	Total (N=650)
RA disease duration	(years)	6.4 (6.7)	7.3 (8.6)	6.9 (7.7)
Replaced and/or fused joint		35 (11%)	23 (7%)	58 (9%)
Swollen joint counts	(out of 66)	16.3 (8.7)	16.1 (9.4)	16.2 (9.1)
Tender joint counts	(out of 68)	25.7 (12.9)	24.7 (13.9)	25.2 (13.4)
Swollen joint counts	(out of 28)	11.9 (5.2)	11.2 (5.7)	11.5 (5.5)
Tender joint counts	(out of 28)	15.8 (6.4)	15.2 (6.9)	15.5 (6.7)
Physician global assessment	(0 – 100mm)	64.2 (16.8)	65.4 (16.2)	64.8 (16.5)
Patient global assessment	(0 – 100mm)	63.9 (23.0)	65.4 (20.7)	64.6 (21.9)
Patient pain assessment	(0 – 100mm)	63.2 (21.6)	63.6 (20.6)	63.4 (21.1)
HAQ-DI	(0 – 3 scale)	1.6 (0.6)	1.6 (0.6)	1.6 (0.6)
CRP	(mg/L)	25.3 (28.4)	25.8 (24.3)	25.6 (26.4)
RF or anti-CCP antibody positive		267 (82%)	249 (77%)	516 (80%)
Methotrexate (MTX) dose	(mg/week)	14.4 (4.5)	14.2 (4.5)	14.3 (4.5)
Prior use of traditional DMARDS (exc MTX) ^a		50 (15%)	57 (18%)	107 (16%)
Number of traditional DMARDS (exc MTX) ^a		0.2 (0.5)	0.2 (0.5)	0.2 (0.5)
Prior use of 1 biologic drugs ^b		3 (1%)	6 (2%)	9 (1%)
Use of anti-malarial drugs		5 (2%)	2 (1%)	7 (1%)
Use of sulfasalazine (>1)		2 (1%)	2 (1%)	4 (1%)
Use of cortisone		192 (59%)	178 (55%)	370 (57%)

[Source: Reviewer]

Counts (percentages relative to N) or means (standard deviation) are presented.

^a: Patients reported previous use of auranofin, azathioprine, chloroquine, gold, hydroxychloroquine, leflunomide, sodium aurothiomalate, sulfasalazine, tofacitinib.

^b: Patients reported previous use of adalimumab, tocilizumab, certolizumab pegol.

Abbreviations: CRP=C-reactive protein; anti-CCP=anti-cyclic citrullinated peptide; DMARDS=disease modifying anti-rheumatic drugs; HAQ-DI=health assessment questionnaire-disability index; MTX=methotrexate;

RA=rheumatoid arthritis; RF=rheumatoid factor

3.2.4 Results and Conclusions

3.2.4.1 Evaluation of ACR response

The 90% confidence limits on differences between treatments were within the pre-specified FDA asymmetric margin of -12% and +15% (Table 8). At week 14, 63.8% of the patients randomized to infliximab-EU and 61.1% of the patients randomized to PF-06438179 remained in the study and achieved ACR20 response at week 14, with an estimated absolute response rate -2.7% lower on the PF-06438179 arm compared to the infliximab-EU arm (90% exact CI: -9.1%, 3.6%).

Additional supportive analysis based on all observed data analysis provided similar findings.

The per protocol analysis with non-responder imputation or all observed data collected similarly ruled out the FDA asymmetric margin.

Table 8 Differences in ACR20 Responses for Various Analysis Datasets Based on Exact Binomial Score Test Conducted at Primary Week of Analysis

	Infliximab-EU (N=326)		PF-06438179 (N=324)		Difference in Response (%)	90% CI Lower (%)	90% CI Upper (%)
	n	Count (%)	n	Count (%)			
ACR20 @ week 14^a							
ITT: NRI	326	208 (63.8%)	324	198 (61.1%)	-2.7	-9.1	3.6
ITT: Observed ^b	316	210 (66.5%)	311	203 (65.3%)	-1.2	-8.4	6.6
PP: NRI	290	196 (67.6%)	279	186 (66.7%)	-0.9	-7.5	5.7
PP: Observed	290	196 (67.6%)	279	186 (66.7%)	-0.9	-7.5	5.7

[Source: Reviewer]

Observed analysis included efficacy data collected regardless of adherence to study treatment.

NRI: Patients who discontinued the study treatment or withdraw from the study were imputed as non-responder.

^a: This differs from original submission (Seq0001) after a data entry was noted by the applicant in a later submission (Seq0022) for subject (b) (6) at week 14.

^b: Two patients from infliximab-EU and five from PF-06438179 discontinued study treatment but were followed up beyond week 14. This analysis was based on data from the original submission and included (b) (6) as an ACR20 responder at week 14.

Abbreviations: ACR=American College of Rheumatology; CI=confidence intervals; ITT=intent to treat; NRI=non-responder imputation; PP=per protocol

The cumulative distribution curves for infliximab-EU and PF-06438179 at week 14 were similar, with a slight trend toward greater efficacy of PF-06438179 at higher ACR thresholds (Figure 2), regardless of adherence to study treatment.

Supportive analyses for ACR20 at other weeks, not accounting for multiplicity but otherwise analyzed using the primary analysis method, did not show notable differences in efficacy (Table 20). The slight trend towards greater efficacy of PF-06438179 at higher ACR thresholds observed at week 14 did not persist at later weeks.

In Figure 3, the difference in response comparing PF-06438179 relative to infliximab-EU using all possible ACR thresholds at all study visits up to week 30 were in general similar and did not exhibit gains or losses in efficacy compared to infliximab-EU. Of note, a supportive analysis that excluded one subject from the infliximab-EU arm ((b) (6)) whose efficacy assessment was conducted on day 274 (outside the pre-specified week 30 window) did not affect the findings at week 30.

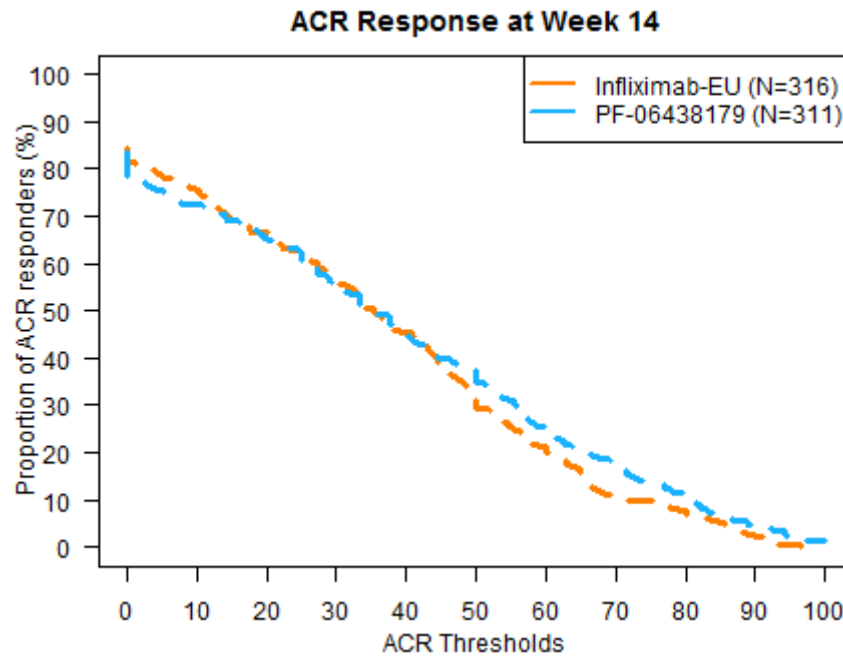
Summary measures based on all observed data for the individual components of ACR, i.e., SJC66, TJC68, patient global, physician global, patient pain, HAQ-DI, and CRP were numerically similar (Table 9 and Table 10). The observed range of the change from baseline in CRP was generally narrower on the PF-0638179 arm than that of the infliximab-EU arm. However, cumulative responder curves were generally similar from the 5% to the 95% quantiles.

Based on the applicant's data revisions submitted on October 26, 2017 and November 07, 2017, sensitivity analyses conducted for SJC66, TJC68, and ACR20, using all observed data regardless of treatment adherence at week 14 and 30, yielded results similar to the above analyses.

Therefore, despite revisions in the data, the results remain highly persuasive in favor of similar efficacy between infliximab-EU and PF-06438179.

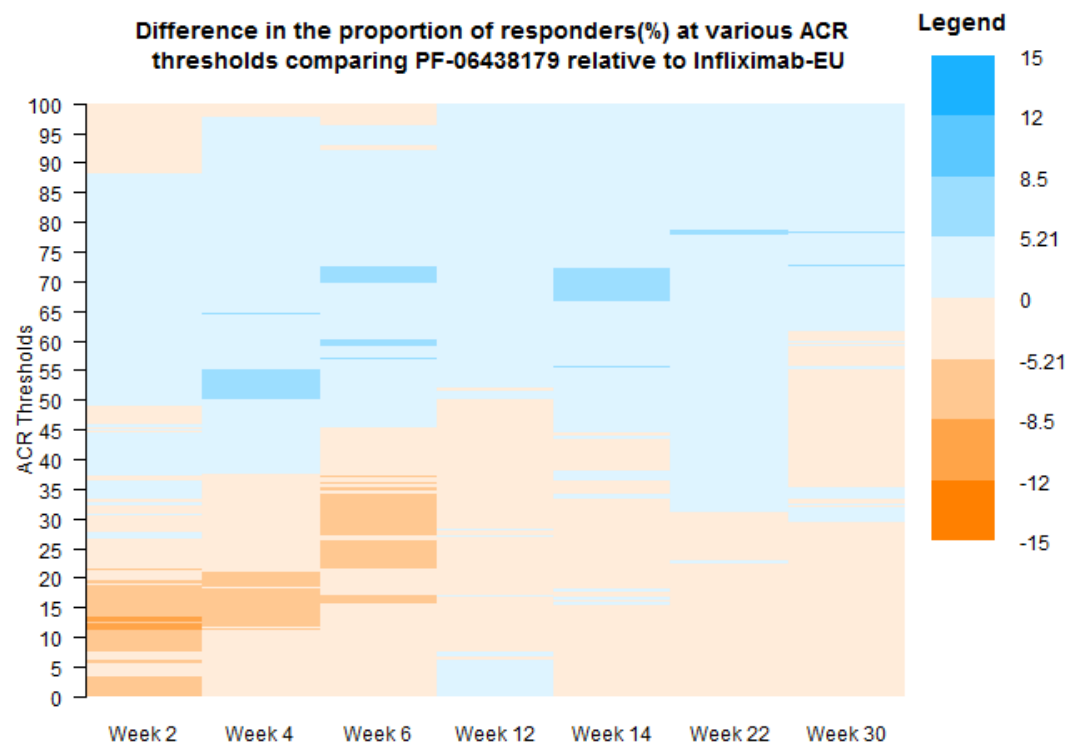
In summary, improvements in ACR20, other thresholds of ACR, and individual components of ACR at protocol-defined primary week of analysis as well as other study weeks were similar between PF-06438179 and infliximab-EU.

Figure 2 ACR Responder Curve, Regardless of Adherence to Study Treatment for Various Thresholds at Week 14 Comparing Infliximab EU vs PF-06438179



[Source: Reviewer]

Figure 3 Differences in ACR Response Comparing PF-06438179 vs Infliximab-EU Using all Possible ACR Thresholds Across Different Study Visits up to Week 30, Regardless of Adherence to Study Treatment



[Source: Reviewer]

Table 9 Change from Baseline for the Individual Components of ACR at Scheduled Visit Weeks Based on all Data Collected Regardless of Study Treatment Discontinuation by Treatment Arm

Visit week	Treatment group	Descriptive statistics	SJC (Out of 66)	TJC (Out of 68)	Patient (0 – 100mm)	Physician (0 – 100mm)	Patient pain (0 – 100mm)	HAQ-DI (0 – 3)	CRP (mg/L)
2	Infliximab-EU (n=324)	Mean (SD)	-5.7 (7.3)	-7.5 (8.4)	-17 (20)	-20 (17)	-15 (19)	-0.33 (0.44)	-16 (24)
		Med (Min, Max)	-4 (-52, 15)	-5.5 (-36, 22)	-15 (-84, 41)	-18 (-89, 35)	-14 (-85, 37)	-0.25 (-1.9, 1.1)	-9.1 (-185, 39)
	PF-06438179 (n=319)	Mean (SD)	-5.5 (6.9)	-5.9 (8.8)	-17 (23)	-22 (19)	-16 (22)	-0.32 (0.41)	-17 (21)
		Med (Min, Max)	-4 (-40, 19)	-5 (-43, 25)	-12 (-91, 64)	-19 (-90, 20)	-12 (-95, 68)	-0.25 (-2.2, 0.8)	-11 (-104, 52)
4	Infliximab-EU (n=321)	Mean (SD)	-7.9 (7.4)	-10 (9.4)	-21 (24)	-28 (18)	-21 (21)	-0.48 (0.49)	-13 (34)
		Med (Min, Max)	-7 (-48, 17)	-9 (-40, 25)	-20 (-85, 45)	-27 (-92, 12)	-20 (-92, 48)	-0.38 (-2.1, 1.1)	-8.5 (-184, 366 ^a)
	PF-06438179 (n=317)	Mean (SD)	-7.8 (7.7)	-9.5 (10)	-23 (23)	-30 (19)	-22 (22)	-0.47 (0.47)	-16 (20)
		Med (Min, Max)	-7 (-37, 15)	-7 (-56, 24)	-21 (-89, 71)	-30 (-83, 22)	-20 (-88, 48)	-0.38 (-2.6, 0.9)	-12 (-100, 50)
6	Infliximab-EU (n=319)	Mean (SD)	-9.0 (7.9)	-12 (10)	-23 (24)	-31 (19)	-23 (23)	-0.52 (0.5)	-13 (28)
		Med (Min, Max)	-7 (-53, 18)	-11 (-54, 21)	-23 (-89, 58)	-31 (-90, 24)	-22 (-91, 49)	-0.5 (-2.2, 0.5)	-8.1 (-174, 95)
	PF-06438179 (n=313)	Mean (SD)	-8.6 (8.0)	-11 (11)	-26 (25)	-33 (20)	-24 (25)	-0.5 (0.55)	-14 (20)
		Med (Min, Max)	-7 (-41, 17)	-8 (-58, 35)	-24 (-95, 67)	-32 (-86, 20)	-23 (-95, 49)	-0.38 (-2.8, 0.9)	-11 (-98, 58)
12	Infliximab-EU (n=318)	Mean (SD)	-9.6 (8.4)	-13 (12)	-27 (26)	-34 (20)	-26 (25)	-0.52 (0.59)	-13 (28)
		Med (Min, Max)	-9 (-49, 29)	-12 (-54, 26)	-27 (-87, 67)	-35 (-96, 28)	-26 (-91, 46)	-0.5 (-2.4, 1.4)	-7.4 (-168, 96)
	PF-06438179 (n=311)	Mean (SD)	-9.6 (8.6)	-12 (12)	-27 (25)	-35 (20)	-25 (26)	-0.53 (0.58)	-12 (24)
		Med (Min, Max)	-8 (-43, 20)	-10 (-56, 28)	-25 (-96, 63)	-34 (-91, 18)	-24 (-96, 46)	-0.5 (-2.8, 1.1)	-10 (-103, 130)
14 ^b	Infliximab-EU (n=316)	Mean (SD)	-9.6 (8.4)	-13 (12)	-25 (27)	-34 (21)	-25 (25)	-0.53 (0.59)	-12 (29)
		Med (Min, Max)	-9 (-52, 31)	-13 (-54, 42)	-24 (-91, 77)	-35 (-94, 33)	-25 (-92, 45)	-0.5 (-2.9, 1.5)	-7.6 (-178, 109)
	PF-06438179 (n=311)	Mean (SD)	-9.3 (8.9)	-12 (12)	-28 (27)	-36 (21)	-26 (27)	-0.57 (0.59)	-13 (23)
		Med (Min, Max)	-8 (-40, 39)	-10 (-63, 42)	-25 (-95, 60)	-36 (-95, 19)	-25 (-95, 48)	-0.5 (-2.8, 1.1)	-9.1 (-102, 82)

[Source: Reviewer]

^a: One patient from the infliximab-EU arm had elevated CRP, contributing to the large change from baseline in CRP at week 4.

^b: SJC66 (value of 1) and TJC68 (value of 8) were revised for subject ^{(b) (6)} at Week 14 based on applicant's revised submission dated October 26, 2017.

Abbreviations: ACR=American College of Rheumatology; CRP=C-reactive protein; Max=maximum; Med=median; Min=minimum; SD=standard deviation; SJC=swollen joint counts; TJC=tender joint counts; IR=information request

Table 10 Change from Baseline for the Individual Components of ACR at Scheduled Visit Weeks Based on all Data Collected Regardless of Study Treatment Discontinuation by Treatment Arm (continued from Table 9)

Visit week	Treatment group	Descriptive statistics	SJC (Out of 66)	TJC (Out of 68)	Patient (0 – 100mm)	Physician (0 – 100mm)	Patient pain (0 – 100mm)	HAQ-DI (0 – 3)	CRP (mg/L)
22	Infliximab-EU (n=311)	Mean (SD)	-10 (7.9)	-15 (13)	-26 (27)	-36 (21)	-26 (25)	-0.57 (0.6)	-11 (31)
		Med (Min, Max)	-9 (-49, 19)	-14 (-60, 33)	-27 (-86, 59)	-37 (-95, 28)	-24 (-79, 59)	-0.5 (-2.8, 0.8)	-6.8 (-186, 136)
	PF-06438179 (n=301)	Mean (SD)	-11 (8.8)	-13 (13)	-29 (28)	-38 (21)	-28 (27)	-0.59 (0.61)	-11 (25)
		Med (Min, Max)	-9 (-43, 15)	-12 (-63, 27)	-26 (-97, 63)	-38 (-88, 11)	-26 (-96, 53)	-0.5 (-2.8, 0.6)	-8.9 (-94, 103)
30	Infliximab-EU (n=298)	Mean (SD)	-11 (8.5)	-16 (13)	-29 (29)	-37 (22)	-29 (27)	-0.61 (0.65)	-12 (30)
		Med (Min, Max)	-10 (-42, 23)	-14 (-59, 22)	-29 (-96, 58)	-38 (-96, 31)	-28 (-93, 56)	-0.5 (-2.8, 0.9)	-7 (-189, 102)
	PF-06438179 (n=294)	Mean (SD)	-11 (9.3)	-14 (13)	-29 (29)	-40 (22)	-29 (28)	-0.62 (0.65)	-12 (26)
		Med (Min, Max)	-9 (-45, 27)	-13 (-66, 39)	-28 (-98, 66)	-40 (-88, 29)	-31 (-96, 71)	-0.5 (-2.8, 1.1)	-9.1 (-109, 108)

[Source: Reviewer]

Abbreviations: ACR=American College of Rheumatology; CRP=C-reactive protein; Max=maximum; Med=median; Min=minimum; SD=standard deviation; SJC=swollen joint counts; TJC=tender joint counts

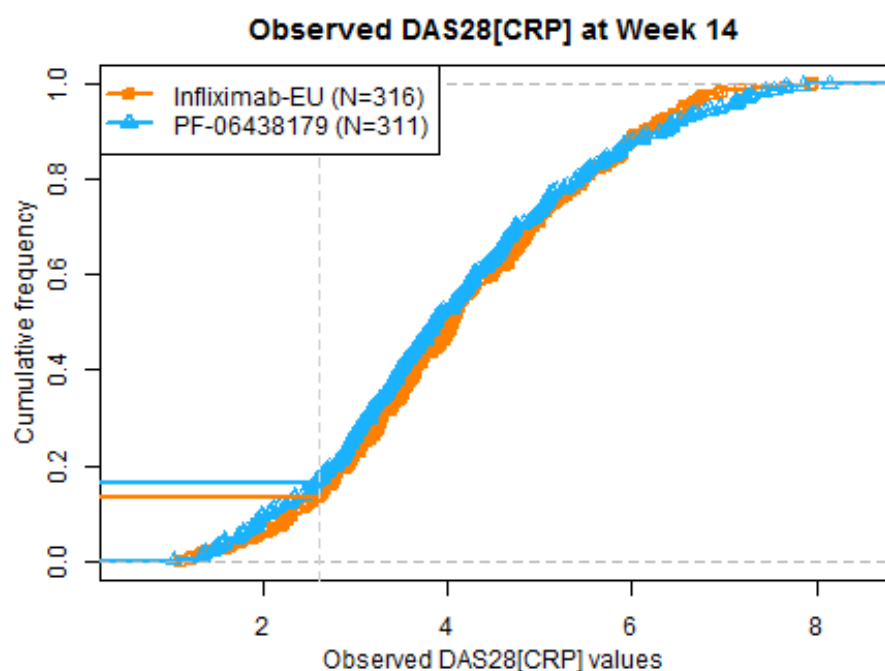
3.2.4.2 Evaluation of DAS28(CRP)

The cumulative distribution curves for all observed DAS28(CRP) at week 14 were in general similar for the two treatments (Figure 4).

The adjusted change from baseline in DAS28(CRP) among all randomized patients regardless of adherence to study treatment and per-protocol population at each visit week were in general similar across study weeks for the two products (Table 11).

Based on the applicant's data revision submitted on October 26, 2017 and November 07, 2017, sensitivity analyses conducted for DAS28(CRP) regardless of treatment adherence were similar in conclusions to the above analyses carried out.

Figure 4 Cumulative Distribution Curves for Patients with Observed DAS28(CRP) at Week 14



[Source: Reviewer]

Vertical dotted gray line represents DAS28(CRP) < 2.8.

Table 11 Regression Analysis of the Change from Baseline in DAS28(CRP) Based on all Patients Regardless of Adherence to Study Treatment and PP Dataset

	Week	Infliximab-EU (N=326)		PF-06438179 (N=323)		Estimated difference (90% Robust CI) ^a	Estimated difference (90% CI) ^b
		Mean (SD)	n	Mean (SD)	n		
ITT (Observed)	2	-1.24(0.89)	324	-1.21(0.93)	317	0.02 (-0.09, 0.14)	0.02 (-0.09, 0.14)
	4	-1.60(1.09)	315	-1.60(1.13)	312	0.01 (-0.14, 0.15)	-0.02 (-0.16, 0.12)
	6	-1.75(1.09)	319	-1.71(1.20)	312	0.04 (-0.11, 0.18)	0.05 (-0.10, 0.19)
	12	-1.89(1.21)	316	-1.90(1.35)	310	-0.02 (-0.19, 0.14)	-0.02 (-0.18, 0.15)
	14	-1.83(1.30)	314	-1.90(1.41)	310	-0.07 (-0.25, 0.10)	-0.08 (-0.26, 0.09)
	22	-2.00(1.30)	307	-2.01(1.42)	301	-0.01 (-0.19, 0.16)	-0.02 (-0.19, 0.15)
	30	-2.12(1.27)	297	-2.14(1.42)	292	-0.04 (-0.21, 0.13)	-0.07 (-0.25, 0.11)
PP	2	-1.26(0.89)	290	-1.25(0.94)	276	0.01 (-0.11, 0.13)	0.01 (-0.11, 0.14)
	4	-1.65(1.08)	282	-1.63(1.14)	273	0.02 (-0.13, 0.18)	-0.00 (-0.15, 0.15)
	6	-1.77(1.10)	289	-1.74(1.18)	277	0.04 (-0.12, 0.19)	0.04 (-0.12, 0.19)
	12	-1.94(1.19)	286	-1.92(1.33)	277	0.02 (-0.15, 0.19)	0.02 (-0.15, 0.19)
	14	-1.85(1.28)	288	-1.92(1.39)	278	-0.05 (-0.23, 0.13)	-0.06 (-0.24, 0.12)
	22	-2.06(1.30)	281	-2.01(1.43)	276	0.04 (-0.14, 0.23)	0.02 (-0.17, 0.20)
	30	-2.21(1.27)	272	-2.18(1.41)	267	0.01 (-0.17, 0.19)	-0.03 (-0.21, 0.16)

[Source: Reviewer]

^a: Linear regression model with Huber-White sandwich errors to the change from baseline in DAS28(CRP) at each week, adjusting for baseline DAS28(CRP), categorical geographic region, and treatment

^b: MMRM model fit to the change from baseline in DAS28(CRP) adjusting for baseline continuous DAS28(CRP), categorical geographic region, categorical treatment group, categorical scheduled visit week, categorical scheduled visit week and treatment interaction, with an unstructured covariance matrix, and Kenward-Roger denominator degrees of freedom

Abbreviations: DAS28(CRP)= Disease Activity Score-28, 4 components based on CRP; MMRM= mixed-effects model repeat measures; ITT=intent-to-treat; PP=per protocol; SD=standard deviation

3.2.4.3 Impact of dose escalation post week 14

These proportions of patients who met criteria to dose escalate were generally similar across arms (Table 12). The proportion of patients who had not met criteria to dose escalate but met criteria to dose escalate at week 22 and those who did not meet criteria to dose escalate at either week 14 or week 22 but were dose escalated at these two weeks were also similar across arms.

A per-protocol analysis of ACR20 after week 14 depending on dose escalation were similar across treatment groups (Table 13). Interpretation of such results based on “groups” defined after having been treated can be affected by differences in baseline characteristics, treatment administered, or discontinuation from the study. In evaluation of biosimilarity between products, the per-protocol effect may be larger and easier to detect after restricting analyses to subsets of patients who adhere. These supportive analyses suggest that, despite dose escalation, the proportion of ACR20 responders were in general similar across the two arms and did not present any numerical imbalance across the groups.

Table 12 Number of Patients who Dose Escalated During the Study after Week 14

	Infliximab-EU (N=326) Counts (%)	PF-06438179 (N=323) Counts (%)
Week 14		
<i>Dose escalated</i>	65 (20%)	58 (18%)
<i>Did not meet criteria but dose escalated</i>	3 (1%)	2 (1%)
Week 22		
<i>Dose escalated</i>	7 (2%)	12 (4%)
<i>Did not meet criteria but dose escalated</i>	8 (2%)	11 (3%)

[Source: Reviewer]

Table 13 Post-hoc Analyses of ACR20 Responses Regardless of Adherence to Study Treatment Based on Whether Patients Dose Escalated at Weeks 14 and 22

	Infliximab-EU (N=326)		PF-06438179 (N=323)^c	
	n/n _{miss}	Count (%)	n/n _{miss}	Count (%)
ACR20 @ week 22 ^a	326/15	213 (68.5%)	323/22	205 (68.1%)
<i>Dose escalated</i>	68/5	27 (39.7%)	60/1	23 (38.3%)
<i>Did not dose escalate</i>	258/10	186 (72.1%)	263/21	182 (69.2%)
ACR20 @ week 30 ^b	326/38	209 (70.1%)	293/29	197 (67.0%)
<i>Dose escalated</i>	83/12	34 (41.0%)	83/5	33 (39.8%)
<i>Did not dose escalate</i>	243/16	175(72.0%)	240/24	164(68.3%)

[Source: Reviewer]

^a: Patients who dose escalated at week 14 were classified as having dose escalated^b: Patients who either dose escalated at week 14 or week 22 were classified as dose escalated. This differs from the applicant's classification in the original clinical report but were consistent in the revised report submitted in Seq0022.^c: Excluded the subject who was randomized twice

Abbreviations: ACR=American College of Rheumatology

3.2.4.4 Potential impact of missing data

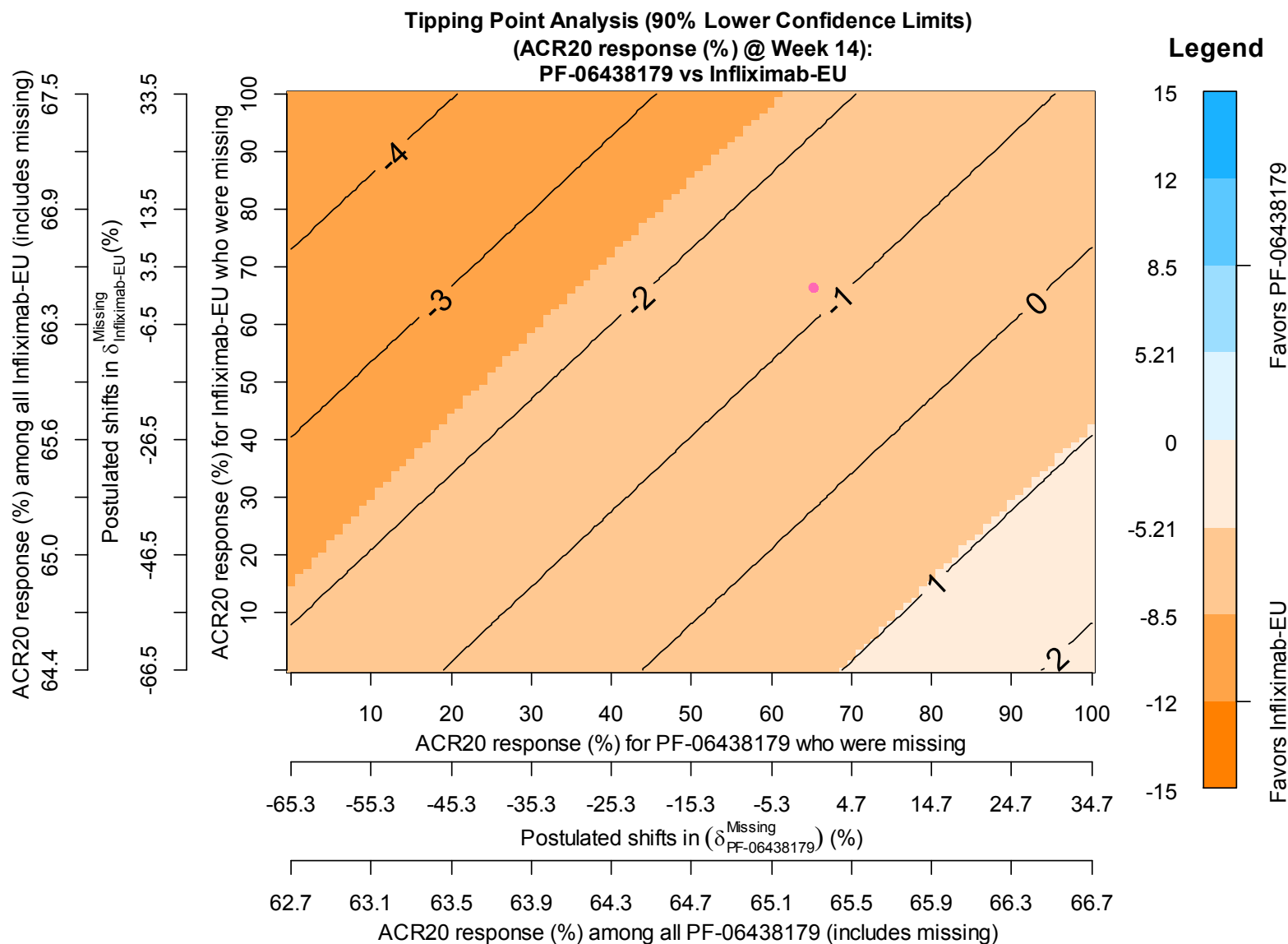
This section addresses the effect of missing data on the reliability of comparative efficacy results from the study. As I noted in Table 2, up to week 14, 23 (4%) patients had withdrawn from the study: 13 patients (4%) from the PF-06438179 treatment group and 10 patients (3%) from the infliximab-EU treatment group. In the primary analysis based on the ITT set, patients who dropped out were considered non-responders, thus the primary endpoint was a composite measure of treatment success defined by adherence to the treatment through week 14 and achieving an ACR20 response at week 14. 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 14 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. Such evaluations are subject to some missing data (because patients who discontinued treatment were not followed up for assessment) and rely on strong and unverifiable assumptions, such as the assumption that outcomes in patients who withdrew early are missing (completely) at random. Therefore, we requested from the applicant, and conducted our own, tipping point analyses to explore the sensitivity of results to violations in assumptions about the missing data (i.e., to various missing-not-at-random assumptions). Moreover, the 2010 National Research Council (NRC) Report *The Prevention and Treatment of Missing Data in Clinical Trials* recommends that “examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting.”

In the tipping point analyses, all points evaluated were within the biosimilarity margins, ruling out a difference between infliximab-EU and PF-06438179 exceeding a 12% absolute gain in probability of ACR20 response (Figure 5) and ruling out a 15% absolute loss (Figure 6).

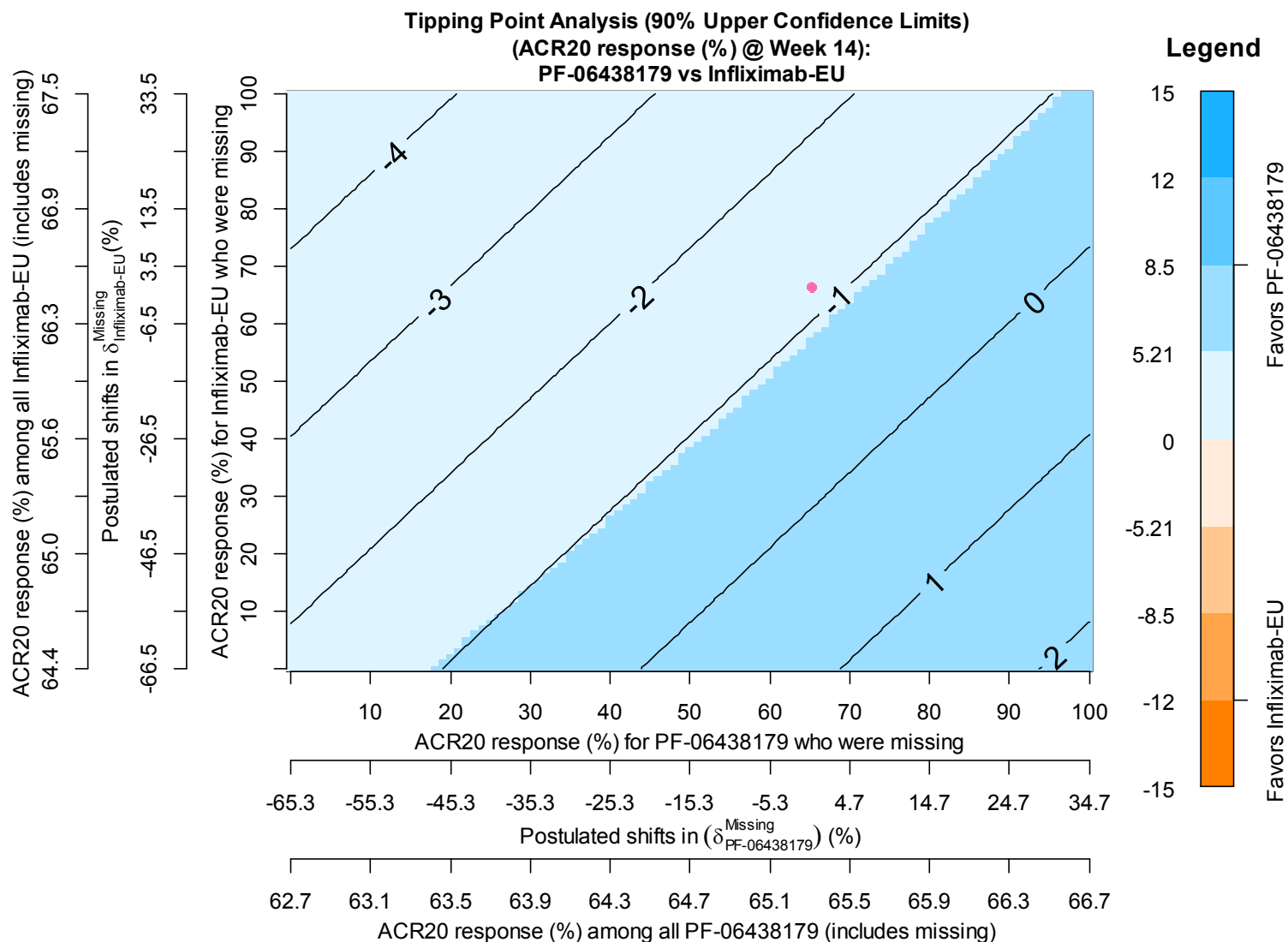
In summary, the efficacy of infliximab-EU was similar to that of PF-06438179 for the primary endpoint, ACR20 at week 14 notwithstanding missing data. There were relatively small proportions of dropouts at week 14, similar distributions of reasons for early withdrawal on the two treatment arms (see Figure 1 and Table 5), and large differences between the outcomes in dropouts on the two arms based on the tipping point analyses seemed implausible (Figure 5 and Figure 6). As a point of reference, the response probabilities among completers on PF-06438179 and infliximab-EU were 65% and 66%, respectively. Therefore, the tipping point sensitivity analyses largely supported the findings of the key efficacy analyses in Study B5371002.

Figure 5 Tipping Point Analyses Results for 90% Asymptotic Lower CL (Heatmap) and Point Estimates (Contour Lines): Inference on the Difference Between PF06438179 and Infliximab-EU in the Probability of Week 14 ACR20 Responses under Varying Assumptions About the Differences on Each Treatment Arm



[Source: Reviewer] Abbreviations: ACR=American College of Rheumatology; CL=confidence limits

Figure 6 Tipping Point Analyses Results for 90% Asymptotic Upper CL (Heatmap) and Point Estimates (Contour Lines): Inference on the Difference Between PF06438179 and Infliximab-EU in the Probability of Week 14 ACR20 Responses under Varying Assumptions About the Differences on Each Treatment Arm



[Source: Reviewer] Abbreviations: ACR=American College of Rheumatology; CL=confidence limits

3.3 Evaluation of Safety

Dr. Erika Torjusen, the Medical Reviewer, conducted a complete safety evaluation, and the reader is referred to Dr. Torjusen's review for more detailed information. I conducted supplementary analyses to compare PF-06438179 and infliximab-EU with respect to the incidence of adverse events of special interest (AESIs) in B5371002 for Treatment Period 1.

3.3.1 Drug Exposure

The duration of use of study drug injections was similar across the two treatment arms. Total exposure to the study drug was balanced across the treatment arms, with less than 5% of the patients missing a study dose. Less than 5% of the patients had their injections interrupted for reasons attributed to adverse events.

Table 14 Exposure to Study Drug during Treatment Period 1

	Infliximab-EU (N=326)	PF-06438179 (N=323)
Exposure (weeks)	22 (0.1 - 26)	22 (0.1 - 27)
Total exposure		
<i>Median (min – max)</i>	1104 (217 - 3091)	1068 (138 - 2691)
<i>Mean (standard deviation)</i>	1137 (377)	1111 (390)
Patients with any missed dose	15 (5%)	14 (4%)
Patients with any infusion interrupted	18 (6%)	15 (5%)
<i>Interrupted due to AE</i>	15 (5%)	12 (4%)
Dose reduction ^a	1 (0.3%)	-

[Source: Reviewer]

^a: None had a dose reduction due to adverse events.

Abbreviations: AE=adverse event

3.3.2 Deaths

There was a total of four deaths through week 30 + 28 days. Two occurred on each treatment arm. Two patients from PF-06438179 and one from infliximab-EU had dose escalated to an infusion dose of 5 mg/kg. Table 21 included the details of the deaths and the any adverse event reported prior to death.

3.3.3 Adverse Events

In Study B5371002, the number of patients with at least 1 adverse events were overall proportion of adverse events were balanced across the two treatment arms during Treatment Period 1 (Table 15). The proportion of patients with at least one AE classified under key system organ class of interest were similar (Table 16). Numerically more patients had bronchitis infections on the PF-06438179 arm.

Table 15 Summary of Adverse Events in Treatment Period 1 through 30 Weeks

Number of patients with:	Infliximab-EU (N=326)	PF-06438179 (N=323)
At least one TEAE	176 (54.0%)	185 (57.3%)
Serious Adverse Events	20 (6.1%)	16 (5.0%)
Grade 3 AE	34 (10.4%)	34 (10.5%)
Grade 4 AE	6 (1.8%)	2 (0.6%)
Grade 5 AE	1 (0.3%)	2 (0.6%)
Discontinued from treatment due to AEs	24 (7.4%)	23 (7.1%)
Discontinued from study due to AEs	14 (4.3%)	16 (5.0%)
Temporary discontinued due to AEs	28 (8.6%)	31 (9.6%)

[Source: Reviewer]

Counts and percentages relative to N in parenthesis.

Abbreviations: AE=adverse events; TEAE=treatment-emergent adverse event

Table 16 Adverse Events by MedDRA System Organ Class and Preferred Term during Treatment Period 1

System Organ Class Preferred Term	Infliximab-EU (N=326)	PF-06438179 (N=323)	Risk Diff ^a	95% CI ^a
Infections and infestations	72 (22.1%)	86 (26.6%)	4.5%	[-2.1%, 11.1%]
<i>Bronchitis ^b</i>	6 (1.8%)	14 (4.3%)	2.5%	[-0.2%, 5.2%]
<i>Nasopharyngitis</i>	13 (4.0%)	14 (4.3%)	0.3%	[-2.7%, 3.4%]
<i>Upper respiratory tract infection</i>	13 (4.0%)	12 (3.7%)	-0.3%	[-3.2%, 2.7%]
<i>Urinary tract infection</i>	9 (2.8%)	6 (1.9%)	-0.9%	[-3.2%, 1.4%]
<i>Urinary tract infection bacterial</i>	1 (0.3%)	0 (0.0%)	-0.3%	[-0.9%, 0.3%]
Serious Infections ^c	9 (2.8%)	6 (1.9%)	-0.9%	[-3.2%, 1.4%]
<i>Pneumonia</i>	2 (0.6%)	2 (0.6%)	0.0%	[-1.2%, 1.2%]
<i>Pneumocystis jirovecii pneumonia</i>	-	1 (0.3%)	0.3%	[-0.3%, 0.9%]
<i>Bronchitis</i>	1 (0.3%)	-	-0.3%	[-0.9%, 0.3%]
<i>Tuberculosis</i>	1 (0.3%)	-	-0.3%	[-0.9%, 0.3%]
Investigations	26 (8.0%)	37 (11.5%)	3.5%	[-1.1%, 8.0%]
<i>ALT increased</i>	15 (4.6%)	19 (5.9%)	1.3%	[-2.1%, 4.7%]
<i>AST increased</i>	11 (3.4%)	14 (4.3%)	1.0%	[-2.0%, 3.9%]
<i>Mycobacterium tuberculosis complex test positive</i>	4 (1.2%)	1 (0.3%)	-0.9%	[-2.3%, 0.4%]
Injury, poisoning, and procedural complications	36 (11.0%)	36 (11.1%)	0.1%	[-4.7%, 4.9%]
<i>Infusion related reactions</i>	21 (6.4%)	19 (5.9%)	-0.6%	[-4.3%, 3.1%]
Neoplasms	2 (0.6%)	1 (0.3%)	-0.3%	[-1.3%, 0.7%]
<i>Colon Cancer</i>	1 (0.3%)	1 (0.3%)	0.0%	[-0.9%, 0.9%]
<i>Lipoma</i>	1 (0.3%)	-	-0.3%	[-0.9%, 0.3%]

[Source: Reviewer]

Counts and percentages relative to N in parenthesis.

^a: Risk difference (Ref=infliximab-EU) with 95% CI based on normal approximation to difference in proportions^b: One case of bronchitis bacterial from the PF-06438179 was not counted in this definition.^c: Defined by AEFDAY>0, flagged as 1 based on AESER, with System Organ Class like "Infections"

Abbreviations: AE = adverse event; ALT=Alanine aminotransferase; AST=aspartate aminotransferase; EU = European Union; n = number of patients with at least one adverse event; N=number of patients in the safety analysis set; SD = standard deviation; TP1 = Treatment Period 1

3.3.4 Adverse Events of Special Interest Determined by Applicant

AEs of special interest, identified prior to week 30 database lock, were numerically balanced across treatment arms (Table 17). Because study follow-up was generally similar on each arm (median of 211 days for each arm), the proportion of patients with adverse events, without adjustment for exposure years, could be used to compare treatments.

Table 17 Summary of Adverse Events of Special Interest Flagged by the Applicant through TP1

System Organ Class Preferred Term	Infliximab-EU (N=326) Count (%)	PF-06438179 (N=323) Count (%)	Risk Diff ^a	95% CI ^a
Infections ^b	73 (22.4%)	87 (26.9%)	4.5%	[-2.1%, 11.2%]
Latent tuberculosis	-	1 (0.3%)	0.3%	[-0.3%, 0.9%]
Tuberculosis	1 (0.3%)	-	-0.3%	[-0.9%, 0.3%]
Mycobacterium tuberculosis complex test positive	4 (1.2%)	1 (0.3%)	-0.9%	[-2.3%, 0.4%]
Bronchitis ^c	6 (1.8%)	14 (4.3%)	2.5%	[-0.2%, 5.2%]
Pneumonia	3 (0.9%)	2 (0.6%)	-0.3%	[-1.6%, 1.0%]
Pneumocystis jirovecii pneumonia	-	1 (0.3%)	0.3%	[-0.3%, 0.9%]
Hypersensitivity ^d	51 (15.6%)	44 (13.6%)	-2.0%	[-7.5%, 3.4%]
Infusion related reactions	21 (6.4%)	19 (5.9%)	-0.6%	[-4.3%, 3.1%]
Neoplasms,	2 (0.6%)	1 (0.3%)	-0.3%	[-1.3%, 0.7%]
Colon Cancer	1 (0.3%)	1 (0.3%)	0.0%	[-0.9%, 0.9%]
Lipoma	1 (0.3%)	-	-0.3%	[-0.9%, 0.3%]

[Source: Reviewer]

Counts and percentages relative to N in parenthesis.

^a: Risk difference (Ref=infliximab-EU) with 95% CI based on normal approximation to difference in proportions.

^b: SOC infections and infestations, high-level terms (HLTs) tuberculous infections and HLT mycobacteria identification and serology. The mycobacterium tuberculosis complex test used to test for tuberculosis may not be further confirmed by chest x-rays or additional tests.

^c: One case of bronchitis bacterial from the PF-06438179 was not counted in this definition.

^d: Hypersensitivity is defined based on MedDRA version 19.0 search for hypersensitivity Standardized MedDRA Query (SMQ, broad and narrow) and anaphylactic reactions SMQ (broad and narrow) and high-level group terms (HLGT) immunology and allergy investigations.

Abbreviations: SOC=systems organ class; HLTs=high-level terms; HLGT=high-level group terms;

SMQ=standardized MedDRA query

3.4 Assay Sensitivity and Constancy Assumption

To reliably evaluate whether there are clinically meaningful differences between two products, a comparative clinical study must have assay sensitivity, or the ability to detect meaningful differences between the products, if such differences exist. In addition, to reliably evaluate whether the experimental treatment retains a certain proportion of the effect of the reference product versus placebo, the constancy assumption must be invoked. This is the assumption that estimates of the effect of the reference product from historical, placebo controlled trials accurately reflect reference product effects in the setting of the biosimilarity 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 are similarly effective or, in the case of poor assay sensitivity, similarly ineffective. As discussed in the ICH E10 guidelines and in the literature, historical evidence of sensitivity to drug effects and appropriate trial conduct may be used to support the presence of assay sensitivity and a conclusion that the treatments are similarly effective.

Table 18 describes key characteristics of trial B5371002 and the five historical randomized, double-blind, parallel-group, placebo-controlled clinical trials of infliximab used to calculate the non-inferiority and no-superiority limits. All patients in Table 18 had active RA despite treatment with methotrexate. Important aspects of the design of the historical studies, including key inclusion/exclusion criteria, permitted concomitant medications, and baseline disease severity, were largely similar if not identical across the six studies. One notable difference was the timing of the ACR20 assessment, which ranged from week 14 to week 30. However, the ATTRACT study demonstrated large treatment effects as early as week 6 (Maini et al), and there was no apparent trend in effect size as a function of the timing of endpoint assessment across the historical studies. Estimated treatment effects with respect to ACR20 for the five historical trials were displayed earlier in Table 2. The estimated effects ranged from 18% to 38% on the absolute difference scale, with an overall estimated effect size of 28%. Thus, the information in Table 1 and Table 18 indicates that (1) the design of the five historical placebo-controlled clinical trials were largely similar to that of comparative clinical Study B5371002; and (2) there were relatively large and consistent treatment effects across the five historical studies. This evidence of historical sensitivity to effects of infliximab in similarly designed clinical trials provides some support for a conclusion that Study B5371002 had adequate assay sensitivity.

Study quality can impact the results of bioequivalence studies. 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 similarity. In the study, 4% of the patients discontinued treatment at week 14 and, on both arms, 13% discontinued treatment prior to week 30. The revised protocol amendment to continue follow-up of patients after study treatment discontinuation reduced the missing data to approximately 10% on each arm. The number of protocol deviations was similar across treatment groups and potential violations in eligibility criteria were limited. In summary, the study conduct was generally acceptable.

We also examined whether the within-group responses in the comparative clinical study were similar to those observed in previous placebo-controlled trials. The ACR20 response rate of 63% on infliximab-EU in Study B5371002 at week 14 is in line with the historical rates, which ranged from 50% to 76%.

In summary, although there were some concerns about study conduct in the submitted trial, including rates of treatment discontinuation and missing data, the study design, general conduct, potential protocol deviations, response rates for B5371002, and per protocol analyses were generally similar to the key characteristics of the five historical trials which had large consistent treatment effects of infliximab over placebo. Therefore, the totality of available information largely supports the assay sensitivity of B5371002 and the constancy assumption.

Table 18 Comparison of Key Characteristics of Historical Randomized, Placebo-Controlled Clinical Trials^a of Infliximab in RA and Comparative Clinical Study B5371002

	Maini	Westhovens	Schiff	Zhang	Abe	B5371002
Selected inclusion/exclusion criteria	SJ≥6, TJ≥6 2 of: morning stiffness ≥45 min, ESR >28 mm/h, CRP >2 mg/dL	SJ≥6, TJ≥6	SJ≥10, TJ≥12, disease duration ≥1 year, CRP≥1 mg/dL	SJ≥3, TJ≥8, 2 of: morning stiffness ≥45 min, ESR >28 mm/h, CRP>1.5 ULN	SJ≥6, TJ≥6, 2 of: morning stiffness ≥45 min, ESR >28 mm/h, CRP>2 mg/dL	SJ≥6, TJ≥6, disease diagnosis ≥4 month, CRP≥1 mg/dL,
Anti-TNF allowed?	No	No	No	Yes	No	No
Concomitant DMARDs	stable MTX	stable MTX + additional DMARDs allowed	stable MTX	stable MTX + additional DMARDs allowed	stable MTX (low-dose)	stable MTX
Region Country	NA, EU	NA, EU, AU, SA	NA, EU, AU, AF, SA	China	Japan	NA, EU (East and West), AS, LA
Baseline characteristics^b	SJ: 19; TJ: 32; disease dur: 8yrs; HAQ: 1.8	SJ: 15; TJ: 22; disease dur: 8yrs; HAQ: 1.5	SJ: 20; TJ: 32; disease dur: 7yrs; HAQ: 1.7	disease dur: 7yrs	SJ: 15; TJ: 19; disease dur: 9yrs	SJ: 16; TJ: 25; disease dur: 7yrs; HAQ: 1.6
Time of ACR20 evaluation	Week 30	Week 22	Week 28	Week 18	Week 14	Week 14
ACR20 response on infliximab	50%	55%	59%	76%	61%	64%
Withdrawal rates	11%	7%	8%	10%	5%	4% @ week 14/ 13% @ week 30

[Source: Reviewer]

^a: Based on best attempts to identify/estimate characteristics from literature review

^b: Means or medians, depending on what was reported in publication

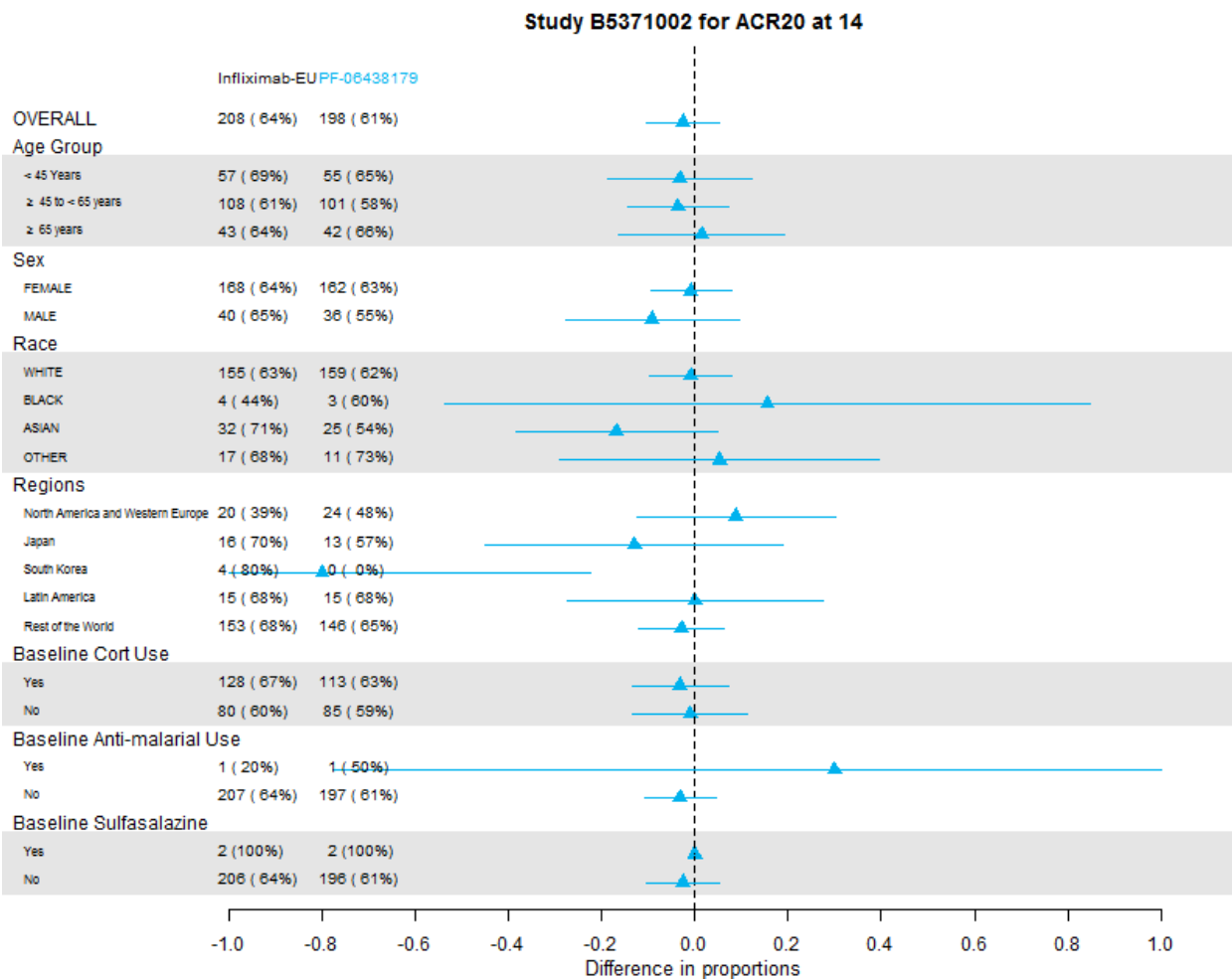
Abbreviations: ACR=American College of Rheumatology; AF=Africa; AS=Asia; AU=Australia; CRP=C-reactive protein; DMARD=disease-modifying anti-rheumatic drug; DUR=duration; ESR=erythrocyte sedimentation rate; EU=Europe; HAQ=health assessment questionnaire-disability index MTX=methotrexate; NA=North America; SA=South America; SJ=swollen joint count; TJ=tender joint count; ULN=upper limit of normal; YRS=years

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, Geographic Region, and Key Baseline RA Characteristics

There were no striking efficacy trends favoring the reference or experimental product base on subgroup analyses by sex, race, age, geographical region, as well as protocol-defined key RA baseline characteristics (Figure 7). Of note, subgroups results based on anti-malarial or sulfasalazine use were limited due to small sample size. In summary, there were no striking trends of non-similarity between products.

Figure 7 Estimated Differences Between Infliximab-EU and PF-06438179 in the Probability of ACR20 Response at Week 14 and on Randomized Treatment, Stratified by Selected Demographic Subgroups and Key RA Baseline Characteristics



[Source: Reviewer]

Differs from the applicant's analysis which was based on all patient regardless of adherence to treatment.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

- Data quality

The applicant submitted a data revision for the primary efficacy endpoint ACR20 at week 14 for a subject who were identified to have a transcription error for joint counts on October 26, 2017 (Seq. # 0022). I submitted an information request for a revised dataset for this subject change. The applicant replied on November 07, 2017 (Seq. #0025) noting additional data revisions for joint counts for five patients at week 30 (PF-06438179: (b) (6); infliximab-EU: (b) (6)). The data analyses reported in the above review were based on all the original data submitted in Seq. #0001 with the revised joint counts for subject (b) (6). Based on the additional sensitivity analyses conducted on the newly submitted datasets with revisions for all patients, submitted on November 07, 2017, I concluded that the additional data revisions for the five patients were numerically minor; they did not impact observed data distributions reported in the descriptive analyses, the sensitivity analyses, and did not affect any of the overall conclusions in this statistical review. In summary, the quality of the data was adequate.

- Margin selection and evidence of similarity

The determination of a similarity margin is a critical aspect of the design of a comparative clinical study because it determines the null hypothesis being tested in the primary analysis, i.e., the differences in efficacy that need to be ruled out at an acceptable significance level. The applicant prespecified a primary analysis to compare the 90% CI for the difference in week 14 ACR20 responses to asymmetric margins of -12% and +15% in response to feedback from FDA. Results from the primary analysis in the intent-to-treat set (90% CI: -9.1%, 3.6%) and a supportive analysis in the full analysis set (90% CI: -8.4%, 6.6%) were well contained within the prespecified asymmetric similarity bounds of [-12%, 15%]. In addition, there were similar improvements from baseline in the components of the composite primary endpoint, as well as in additional important secondary endpoints, on the two treatment arms across time during the 30-week study. Therefore, the totality of the evidence from the comparative clinical studies supports a demonstration of no clinically meaningful differences between PF-06438179 and infliximab-EU.

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

The impact of missing data on the reliability of the results was addressed in 3.2.4.4. At the prespecified primary week of analysis at week 14, a total of 4% discontinued study. The proportion of missing data was relatively small at week 14. Despite that, missing data can result in potential bias towards the alternate hypothesis of biosimilarity. In a supportive tipping point analyses, we explored the sensitivity of the results to violations of this assumption. Confidence intervals for the difference between PF-06438179 and infliximab-EU did not rule out any losses or gains of concern in efficacy under various scenarios. Given that the similar proportions of patients and distributions of reasons for early withdrawal on the two treatment arms were similar, an assumption of large differences between outcomes in dropouts on the two treatments appears

implausible. Therefore, the tipping point sensitivity analyses were supportive of the efficacy findings in B5371002.

- Assay sensitivity and constancy assumption

This issue was discussed in detail in 3.4. It is critical that a comparative clinical study has assay sensitivity, or the ability to detect meaningful differences between products, if such differences exist. In addition, the constancy assumption should be reasonable. This is the assumption that estimates of the reference product effect from historical, placebo-controlled trials are unbiased for the setting of the comparative study. Our evaluation of the literature indicated historical sensitivity to effects of infliximab over placebo in five clinical trials with similar designs to that of comparative clinical study B5371002. Within-group responses in the study were also similar to those of historical trials. It is also important that a study designed to evaluate similarity has appropriate conduct because conduct issues tend to bias results toward the alternative hypothesis of biosimilarity. There were concerns regarding increasing rates of treatment discontinuation, dose escalation of patients, and missing data. However, despite these issues, the totality of available information largely supports the sufficiency of the assay sensitivity of study B5371002, in addition to the constancy assumption.

5.2 Collective Evidence

The collective evidence from this comparative clinical study in rheumatoid arthritis supports a conclusion of no clinically meaningful differences between PF-06438179 and infliximab-EU. At week 14, 63.8% of the patients randomized to infliximab-EU and 61.1% of the patients randomized to PF-06438179 were ACR20 responders who stayed on randomized treatment, and the estimated difference on the absolute scale comparing PF-06438179 relative to infliximab-EU was -2.7% (90% exact CI: -9.1, 3.6). The limits of this 90% CI successfully ruled out the prespecified asymmetric margin of -12% and 15% agreed upon with the Agency. ACR response probabilities at other thresholds, individual components of the changes from baseline in ACR composite endpoint, and disease activity scores were similar between the treatment arms across time. There were protocol violations and deviations noted. Per-protocol analysis of the above primary endpoint yielded similar findings as the ITT analysis. Because there were amendments to the protocol to continue patient follow-up for important efficacy endpoints, missing data at the primary week of analyses were limited, and tipping point analyses largely supported the key efficacy findings. In summary, the totality of available information largely supports the assay sensitivity of the study B5371002, in addition to the constancy assumption, lending credence to the conclusion that PF-06438179 is biosimilar to infliximab-EU.

6 APPENDICES

6.1 Additional Tables and Figures

Table 19 Number of Patients in each Specific Dataset for Analysis of ACR20 at Week 14

	Infliximab-EU (N=326)		PF-06438179 (N=324)	
	n	%	n	%
ITT: NRI	326	100	324	100
ITT: Observed	324	99%	320	99%
PP: NRI	290	89%	279	86%
PP: Observed	290	89%	278	86%

[Source: Reviewer]

Observed included all patients who had any observation at week 14

Abbreviations: ITT=intent-to-treat; NRI=non-responder imputation; PP=per protocol

Table 20 Probability of ACR20 Response at Other Study Visit Weeks

	Infliximab-EU (N=326)		PF-06438179 (N=324)		Difference in Response (%)	90% CI Lower (%)	90% CI Upper (%)
	N	Count (%)	N	Count (%)			
ACR20 @ week 2							
ITT: NRI	326	121 (37.1%)	324	104 (32.1%)	-5.0	-11.2	1.2
ITT: Observed	324	121 (37.3%)	320	105 (32.9%)	-4.4	-11.8	3.1
PP: NRI	290	108 (37.6%)	279	93 (33.3%)	-4.3	-11.0	2.4
PP: Observed	290	109 (37.6%)	278	93 (33.5%)	-4.1	-10.9	2.5
ACR20 @ week 4							
ITT: NRI	326	186 (57.1%)	324	167 (51.5%)	-5.5	-12.0	1.0
ITT: Observed	321	190 (59.2%)	317	170 (53.6%)	-5.6	-13.3	2.2
PP: NRI	290	175 (60.3%)	279	151 (54.1%)	-6.2	-13.1	0.7
PP: Observed	287	175 (61.0%)	277	151 (54.5%)	-6.5	-13.3	0.5
ACR20 @ week 6							
ITT: NRI	326	198 (60.7%)	324	185 (57.1%)	-3.6	-10.0	2.8
ITT: Observed	320	201 (63.0%)	314	187 (59.7%)	-3.3	-10.9	4.4
PP: NRI	290	184 (63.4%)	279	168 (60.2%)	-3.2	-10.0	3.6
PP: Observed	289	184 (63.7%)	278	168 (60.4%)	-3.2	-10.0	3.6
ACR20 @ week 12							
ITT: NRI	326	209 (64.1%)	324	203 (62.7%)	-1.5	-7.8	4.8
ITT: Observed	318	214 (67.3%)	312	210 (67.5%)	0.2	-7.2	7.6
PP: NRI	290	199 (68.6%)	279	191 (68.5%)	-0.2	-6.7	6.3
PP: Observed	288	199 (69.1%)	278	191 (68.7%)	-0.4	-6.9	6.1
ACR20 @ week 22							
ITT: NRI	326	208 (63.8%)	324	198 (61.4%)	-2.4	-8.7	4.0
ITT: Observed	311	213 (68.5%)	301	205 (68.1%)	-0.4	-7.9	7.0
ACR20 @ week 30							
ITT: NRI	326	207 (63.5%)	324	190 (59.0%)	-4.6	-10.9	1.9
ITT: Observed	298	209 (70.1%)	294	197 (67.0%)	-3.1	-10.6	4.4

[Source: Reviewer]

Results and 90% CI were based on exact binomial score test conducted at each visit week.

PP analysis for weeks 22 and 30 were not included since the per-protocol population was defined up to week 14.

Abbreviations: ACR=American College of Rheumatology; CI=confidence intervals; ITT=intent to treat; NRI=non-responder imputation; PP=per protocol; @=at

Table 21 Listings of Patients who Died during Treatment Period 1 (+ 28 Days)

Subject ID	Treatment	Last Infusion Day	Last Infusion Dose	Death Day	AE Closest to Death	SOC/PT
(b) (6)	PF-06438179	155	5	199	199	Cardiac disorders/MI
	PF-06438179	98	5	148	148	Cardiac disorders/Acute MI
	Infliximab-EU	149	5	190	190	General disorders/Multi-organ disorder Vascular disorder/Shock
	Infliximab-EU	155 ^a	3	238	187	chronic pyelonephritis (Grade 3), chronic kidney disease (Grade 3), arterial hypertension (Grade 1), and heart failure (Grade 1) ^b

[Source: Reviewer]

The data cut-off date was June 22, 2017. Patients whose death event were prior to the cut-off date and followed up to day 238 were included.

^a: Subject did not have week 30 study treatment.

^b: Clinical study report pp 505

Abbreviations: AE=adverse events; MI=myocardial infarction; PT=preferred term; SOC=system organ class

6.2 References

- Maini, R., St Clair, E. W., Breedveld, F., Furst, D., Kalden, J., Weisman, M., et al & ATTRACT Study Group. (1999). Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomized phase III trial. *The Lancet*, 354(9194), 1932-1939.
- Westhovens, R., Yocum, D., Han, J., Berman, A., Strusberg, I., Geusens, P., & Rahman, M. U. (2006). The safety of infliximab, combined with background treatments, among patients with rheumatoid arthritis and various comorbidities: a large, randomized, placebo-controlled trial. *Arthritis & Rheumatism*, 54(4), 1075-1086.
- Schiff, M., Keiserman, M., Coddig, C., Songcharoen, S., Berman, A., Nayaiger, S., ... & Dougados, M. (2008). Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. *Annals of the rheumatic diseases*, 67(8), 1096-1103.
- Zhang, F. C., Hou, Y., Huang, F., Wu, D. H., Bao, C. D., Ni, L. Q., & Yao, C. (2006). Infliximab versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a preliminary study from China. *APLAR Journal of Rheumatology*, 9(2), 127-130.
- Abe, T., Takeuchi, T., Miyasaka, N., Hashimoto, H., Kondo, H., Ichikawa, Y., & Nagaya, I. (2006). A multicenter, double-blind, randomized, placebo controlled trial of infliximab combined with low dose methotrexate in Japanese patients with rheumatoid arthritis. *The Journal of rheumatology*, 33(1), 37-44.
- Chan, I. S. F. and Zhang, Z. "Test-Based Exact Confidence Intervals for the Difference of Two Binomial Proportions," *Biometrics*, 1999, 55:1202-1209.
- Santner, T. J. and Snell, M. K. "Small-Sample Confidence Intervals for $p_1 - p_2$ and p_1/p_2 in 2 x 2 Contingency Tables," *Journal of the American Statistical Association*, 1980, 75:386-394.

8. Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness: Guidance for Industry. 2016
9. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. E9: Statistical principles for clinical trials, 1998.
10. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat 2013; 23(6):1352-71.
11. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons 1987; p. 166-7.

6.3 Tipping Point Analysis Methodology³

The goal is to evaluate the potential effect of violations in assumptions about missing data on the reliability of conclusions. Suppose that outcomes Y are independently distributed on the control and test drug arms. The parameter of interest is the difference in means θ . Consider the following parameterization and notation to describe the probabilities of completing the study (non-missingness), the true means in completers and dropouts, and the numbers of completers and total patients on the two treatment arms:

Table 22 Parameters and Notation for Tipping Point Analysis in Presence of Missing Data

Arm	Probability of non-missing	Mean among completers	Mean among dropouts	Number of completers	Sample size per arm
Placebo	π_c	μ_c	$\mu_c + \delta_c$	N_c	n_c
Treated	π_t	μ_t	$\mu_t + \delta_t$	N_t	n_t

Given this parameterization, the target of inference is $\theta = [\pi_t \mu_t + (1 - \pi_t)(\mu_t + \delta_t)] - [\pi_c \mu_c + (1 - \pi_c)(\mu_c + \delta_c)] \equiv \mu_t + (1 - \pi_t)\delta_t - [\mu_c + (1 - \pi_c)\delta_c]$. An analysis based on completers will provide reliable inference on θ if the missing-at random assumption, i.e., the assumption that $\delta_c = \delta_t = 0$, is valid. We will perform sensitivity analyses that allow for the possibility that outcomes among dropouts are not missing-at-random by performing inference under different assumed values of the parameters δ_c and δ_t .

Denote M_{ij} to be an indicator that patient j on treatment i is a completer, i.e., his or her outcome is observed where $i = c, t$, and $j = 1, \dots, n_i$. By assuming fixed values of sensitivity parameters δ_c and δ_t , an estimator of θ can be represented by

$$\hat{\theta} = \hat{\mu}_t + (1 - \hat{\pi}_t)\delta_t - [\hat{\mu}_c + (1 - \hat{\pi}_c)\delta_c]$$

where $\hat{\mu}_i = \frac{1}{N_i} \sum_{k=1}^{n_i} Y_{ik} | M_{ik} = 1$ is the sample mean in the completers and $\hat{\pi}_i = \frac{N_i}{n_i} \equiv \sum_{k=1}^{n_i} M_{ik} / n_i$ is the sample proportion of completers on the treatment arm i , with i taking values c or t .

The test statistic can be constructed as follows:

³ Source: Gregory Levin

$$\frac{\hat{\theta} - \theta}{\sqrt{\frac{s_t^2}{N_t} + \frac{s_c^2}{N_c} + \frac{\delta_t^2 \hat{\pi}_t(1 - \hat{\pi}_t)}{n_t} + \frac{\delta_c^2 \hat{\pi}_c(1 - \hat{\pi}_c)}{n_c}}}$$

where s_i^2 is the sample variance of the outcome. Under suitable conditions, the sampling test statistic is asymptotically normal with mean 0 and standard deviation 1.

The Wald-based 100 (1- α) % confidence interval of the form $\hat{\theta} \pm z_{1-\alpha/2}$

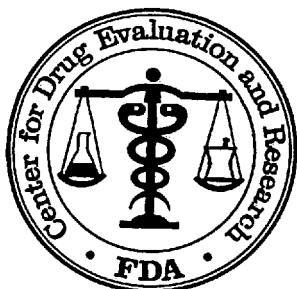
$\sqrt{\frac{s_t^2}{N_t} + \frac{s_c^2}{N_c} + \frac{\delta_t^2 \hat{\pi}_t(1 - \hat{\pi}_t)}{n_t} + \frac{\delta_c^2 \hat{\pi}_c(1 - \hat{\pi}_c)}{n_c}}$ can be constructed where z_q is the q quantile of the standard normal distribution.

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

/s/

WILLIAM J KOH
11/09/2017

ROBERT ABUGOV
11/09/2017



STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA No.:	761072
SERIAL No.:	0000
DATE RECEIVED BY THE CENTER:	February 13, 2017
DRUG NAME:	IXIFI (proposed biosimilar to US-licensed Remicade)
DOSAGE FORM:	100 mg lyophilized powder for infusion
INDICATIONS:	Same indications of use as approved for US-licensed Remicade
APPLICANT:	Pfizer Inc.
REVIEW FINISHED:	October 31, 2017
NAME OF PRIMARY STATISTICAL REVIEWER:	Yu-Ting Weng
NAME OF SECONDARY STATISTICAL REVIEWER:	Meiyu Shen
NAME OF PROJECT MANAGER:	Sadaf Nabavian

Reviewer: Yu-Ting Weng, Ph.D., Mathematical Statistician, CDER/OTS/OB/DB VI

Secondary reviewer: Meiyu Shen, Ph.D., Team Leader, CDER/OTS/OB/DB VI

Concur:

Meiyu Shen, Ph.D., Team Leader, CDER/OTS/OB/DB VI

Distribution: CDER/OTS/OB/DB VI/Yi Tsong
CDER/OTS/OB/DB VI/Meiyu Shen
CDER/OTS/OB/Lillian Patrician
CDER/OBP/Yanming An
CDER/OBP/Christopher Downey
CDER/OMPT/CDER/OND/ ODEII/DPARP/Sadaf Nabavian
CDER/TBBS

TABLE OF CONTENT

1	<i>Executive summary and recommendation.....</i>	4
2	<i>Introduction.....</i>	5
3	<i>Applicant's statistical equivalence testing.....</i>	6
4	<i>FDA statistical analyses.....</i>	6
4.1	Data analyzed.....	7
4.2	Statistical method	8
4.3	FDA statistical equivalence testing for relative potency assay	9
4.4	FDA statistical equivalence testing for sTNF- α binding assay	10
5	<i>Conclusion and recommendation.....</i>	12
6	<i>Reference.....</i>	12

List of Tables

Table 1. Results of Equivalence Testing for the Relative Potency Assay	4
Table 2. Results of Equivalence Testing for the sTNF- α Binding Assay	4
Table 3. Summary of PF-06438179 Lots for Tier 1 Analysis	7
Table 4. Descriptive Statistics for the Relative Potency Assay Data	10
Table 5. Results of Equivalence Testing for the Relative Potency Assay	10
Table 6. Descriptive Statistics for the sTNF- α Binding Assay Data	11
Table 7. Results of Equivalence Testing for the sTNF- α Binding Assay	11

List of Figures

Figure 1. Scatter Plot of the Relative Potency Assay for PF-06438179, US-licensed Remicade, and EU-approved Remicade	9
Figure 2. Scatter Plot of the sTNF- α Binding Assay for PF-06438179, US-licensed Remicade, and EU-approved Remicade	11

1 EXECUTIVE SUMMARY AND RECOMMENDATION

The CMC statistical reviewer in the Office of Biostatistics analyzed the comparative results of two critical Quality Attributes (QAs): the inhibition of apoptosis potency assay and the sTNF- α binding assay, which were recommended for equivalence testing analysis by the Office of Biotechnology Products (OBP). Tier 1 statistical equivalence testing was conducted using equivalence margins of $\pm 1.5 \sigma_R$, where R represents the US-licensed reference product variability or the comparator variability.

Eleven batches of PF-06438179 (PFE), 53 batches of US-licensed Remicade (US), and 60 batches of EU-approved Remicade (EU) were used for equivalence testing of apoptosis inhibition (relative potency) assay. The results are summarized in Table 1.

Table 1. Results of Equivalence Testing for the Relative Potency Assay

Test Product (Number of lots)	Reference Product (Number of lots)	Mean Difference, %		Equivalence Margin, %	Statistical Equivalence
		Estimate	90% CI ¹		
PFE ² (11)	US ² (53)	-1.93	(-5.92, 2.06)	(-11.68, 11.68)	Pass
PFE ² (11)	EU ² (60)	-3.73	(-7.88, 0.42)	(-12.46, 12.46)	Pass
EU ² (60)	US ² (53)	1.80	(-0.71, 4.31)	(-11.68, 11.68)	Pass

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

2. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

Eleven PFE batches, 11 US batches, and 11 EU batches were included in the sTNF- α binding dataset for the statistical equivalence testing. The results are summarized in Table 2.

Table 2. Results of Equivalence Testing for the sTNF- α Binding Assay

Test Product (Number of lots)	Reference Product (Number of lots)	Mean difference, %		Equivalence Margin, %	Statistical Equivalence
		Estimate	90% CI		
PFE ² (11)	US ² (11 ¹)	-0.82	(-5.04, 3.40)	(-6.13, 6.13)	Pass
PFE ² (11)	EU ² (11 ¹)	0.82	(-3.40, 5.04)	(-6.16, 6.16)	Pass
EU ² (11 ¹)	US ² (11 ¹)	-1.64	(-4.65, 1.38)	(-6.13, 6.13)	Pass

1. Only include those 11 selected lots of all provided US-licensed and EU-approved lots for sTNF- α binding

2. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

As shown in Tables 1 and 2, the results from statistical equivalence testing of the relative potency assay and the sTNF- α binding assay support a demonstration that the proposed

biosimilar PFE is highly similar to US and also support the analytical portion of the scientific bridge to justify the relevance of EU data from the comparative clinical study.

2 INTRODUCTION

On February 13, 2017, Pfizer (the applicant) submitted to the US Food and Drug Administration (FDA) a 351(k) BLA which included analytical similarity assessments of comparing PFE, US and EU.

Two Tier 1 QAs are the relative potency assay and the sTNF- α binding assay. Two review issues are the selection of a subset of all provided US and EU lots for testing sTNF- α binding and different lot selection for testing different Tier 1 QAs.

On May 9, 2017, the FDA sent the first Information Request (IR) to the applicant for the clarification. This IR included the following questions.

“Question 1. For the analytical similarity assessment, you tested sTNF- α binding for a subset of the US-licensed Remicade and EU-approved Remicade lots that were tested for other attributes (e.g. apoptosis). Provide the justification for the lot selection of US-licensed Remicade and EU-approved Remicade for testing of sTNF- α binding. In addition, to enable a more rigorous statistical comparison to PF-06438179, provide sTNF- α binding data from all additional of US-licensed and EU-approved Remicade lots that may be available.

Question 2. The analytical similarity assessment includes data from Drug Product (DP) lots Z09719, A03038, and A02637 for the apoptosis assay, whereas data from the respective constituent Drug Substance (DS) lots 94000, 94001, and 94002 are included for the sTNF- α binding assay. Provide your rationale for testing these three DS lots for the sTNF- α binding assay but testing the DP lots manufactured from these DS lots for the apoptosis assay.”

On May 22, 2017, the applicant provided the following clarifications:

- No additional US and EU lots have been tested in the sTNF- α binding assay.
- The FDA agreed this approach in a subsequent post-meeting communication on Jan 11, 2017.

Because sTNF- α binding assay failed equivalence testing at 0.05 significance level (recommended by the FDA) and the number of the proposed biosimilar lots was seven in the first-time analysis, the applicant proposed to use a larger significance level (b) (4) for equivalence testing. The CMC statistical team provided the following comments in the mid-cycle meeting on July 17, 2017.

- (b) (4) % confidence level is not acceptable when the number of the proposed biosimilar lots is less than or equal to ten unless the product is for rare diseases.
- For sTNF- α binding, the small mean differences between the proposed biosimilar and the reference products but the larger variability of the proposed biosimilar product might be from a small number of the selected lots. Incorporation of additional lots of all three products in the analytical similarity assessment might address these concerns.

On Aug 9, 2017, the FDA sent the second IR to the applicant for the clarification.

“Question. Both Tier 1 analytical similarity assessments include data from only 7 lots of PF-06438179, and the soluble TNF (sTNF) binding assay includes data from 11 lots each of US-licensed and EU-approved Remicade. The data for the sTNF binding assay fail to meet Tier 1 statistical equivalence criteria for both the PF-06438179 versus US-licensed Remicade and PF-06438179 versus EU-approved Remicade comparisons. The standard deviation of the PF-06438179 sTNF binding data is twice as large as the standard deviation of US-licensed Remicade and EU-approved Remicade data. To enable more precise standard deviation estimates and to support a more rigorous statistical comparison between PF-06438179, US-licensed Remicade, and EU-approved Remicade, provide data for both Tier 1 assays and for all lot release tests for all independent lots of PF- 06438179 manufactured to date, including lots manufactured since submission of the BLA. Additionally, provide data for the Tier 1 assays for any additional untested US- licensed Remicade and EU-approved Remicade lots that you may have available or can purchase. Repeat equivalence testing using a 90% confidence level for all three pairwise comparisons for both attributes assigned to Tier 1 testing and report the results by September 22, 2017.”

On September 18, 2017, the applicant provided the following information:

- Four additional DS batches manufactured at the (b) (4) manufacturing facility since the submission of the BLA.
- Repeated equivalence testing using a 90% confidence level for all three pairwise comparisons for both Tier 1 quality attributes.

The CMC statistical reviewer carefully evaluated data for the relative potency assay and the sTNF- α binding assay provided in the initial BLA submission and the applicant’s responses to the FDA’s information requests. The applicant’s Tier 1 statistical equivalence testing is provided in Section 3, and the CMC statistical reviewer’s independent statistical equivalence testing analyses are present in Section 4.

3 APPLICANT’S STATISTICAL EQUIVALENCE TESTING

In this submission, the applicant followed the FDA’s recommendation to conduct Tier 1 statistical equivalence testing with the margin defined as $1.5\hat{\sigma}_R$, where $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots, for relative potency assay and sTNF- α binding assay. Hence, the applicant’s analyses are valid.

The CMC statistical reviewer performs an independent statistical analysis to confirm the applicant’s analyses in the next section.

4 FDA STATISTICAL ANALYSES

To evaluate analytical similarity, the FDA recommended the applicant to apply a tiered approach in the FDA responses to IND meetings with the applicant. That is, product QAs amenable to statistical evaluation are assigned to three tiers based on their criticality. The quality attributes with potential highest risk in product quality, efficiency, safety, and PK/PD are generally assigned to Tier 1, in which analytical similarity is assessed by statistical equivalence

test. QAs with lower impact are generally assigned to Tier 2 and their analytical similarity is evaluated by Quality Range approach. That is, a high percentage of the biosimilar data should be covered by $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is the sample mean, $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots, and the multiplier X typically ranges from 2 to 4. The QAs with the lowest risk are generally assigned to Tier 3 and their analytical similarity is evaluated by side-by-side comparison using graphic display. More details are described in the FDA Draft Guidance on Analytical Similarity (2017)

However, this review focuses on the Tier 1 statistical equivalence testing.

4.1 Data analyzed

The applicant submitted the analytical data on February 13, 2017. Then, on September 18, 2017, the applicant provided four additional DS batches manufactured at the (b) (4) manufacturing facility since the submission of the BLA. The summary of PFE lots for Tier 1 analysis is provided in Table 3. Although the applicant selected different lots for different Tier 1 QAs, the OBP CMC reviewers agreed this approach in a subsequent post-meeting communication on Jan 11, 2017. Thus, the CMC statistical reviewer conducted Tier 1 statistical equivalence testing based on these 11 selected independent lots for each QA.

Table 3. Summary of PF-06438179 Lots for Tier 1 Analysis

DP lot number	DS lot number	Tested lots for Tier 1 QAs	
		Relative Potency	sTNF- α binding
	94200	X	X
	94000		X
	94001		X
	94002		X
Z09719	94000	X	
A03038	94001	X	
A02637	94002	X	
L07814	94002		
	94003		
	94004		
	94005		
M01013	94003	X	X
M29852	94004	X	X
M29851	94005	X	X
M29853	94003, 94004, 94005		
Four additional DS batches since the submission of the BLA			
	19200000	X	X
	19200001	X	X
	19210000	X	X
	19210001	X	X

X: tested independent lots included in the analysis

4.2 Statistical method

Let μ_T and μ_R be the population mean of the QA for the test product and the population mean of the QA for the reference product, respectively. Let σ_R be the standard deviation of the QA of interest for the reference product. To conclude the equivalence in the QA of interest between the test product and the reference product, we aim to reject the null hypothesis of the following null and alternative hypotheses:

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

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

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

We reject H_0 if 90% Confidence Interval (CI) for the mean difference in the QA of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. In other words, we conclude that the equivalence in the QA of interest between the test product and the reference product if 90% CI for the mean difference in the QA of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. This specific equivalence margin was set as 1.5 times the standard deviation of the quality attribute for the reference product to ensure an adequate power for the case in which a small but sufficient number of lots are available for testing. For example, the probability of rejecting H_0 in the above two one-sided tests procedure with the equivalence margin being $\pm 1.5\sigma_R$ is 87% if the true mean difference is $0.125\sigma_R$ for a sample size of 10 test product lots and 10 reference product lots with equal variability.

First, we estimate σ_R by the sample variability of the reference product, and then θ_1 and θ_2 are treated as a constant but not a random variable in the statistical analysis.

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

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

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

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

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

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

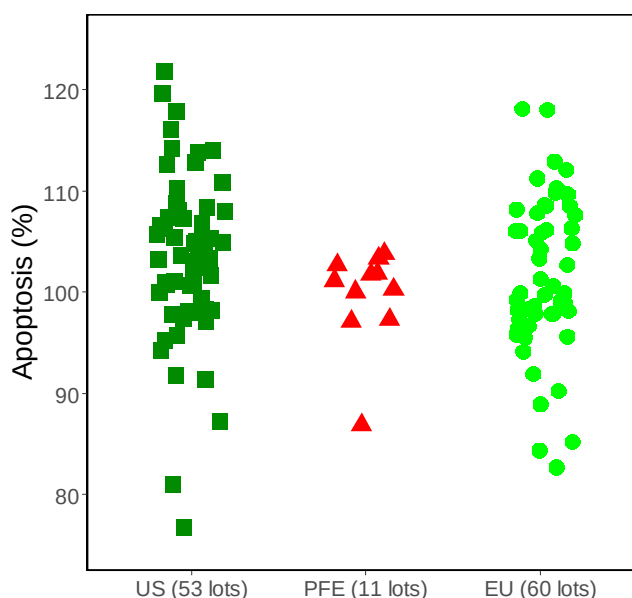
$$\text{where } n_R^* = \min(n_R, 1.5n_T) \text{ and } v^* = \frac{\left(\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R^*} \right)^2}{\frac{1}{n_T - 1} \left(\frac{S_T^2}{n_T} \right)^2 + \frac{1}{n_R - 1} \left(\frac{S_R^2}{n_R^*} \right)^2}.$$

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

4.3 FDA statistical equivalence testing for relative potency assay

The relative potency assay's data points of PFE, US, and EU are displayed in Figure 1. The sample mean of PFE is smaller than the sample means of US and EU. The sample variability of PFE is smaller than the sample variabilities of US and EU.

Figure 1. Scatter Plot of the Relative Potency Assay for PF-06438179, US-licensed Remicade, and EU-approved Remicade



Eleven PFE lots, 53 US lots, and 60 EU lots were included for statistical equivalence testing for the relative potency assay. Descriptive statistics for the relative potency assay's data of three products are listed in Table 4.

Table 4. Descriptive Statistics for the Relative Potency Assay Data

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
PFE ¹	11	99.64	4.78	87	104
US ¹	53	101.57	7.79	83	118
EU ¹	60	103.37	8.30	77	122

1. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

In Table 5, the result shows that the three-way pairwise comparisons for the relative potency assay pass equivalence testing.

Table 5. Results of Equivalence Testing for the Relative Potency Assay

Test Product (Number of lots)	Reference Product (Number of lots)	Mean Difference, %		Equivalence Margin, %	Statistical Equivalence
		Estimate	90% CI ¹		
PFE ² (11)	US ² (53)	-1.93	(-5.92, 2.06)	(-11.68, 11.68)	Pass
PFE ² (11)	EU ² (60)	-3.73	(-7.88, 0.42)	(-12.46, 12.46)	Pass
EU ² (60)	US ² (53)	1.80	(-0.71, 4.31)	(-11.68, 11.68)	Pass

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

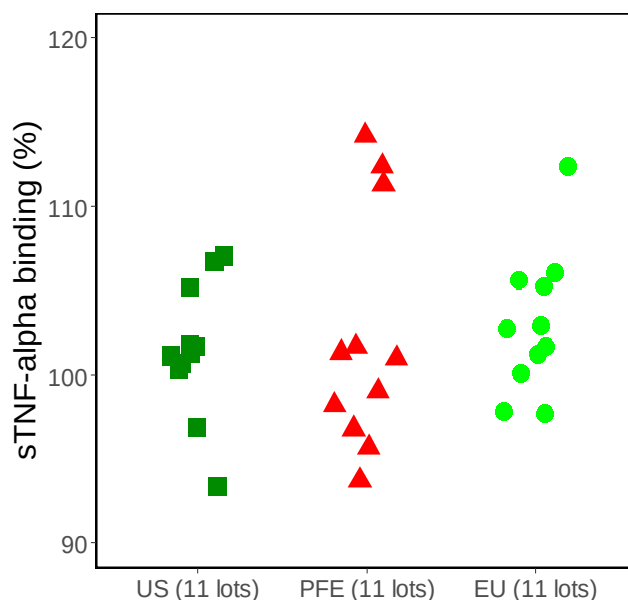
2. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

4.4 FDA statistical equivalence testing for sTNF- α binding assay

The sTNF- α binding assay's data points of PFE, US, and EU are displayed in Figure 2. There appears a small sample mean difference among three products. The sample variability of PFE is larger than the sample variabilities of US and EU.

Eleven PFE lots, 11 US lots, and 11 EU lots were included in the sTNF- α binding assay's dataset for statistical equivalence testing. Descriptive statistics for the sTNF- α binding assay's data of three products are listed in Table 6.

In Table 7, the result shows that the three-way pairwise comparisons for the sTNF- α binding assay pass equivalence testing.

Figure 2. Scatter Plot of the sTNF- α Binding Assay for PF-06438179, US-licensed Remicade, and EU-approved Remicade**Table 6. Descriptive Statistics for the sTNF- α Binding Assay Data**

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Minimum, %	Maximum, %
PFE ²	11	102.27	6.90	94	114
US ²	11 ¹	103.09	4.09	98	112
EU ²	11 ¹	101.45	4.11	93	107

1. Only include those 11 selected lots of all provided US-licensed and EU-approved lots for sTNF- α binding

2. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

Table 7. Results of Equivalence Testing for the sTNF- α Binding Assay

Test Product (Number of lots)	Reference Product (Number of lots)	Mean difference, %		Equivalence Margin, %	Statistical Equivalence
		Estimate	90% CI		
PFE ² (11)	US ² (11 ¹)	-0.82	(-5.04, 3.40)	(-6.13, 6.13)	Pass
PFE ² (11)	EU ² (11 ¹)	0.82	(-3.40, 5.04)	(-6.16, 6.16)	Pass
EU ² (11 ¹)	US ² (11 ¹)	-1.64	(-4.65, 1.38)	(-6.13, 6.13)	Pass

1. Only include those 11 selected lots of all provided US-licensed and EU-approved lots for sTNF- α binding

2. PFE = PF-06438179; US = US-licensed Remicade; EU = EU-approved Remicade

5 CONCLUSION AND RECOMMENDATION

The results from the statistical equivalence analyses for the relative potency assay and the sTNF- α binding assay support a demonstration that the proposed biosimilar PFE is highly similar to US. In addition, the results support the analytical portion of the scientific bridge to justify the relevance of EU data from the comparative clinical study.

6 REFERENCE

Statistical Approaches to Evaluate Analytical Similarity Guidance for Industry (2017)
<https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM576786.pdf>

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

/s/

YU-TING WENG
11/09/2017

MEIYU SHEN
11/09/2017