Drug Design, Development and Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

A critical review of the United States regulatory pathways for determining the equivalence of efficacy between CT-P13 and original infliximab (Remicade®)

This article was published in the following Dove Press journal:
Drug Design, Development and Therapy

Soohyun Kim[1,2,*]
Siun Kim[1,2,*]
Howard Lee[1–4]

[1]Department of Transdisciplinary Studies, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea; [2]Center for Convergence Approaches in Drug Development, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea; [3]Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, Seoul, Korea; [4]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea

*These authors contributed equally to this work

Abstract: We evaluated the appropriateness of various equivalence margins for CT-P13, an infliximab biosimilar, in the PLANETRA clinical trial. The 95–95% method was used to independently determine an equivalence margin by pooling the historical clinical trials with original infliximab versus placebo, identified in a systematic literature search. The constancy assumption with the PLANETRA trial was assessed for each identified historical clinical trial to decide which study was scientifically justifiable to be pooled. A sensitivity analysis was performed for each study-pooling scenario. As a result, we identified two historical clinical trials that were deemed appropriate, whereas the PLANETRA trial pooled three additional studies to determine an equivalence margin, which was accepted by the United States Food and Drug Administration. However, those extra clinical trials did not meet the constancy assumption in baseline characteristics, methotrexate dose, and efficacy assessment time. The clinically more appropriate equivalence margin was 5.7 percentage points, which was much narrower than the 12 percentage points applied in the approval of CT-P13. In conclusion, the equivalence claim for CT-P13 to original infliximab in patients with rheumatoid arthritis did not appear to be supported when the constancy assumption was strictly assessed. The equivalence margin for biosimilars could be determined more conservatively.

Keywords: biosimilar, equivalence margin, rheumatoid arthritis, infliximab, CT-P13

Introduction

CT-P13 (Celltrion Inc., South Korea) is a follow-on biological product claimed to be biosimilar to original infliximab (Remicade®), the original reference product manufactured by Centocor. Equivalence between CT-P13 and the original infliximab was concluded in the PLANETRA trial, a randomized, double-blind, parallel study of CT-P13 in patients with rheumatoid arthritis (RA) who had poorly responded to methotrexate (NCT01217086).[1] The equivalence conclusion was made because the 95% confidence interval of the difference in the American College of Rheumatology 20% response rate (ACR20) between the two drugs, assessed after treatment for 30 weeks, fell within the equivalence margin of 15 percentage points or [−15,15].[1] However, the equivalence margin in the PLANETRA trial was voluntarily lowered to 13 percentage points by Celltrion, which was further reduced to 12 percentage points by the United States Food and Drug Administration (FDA) in the approval process of CT-P13.[2,3] Afterwards, the

Correspondence: Howard Lee
Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, 103 Daehak-Ro, Jongno-Gu, Seoul 110-799, Korea
Tel +82 2-740-7602
Fax +82 31-888-9575
Email howardlee@snu.ac.kr

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/DDDT.S254776

same equivalence margin of 12 percentage points has been consistently recommended by the FDA in the application of other infliximab biosimilars including SB2 (Renflexis ®, Samsung Bioepis Co., Ltd.), PF-06438179 (Ixifi ®, Pfizer, Inc.), and ABP 710 (Avsola ®, Amgen, Inc.).[4,5]

As previously indicated,[6] the equivalence margin in clinical trials with biosimilars could be determined by applying the same principle proposed by the FDA to set the non-inferiority margin.[7] Based on this understanding, the objective of this study was to evaluate if the equivalence margins chosen by Celltrion and the FDA for the first infliximab biosimilar CT-P13 were scientifically valid and clinically justifiable. To this end, the study design and baseline characteristics of the participants in the historical clinical trials with original infliximab were compared with those of the equivalence trial with CT-P13. Furthermore, an equivalence margin was independently determined for the infliximab biosimilar in RA. Then, a sensitivity analysis was performed to show why the equivalence margins previously chosen for CT-P13 could be inappropriate, i.e., too forgiving.

Materials and Methods

Determination of the Equivalence Margin

The equivalence clinical trial shares some features in common with the non-inferiority clinical trial. An equivalence clinical trial is performed to show that the efficacy of the test drug is not different from that of a reference drug while a non-inferiority clinical trial has to show that the efficacy of the test drug is not worse than that of a reference drug, in each of which the difference or inferiority should be smaller than an amount that is clinically unimportant.[8,9] In these trials, the largest amount of the clinically acceptable difference or inferiority is called the equivalence or non-inferiority margin, respectively.

The 95–95% method has been recommended by the FDA as the main approach to determine the non-inferiority margin and to test if the test drug is non-inferior to a reference drug.[7] The first 95% refers to the 95% confidence interval for the placebo-adjusted efficacy of a reference drug, which can be estimated by pooling the historical studies with the reference drug. Then, two fractions, M_1 and M_2, are applied to derive the non-inferiority margin. M_1 is the entire effect of a reference drug, which is assumed to be fully present in a non-inferiority study (i.e., assay sensitivity). Typically, the lower 95% confidence bound of the pooled estimate for the placebo-adjusted efficacy is used as M_1, which minimizes the risk to erroneously conclude the non-inferiority of the test drug that is actually inferior to a reference drug. On the other hand, M_2 is the largest loss of efficacy for the test drug that is clinically acceptable compared with a reference drug. M_2 is clinically determined as a fraction of M_1, and 50% is frequently used for M_2. The non-inferiority margin is determined as M_2.

The second 95% in the 95–95% method denotes the 95% confidence interval constructed using the results of a non-inferiority study. If the 95% confidence interval of the difference in efficacy between the test drug and a reference drug (i.e., test–reference) stays above the non-inferiority margin, non-inferiority of the test drug can be declared.

In the equivalence trial with a biosimilar, the same 95–95% method can be used to determine an equivalence margin and to assess if the biosimilar is equivalent to a reference biologic.[6] The only difference is that the 95% confidence interval of the difference in efficacy between the biosimilar and a reference biologic, derived in the equivalence study, should be entirely contained within the equivalence margin to conclude their equivalence (Figure 1).

Literature Search

We performed a systematic literature search to identify the historical clinical trials that could be pooled to determine an equivalence margin for CT-P13 in comparison with its reference biologic, Remicade®. Two meta-analysis papers that reviewed the efficacy of TNF-α inhibitors including original infliximab[10,11] were used as a starting point for the literature search. We identified the clinical trials that met all of the following criteria in order to mimic the design of the PLANETRA trial. First, clinical trials had to be carried out using a randomized, double-blind, parallel design in patients with RA. Clinical trials conducted in patients with early RA were excluded. Second, original infliximab 3 mg/kg was administered intravenously every 8 weeks as an add-on intervention to methotrexate, and the combination of placebo and methotrexate was used as control. Third, efficacy was assessed after treatment for 30 ± 4 weeks. Fourth, ACR20 was an efficacy endpoint though it did not need to be primary. Lastly, the manuscript was written in English and had to be available for review.

We also identified additional historical clinical trials that Celltrion and the FDA included when estimating an equivalence margin for CT-P13.[1,3] However, these clinical trials[12–14]

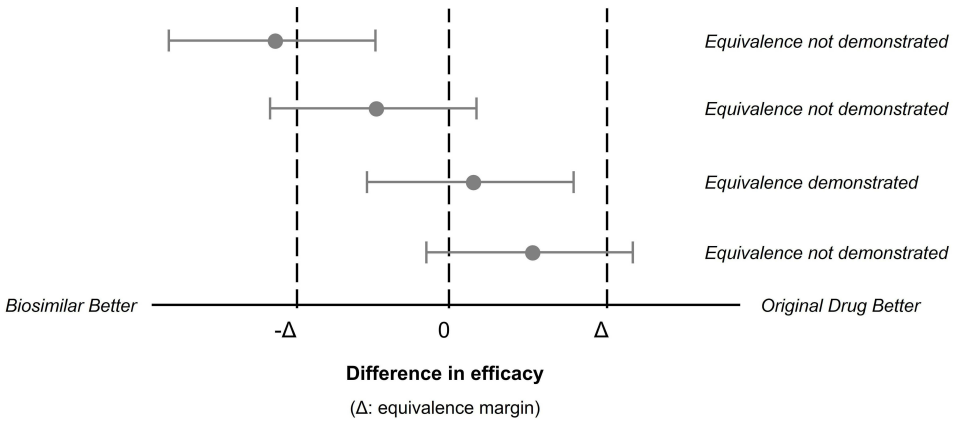

Figure 1 Illustration of equivalence in efficacy, or lack thereof, between original drug and its biosimilar product using the 95–95% method. Equivalence can be demonstrated when the 95% CI of difference between original drug and a biosimilar falls entirely within the range of [- Δ, Δ], where Δ is an equivalence margin. The solid circles denote the point estimate of the difference in efficacy between original drug and its biosimilar.

did not meet all of the literature search criteria set in the present study, particularly for assessment time, i.e., 30 ± 4 weeks.

Evaluation of Constancy Assumption

We evaluated if the historical clinical trials with original infliximab identified in the previous section were similar enough to the PLANETRA trial so that they could be pooled together to estimate M_1 (i.e., the constancy assumption).[7] To this end, we compared the following study characteristics between the PLANETRA trial and each of the identified clinical trials: patient population and disposition, main eligibility criteria, primary efficacy endpoints and their times of assessment, infliximab dose, and concomitant medications and their doses. The baseline characteristics of the study participants were also examined. Based on the constancy assumption evaluation, several pooling scenarios were constructed to estimate M_1.

Furthermore, we evaluated if the placebo-adjusted ACR20 by the original infliximab seen at an earlier assessment time such as 8 weeks after treatment remained stable beyond that time. To do so, ACR20s at various assessment times were digitized from the graphs reported in the literature by using the GetData Graph Digitizer (version 2.26, http://getdata-graph-digitizer.com). ACR20 was then regressed on assessment time so see if the slope of a regression line was insignificant (i.e., remained stable).

Estimation of the Equivalence Margins for CT-P13 by Study-Pooling Scenario

Using the meta-analysis library meta in R (version 3.5.1),[15] M_1 was determined as the lower bound of the 95% confidence interval for the pooled estimate of the placebo-adjusted ACR20 of original infliximab by the study-pooling scenario. The inverse variance and the DerSimonian and Laird's methods were used to combine multiple studies for fixed- and random-effects models, respectively, while the random-effects model was used to account for heterogeneity between studies.[16] Furthermore, 50% was used for M_2. The intention-to-treat (ITT) population was used to determine the sample size of each study to increase data availability.

Results

Literature Search

Two historical clinical trials were identified that met the search criteria: Maini et al[17] and Schiff et al.[18] In contrast, Celltrion and the FDA pooled three additional placebo-controlled clinical trials besides the two studies we identified to determine the equivalence margin for CT-P13 to the original infliximab: Westhovens et al.[12] Abe et al[13] and Zhang et al.[14]

Evaluation of Constancy Assumption

Not all of the five historical clinical trials met the constancy assumption, particularly the three additional clinical trials that Celltrion and FDA included in the following aspects: baseline characteristics, methotrexate (MTX) dose, and efficacy assessment time (Table 1). First, the study by Abe et al enrolled slightly milder patients based on the finding that the average number of tender joints was smaller than that in the other trials (19 vs 22–32).[13] Moreover, Zhang et al did not report the RA severity of the patients at baseline such as tender joint counts, swollen joint counts, HAQ score, and serum CRP.[14] Second,

Table 1 Study Design and Baseline Characteristics of Subjects in Placebo-Controlled Original Infliximab (Remicade) Trials vs The PLANETRA Study

Characteristics	Study					
	Maini et al.[17]	**Schiff et al.**[18]	**Westhovens et al.**[12]	**Abe et al.**[13]	**Zhang et al.**[14]	**PLANETRA**[1]
# of patients	428	431	1084	147	173	606
Study design	DB RCT	DB RCT	DB RCT, conditional dose increase after week 22	DB RCT, open-label extension after week 14	DB RCT	DB RCT
Treatment and patient disposition (# of patients)	P + MTX (n=88), INX (3mg/kg, 8wk) + MTX (n=86), INX (10mg/kg, 8wk) + MTX (n=86), INX (3mg/kg, 4wk) + MTX (n=87), INX (10mg/kg, 4wk) + MTX (n=81)	P + MTX (n=110), ABT (10mg/kg, 4wk) + MTX (n=156), INX (3mg/kg, 8wk) + MTX (n=165)	P + MTX (n=363), INX (3mg/kg, 8wk) + MTX (n=360), INX (10mg/kg, 8wk) + MTX (n=361)	P + MTX (n=47), INX (3mg/kg, 8wk) + MTX (n=49), INX (10mg/kg, 8wk) + MTX (n=51)	P + MTX (n=86), INX (3mg/kg, 8wk) + MTX (n=87)	INX (3mg/kg, 8wk) + MTX (n=304), CT-P13 (3mg/kg, 8wk) + MTX (n=302)
Eligibility criteria	≥6 SJCs, ≥6 TJCs despite receiving MTX	≥10 SJCs, ≥12 TJCs, CRP levels≥1mg/dl	≥6 SJCs, ≥6 TJCs despite receiving MTX	≥6 SJCs, ≥6TJCs despite MTX therapy	≥3 SJCs, ≥8 TJCs despite treatment with MTX	≥6 SJCs, ≥6TJCs despite MTX therapy
Primary endpoint	ACR20	DAS28	ACR20	ACR20	ACR20	ACR20
Primary ACR20 assessment time (wk)	30	28	22	14	18	30
Other ACR20 assessment times (wk)	2,4,6,12 and 22	Every 8 wks until wk 52	54	2, 6 and 10 (open-label extension: every 4 wks until wk 36)	2	14 and 52
Secondary endpoints	SJC, TJC, HAQ-DI, CRP, DAS28	ACR20, HAQ-DI, SF-36	DAS28, the presence of ATI or ANAs	N/A	SJC, TJC, HAQ-DI, CRP, ESR, duration of morning stiffness	SJC, TJC, HAQ-DI, ESR, CRP, DAS28, EULAR, CDAI, SDAI, SF-36
Allowed concomitant drugs excluding MTX	Other DMARs, corticosteroids and NSAIDs	Oral corticosteroids and NSAIDs	Study-approved DMARDs, oral corticosteroids and NSAIDs	None	N/A	Oral glucocorticoids and NSAIDs.
Age (year)	56 (25–74)†	49.1 ± 12.0	53.0 (45–61)	55.2 ± 10.9	47.9 ± 10.1	50 (21–74)†
Females (%)	81.0	82.4	80.0	81.6	85.1	84.2
Disease duration (yr.)	8.4 (0.7–45.0)†	7.3 ± 6.2	7.8 (3–15)	9.1 ± 7.4	7.1±6.2	N/A
MTX dose (mg/week)	15 (12.5–17.5)	16.3 ± 3.6	15.0 (10–18)	7.1 ± 1.9	N/A	15.6 ± 3.2
No. tender joints	32 (16–46)	31.7 ± 14.5	22 (15–31)	19.0 ± 11.8	N/A	24.0 ± 12.9

(*Continued*)

Table 1 (Continued).

Characteristics	Study					
	Maini et al.[17]	Schiff et al.[18]	Westhovens et al.[12]	Abe et al.[13]	Zhang et al.[14]	PLANETRA[1]
No. swollen joints	19 (13–30)	20.3 ± 8.0	15 (11–21)	15.1 ± 9.0	N/A	15.2 ± 8.3
HAQ score	1.8 (1.4–2.3)	1.7 ± 0.7	1.5 (1–2)	N/A	N/A	1.6 ± 0.6
Serum CRP (mg/dL)	3.1 (1.3–5.3)	3.3 ± 3.2	1.6 (1–3)	N/A	N/A	1.9±2.2

Notes: Values are "mean±SD" or "mean (interquartile range)" except where indicated otherwise. †mean (range).
Abbreviations: DB, double blind; RCT, randomized controlled trial; RA, rheumatoid arthritis; P, placebo; SJC, swollen joint count; TJC, tender joint count; ABT, abatacept; INX, infliximab; MTX, methotrexate; ACR20, the American College of Rheumatology 20% improvement criteria; HAQ-DI, Health Assessment Questionnaire-Disability Index; DAS28, Disease Activity Score 28; ATI, antibody to infliximab; ANA, antinuclear antibodies; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatism; CDAI, Clinical Disease Activity Index; SDAI, Simplified Disease Activity Index; SF-36, Short Form 36; DMARDs, disease-modifying antirheumatic drugs; NSAIDs, non-steroidal anti-inflammatory drugs; N/A, not available; MTX, methotrexate; N/A, not available; HAQ, Health Assessment Questionnaire; CRP, C-reactive protein; wk, week; wks, weeks.

a much smaller MTX dose was administered in Abe et al[13] (7.1 vs ~15 mg/week). Again, no information was available for the MTX dose in Zhang et al.[14] Finally, ACR20 was assessed at week 30 ± 4 after treatment only in Maini et al[17] and Schiff et al[18] as was in the PLANETRA trial. However, ACR20 was determined much earlier in the additional three clinical trials, ranging from 14 to 22 weeks. In a regression analysis of ACR20 on assessment time >4 weeks, the slope was significantly different from 0 for original infliximab, which was −0.31 (95% confidence interval (CI) [−0.44, −0.18], p-value = 0.00031) in Schiff et al[18] and 0.47 (95% CI [0.09, 0.84], p-value = 0.02) in Maini et al[17] (Figure 2). In contrast, the slope for placebo was 0.14, which was not statistically significant (95% CI [−0.03, 0.32], p-value = 0.09, Figure 2) in Maini et al.[17]

In all of the five historical clinical trials, original infliximab was significantly better than placebo to treat RA in terms of ACR20 (all p-values <0.01, Table 2). However, the placebo-adjusted original infliximab efficacy varied widely by study, ranging from 17.6% (Schiff et al) to 37.8% (Abe et al)[13,18]

Study-Pooling Scenarios to Estimate the Equivalence Margin for CT-P13

Six scenarios were constructed as to which studies to be pooled out of the five historical clinical trials to estimate M_1 for CT-P13 (Table 3). The first three scenarios (i.e., 1–3) recapitulated what Celltrion and the FDA used. No study was excluded in scenarios 1 or 3, each of which was adopted by Celltrion in the PLANETRA trial[1] and by the FDA in the approval decision for CT-P13,[19] respectively. Celltrion used the point estimate of the placebo-adjusted ACR20 as M_1 (scenario 1), while the FDA *appropriately* determined M_1 as the lower bound of the 95% confidence interval (scenario 3).

Scenarios 4–6 reflected the results of the constancy assumption evaluation in the previous section. The study by Zhang et al[14] was excluded from all of the scenarios 4–6 because we could not identify the disease severity of RA patients at baseline and the dose of MTX solely based on the published information.[14] Likewise, the study by Abe et al[13] was removed in scenarios 5 and 6 because

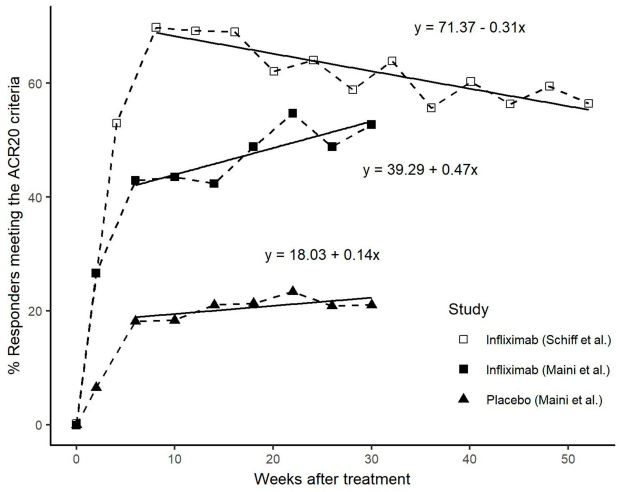

Figure 2 Time trend in ACR 20 in selected placebo-controlled original infliximab trials. We performed a linear regression analysis to evaluate the time trend in ACR20 in the historical clinical trials with original infliximab. Because the time trend of ACR20 dramatically changed around week 4, we only analyzed ACR20 assessed after week 4. The slopes of the regression line were significantly different from 0 in both Schiff et al and Maini et al (p-values: 0.00031 and 0.002), respectively.

Table 2 Summary of ACR20 by Original Infliximab (Remicade®)

Study	ACR20 (%) By Treatment		Placebo-Adjusted ACR20* (Percentage Points)	p-value†
	Placebo + MTX	Original Infliximab + MTX		
Maini et al.[17]	20.0	50.0	30.0	<0.001
Schiff et al.[18]	41.8	59.4	17.6	0.006
Westhovens et al.[12]	25.5	58.0	32.5	<0.0001
Abe et al.[13]	23.4	61.2	37.8	<0.001
Zhang et al.[14]	48.9	75.9	27.0	0.0003

Notes: *Difference between the original infliximab and placebo in the proportion of patients who met the ACR20 criteria. †Original infliximab vs placebo.
Abbreviations: ACR20, the American College of Rheumatology 20% response rate; MTX, methotrexate.

the mean MTX dose was much lower than the other studies. In scenario 6, we further excluded the study by Westhovens et al[12] because ACR20 was assessed much earlier at week 22 after treatment in contrast to week 28–30 in Maini et al.[17] Schiff et al[18] and the PLANETRA trial.[1]

Estimation of the Equivalence Margins for CT-P13

M_1, the lower bound of the 95% confidence interval for the placebo-adjusted effect of INX, ranged from 11 to 26 percentage points (Figure 3). Taking the half of M_1 as M_2, the appropriate equivalence margin was 12.8, 11.3, 10.5, 9.4, and 5.7 percentage points in scenarios 2–6, respectively (Table 4). In scenario 1, the half of the pooled point estimate (14.2%) instead of the half of its lower 95% bound was the equivalence margin.

The equivalence of CT-P13 to the original infliximab could not be claimed in scenarios 5 and 6, while it was barely met in scenario 4 (Table 4). In the other scenarios, equivalence was concluded.

Discussion

The equivalence conclusion for CT-P13 in the PLANETRA trial did not appear to be supported by the equivalence margins we independently derived in the present study, which was narrower than the 12 percentage points that the FDA employed (9.3 or 5.7 percentage points for scenarios 5 and 6, respectively, Table 4). We also found out that the equivalence margin values in scenarios 1–3 did not match with what Celltrion and the FDA used (i.e., 14.2 vs 15, 12.8 vs 13, and 11.3 vs 12 percentage points, respectively) even though we pooled studies the same way as they did (Table 3). It was because we consolidated historical data according to the ITT principle that gave the different sample sizes from theirs.

Not all of the historical studies pooled by Celltrion and the FDA were sufficiently similar to the PLANETRA trial in terms of baseline characteristics of patients, MTX dose, and efficacy assessment time. This lack of sufficient similarity in the study design and patient population could weaken the constancy assumption. When the information on the important study design and patient characteristics is not fully available for a trial, it might be risky to use its efficacy data

Table 3 Study-Pooling Scenarios to Estimate the Equivalence Margin for CT-P13

Scenario No.	Studies Excluded*	M_1 Based on	Used In/By
1	None	Point estimate	The PLANETRA trial[1]
2	Schiff et al.[18]	Lower bound of 95% CI	Celltrion[2]
3	None	Lower bound of 95% CI	The US FDA[3]
4	Zhang et al.[14]	Lower bound of 95% CI	–
5	Zhang et al.,[14] Abe et al.[13]	Lower bound of 95% CI	–
6	Zhang et al.,[14] Abe et al.,[13] Westhovens et al.[12]	Lower bound of 95% CI	We propose this scenario as the most appropriate one.

Notes: *Out of the following five placebo-controlled original infliximab trials: Maini et al.,[17] Westhovens et al.,[12] Schiff et al.,[18] Zhang et al.,[14] and Abe et al.[13]
Abbreviations: CI, confidence interval; FDA, Food and Drug Administration.

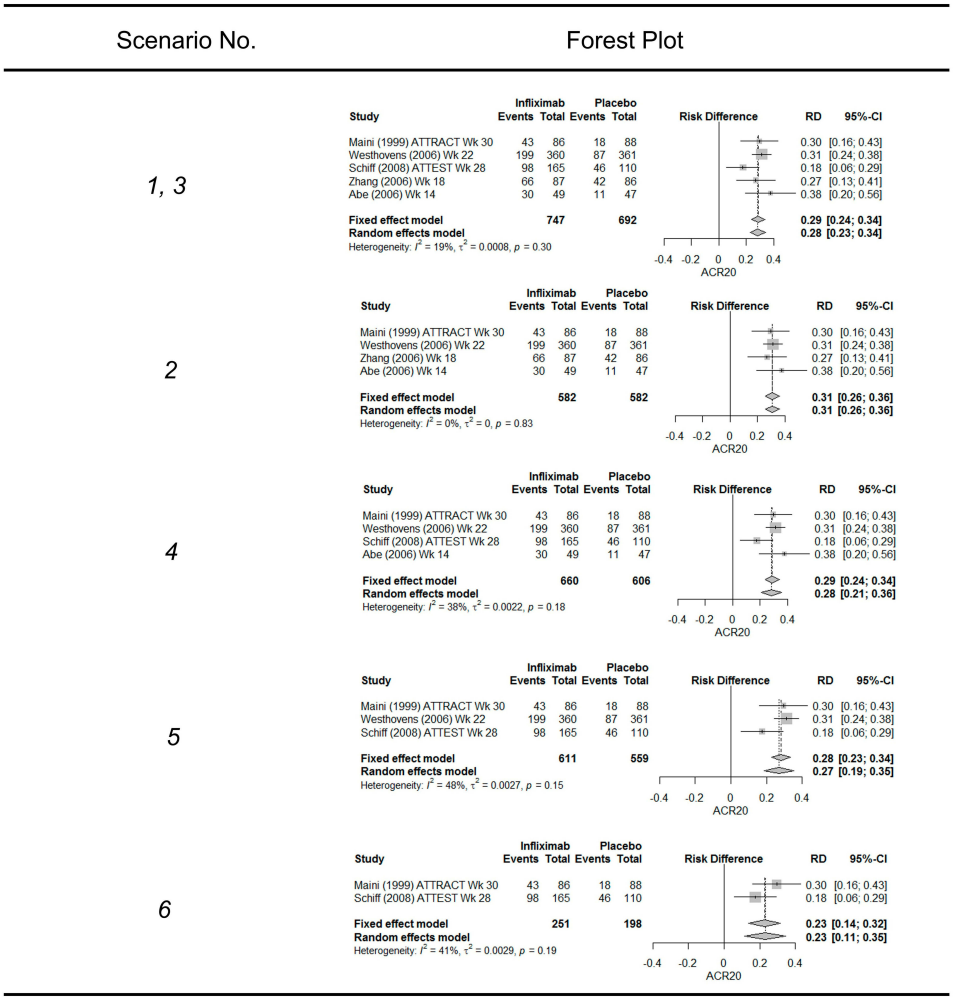

Figure 3 Forest plots of the differences in ACR20 between the original infliximab and placebo by scenario. Original infliximab was used as an add-on treatment to methotrexate. 'Events' denotes the number of patients who met the ACR20 criteria in each treatment group. "RD" represents risk difference, where risk means meeting the response criteria. Each scenario used a different set of historical clinical trials with original infliximab.

as a basis of judgments. This is why we excluded the study by Zhang et al in scenarios 4–6 (Table 3), which neither provided the disease severity of RA patients at baseline nor the dose of MTX.[14] Likewise, if a clinical trial was performed only in a specific population, it would be possible that the data may be less comparable. For example, RA patients in the study by Abe et al received only half (i.e., 7.1 mg/week) of the MTX dose administered in the other historical clinical trials with original infliximab and the PLANETRA trial because it was close to the maximum dose (i.e., 8 mg/week) permitted in Japan.[13,20] Furthermore, ACR20 was assessed much earlier than 30 weeks of treatment in the studies by Westhovens et al.[12] Abe et al[13] and Zhang et al[14] However, efficacy assessment at earlier time points could incorrectly represent efficacy later times as we clearly showed (Figure 2).

In this sense, scenario 6 may be a more scientifically appropriate and clinically justifiable pooling strategy to estimate M_1, which should have been adopted to derive the equivalence margin for CT-P13 in the PLANETRA trial. Interestingly, the 5 percentage point was chosen as the non-inferiority margin in a meta-analysis, which compared the effect of certolizumab pegol with that of other anti-cytokines in RA.[10] Although this margin was clinically determined, the value was similar to those we thought were more appropriate (5.7 percentage points for scenario 6, Table 4). This similarity suggests that the equivalence margins scientifically derived in the present study could have some clinical meanings. Likewise, the margin was only 7.5 percentage points in a non-inferiority trial of subcutaneous abatacept to intravenous abatacept.[21] Sometimes, however, a narrower equivalence margin may

Table 4 Equivalence Assessment in ACR20 Between CT-P13 and the Original Infliximab by Scenario

ACR20 (%) at Week 30 in the PLANETRA Trial, ITT Analysis		Difference [CI] (Percentage Points)*	Scenario No.†	Equivalence Margin (Percentage Points)	Equivalence to the Original Infliximab
CT-P13	Original Infliximab				
60.9	58.6	2 [−6, 10]	1	14.2	Yes
			2	12.8	Yes
			3	11.3	Yes
			4	10.5	Yes
			5	9.3	No
			6	5.7	No

Notes: *Difference between CT-P13 and the original infliximab in the proportion of patients who met the ACR20 criteria. Positive numbers mean CT-P13 was better than the original infliximab. †See Table 3.
Abbreviations: ACR20, the American College of Rheumatology 20% response rate; ITT, intention-to-treat; CI, confidence interval.

not be feasible because a larger sample size is required to prove equivalence with sufficient power. In this regard, a compromised equivalence margin may be considered, for example, 9.3 percentage points (scenario 5) in the case of CT-P13.

We also want to express concern about inappropriately cherry-picking a study to exclude when estimating a pooled estimate of the placebo-adjusted efficacy of the reference drug. If a study is excluded for the reason that it showed a smaller effect of the reference drug than other studies, a falsely large M_1 will be estimated, which should be avoided.[7] Undoubtedly, this practice wrongly overestimates the equivalence margin, which eventually increases the consumer risk of erroneously regarding the follow-on product as equivalent to the original product when, in fact, it is not. For example, when the study by Schiff et al was excluded in scenario 2 probably because the placebo-adjusted efficacy in this study was the smallest, the equivalence margin was widened by 13.3% (12.8 vs 11.3 percentage points, Table 4).

The current FDA's non-inferiority guidance recommends M_2 be determined based on clinical judgments about how much the original reference drug effect needs to be retained to show sufficient benefit for drug approval as well as on practical consideration for the study size.[7] However, detailed guidelines are still missing in the FDA guidance on how to determine an M_2 while 50% of M_1 is introduced as a starting point. Although we adopted the same approach in this study, the clinical justification of 50% in determining M_2 should be further investigated in future studies. The seriousness of the outcome, the benefit of the active control, and relative safety profiles of test and comparator should be considered to determine M2.

The limitations of this study include the following. First, the equivalence margin based on the per-protocol (PP) principle could not be determined because two of the five historical studies reported only the number of patients at baseline with no number at efficacy assessment. The PP approach is generally more conservative than the ITT approach in equivalence trials, i.e., equivalence claim becomes more difficult to make due to a narrower equivalence margin.[22] Therefore, our analysis still holds valid even though the ITT analysis approach was used.

Second, some numbers of responders in our meta-analysis might be slightly different from the exact numbers because some studies reported only the proportion of responders, not their numbers. Thus, we obtained the number of responders by multiplying the reported proportion and the sample size. Although such numbers were agreeable, they might be less accurate than the actual numbers in each trial. In the PLANETRA trial, both the number of patients enrolled and the number of patients completed as well as the responder proportions by both the ITT and the per-protocol principle were available. Therefore, our assessment could have been more exhaustive and exact if the historical data were more as in the PLANETRA trial.

In conclusion, the equivalence margin Celltrion chose for CT-P13 in the PLANETRA trial was much larger (i.e., more forgiving) than the one estimated independently in this exercise using a scientifically valid, but stricter, approach that adheres to the principle of the regulatory guideline. When estimating the placebo-adjusted efficacy of the reference product using historical clinical trials, the constancy assumption should be carefully checked and

ascertained. Given the lack of experience with the biosimilar products of such complex molecules as the monoclonal antibody, the smaller equivalence margin could be one mechanism to ensure similarity between the biosimilar and its reference product. This approach can guarantee that patients are protected from any remaining uncertainty or risks with the use of follow-on biological products.

Funding

This research was supported by the BK21 Plus Program of the National Research Foundation of Korea (NRF) (10Z20130000017).

Disclosure

The authors report no conflicts of interest in this work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Yoo DH, Hrycaj P, Miranda P, et al. A randomised, double-blind, parallel-group study to demonstrate equivalence in efficacy and safety of CT-P13 compared with innovator infliximab when coadministered with methotrexate in patients with active rheumatoid arthritis: the PLANETRA study. *Ann Rheum Dis*. 2013;72(10):1613. doi:10.1136/annrheumdis-2012-203090
2. Celltrion. CT-P13 briefing document for the arthritis advisory committee. [updated February 9, 2016; cited Feb, 2016]. Available from: https://www.fda.gov/media/95998/download.
3. Food and Drug Administration. Statistical review of BLA125544. [cited Oct, 2015]. Available from: www.accessdata.fda.gov/drugsatfda_docs/nda/2016/125544Orig1s000StatR.pdf. Accessed May 7, 2020.
4. Cohen SB, Alten R, Kameda H, et al. A randomized controlled trial comparing PF-06438179/GP1111 (an infliximab biosimilar) and infliximab reference product for treatment of moderate to severe active rheumatoid arthritis despite methotrexate therapy. *Arthritis Res Ther*. 2018;20(1):155. doi:10.1186/s13075-018-1646-4
5. Food and Drug Administration. Biosimilar multi-disciplinary evaluation and review: BLA 761086 ABP 710, a proposed biosimilar to U.S.-licensed remicade (infliximab). [cited Dec, 2018]. Available from: https://www.fda.gov/media/134460/download. Accessed May 7, 2020.
6. Lee H. Equivalence margin of the biosimilar product. *J Korean Soc Clin Pharmacol Ther*. 2012;20(1):17–33. doi:10.12793/jkscpt.2012.20.1.17
7. Food and Drug Administration. Guidance for industry: non-inferiority clinical trials to establish effectiveness. [cited Nov, 2016]. Available from: https://www.fda.gov/media/78504/download. Accessed May 7, 2020.
8. Njue C. Statistical considerations for confirmatory clinical trials for similar biotherapeutic products. *Biologicals*. 2011;39(5):266–269. doi:10.1016/j.biologicals.2011.06.006
9. Hwang IK, Morikawa T. Design issues in noninferiority/equivalence trials. *Drug Inf J*. 1999;33(4):1205–1218. doi:10.1177/009286159903300424
10. Launois R, Avouac B, Berenbaum F, et al. Comparison of certolizumab pegol with other anticytokine agents for treatment of rheumatoid arthritis: a multiple-treatment bayesian metaanalysis. *J Rheumatol*. 2011;38(5):835–845. doi:10.3899/jrheum.100665
11. Aaltonen KJ, Virkki LM, Malmivaara A, Konttinen YT, Nordstrom DC, Blom M. Systematic review and meta-analysis of the efficacy and safety of existing TNF blocking agents in treatment of rheumatoid arthritis. *PLoS One*. 2012;7(1):e30275. doi:10.1371/journal.pone.0030275
12. Westhovens R, Yocum D, Han J, et al. The safety of infliximab, combined with background treatments, among patients with rheumatoid arthritis and various comorbidities: a large, randomized, placebo-controlled trial. *Arthritis Rheum*. 2006;54(4):1075–1086. doi:10.1002/art.21734
13. Abe T, Takeuchi T, Miyasaka N, et al. A multicenter, double-blind, randomized, placebo controlled trial of infliximab combined with low dose methotrexate in Japanese patients with rheumatoid arthritis. *J Rheumatol*. 2006;33(1):37–44.
14. Zhang FC, Hou Y, Huang F, et al. Infliximab versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a preliminary study from China. *APLAR J Rheum Dis*. 2006;9(2):127–130. doi:10.1111/j.1479-8077.2006.00186.x
15. R Foundation for Statistical Computing. R: a language and environment for statistical computing. [cited 2018]. Available from: https://www.R-project.org.
16. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–188. doi:10.1016/0197-2456(86)90046-2
17. Maini R, St Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised Phase III trial. *Lancet*. 1999;354(9194):1932–1939. doi:10.1016/S0140-6736(99)05246-0
18. Schiff M, Keiserman M, Codding C, et al. 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. *Ann Rheum Dis*. 2008;67(8):1096–1103. doi:10.1136/ard.2007.080002
19. Food and Drug Administration. FDA briefing information for the february 9, 2016 meeting of the arthritis advisory committee. [cited Feb 9, 2016]. Available from: https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/UCM484859.pdf. Accessed May 7, 2020.
20. Kameda H, Fujii T, Nakajima A, et al. Japan college of rheumatology guideline for the use of methotrexate in patients with rheumatoid arthritis. *Mod Rheumatol*. 2019;29(1):31–40. doi:10.1080/14397595.2018.1472358
21. Genovese MC, Covarrubias A, Leon G, et al. Subcutaneous abatacept versus intravenous abatacept: a phase IIIb noninferiority study in patients with an inadequate response to methotrexate. *Arthritis Rheum*. 2011;63(10):2854–2864. doi:10.1002/art.30463
22. ICH Expert Working Group. ICH harmonised tripartite guideline, E9: statistical principles for clinical trials. [cited 1998]. Available from: https://database.ich.org/sites/default/files/E9_Guideline.pdf. Accessed May 13, 2020.