16.1.9 Documentation of Statistical Methods

The document listed below is provided in this section.

Statistical Analysis Plan (NI071F2) Version 5.0 dated 27 June 2019

STATISTICAL ANALYSIS PLAN

Nichi-Iko Pharmaceutical Co., Ltd.
1-6-21 Sogawa, Toyama City,
Toyama Prefecture, Japan

A Randomized, Double-Blind, Multicenter, 3-Stage, Efficacy and Safety Study of NI-071 and US-Licensed Remicade® (Infliximab) for the Treatment of Patients with Rheumatoid Arthritis

Protocol No: NI071F2

SAP Version 5.0

SAP Date: 27 June 2019

Prepared by:

PPD

Syneos Health

STATISTICAL ANALYSIS PLAN

SIGNATURE PAGE

Nichi-Iko Pharmaceutical Co., Ltd.
1-6-21 Sogawa, Toyama City,
Toyama Prefecture, Japan

A Randomized, Double-Blind, Multicenter, 3-Stage, Efficacy and Safety Study of NI-071 and US-Licensed Remicade® (Infliximab) for the Treatment of Patients with Rheumatoid Arthritis

Protocol No: NI071F2

PPD , PhD, PPD
Nichi-Iko Pharmaceutical Co., Ltd.
PPD
Approval Date
PPD

PPD , PhD, PPD
Nichi-Iko Pharmaceutical Co., Ltd.
PPD
Approval Date
PPD

PPD
Syneos Health PPD
PPD
Approval Date
PPD

STATISTICAL ANALYSIS PLAN
Nichi-Iko Pharmaceutical Co., Ltd.
NI071F2

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
ACPA	Anti-citrullinated protein antibody
ACR	American College of Rheumatology
ACR20 (50, 70)-CRP	20% (50%, 70%) improvement from baseline in the ACR core set criteria using C-reactive protein as the acute-phase reactant
ACR20 (50, 70)-ESR	20% (50%, 70%) improvement from baseline in the ACR core set criteria using erythrocyte sedimentation rate as the acute-phase reactant
AE	Adverse event
Anti-dsDNA	Anti-double-stranded deoxyribonucleic acid
ANCOVA	Analysis of co-variance
ANOVA	Analysis of variance
AUC	Area under the concentration-time curve
CI	Confidence interval
C_{max}	Maximum concentration
CRP	C-reactive protein
CV	Coefficient of variation
DAS28-CRP	Disease activity score based on 28 joints calculated with CRP value
DAS28-ESR	Disease activity score based on 28 joints calculated with ESR value
DMARD	Disease-modifying anti-rheumatic drug
DNA	Deoxyribonucleic acid
ESR	Erythrocyte sedimentation rate
FAS	Full analysis set
GCP	Good Clinical Practice
HAQ-DI	Health Assessment Questionnaire Disability Index
HbA1c	Glycated hemoglobin
HBcAb	Hepatitis B core antibody
HBsAb	Hepatitis B surface antibody
HBsAg	Hepatitis B surface antigen
HCV Ab	Hepatitis C antibody
HIV Ab	Human immunodeficiency virus antibody

ICH	International Conference on Harmonization
IDMC	Independent data monitoring committee
ITT	Intent-to-treat
NRI	Non-responder imputation
LOCF	Last observation carried forward
MDHAQ	Multidimensional health assessment questionnaire
MedDRA	Medical Dictionary for Regulatory Activities
MTX	Methotrexate
NRI	Non-responder imputation
PK	Pharmacokinetic
PP	Per protocol
PT	Preferred term
RA	Rheumatoid arthritis
RAPID3	Routine assessment of patient index data 3
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SF-36	Medical Outcomes Study 36-Item Short-Form Health Survey
SOC	System organ class
TB	Tuberculosis
TEAE	Treatment-emergent adverse event
t_{max}	Time at maximum concentration
TNF-α	Tumor necrosis factor alpha
TOST	Two 1-sided test analysis
US	United States
VAS	Visual analog scale
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

1. INTRODUCTION

Rheumatoid arthritis (RA), the most common autoimmune inflammatory arthritis in adults, is characterized by the persistent and progressive inflammation of multiple synovial membranes, causing symptoms of joint pain, swelling, stiffness, and redness, and often ultimately leading to the erosion and destruction of cartilage and bone. RA causes increasing disability and decreases patients' ability to perform daily activities, thus reducing their economic and social quality of life. RA is associated with increased mortality and comorbidities (most commonly, cardiovascular disease, infections, mental health conditions, and malignancies). There is no cure, but effective treatments are increasingly available to treat RA. In patients with early RA, treatment usually starts with non-biologic disease-modifying antirheumatic drugs (DMARDs) that help reduce disease activity and prevent joint deformity. Over time as the disease activity increases, combination treatment with non-biologic and biologic DMARDs are introduced.[1,2]

Biologic therapies are designed to modify or regulate specific mechanisms involved in the proinflammatory cytokine overproduction that plays a key role in inflammatory diseases. Tumor necrosis factor alpha (TNF-α) is a proinflammatory cytokine that is involved in normal inflammatory and immune response, but overexpressed TNF-α has been implicated in the pathogenesis of RA and other immunological diseases. Elevated TNF-α levels have been found in the joints of RA patients and correlate with elevated disease activity; elevated levels are also found in patients with Crohn's disease, ulcerative colitis, ankylosing spondylitis, psoriatic arthritis, and plaque psoriasis.[3,4]

Remicade® (infliximab), one of the biologics used to treat RA, is a chimeric human-murine IgG1 monoclonal antibody specific for human TNF-α produced in murine hybridoma cells by recombinant deoxyribonucleic acid (DNA) technology. Infliximab neutralizes the biological activity of TNF-α by binding with high affinity to the soluble and transmembrane forms of TNF-α and inhibits binding of TNF-α with its receptors. Infliximab inhibits the functional activity of TNF-α in a wide variety of in vitro bioassays. Also, infliximab prevented disease in transgenic mice that develop polyarthritis as a result of constitutive expression of human TNF-α and when administered after disease onset, it allowed eroded joints to heal. In vivo, infliximab rapidly forms stable complexes with human TNF-α, a process that parallels the loss of TNF-α bioactivity. In RA, treatment with infliximab reduced infiltration of inflammatory cells into inflamed areas of the joint as well as expression of molecules mediating cellular adhesion, chemo-attraction, and tissue degradation. Also after infliximab treatment, patients exhibited decreased levels of serum interleukin 6 and C-reactive protein (CRP).[3,4]

Remicade is approved for use in the United States (US) for the treatment of adult patients with moderately to severely active RA, in combination with methotrexate (MTX), for reducing signs and symptoms, inhibiting the progression of structural damage, and

improving physical function. In addition, Remicade is approved for the treatment of the following other indications in the US:[3]

- Adult Crohn's Disease (indicated for reducing signs and symptoms and inducing and maintaining clinical remission in adult patients with moderately to severely active disease who have had an inadequate response to conventional therapy and for reducing the number of draining enterocutaneous and rectovaginal fistulas and maintaining fistula closure in adult patients with fistulizing disease)
- Pediatric Crohn's Disease (indicated for reducing signs and symptoms and inducing and maintaining clinical remission in pediatric patients with moderately to severely active disease who have had an inadequate response to conventional therapy)
- Adult Ulcerative Colitis (indicated for reducing signs and symptoms, inducing and maintaining clinical remission and mucosal healing, and eliminating corticosteroid use in adult patients with moderately to severely active disease who have had an inadequate response to conventional therapy)
- Pediatric Ulcerative Colitis (indicated for reducing signs and symptoms and inducing and maintaining clinical remission in pediatric patients with moderately to severely active disease who have had an inadequate response to conventional therapy)
- Ankylosing Spondylitis (indicated for reducing signs and symptoms in patients with active disease)
- Psoriatic Arthritis (indicated for reducing signs and symptoms of active arthritis, inhibiting the progression of structural damage, and improving physical function)
- Plaque Psoriasis (indicated for treatment of adult patients with chronic severe [i.e., extensive and/or disabling] plaque psoriasis who are candidates for systemic therapy and when other systemic therapies are medically less appropriate)

Nichi-Iko Pharmaceutical Co., Ltd., (the Sponsor) is developing NI-071, a biosimilar of Remicade containing the active substance infliximab. The indications for NI-071 are proposed to be the same as currently approved for US-licensed Remicade, and the dosage forms, route of administration, and dosing regimens for NI-071 are to be the same as for Remicade for each indication.[5]

2. STUDY OBJECTIVES

2.1 Study Objectives

2.1.1 Stage 1 (Biosimilarity Stage) Objectives

Primary:

- To investigate the biosimilarity of NI-071 (infliximab biosimilar) and US-licensed Remicade (infliximab) in terms of efficacy in patients with RA not adequately responding to MTX, by determining the percentages of patients achieving ACR20-CRP (a 20% improvement from baseline in the American College of Rheumatology [ACR] core set criteria using CRP as the acute-phase reactant)

Secondary:

- To evaluate safety and immunogenicity of NI-071 and Remicade-US
- To evaluate efficacy based on other rates of improvement in the ACR criteria, on the DAS28 (disease activity score involving 28 joints), and the RAPID3 (routine assessment of patient index data 3), and on changes in health-related quality of life

2.1.2 Stage 2 (Interchangeability Stage) Objectives

Primary:

- To investigate the interchangeability between NI-071 and Remicade-US by evaluating key pharmacokinetic (PK) parameters in each treatment group following treatment switches

Secondary:

- To evaluate the safety, efficacy, and immunogenicity of NI-071 and Remicade-US

2.1.3 Stage 3 (Safety Follow-up Stage) Objective

Primary:

- To evaluate the safety of long-term treatment (follow-up through Week 62)

3. STUDY DESIGN

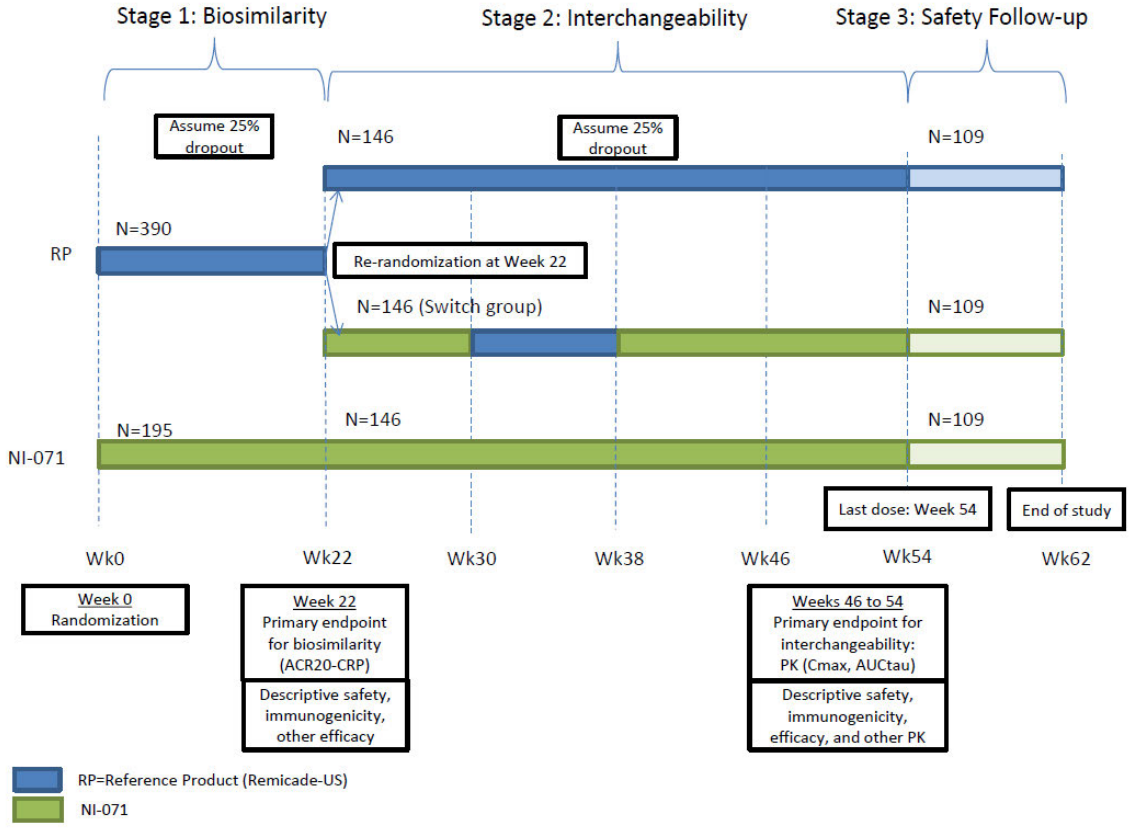
This Phase 3, randomized, double-blind, multicenter, biosimilarity, interchangeability, and safety study is designed to evaluate NI-071 versus Remicade-US for the treatment of patients with RA. Following a screening period, there will be 3 stages: the biosimilarity stage (Stage 1), the interchangeability stage (Stage 2), and the safety follow-up stage (Stage 3). A schematic of the study design is provided in Figure 1.

The screening period, approximately 4 weeks long for each patient, will be used to assess eligibility. At baseline (Week 0), approximately 585 patients who are eligible will be randomized in a 2:1 ratio to the Remicade-US and NI-071 groups, respectively. This initial randomization will be stratified by CRP level (≤2, >2 mg/dL). During Stage 1, patients will receive their randomized treatment at Weeks 0, 2, 6, and 14. Primary biosimilarity efficacy assessments (specifically ACR20-CRP) will be performed at Week 22.

During Stage 2, at Week 22, patients in the original Remicade-US group are re-randomized (1:1) to a Remicade-US group and a Switch group. The Switch group will receive NI-071 at Week 22, Remicade-US at Week 30, and NI-071 at Weeks 38, 46, and 54; the Remicade-US group will continue to receive Remicade-US from Week 22 through Week 54 every 8 weeks. The original NI-071 group will continue to receive NI-071 from Week 22 through Week 54 every 8 weeks. To evaluate interchangeability, a full PK similarity analysis (specifically AUC_{tau} and C_{max}) will be performed for all treatment groups from samples drawn at 11 time points from pre-dose Week 46 to pre-dose Week 54. During Stage 3, the last dose of study treatment will be given at Week 54, and patients will return for 2 additional safety follow-up visits through Week 62.

Efficacy and safety evaluations, including assessments of ADAs, will be performed throughout the study. In addition, samples for determining trough concentrations will be collected prior to each infusion in the study and at the end of the study.

Figure 1 NI071F2 Study Design



4. ANALYSIS POPULATIONS

4.1 Intent-To-Treat Population

The intent-to-treat (ITT) population will include all patients who are randomized, irrespective of any deviation from the protocol or premature discontinuation. Patients will be analyzed according to the treatment group assigned at the Stage 1 randomization. The ITT population will serve as the basis for the primary efficacy analysis in Stage 1.

4.2 Safety Population

The safety population will include all ITT patients who receive at least 1 dose of study drug. Patients will be analyzed according to the actual treatment they receive rather than as randomized. This population will be used for the analysis of safety.

4.3 Full Analysis Set

The full analysis set (FAS) will include all patients in the safety population with at least 1 post-baseline ACR20 efficacy assessment. Patients will be analyzed according to the treatment group assigned at the most recent randomization. This population will be used for the secondary efficacy analyses, as well as a sensitivity analysis for the primary efficacy endpoint.

4.4 Per-Protocol Population

The per-protocol (PP) population will include all patients in the FAS with no major protocol deviations in Stage 1 that have an impact on the primary efficacy endpoint, ie, ACR20-CRP response at Week 22, and sufficient time in study. Sufficient time will be defined as having at least the Week 22 ACR20-CRP efficacy assessment and all study drug administrations up to and including the Week 14 study drug administration. Since this population excludes patients with such major protocol deviations as defined above, patients will be analyzed according to the treatment they were randomized to and received during Stage 1. This population will be used for a sensitivity analysis of the primary efficacy analysis.

4.5 Pharmacokinetic Set

The PK population will include all patients in the FAS population in Stage 1 who have valid PK assessments through Stage 2. This includes only patients who have received all and complete treatments required per protocol through Stage 2. This population will be used for the analysis of interchangeability in Stage 2.

5. STUDY ENDPOINTS

5.1 Clinical Efficacy Endpoints

Primary:

- ACR20-CRP at Week 22

Secondary:

- Change from baseline in DAS28-CRP and DAS28-ESR scores (DAS28 scores calculated with CRP and erythrocyte sedimentation rate (ESR) values, respectively)
- ACR20-CRP (at time points other than Week 22), ACR20-ESR, ACR50-CRP, ACR50-ESR, ACR70-CRP, and ACR70-ESR
- Change from baseline in each item of the ACR core set
- Change from baseline in RAPID3 scores
- Change from baseline in the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) domain subscores and summary scores

5.2 Pharmacokinetic Endpoints

Primary:

- AUC_{tau} and C_{max} during the dosing interval of Weeks 46 to 54

Secondary:

- Minimum concentration (C_{min}) and time at maximum concentration (t_{max}) during the dosing interval of Weeks 46 to 54

5.3 Safety Endpoints

The safety endpoints of this study are as follows:

- Incidence of TEAEs and serious AEs (SAEs)
- Incidence and risk difference between treatment groups of AEs of special interest, including serious infections, malignancies, infusion-related reactions, hepatotoxicity, heart failure, and cytopenias
- Incidence of ADAs and neutralizing antibodies
- Clinically significant changes in laboratory values, vital signs, weight, ECG, and physical examination results

6. DEFINITIONS AND DERIVED VARIABLES

Derivation of visit windows can be found in Appendix A. Other definitions of derived variables not listed below can be found in applicable subsections of the statistical methodology section (Section 7) of this document.

6.1 ACR20/50/70

ACR20 will be defined as a dichotomous variable, for which a patient is defined as either a responder or a non-responder. Patients will be defined as a responder if all of the following criteria are met:

- A 20% improvement from baseline in the tender/painful joint count.
- A 20% improvement from baseline in the swollen joint count.
- A 20% improvement from baseline in *at least 3* of the following 5 variables:
    - Patient global assessment of disease activity (VAS)
    - Physician global assessment of disease activity (VAS)
    - Patient pain assessment (VAS)
    - HAQ-DI
    - CRP when calculating ACR20-CRP, or ESR when calculating ACR20-ESR

ACR50 and ACR70 are similarly defined, except 50% and 70% improvement is required, respectively.

6.2 DAS28

DAS28-CRP and DAS-ESR will be defined as a continuous variable based on 28-count subsets of tender/painful joints and swollen joints, together with either CRP or ESR, respectively. These variables will be calculated as follows:

- If the patient global assessment of disease activity on the 100 mm VAS is not missing at the visit, then the following calculations are used:
    - DAS28(ESR) = (0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*ln(ESR) + 0.014*GH
    - DAS28(CRP) = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1) + 0.014*GH + 0.96
- If the patient global assessment of disease activity on the 100 mm VAS is missing at the visit, then the following calculations are used:
    - DAS28(ESR) = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*ln(ESR)]*1.08 + 0.16

- DAS28(CRP) = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1)] * 1.10 + 1.15,

where TJC28 is number of painful joints out of 28 joints, SJC28 is number of swollen joints out of 28 joints, GH is the score of the patient global assessment of disease activity, ln is the natural logarithm, ESR is in mm/first hour, and CRP is in mg/L.

7. STATISTICAL METHODOLOGY

7.1 Statistical and Analytical Issues

7.1.1 Statistical Methods

The statistical evaluation will be performed using the Statistical Analysis Software (SAS®) Version 9.3 or higher (SAS Institute, Cary, NC). For all inferential analyses and all modeling type analyses, original SAS outputs will be provided to be included as appendices to the CSR.

All data will be listed, and summary tables will be provided. The data listings will be sorted by treatment group, site number, patient screening number and visit. Any data collected on patients who were screened but subsequently not randomized will be presented within a non-randomized group.

Baseline is generally defined as the last assessment taken prior to treatment. This typically occurs at the baseline visit. If the value from the baseline visit is missing for any given parameter, values from the screening visits (or any unscheduled visits that occurred later, but still before the first administration of study medication) will be applied as baseline. In no case, will baseline be imputed with assessments performed after the first application of study medication. If no pre-treatment assessments are available, baseline will be missing.

As shown in Figure 1, the treatment groups according to study stages are as follows: 2 treatment groups in Stage 1 (i.e., the Remicade-US and NI-071 groups), 3 treatment groups in Stage 2 (the NI-071 group, continuing to receive NI-071; the Remicade-US group, continuing to receive Remicade-US; and the Switch group, receiving NI-071 at Week 22, Remicade-US at Week 30, and NI-071 at Weeks 38 and 46); and the same 3 treatment groups in Stage 3 (the NI-071 group, the Remicade-US group, and the Switch group [receiving NI-071 at Week 54]).

The treatment group labels for all tables, figures and listings will be: NI-071, Remicade-US, Switch, or Overall.

Data will be summarized by treatment group; a total/overall column will be included to summarize all patients, where appropriate. Also, data will be summarized by visit, where appropriate.

P-values will be rounded to 4 decimal places. P-values less than 0.0001 will be reported as "<0.0001" in tables. Unless otherwise specified, two-sided statistical tests will be performed with a significance level of 0.05, and one-sided tests will be performed with a significance level of 0.025.

Categorical variables will be summarized using frequencies and percentages. Percentages will be presented to 1 decimal place. In cases where the percentage calculated is >0% and <0.1%, "<0.1%" will be presented. Percentage of 100% will be displayed without a decimal place, zero-incidences (i.e., where n=0) will not show a percentage value.

Continuous variables will be summarized using descriptive statistics (number of patients with an observation [n], mean, standard deviation [SD], median, minimum [min] and maximum [max]) where n ≥2. Unless otherwise specified, the mean, SD, and median for a continuous variable will be presented with 1 more decimal place than the original (raw) values. The min and max will be presented with the same number of decimal places as the original values.

7.1.2 Handling of Dropouts and Missing Data

Patients who drop out will not be replaced in this study.

As a general rule, unless specified otherwise, missing data for efficacy and safety assessments will not be imputed, assuming that these values are missing at random. Furthermore, all available data collected from patients who drop out of the study or discontinue study drug treatment will be included in the analyses.

7.1.2.1 ACR20, ACR50, and ACR70

Because the ACR20, ACR50, and ACR70 variables (determined with either CRP or ESR) are based on several component variables (the ACR core set criteria), it is possible that the values may still be calculated even if the component variables have some missing values. In this case, no imputation method is needed. If the ACR responder status still cannot be determined with the available visit specific data, an imputation method will be applied, except for the ACR20-CRP analyses done at Week 22 (the primary endpoint).

If the ACR responder status is missing due to some missing (but not all) ACR core set criteria values, the method of last observation carried forward (LOCF) will be used to carry forward any of the individually missing component values, and from that mix of actual and carried-forward values, the ACR responder status will be determined.

This type of LOCF method will be known as "LOCF mixed components," since it is based on determining the ACR responder status with a mixture of values at the visit and values carried forward from previous visits.

After the LOCF imputation has been applied, *missing values due to a patient dropping from the study for any reason will be handled by setting the ACR responder status (*ACR20, ACR50, and ACR70) *to nonresponsive* for future visits onward. For the

primary analysis performed, a patient who stays in the study but discontinues study drug treatment on or before Week 14 will also be regarded as nonresponsive. Note that this method also goes by the name non-responder imputation, NRI. NRI is the term to be used in the study report.

If for some reason the ACR responder status still cannot be determined, including the case where baseline data is missing, then its value will be set to non-responder.

Of note, the ACR responder status is aligned to a specific study week (defined in Appendix A); although the individual components might be assessed on different individual days, as long as the assessments are assigned to the specific visit/week, the ACR responder status will be determined with all assessments at that visit/week and for that specific visit/week.

7.1.2.2 DAS28

The DAS28-CRP and DAS28-ESR endpoints are calculated from several components (Section 6.2). As with ACR20/50/70, the DAS28 endpoints can be calculated even when one of the components is missing, specifically when the patient's global assessment is missing. If more than the patient's global assessment is missing, then the LOCF mixed component method will be used to calculate the endpoints. DAS28 scores that cannot be calculated from the LOCF mixed component method will not be imputed, but rather they will be set to missing; it is assumed that these values are missing at random.

Of note, the DAS28 is aligned to a specific study week; although the individual components might be assessed on different individual days, as long as the assessments are assigned to the specific visit/week, the DAS28 will be calculated with all assessments at that visit/week and for that specific visit/week.

7.1.2.3 HAQ-DI and SF-36

The Health Assessment Questionnaire-Disability Index (HAQ-DI) and SF-36 endpoints are calculated from validated questionnaires. Each endpoint has a unique scoring method. The methods that will be employed for handling missing data from these assessments are intricate and related the unique scoring method for each. The rules for handling the missing data related to these questionnaires are presented in Section 7.3.2.4 for HAQ-DI and in Section 7.3.2.7 for SF-36.

7.1.2.4 Tender and Swollen Joint Counts

Handling of missing/unevaluable joint data is dependent on the reason (missing versus unevaluable) and the time point (baseline versus follow-up).

- If baseline, and the joint is marked as "unevaluable", then it is NOT counted as tender/swollen.

- If follow-up, and the joint was evaluable at baseline, then it IS counted as tender/swollen.
- If follow-up, and the joint was unevaluable at baseline, then it is NOT counted as tender/swollen.
- If follow-up, and the evaluation is missing (not marked as "unevaluable"), then the classification of the joint (tender/swollen) will be based on LOCF, except for the ACR20-CRP analyses done at Week 22 (the primary endpoint).

7.1.2.5 Visual Analog Scales

Patients with missing scores at baseline will not be included in the analysis. Missing scores at follow-up will be based on LOCF, except for the ACR20-CRP analyses done at Week 22 (the primary endpoint).

7.1.2.6 Safety

As stated above, missing data will not be imputed for any safety analyses, except for missing AE relationship to study drug, AE severity, and AE seriousness. If AE relationship to study drug is missing, the relationship will be imputed to be "related" in any pertinent summary tables, severity will be imputed to be "severe", and seriousness to be "serious." This will be footnoted in the listings.

7.1.2.7 Missing and Partially Missing Dates

For any scheduled assessment at a specific study visit/week, if the date is completely missing, then the actual visit date (if available) or the target visit date (if actual is unavailable) will be used (see Appendix A for target date definitions). If the date is partially missing, then the date will be imputed to using the date of the actual visit (if available) or target date (if actual is unavailable). If a partial date is given that contradicts the actual/target date by specifying a different month/year, the imputed date will be set to the closest day to the target visit date including the given information.

Missing start or end dates for events not aligned to a specific visit/week such as adverse events, medications reported, etc., will be imputed using the following rules:

- A completely missing start date of an event will be set to the date of the first dose of study medication, unless it is clearly indicated that this is an event happening pretreatment. In this case, it will be set to the day before the first dose of study medication.
- A completely missing stop date will not be imputed, but the event will be assumed to be ongoing.
- Partially missing dates where the year is known will be set to 01 January of that year or the first dose of study medication date, if the first dose of study medication occurred in that year.

- Partially missing dates where the year and month are known will be set to the first day of that month or the first dose of study medication date, if the first dose of study medication occurred in that month.

If, by applying the above rules, the stop date is imputed before the start date, then the stop date will be set to be the same as the start date.

7.1.2.8 PK/PD

Only observed data will be used in the PK analysis except for concentration values below the lower limit of quantification (BLQ) as described in Section 7.6.1. No attempt will be made to extrapolate or interpolate estimates for missing data.

7.1.3 Pooling of Investigative Sites

This is a multicenter, multinational study. To analyze the potential impact of different regions on study results, sites will be pooled into 4 regions:

- US
- Puerto Rico
- Eastern Europe
- Western Europe

7.1.4 Determination of Sample Size

The study is stratified into 3 stages. Stage 1 runs from Week 0 to Week 22 to collect data to evaluate biosimilarity, Stage 2 runs from Week 22 to Week 54 to collect data to evaluate interchangeability, and Stage 3 runs from Week 54 through Week 62 to collect safety data. The sample size required to be randomized in Stage 1 was derived with consideration of the required sample size needed for PK analysis at the end of Stage 2 and assumes a 25% drop-out of patients during Stages 1 and 2.

To show PK similarity, the 90% geometric CIs of the ratio (Switch/Remicade-US) of the least squares means from the analysis of variance (ANOVA) of the ln-transformed AUC_{tau} and C_{max} must be within 80.00% and 125.00%. For the PK interchangeability analysis at the end of Stage 2, assuming coefficients of variation (CVs) of 49% and 33% for AUC_{tau} and C_{max}, respectively, and CIs of the ratios between 95% and 105%, at least 109 patients per treatment group are required for an 80% powered trial. [redacted]

[redacted]

In order to meet this Stage 2 requirement, in Stage 1 (at Week 0), approximately 585 patients are expected to be randomized in a ratio of 2:1 to Remicade-US (390 patients) and NI-071 (195 patients). This sample size will provide over 90% power to test for equivalence, using an asymmetric margin (-12%, 15%) and a two 1-sided test (TOST) analysis with α=0.05 for each 1-sided statistical test. Greater than 90% power is achieved when the ACR20-CRP response rate for Remicade at Week 22 is assumed to be 79.4% (the Remicade-US ACR20-CRP result at Week 22 from Study NI071F1 using the study's modified ITT population).

[redacted]

The following provides a justification for margin selection.

The lower -12% margin is justified for the following reasons:

- Combest et al. (2014) indicated "From a clinical perspective, an equivalence margin of [±]15% or less is not a clinically meaningful difference based on a meta-analysis of both infliximab and adalimumab trials with a focus on Phase III trials."[6]
- Smolen et al. (2014), in a review that considered the ACR20 and other responses across a broad range of therapies, suggested that when comparing treatment effects between products, any differences in the effects of one RA treatment on the ACR20 endpoint that are -12% or less (i.e., -12% to -1%) than effects seen with the comparator product would not be considered to represent clinically meaningful differences in effects.[8]
- A lower margin of -12% is more conservative than a -15% lower margin, which was used to test equivalence between a study product and infliximab in patients with RA, as reported by Yoo et al. (2013)[9], and more conservative than a -18% lower margin, which was used to test noninferiority in patients with RA, as reported by Ogata et al. (2014).[7]

The upper 15% margin is justified for the following reasons:

- Combest et al. (2014) indicated "From a clinical perspective, an equivalence margin of [±]15% or less is not a clinically meaningful difference based on a

meta-analysis of both infliximab and adalimumab trials with a focus on Phase III trials."[6]

- Van Vollenhoven et al. (2012) reported results from testing a product against a placebo looking for 20% improvement in ACR20.[10] This suggests that differences in ACR20 that are not at least 20% or more should not be considered to be clinically meaningful. Therefore, a product having a treatment effect that is not more than 15% greater than that observed with a reference comparator product could be considered to be equivalent with the comparator product.

[redacted]

7.1.5 Adjustment for Multiplicity

This study has 2 primary objectives, i.e., to prove biosimilarity by demonstrating equivalent ACR20-CRP response rates between the reference product (Remicade-US) and the investigational product (NI-071), and to prove interchangeability by demonstrating bioequivalence of PK parameters between patients in the Switch group and patients in the Remicade-US group.

As the objectives have a clear hierarchy where biosimilarity is needed first, and only if biosimilarity is demonstrated, demonstration of interchangeability is needed, the multiplicity will be taken care of by a strict hierarchical testing procedure, where the biosimilarity is tested at a significance level of 0.05 first, and only if successful, interchangeability (tested at a significance level of 0.05) will be considered. This procedure controls the overall significance level of 0.05.[11]

No adjustments for multiplicity are necessary following the interim analysis because the interim analysis will be checking for a stop for futility reasons only. Further, the PK analysis for the interchangeability assessment in Stage 2 is based solely on objective variables (plasma concentration data); hence, the introduction of additional bias can be excluded.

7.2 Patient Characteristics

7.2.1 Patient Disposition

A listing containing the analysis population assignment for each patient will be presented by stage, with each patient's ID and a Yes/No column for each of the analysis populations. Stages are defined in Appendix A.

Patient disposition information will be tabulated by stage, summarizing the number of patients screened, randomized, initiated treatment, completed, and discontinued. Also, patient counts in each analysis population will be presented: ITT (Stage 1 only), safety, FAS, PP, and PK (Stage 2 only) populations. This disposition summary will be presented overall and by region.

Patient discontinuation information will be presented by stage, summarizing the reasons for discontinuation (e.g., AE, patient consent withdrawal). This discontinuation table will be presented overall and by region.

7.2.2 Protocol Deviations

As specified in the definition of the PP population, individuals with major deviations concerning the primary efficacy endpoint ACR20-CRP response at Week 22 will be excluded from this analysis set. All decisions regarding major deviations will be discussed and finalized during a Blinded Data Review Meeting (BDRM) prior to breaking the treatment blind and commencing the final analysis on the locked database. Participants of the BDRM will include as a minimum the lead CRA, project manager, statistician, data manager, and the study physician. The decisions will be documented in the BDRM minutes and signed by the Sponsor and the CRO.

For the BDRM, a blinded listing of possible protocol deviations (events sorted by patients and dates of events) will be created to include descriptions of the possible protocol deviation events. At least the following programmatic checks will be made to identify individuals deviating from the protocol: checks for patients not meeting all inclusion criteria and for meeting any exclusion criteria, checks on the use of prohibited concomitant medications, checks on visit window deviations. Other types of protocol deviations that are detected during the course of the study may be added. This listing will be reviewed in the BDRM immediately prior to data lock. The listing will be used to mark the review committee's decision of each possible protocol deviation event as a major or minor protocol deviation. The BDRM determinations of each event will be used by the biostatistician to create population flags for the analysis populations described in Section 4.

A frequency table of the major protocol deviations will be presented by stage for the ITT population. In addition, a by-patient listing of all protocol deviations (identified as major and minor) will be provided.

7.2.3 Background and Demographic Characteristics

Demographic data including age, gender, race, ethnicity, body weight and height, and BMI will be listed and summarized descriptively by treatment group and overall using the ITT, safety, FAS, PP, and PK populations.

Other baseline characteristic data such as disease history, medical conditions, HAQ-DI, CRP (as continuous and categorized parameters), smoking status, and DAS28 (CRP and

ESR, as continuous and categorized parameters) will be summarized descriptively using the FAS and ITT populations.

7.2.4 Treatment Exposure and Compliance

For patients in the ITT population, duration of study treatment exposure (days), the number of infusions, and the percent compliance (derived as the number of infusions received at applicable study visits divided by the number of applicable study visits) will be listed and summarized separately by stage, treatment group, and overall.

Duration of study treatment exposure (days) will be derived as (date of last infusion) – (date of first infusion) + 1.

In addition, the percent compliance will be categorized (>=120%, 80-120%, 60-<80%, 40-<60%, 20-<40% and <20%) and summarized with frequency counts and percentages, by treatment group and overall.

7.2.5 Prior and Concomitant Medications and Therapies

All prescription medications and over-the-counter drugs including vitamins and blood transfusions taken prior to the start of and throughout the study will be recorded and coded using the WHO Drug Dictionary (WHO-DD) version March 2016. Prior medications include medications started before enrollment and stopped prior to the start of study treatment. Concomitant medications include medications taken on or after the start of study treatment.

Prior and concomitant medications will be summarized according to the Anatomical Therapeutic Chemical (ATC) classification system, and tables of descriptive statistics (frequency counts and percentages of patients using these medications) will be provided for the ITT population.

All prior and concomitant medications will be listed by patient. The listing will include verbatim descriptions and coded terms, and flags for prior medications.

7.2.6 Medical Histories

A listing of medical history will be presented by patient.

7.3 Efficacy Analysis

7.3.1 Primary Efficacy Analysis

The primary efficacy analysis will be performed using the ITT population. The primary efficacy endpoint is the ACR20-CRP response at Week 22, and will be calculated as described in Section 6.1. Patients with a missing ACR20-CRP assessment at Week 22, as well as patients who did not receive all study drug treatments thru Week 14, will be considered as having no response (NRI). Because these patients are considered Non-

responders for this analysis, the primary endpoint is considered a composite of both ACR20-CRP response and adherence to study drug treatment.

To determine biosimiliarity, 2 asymptotic one-sided Wald tests of the difference between NI-071 and Remicade-US response rates will be performed. If the null hypotheses of both tests are rejected at $\alpha = 0.05$, the equivalence in efficacy is achieved. The hypotheses that will be evaluated are stated, as follows, where P_{NI-071} is the response rate in the NI-071 arm and $P_{Remicade}$ is the response rate in the Remicade-US arm.

H_o: $P_{NI-071} - P_{Remicade} \leq -12\%$

H_o: $P_{NI-071} - P_{Remicade} \geq 15\%$

H_a: $-12\% < (P_{NI-071} - P_{Remicade}) < 15\%$

The z test statistic for the Wald tests will be calculated as z=(d+margin)/SE(d), where $d = P_{NI-071} - P_{Remicade}$, and the standard error (SE) of d is calculated as:

$SE(d) = \sqrt{(P_{NI-071}(1- P_{NI-071})/n_1 + P_{Remicade}(1- P_{Remicade})/n_2)}$

7.3.2 Secondary Efficacy Analysis

The secondary efficacy analysis will be performed using the FAS population. In addition to the analyses described in the following subsections, all data will be listed.

7.3.2.1 ACR20, ACR50, and ACR70

Response according to ACR20/50/70 will be derived as described in Section 6.1 and analyzed as a dichotomous, categorical variable. Of note, ACR20-CRP at Week 22 will not be included here, but is analyzed using the same method for a sensitivity analysis (described in Section 7.3.4).

Summary statistics (including frequency count, percentage, and 95% CIs) for these items will be presented by visit and treatment group. In addition, these secondary efficacy items will be analyzed using a logistic model, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates for response rates and risk ratios will be calculated and presented by visit (along with associated 95% CIs for each). These will be presented both in table format and a figure. The Hosmer-Lemeshow test will be performed to assess goodness of fit of the logistic model; fit will not be considered good if the p-value is rejected at the 5% level.

Histograms will be provided for the above to visualize ACR20/50/70 response over time.

7.3.2.2 Tender and Swollen Joint Counts

Counts of tender (ranging 0-68) and swollen (ranging 0-66) joints will be analyzed, using continuous and categorical methods.

Summary statistics for these items and change from baseline will be calculated and presented by visit and treatment group. In addition, these secondary efficacy items will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean count for each treatment group and the difference in mean count between treatment groups with associated 95% CIs will be calculated and presented by visit.

7.3.2.3 Visual Analog Scales

Three VAS items will be analyzed as continuous variables.

- Patient global assessment of disease activity VAS (ranging from very well to very poor; 0-100)
- Physician global assessment of disease activity VAS (ranging from none to extremely active; 0-100)
- Patient pain assessment VAS (ranging from no pain to pain as bad as it can be; 0-100)

Summary statistics for these scales and change from baseline will be calculated and presented by visit and treatment group. In addition, these secondary efficacy items will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean score (least squares mean) for each treatment group and the difference in mean score between treatment groups with associated 95% CIs will be calculated and presented by visit.

7.3.2.4 Disability measured with the HAQ-DI

Disability will be analyzed as a continuous variable based on the HAQ-DI which contains 20 questions that measure the patient's functional ability on a 4-level difficulty scale (0 to 3, with 0 representing normal or no difficulty, and 3 representing inability to perform). Eight categories of functioning are included: dressing/grooming, arising, eating, walking, hygiene, reach, grip, and usual activities. Analysis will be performed for the overall index score.

The score for each of the eight functional categories will be the single response within the category with the highest score (greatest difficulty). For example, in the "Walking" category, there are 2 component questions (1 for each item). If "Walk outdoors on flat ground" is marked as "3" and "Climb up five steps" is marked as "0", then the score for the "Walking" category would be "3" (the response indicating the greatest difficulty within the category). If a component question is left blank or the response is too ambiguous to assign a score, then the score for that category is determined by the remaining completed question(s), per HAQ-DI scoring instructions. If responses to all component questions are missing, then scores will be based on LOCF.

The score for the overall disability index is the mean of the eight category scores. If 2 or more of the categories are missing, then the mean score will be based on LOCF. If only 1 of the categories is missing, then the mean is calculated by dividing the sum of the categories by 7 (instead of 8). Higher scores indicate greater disability.

Summary statistics for the overall disability index and change from baseline will be calculated and presented by visit and treatment group. In addition, this secondary efficacy item will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean index (least squares means) for each treatment group and the difference in mean index between treatment groups with associated 95% CIs will be calculated and presented by visit.

7.3.2.5 DAS28-CRP and DAS28-ESR

The DAS28 variables will be calculated as described in Section 6.2 and will be analyzed using continuous and categorical methods.

Summary statistics for the DAS28 variables and change from baseline will be presented by visit and treatment group. In addition, these secondary efficacy variables will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean value (least squares means) for each treatment group and the difference in mean value between treatment groups with associated 95% CIs will be calculated and presented by visit.

In addition to providing the continuous-based analysis (above), values will be stratified into 4 disease activity categories: minimum (defined as values <2.6), low (defined as ≥2.6 to <3.2), moderate (defined as values ≥3.2 to ≤5.1), and high (defined as >5.1). A frequency with counts and percentages will be presented for each category. These analyses will be presented by visit and treatment group. Additionally, shift tables from Baseline will be provided by visit and treatment group.

7.3.2.6 RAPID3

The RAPID3 score will be calculated by summing responses from the 3 parts: multidimensional HAQ (MDHAQ) physical function (range 0–10), pain intensity score (range 0–10), and patient global assessment of disease activity score (range 0–10). Thus, the range of the RAPID3 score is 0-30, and disease activity categories are as follows: remission 0 to 3, low >3.1 to 6, moderate >6.1 to 12, and high >12.

This variable will be analyzed using continuous and categorical methods. Summary statistics for the variable and change from baseline will be presented by visit and treatment group. In addition, this secondary efficacy variable will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean values (least

squares means) for each treatment group and the difference in mean values between treatment groups with associated 95% CIs will be calculated and presented by visit.

In addition to providing the continuous-based analysis (above), values will be stratified into 4 disease activity categories (remission, low, moderate, and high). A frequency with counts and percentages will be presented for each category. These analyses will be presented by visit and treatment group. Additionally, shift tables from Baseline will be provided by visit and treatment group.

7.3.2.7 SF-36

Patient health-related quality of life will be analyzed as a continuous variable based on the SF-36 which contains 36 questions that measure how patients' physical and psychological states affect their lives. Eight quality of life sub-scales are included: physical function, role physical, bodily pain, global health, vitality, social function, role emotional, and mental health. Analysis will be performed for the overall score, for 2 higher-order summary scores (i.e., the physical component summary and the mental component summary), and for each of the eight sub-scales. Derivation of these scores will be done by reverse coding some question responses so that high scores indicate better health, summing scores across responses in the same sub-scale (raw scale score), and transforming the raw scale scores to a 0-100 scale .[12]

Summary statistics for the scores and change from baseline will be presented by visit (Weeks 14, 22, 38, and 62) and treatment group. In addition, these secondary efficacy variables will be analyzed using an ANCOVA, with CRP value at baseline (as a continuous variable), region, and treatment arm as covariates. From this model, adjusted point estimates of the mean score for each treatment group and the difference in mean value between treatment groups with associated 95%CIs will be calculated and presented by visit. Mean changes from baseline by treatment group will also be represented in a bar graph or a figure for the overall and sub-scale scores.

Data from SF-36 questionnaires collected at unscheduled visits will be listed.

Regarding missing data, if an individual question response is missing, then the response will be based on LOCF. After individual LOCF, if more than 20% of the individual questions used to derive a sub-scale score are missing, then the sub-scale sore will be based on LOCF. If 2 or more of the sub-scale scores are missing, then the overall mean score, as well as the 2 higher-order summary scores, will be based on LOCF.

7.3.3 Subgroup Analysis of Efficacy Endpoints

Subgroup analyses will be performed for the primary efficacy analysis of ACR20-CRP at Week 22. These analyses will be performed using the FAS population. Summary statistics (including frequency counts, percentages, and 95% CIs) will be presented by treatment group. Furthermore, the difference between treatment groups and 95% CIs for the difference will be presented. A forest plot will be provided. For subgroups with more

than 2 categories, dummy categorizations, i.e., respective category yes/no, will be used.

[redacted]

7.3.4 Sensitivity Analysis of Efficacy Endpoints

Five types of sensitivity analyses will be performed for the primary efficacy analysis of ACR20-CRP at Week 22.

The first type of sensitivity analysis will be repeating the analysis described in Section 7.3.1 in a descriptive manner using the ITT population. However rather than applying the non-responder imputation (NRI) for patients who did not receive all treatments through Week 14, data actually collected will be used to determine response status at Week 22. This sensitivity analysis addresses the composite nature of the primary endpoint.

The second type of sensitivity analysis will be repeating the analysis described in Section 7.3.1 in a descriptive manner using the FAS and PP populations, rather than the ITT population. For the FAS-based analysis, both types of approaches will be presented, i.e., with and without NRI for patients who discontinue study drug on or before Week 14.

The third type of sensitivity analysis will be repeating the analysis described in Section 7.3.1 in a descriptive manner using the ITT, FAS, and PP populations. However, rather than not allowing for imputation of missing ACR20 components, the mixed component LOCF method for calculating ACR20 response status will be employed.

The fourth type of sensitivity analysis will be modeling to evaluate if there is any impact on the primary analysis from baseline disease status or geographical region, using the ITT, FAS, and PP populations. To adjust for any impact due to baseline disease status and/or region, a logistic regression model will be used with ACR20-CRP at Week 22 as a dependent variable, and treatment group, baseline CRP values, and region as covariates. From this model, adjusted point estimates for ACR20-CRP response rates and risk ratio will be calculated and presented by visit (along with associated 95% CIs for each). The Hosmer-Lemeshow test will be performed to assess goodness of fit. The fit will not be considered good if the p-value is rejected at the 5% level. For this analysis, the NRI will be applied as done for the primary analysis.

The fifth type of sensitivity analysis will be a tipping point analysis to be performed if the primary analysis in the ITT population concludes biosimilarity (i.e., rejection of both Wald tests using asymmetric margins -12% and 15%). This analysis will be performed to identify the hypothetical response rates required among the patients with missing data in order to produce statistics that no longer support similarity (patients who stop treatment but continue to be followed and have an efficacy evaluation at Week 22 are not considered to have missing data). In the planned primary analysis, a 0% response rate is assumed in both treatment groups of all patients randomized with missing data at Week 22 (i.e., all patients with missing data are considered non-responders). For the first step of this sensitivity analysis, instead of assuming a 0% response rate, the observed response rate for the Remicade-US arm in the PP population will be assumed for both treatment groups. Using the new Remicade-US response rate assumption, the primary analysis will be conducted, and the Wald p-values will be recorded; it is expected that this initial step will result in 2 significant p-values to accept the alternative hypothesis (i.e., difference within -12% and 15%), consistent with the finding of the primary

analysis. The second step of this sensitivity analysis will assign non-response status to one more NI-071 patient than in the first step. Using this unbalanced response rate, the primary analysis will be conducted, and the Wald p-values will be recorded. The remaining steps of this sensitivity analysis will iterate as described in the second step (each time re-assigning non-response status to one more NI-071 patient, considering the patient to be a non-responder rather than a responder). These steps will continue until the maximum number of NI-071 patients is achieved, or the Wald test shows non-similarity ($p > 0.05$). The maximum number of patients will be defined as the total number of missing patients in the NI-071 arm. For each iteration of this sensitivity analysis, the following statistics will be presented: frequency count and response rate percentage by treatment arm at Week 22, both overall and by missing-status; Wald p-values; and the percent difference in response rates within the original missing subset. Percent difference will be calculated as: (response rate in the Remicade-US arm minus response rate in the NI-071 arm) divided by the response rate in the Remicade-US arm. An example of interpretation of this percentage follows: if Remicade-US response rate is 79% and NI-071 response rate is 50% at the tipping point, then it may be stated that if the NI-071 response was 36.7% lower (i.e., (79-50)/79 = 0.367) than the Remicade-US arm among the patients with missing values at Week 22, biosimilarity could not be concluded. Also, the difference in response rate between patients with missing data and patients with non-missing data will be presented for each iteration.

The following tabulation summarizes the above mentioned sensitivity analyses:

	Description	Purpose	Population	Imputation of missing data
1	Primary analysis (Wald-test)	Repeat for different population / use of collected data from patients who discontinued study drug	ITT	No NRI for patients who discontinued study drug.
2	Primary analysis (Wald-test)	Repeat for different population	FAS / PP	With and without NRI for FAS
3	Primary analysis (Wald-test)	Applying new imputation method	ITT / FAS / PP	Mixed component LOCF method
4	Logistic regression model	To adjust for any impact due to baseline disease status and/or region	ITT / FAS / PP	NRI as for the primary analysis
5	Tipping point analysis	Assess where the conclusion switches direction	ITT	NRI as for the primary analysis / Based on the Remicade-US PP population response rate

7.4 Safety Analysis

The analysis planned for each safety evaluation/endpoint is described in the following subsections and will be performed using the safety population.

7.4.1 Adverse Events

All reported AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA version 19.0). The incidence of TEAEs (events with onset dates on or after the start of the study drug) and treatment-related TEAEs will be summarized by system organ class (SOC) and preferred term (PT). Events with missing onset dates will be included as treatment-emergent, assuming onset on the first day of study treatment. If a patient experiences more than 1 occurrence of the same AE, the occurrence with the greatest severity and the closest association with the study drug will be used in the summary tables. SAEs and AEs that result in treatment discontinuation will be summarized. AEs of special interest will be summarized by SOC and PT; categories of

AEs of special interest include serious infections, malignancies, infusion-related reactions, hepatotoxicity, heart failure, and cytopenias. In addition, CIs will be calculated for the difference in incidence of these AEs to determine the risk difference between the NI-071 and Remicade-US groups.

All AEs will be listed by patient, along with information regarding onset, duration, severity, relationship with study drug, action taken with study drug, treatment of event, and outcome.

In addition to the overall analysis described above, TEAEs and AEs of special interest will be summarized by system organ class and preferred term separately for each Stage (Stage 1, Stage 2, and Stage 3). These by-stage analyses will be stratified into 3 timing/severity categories: (1) stage-emergent (onset date within the stage), (2) persistent without increasing severity within the stage, and (3) persistent with increasing severity within the stage. One AE may be reported as "stage-emergent" in the analysis of Stage 2 and also be reported as "persistent with increasing severity" in the analysis of Stage 3. Stages (for this safety analysis of AEs) will be defined according to the following table:

Stage	Beginning Date	Ending Date
Stage 1	First treatment in Stage 1	One day prior to first treatment in Stage 2
Stage 2	First treatment in Stage 2	One day prior to first treatment in Stage 3
Stage 3	First treatment in Stage 3	Day of study completion

7.4.2 Physical Examination

Physical examinations will be performed at visits according to the assessment schedule (Appendix B). A detailed data listing of abnormal physical examination findings will be provided.

7.4.3 Vital Signs and Body Weight

Vital signs data (oral body temperature, pulse rate, systolic and diastolic blood pressure measurements) and body weight, as well as change from baseline in vital sign and weight data, will be presented using the summary statistics for continuous variables, by treatment group and overall.

7.4.4 Screening Tests

At screening, Week 30, and Week 62 (or at the early discontinuation visit), a tuberculosis test will be performed, and a chest x-ray will be performed when applicable. Summary statistics (patient count and frequency) for the TB results (positive/negative) will be presented by treatment group and overall.

For women of childbearing potential, a serum pregnancy test will be performed at screening, and urine pregnancy tests will be performed at all dosing visits and at Week 62 (or at the early discontinuation visit). Summary statistics (patient count and frequency)

for the pregnancy results (positive/negative) will be presented by treatment group and overall.

7.4.5 Electrocardiogram

At screening and at the early discontinuation visit (if applicable), patients will have an electrocardiogram. Summary statistics (patient count and frequency) for results (normal/abnormal) will be presented by treatment group and overall.

7.4.6 Laboratory Parameters

Samples for assessing the clinical laboratory parameters listed in the following table are to be obtained at the visits designated in the schedule of assessments.

Hematology	**Serum chemistry**	**Urine analysis (dipstick)**
Full and differential blood count Hematocrit (Hct) Hemoglobin (Hb) Mean corpuscular hemoglobin (MCH) Mean corpuscular hemoglobin concentration (MCHC) Mean corpuscular volume (MCV) Platelet count Red blood cell (RBC) count White blood cell (WBC) count with differential Glycated hemoglobin (HbA1C) (at screening only)	Albumin Alanine aminotransferase (ALT) Alkaline phosphatase (ALP) Aspartate aminotransferase (AST) Blood urea nitrogen (BUN) or urea Creatinine Electrolytes (Na, K, Cl, Ca, P) Gamma-glutamyl transpeptidase (GGT) Glucose Lactate dehydrogenase (LDH) Total bilirubin Direct bilirubin	Appearance pH Protein Glucose Ketone bodies Indicators of blood and WBCs Urobilinogen
Serum immunology tests: At screening, HBsAg, HBsAb, HBcAb, HCV Ab, HIV Ab, anti-dsDNA, and β-D-glucan (using the Fungitell® assay) will be measured.		
RA tests: Rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA) tests will be done at screening.		
Pregnancy test: For women of childbearing potential, a serum pregnancy test will be done at screening, and urine dipstick tests will be done at all dosing visits and at Week 62 (or the early discontinuation visit).		
QuantiFERON tuberculosis (TB) test: This test will be performed for all patients at screening, Week 30, and Week 62 (or the early discontinuation visit). A dedicated incubator must be available at the site for storing samples at 37ºC prior to shipping them to the central laboratory (at room temperature or refrigerated) within 3 days.		
C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) tests: CRP and ESR levels will be analyzed at every visit. CRP will be analyzed by the central laboratory, and ESR will be analyzed locally.		

Data for each hematology and serum chemistry parameter will be summarized using the summary statistics for continuous data, by visit and treatment group. The change from baseline to each visit will be calculated and summarized. Urinalysis data will be summarized using counts and percentage, by visit and treatment group.

Shift tables from baseline to final visit (indicating any shifts) will be produced.

For hematology and serum chemistry, by-patient listings of all laboratory results will be produced (abnormal values and clinically significant values as rated by the investigator will be flagged).

In addition, listing of abnormal laboratory results will be produced, clinically significant values as rated by the investigator will be flagged.

7.5 Anti-drug Antibodies and Neutralizing Antibodies

Positive immunogenicity results will be listed. The incidence of ADAs and the incidence of neutralizing antibodies will be summarized, and the titer of ADA will be summarized over time by treatment group.

7.6 PK/PD Analysis

The PK population will be used for the analysis of PK data.

7.6.1 Handling of the BLQ and the No Reportable Concentration Values

All concentration values BLQ will be set to zero. Samples with no reportable value will be set to missing for tabulation, graphical representation and calculation purposes.

7.6.2 Handling of the Difference between the Scheduled and the Actual Sampling Times

The actual clock time for dosing and the actual clock time for each collection time for the PK samples will be recorded in the eCRF. For all sampling times, the actual sampling times will be calculated as the difference between the sample collection actual clock time and the actual clock time of dosing. The actual post-dose sampling times expressed in hours and rounded off to three decimal digits will be used to calculate the PK parameters, except for pre-dose samples occurring prior to dosing, which will always be reported as zero (0.000), regardless of the time difference. Scheduled sampling times will be presented in concentration tables and mean graphs, while actual sampling times will be presented in the individual graphs in the PK section of the report.

A listing of the actual times for PKs will be provided for PK samples. Actual sampling times will be used in all cases for PK analyses; however samples taken out of windows will be listed as deviations in the report.

7.6.3 Pharmacokinetic Parameters

Trough PK samples will be collected at Weeks 0, 2, 6, 14, 22, 30, 38, 46, 54, and 62. Trough PK sampling should be done within 1 h prior to dosing.

Full PK sampling during Weeks 46 to 54 includes the following time points: pre-dose, 1 h after infusion start, at the end of infusion (EOI), at 4 h and 24 h after infusion start, and at 4 days, 7 days (Week 47), 14 days (Week 48), 28 days (Week 50), 42 days (Week 52), and 56 days (Week 54) post-dose. Sampling windows for the blood draws are ±15 minutes for the 1 h time point, +15 minutes for the EOI, and ±15 minutes for the 4 h time point; ±2 h for the 24 h time point; and ±24 h for the subsequent time point.

Infliximab serum concentrations collected on Weeks 46 to 54 will be used to calculate the following parameters by standard non-compartmental methods:

AUC_{tau}: area under the concentration-time curve for one dosing interval (tau=8 weeks)

C_{max}:	maximum concentration
C_{min}:	minimum concentration
$T_{1/2}$:	terminal half life
K_{el}:	elimination rate constant
CL:	total clearance
Vd:	volume of distribution
t_{max}:	time of observed Cmax ss at steady-state

Some PK parameters may not be calculated for all or some patients, at the discretion of the Syneos Health pharmacokineticist, if the concentration data is not deemed to be amenable to evaluation. Explanations for PK parameters that could not be estimated will be provided in the CSR report.

7.6.4 Statistical Analyses

Individual and mean concentration versus time curves will be presented using linear and semi-log scales for infliximab. Descriptive statistics of the serum concentrations versus time will be presented, including trough concentrations, as well as for all PK parameters (AUC_{tau}, C_{max}, C_{min}, K_{el}, CL, Vd, and t_{max}) by treatment group.

Using GLM procedures in SAS, an ANCOVA will be performed on ln-transformed AUC_{tau} and C_{max} with TREATMENT as fixed effect and REGION as categorical covariate at the alpha level of 0.05. Based on pairwise comparisons of the ln-transformed AUC_{tau} and C_{max} data, the geometric least-squares means for each treatment, the ratios (Switch/Remicade, Switch/NI-071, and NI-071/Remicade) of the geometric least-squares means, calculated according to the formula "$e^{(X-Y)} * 100$", as well as the corresponding 90% geometric confidence intervals will be determined. Coefficient of variation will be estimated.

Wilcoxon rank sum test will be performed on t_{max} for each comparison.

The analysis for each comparison (Switch versus Remicade, Switch versus NI-071, and NI-071 versus Remicade) will be conducted excluding the data from the treatment that is not relevant for the comparison.

Criteria for Interchangeability Assessment

The 90% geometric confidence intervals of the ratio (Switch/Remicade) of least-squares means from the ANCOVA of the ln-transformed AUC_{tau} and C_{max} must be within 80.00% to 125.00%. ANCOVA will be performed with TREATMENT as fixed effect and REGION as categorical covariate.

7.7 Analysis of Other Assessments

No other analyses are planned.

7.8 Interim Analysis

An interim analysis is planned in order to decide if the study should continue or be stopped for futility after Stage 1.

This study was designed using a hierarchical approach, where the results of the interchangeability assessment (PK analysis) performed at the end of Stage 2 will be regarded as confirmatory in nature only if the results of the biosimilarity assessment at the end of Stage 1 have shown significant results. An interim analysis is planned in order to decide if the study should continue or be stopped for futility after Stage 1.

The interim analysis of the primary clinical efficacy endpoint, the ACR20-CRP response rate, will be performed after the last patient has completed the last visit in Stage 1, the Week 22 visit (the visit of the assessments for the primary efficacy endpoint). This analysis will be performed by an unblinded statistician not otherwise affiliated with the study. An independent data monitoring committee (IDMC) will be established to assess the results and determine whether the study should continue or be stopped for futility. An IDMC charter will be prepared to describe the details of the outputs available to the IDMC as well as to describe the decision tree for the IDMC meeting.

7.9 Data Monitoring Committee

An IDMC will be established by the Sponsor or its designee to evaluate the results of a formal interim efficacy analysis. Members will include experts in RA and biostatistics who are not participating in this study and do not have affiliation with the Investigators or the Sponsor. The IDMC can recommend in writing to the Sponsor whether to continue or stop the clinical study on the basis of the results of the interim analysis. The IDMC's specific duties as well as statistical monitoring guidelines and procedures will be fully described in an IDMC charter.

7.10 Changes to Methods Planned in the Protocol

No changes to the methods described in the protocol are planned.

However, the definition of the per protocol population in Section 4.4 have been updated to only exclude such patients from the per protocol population that have major protocol deviations that are classified as having an impact on the primary efficacy outcome ACR20-CRP at Week 22 as defined during the BDRM before unblinding of the study.

8. TABLES, LISTINGS, AND FIGURES

The mock tables, listings, and figures will be presented in a separate document.

The general layout of the tables, figures and listing (TFLs) will be as follows.

Orientation	Landscape
Paper Size	A4
Margins	Top: 3.2 cm Bottom: 2.5 cm Left: 2.5 cm Right: 2.5 cm
Font	Courier New 9pt
Headers (alignment)	Sponsor name and Protocol number (Center) Page X of Y (Right) TFL Number and Title (Center)
Footers (alignment)	SAS program path and file name (Left) Date output generated (Right)

The font size may be reduced as necessary to allow additional columns to be presented, but not at the expense of clarity. Also, the orientation may be changed to portrait, if appropriate.

All tables and data listings will be generated using SAS® v9.3 or higher.

The date format for all presentations will be ddMmmyyyy.

8.1 Preparation of Tables

Certain data will be collated into summary tables; however, in general, most tables will present summary data stratified by treatment group. Some tables may be stage-specific, while others will present comprehensive data spanning all stages.

8.2 Preparation of Data Listings

The data listings will be sorted by treatment group, site number, patient screening number, and visit.

9. REFERENCES

1. Singh JA, Saag KG, Bridges SL, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis and Rheum. 2016;68(1):1-26. Available at www.rheumatology.org/Portals/0/Files/ACR%202015%20RA%20Guideline.pdf
2. Centers for Disease Control and Prevention website: Rheumatoid arthritis. Available at http://www.cdc.gov/arthritis/basics/rheumatoid.htm Accessed 03 Feb 2016.
3. Remicade (US package insert), Horsham, PA: Janssen Biotech; Oct 2015.
4. Remicade (EMA European Public Assessment Report [EPAR] Product Information), Leiden, The Netherlands: Janssen Biologics BV; Nov 2015.
5. NI-071 Investigator's Brochure, Version 1.0. Nichi-Iko; March 2016.
6. Combest AJ, Wang S, Healey BT, et al. Alternative statistical strategies for biosimilar drug development. GaBI Journal. 2014;3(1):13-20.
7. Ogata A, Tanimura K, Sugimoto T, et al. Phase III study of the efficacy and safety of subcutaneous versus intravenous tocilizumab monotherapy in patients with rheumatoid arthritis. Arthritis Care and Research. 2014;66(3):344-354.
8. Smolen JS, Emery P, Fleischmann R, et al. Adjustment of therapy in rheumatoid arthritis on the basis of achievement of stable low disease activity with adalimumab plus methotrexate or methotrexate alone: the randomised controlled OPTIMA trial. 2014;383:321-332.
9. Yoo DH, Hrycaj P, Miranda P. 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):1613-1620.
10. Van Vollenhoven RF, Fleischmann R, Cohen S, et al. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. N Engl J Med. 2012;367:508-519.
11. Koch G and Gansky S. Statistical considerations for multiplicity in confirmatory protocols. Drug Information Journal. 1996;30:523–533.
12. Ware JE, Kosinski M, Dewey JE, and Gandek B. SF-36 health survey: manual and interpretation guide. 2000; Quality Metric Inc.
13. Remsima (EMA European Public Assessment Report). Budapest, Hungary: Celltrion Healthcare; 27 Jun 2013.
14. 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;67:1096-1103.

10. APPENDICES

Appendix A. Definitions of Visit Windows

As described in the protocol, for visits at Weeks 2 through 62, there will be a target visit window of ±3 days.

The following visit windows will be used for efficacy analysis, where day is defined as Visit Date minus date of first infusion at Stage 1. For analysis, the window definitions are derived to provide non-overlapping, consecutive windows. For example, Week 16 is the midpoint between Week 14 and Week 18; therefore, the end of the Week 14 window is 7 days multiplied by 16, and the beginning of the next window is +1 day.

Window Label	Target Day	Target Window (±3 days)	Window Definition (days)
Baseline	1	Not applicable	(-3) – 1
Week 2	14	11-17	2-28
Week 6	42	39-45	29-70
Week 14	98	95-101	71-112
Week 18	126	123-129	113-140
Week 22	154	151-157	140- 168
Week 26	182	179-185	169-196
Week 30	210	207-213	197-224
Week 34	238	235-241	225-252
Week 38	266	263-269	253-280
Week 42	294	291-297	281-308
Week 46	322	319-325	309-336
Week 50	350	347-353	337-364
Week 54	378	375-381	365-392
Week 58	406	403-409	393-420
Week 62	434	431-437	>420

Appendix B. Schedule of Assessments

Schedule of Assessments

	Screening	Stage 1						Stage 2								Stage 3	
Visit at Week (±3 days)	**(Day -28 to Day -7)**	**0 (Day 1)**	**2**	**6**	**14**	**18**	**22**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**	**62[a]**
Informed consent	X																
Inclusion/exclusion criteria	X	X															
Medical history	X																
Demographics	X																
NYHA classification	X																
Vital signs, weight, height[o]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Chest x-ray	X[b]								X[b]								X[b]
Physical examination	X	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]	X[c]
Electrocardiogram	X																X[l]
Clinical laboratory tests[d]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum immunology tests[e]	X																
Pregnancy test[f]	X	X	X	X	X		X		X		X		X		X		X
QuantiFERON TB test	X								X								X
RF and ACPA tests[g]	X																
Joint assessment[g,h,i,m]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2010 ACR/EULAR RA classification	X																
CRP test[g,h,i]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ESR test[g,h,i]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

	Screening	Stage 1						Stage 2								Stage 3	
Visit at Week (±3 days)	**(Day -28 to Day -7)**	**0 (Day 1)**	**2**	**6**	**14**	**18**	**22**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**	**62[a]**
HAQ-DI[h]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Patient global assessment[h,i,j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Patient pain assessment[h,j]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RAPID3		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physician global assessment[h,j]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SF-36		X			X		X				X						X
Concomitant medication review	X[n]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Randomization		X*					X**										
Administration of study drug		X	X	X	X		X		X		X		X		X		
MTX and folic acid reminder	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Adverse event monitoring		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Anti-drug antibodies (pre-dose)		X	X	X	X		X		X		X		X		X		X
Trough PK sample[k] (pre-dose)		X	X	X	X		X		X		X		X		X		X
Full PK sample[k] (Weeks 46 to 54)													X	X	X		

Abbreviations: ACPA=anti-citrullinated protein antibody; ACR=American College of Rheumatology; ACR20(50, 70)=a 20%(50%, 70%) improvement from baseline in the ACR core set criteria; anti-dsDNA=anti-double-stranded deoxyribonucleic acid; CRP=C-reactive protein; DAS28=disease activity score involving 28 joints; EOI=end of infusion; ESR=erythrocyte sedimentation rate; EULAR=European League Against Rheumatism; HAQ-DI=Health Assessment Questionnaire-Disability Index;HbA1c=glycated hemoglobin; HBcAb=hepatitis B core antibody; HBsAb=hepatitis B surface antibody; HBsAg=hepatitis B surface antigen; HCV Ab=hepatitis C antibody; HIV Ab=human immunodeficiency virus antibody; NYHA=New York Heart Association; PK=pharmacokinetic; RA=rheumatoid arthritis; ; RAPID3=routine assessment of patient index data 3; RF=rheumatoid factor; SF-36=Medical Outcomes Study 36-Item Short-Form Health Survey; TB=tuberculosis

[a]Or early discontinuation visit.

[b]Chest x-rays (anterior-posterior and lateral views) are to be done only if the TB test is positive; also, at screening, only if the patient has not had x-rays within 12 weeks prior to screening.

	Screening	Stage 1						Stage 2								Stage 3	
Visit at Week (±3 days)	**(Day -28 to Day -7)**	**0 (Day 1)**	**2**	**6**	**14**	**18**	**22**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**	**62[a]**

[c]A limited physical examination to verify continued patient eligibility and to follow up any changes (symptom-driven) will be performed at these visits.
[d]Standard hematology, blood chemistry, and urinalysis tests; at screening, also HbA1C.
[e]Screening serum immunology tests include HBsAg, HBsAb, HBcAb, HCV Ab, HIV Ab, anti-dsDNA, and β-D-glucan.
[f]Only in women of childbearing potential. Serum test is to be done at screening; urine test is to be done locally at other specified time points.
[g]Assessment contributes to the 2010 ACR and EULAR RA classification at screening.
[h]Assessment contributes to the ACR core set criteria for determining ACR20, ACR50, and ACR70 at all post-screening visits.
[i]Assessment contributes to the DAS28 scores at all post-screening visits.
[j]Measured on a visual analog scale (VAS) of 0 to 100 mm. VAS for patient assessment of pain goes from "no pain" to "pain as bad as it can be"; VAS for patient global assessment of disease activity goes from "very well" to "very poor"; VAS for physician global assessment of disease activity goes from "none" to "extremely active."
[k]Trough PK sampling should be done within 1 h prior to dosing. Full PK sampling during Weeks 46 to 54 includes the following time points: pre-dose, 1 h after infusion start, at the end of infusion (EOI), at 4 h and 24 h after infusion start, and at 4 days, 7 days (Week 47), 14 days (Week 48), 28 days (Week 50), 42 days (Week 52), and 56 days (Week 54) post-dose. Sampling windows for the blood draws are ±15 minutes for the 1 h time point, +15 minutes for the EOI, and ±15 minutes for the 4 h time point; ±2 h for the 24 h time point; and ±24 h for the subsequent time points.
[l]At early discontinuation visit only.
[m]Joint assessment includes determination of tender joints and swollen joints. At screening, counts of tender and swollen joints out of 28 selected joints will be performed (for the 2010 ACR and EULAR classification criteria). At all subsequent visits, counts of tender and swollen joints out of 68 and 66 selected joints, respectively, will be performed (for the ACR core set criteria and the DAS28).
[n]At screening, prior medications will also be reviewed.
[o]Record height at screening only.
*Initial blinded randomization is to be performed after all inclusion/exclusion criteria are reconfirmed at Week 0.
**Blinded re-randomization. At Week 22, the original Remicade-US group is to be randomized 1:1 to a Switch group and a Remicade-US group. The Remicade-US group will continue on Remicade-US from Week 22 through Week 54 every 8 weeks. At Week 22, the Switch group will receive NI-071; at Week 30, this group will receive Remicade-US; and at Weeks 38, 46, and 54, this group will receive NI-071. The original NI-071 group will continue on NI-071 from Week 22 through Week 54 every 8 weeks.