Statistical Analysis Plan for Study M16-063

A Phase 2 Study to Investigate the Safety and Efficacy of ABBV-105 Given Alone or in Combination with Upadacitinib (ABBV-599 Combination) with a Background of Conventional Synthetic DMARDs in Subjects with Active Rheumatoid Arthritis with Inadequate Response or Intolerance to Biologic DMARDs

Date: 23 January 2020

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for ABBV-599 Study M16-063.

Study M16-063 examines the efficacy and safety of ABBV-105 given alone or in combination with upadacitinib (ABBV-599 combination) on a background of conventional synthetic disease modifying anti-rheumatic drugs (csDMARDs) in subjects with active rheumatoid arthritis with inadequate response or intolerance to biologic DMARDs (bDMARDs).

The analyses of pharmacokinetics and biomarkers, if applicable, will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The main objective of this dose-exploratory study is to evaluate the safety and efficacy of ABBV-105 and ABBV-599 (ABBV-105 plus upadacitinib) vs placebo on a background of csDMARDs for the treatment of signs and symptoms of RA at 12 weeks in bDMARD-inadequate response (bDMARD-IR) or bDMARD-intolerant subjects with moderately to severely active RA and to define optimal dose(s) for further development.

Clinical Hypothesis

ABBV-105 given alone or in combination with upadacitinib (as the ABBV-599 combination therapy) should be effective in decreasing signs and symptoms

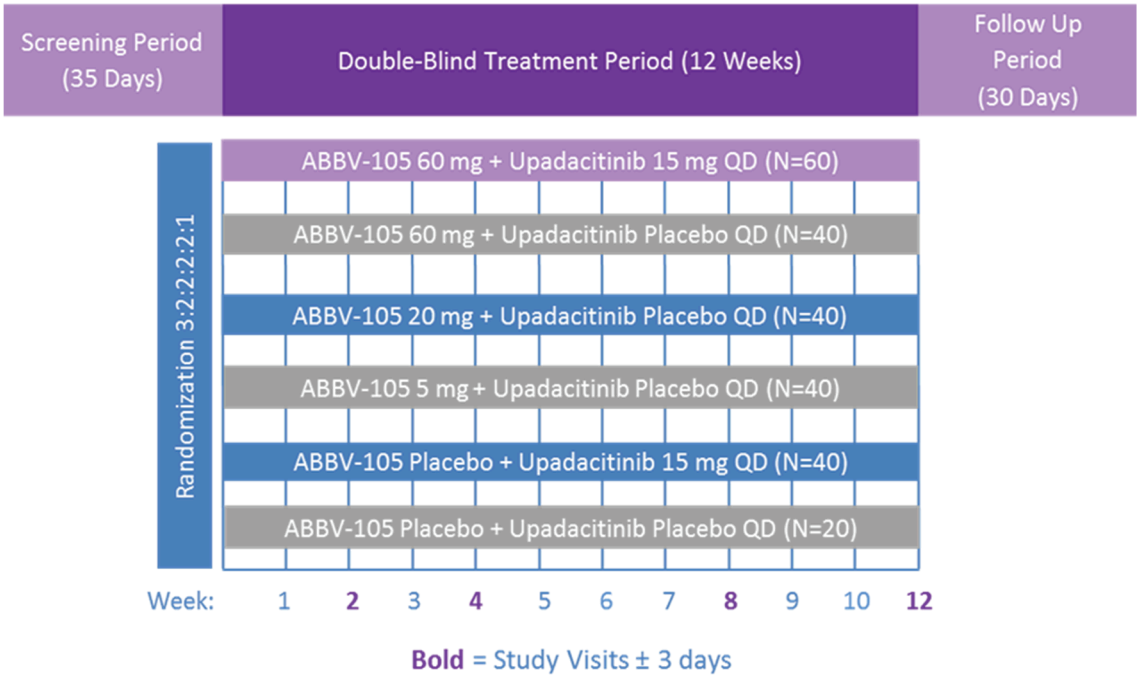
associated with active RA in patients with inadequate response or intolerance to bDMARDs by interfering with the JAK1/Btk pathways. Concurrent inhibition of JAK1/Btk pathways in RA may increase the proportion of those responding as well as the depth of response, relative to inhibiting either pathway alone, while maintaining an acceptable safety profile.

2.2 Study Design Overview

This is a 12-week randomized, double-blind, parallel-group Phase 2, dose-exploratory, multicenter study designed to assess the safety and efficacy of ABBV-105 and ABBV-599 in subjects with active RA who have had inadequate response to bDMARD therapy and are on stable background csDMARD treatment.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



QD = once daily

2.3 Treatment Assignment and Blinding

Subjects who meet eligibility criteria will be randomized in a 3:2:2:2:2:1 ratio to 1 of 6 treatment groups:

- Group 1: Upadacitinib 15 mg and ABBV-105 60 mg once daily (QD) (n = 60)
- Group 2: ABBV-105 60 mg and upadacitinib placebo QD (n = 40)
- Group 3: ABBV-105 20 mg and upadacitinib placebo QD (n = 40)
- Group 4: ABBV-105 5 mg and upadacitinib placebo QD (n = 40)
- Group 5: Upadacitinib 15 mg and ABBV-105 placebo QD (n = 40)
- Group 6: ABBV-105 placebo and upadacitinib placebo QD (n = 20)

The study duration will include a 35-day maximum screening period; a 12-week randomized, double blind, parallel-group treatment period with 30-day follow-up. Study visits will be conducted at Screening, Baseline, Week 2, Week 4, Week 8 and Week 12. For post-baseline visits, a visit window of ± 3 days will be allowed.

Randomization will be stratified by number of prior bDMARDs used (failed 1 or 2 biologics with the same mechanism of action; failed ≥ 3 biologics with the same mechanism of action and/or ≥ 2 biologics with multiple mechanisms of action). Once approximately 35% of the total subjects have been randomized to the group who failed ≥ 3 biologics with the same mechanism of action and/or ≥ 2 biologics with multiple mechanisms of action, further screening of those subjects may be suspended.

2.4 Sample Size Determination

The planned total sample size is 240. Forty (40) patients per arm of ABBV-105 can provide approximately 83% power to detect a –0.88 difference in DAS28 (CRP) change from baseline for ABBV-105 vs placebo combined with 20 borrowed historical placebo subjects (assuming a placebo mean change from baseline –0.77, and ABBV-105: –1.65 with common SD = 1.5) using two group t-test with one-sided significant level $\alpha = 0.05$. It can provide a minimum of 85% power to detect a dose-response model for ABBV-105 and combined placebo from 6 candidate models: Linear,

EMax (ED50 = 5 mg), Exponential (δ = 20), Logistic (ED50 = 5 mg, δ = 10), SigEMax (ED50 = 5 mg, h = 2), and Quadratic (δ = –0.01) using MCP-Mod with one-sided significant level α = 0.05.

The planned sample size of 60 patients treated with ABBV-599 provides approximately 94% power to detect a –1.0 difference in DAS28 (CRP) LS mean change from baseline for ABBV-599 vs the group of 40 patients treated with ABBV-105 (assuming ABBV-105 mean change from baseline: –1.65 and ABBV-599: –2.65 with common SD = 1.5) using two group t-test with one-sided significance level α = 0.05. The sample size of 60 for ABBV-599 vs 20 for placebo is over-powered due to the necessary power for ABBV-105 over placebo.

The sample size of 60 patients for ABBV-599 vs 40 for upadacitinib could produce the half width of 0.5 in the two-sided 90% CI of the difference in mean change from baseline, assuming the common SD = 1.5. Assuming the observed difference for ABBV-599 vs upadacitinib is –0.45, the corresponding 90% confidence interval would be (–0.95, 0.05).

In the event that placebo variability precludes historic borrowing, comparison using only the in-study 20 placebo patients is estimated to still provide 68% power for pairwise comparison of ABBV-105 vs placebo and a minimum of 70% power in detecting a dose-response model for ABBV-105 and placebo.

3.0 Endpoints

3.1 Primary Efficacy Endpoint

The primary endpoint is the change from baseline in disease activity score (DAS)28 (C-reactive protein [CRP]) at Week 12.

3.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

1. Change from baseline in clinical disease activity index (CDAI) and simplified disease activity index (SDAI) at all visits;
2. Proportion of subjects achieving Clinical Remission (CR) at all visits. CR is defined as DAS28 CRP < 2.6;
3. Proportion of subjects achieving low disease activity (LDA) at all visits. LDA is defined as DAS28 CRP ≤ 3.2;
4. Proportion of subjects achieving LDA but not CR at all visits. LDA but not CR is defined as DAS28 CRP ≤ 3.2 but ≥ 2.6;
5. Proportion of subjects achieving CR based on CDAI criteria at all visits. CR is defined as CDAI ≤ 2.8;
6. Proportion of subjects achieving LDA based on CDAI criteria at all visits. LDA is defined as CDAI ≤ 10;
7. Proportion of subjects achieving LDA but not CR based on CDAI criteria at all visits. LDA but not CR based on CDAI criteria is defined as CDAI ≤ 10 but > 2.8;
8. ACR20/50/70 response rates at all visits;
    a. ACR20/50/70 response rate will be determined based on 20%/50%/70% or greater improvement in Tender Joint Count (TJC) and Swollen Joint Count (SJC) and ≥ 3 of the 5 measures of Patient's Assessment of Pain (Visual Analog Scale [VAS]), Patient's Global Assessment of Disease Activity (PGA), Physician's Global Assessment of Disease Activity (PhGA), Health Assessment Questionnaire Disability Index (HAQ-DI), or high-sensitivity C-reactive protein (hsCRP);
9. Change from baseline in individual components of ACR response at all visits;
10. Change from baseline in DAS28 (CRP) and DAS28 (erythrocyte sedimentation rate [ESR]) at all visits;
11. Change from baseline in morning stiffness severity (11 point scale) at all visits;
12. Change from baseline in morning stiffness duration (in minutes) at all visits;

13. Proportion of subjects achieving minimal clinically important difference (MCID) in change from baseline in HAQ-DI (defined as change from baseline in HAQ-DI ≤ –0.22) at all visits

14. Proportion of subjects achieving American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) Boolean remission at all visits. Boolean remission is defined as TJC28 ≤ 1, SJC28 ≤ 1, CRP ≤ 1 mg/dl and PGA ≤ 1 (in cm, 0 – 10).

3.3 Other Efficacy Endpoints

The primary and/or secondary efficacy endpoints are listed in Section 3.1 and/or Section 3.2, respectively. There are no additional efficacy endpoints.

3.4 Safety Endpoints

Safety evaluations include adverse event (AE) monitoring, physical examinations, vital sign measurements, electrocardiograms (ECGs), and clinical laboratory testing (hematology, chemistry, and urinalysis) as a measure of safety and tolerability for the entire study duration.

Serious AEs (SAEs) will be assessed at any dose that results in a death, inpatient hospitalization or prolongation of existing hospitalization, persistent or significant disability/incapacity or a congenital anomaly.

Safety data will be assessed over the course of the study. An independent internal Data Monitoring Committee (DMC) will be created and unblinded safety assessments will be conducted. This is to assess unanticipated safety signals with regards to novel agents ABBV-105 and ABBV-599. The study team, study sites and subjects will remain blinded for the duration of the study.

An independent external Cardiovascular Adjudication Committee (CAC) will adjudicate blinded cardiac, cerebrovascular and thromboembolic events. A CAC charter will be

prepared separately from the protocol that will define objective, scope, frequency, and triggers of data reviews.

3.5 Additional Endpoint(s)

No additional endpoints will be analyzed in the SAP. Pharmacokinetic endpoints and biomarker endpoints will be analyzed separately.

4.0 Analysis Populations

The following population sets will be used for the analyses.

The Full Analysis Set (FAS) includes all randomized subjects who received at least 1 dose of study drug. Subjects will be grouped according to the treatment as randomized. The FAS will be used for all efficacy and baseline characteristics analyses.

The Safety Analysis Set consists of all enrolled subjects who received at least 1 dose of study drug. For the Safety Analysis Set, subjects are assigned to a treatment group based on the treatment received, regardless of the treatment randomized. The Safety Analysis Set will be used for all safety analysis.

5.0 Subject Disposition

The number of subjects will be tabulated by country, investigator site and overall for all randomized subjects.

The total number of subjects who were randomized, and treated will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects randomized in the study;
- Subjects who took at least one dose of study drug;
- Subjects who completed protocol-specified treatment;

- Subjects who prematurely discontinued study drug (all reasons and primary reason);
- Subjects in each analysis population, as applicable;
- Number of subjects who elect to participate in the open label extension Study M16-763.

In addition, the reasons for premature discontinuation (primary reason and all reasons) from the trial and/or from the medication collected from CRF by the following categories will be summarized with frequencies and percentages. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations ("Premature Discontinuation").

- Adverse events,
- Lost to follow-up,
- Withdrew consent by subject,
- Lack of efficacy,
- Others.

6.0 Study Drug Duration and Compliance

A summary of study drug duration (days) will be provided by each treatment arm for the Safety populations.

The duration of exposure to study drug will be summarized for each group as specified above, with the number of subjects, mean, standard deviation, median, minimum and maximum values. In addition, the number and percentage of subjects exposed to study drug will be summarized for the following cumulative duration intervals.

- ≥ 1 day
- ≥ 15 days
- ≥ 29 days
- ≥ 57 days
- ≥ 85 days

The exposure to study drug is calculated as:

Exposure = (date of last study medication – date of first study medication + 1)

Compliance

Study drug compliance will be summarized for each treatment group for the FAS population. The compliance is defined as the number of tablets/capsules taken (i.e., the difference between the number of tablets/capsules dispensed and the number of tablets/capsules returned) divided by the number of tablets/capsules that should have been taken during the treatment period. Compliance with each study drug will be calculated for each subject and summarized with the mean, median, standard deviation, minimum, and maximum.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the FAS overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

The following demographic and baseline characteristics, as measured at baseline of the study, will be summarized.

Demographic Characteristics

- Age (years)
- Age category [18 – < 40 years old, 40 – < 65 years old, ≥ 65 years old]
- Sex [male/female]
- Race [White, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other]
- Ethnicity [Hispanic/Latino, Non-Hispanic/Latino]
- Geographic region (North America, South/Central America, Western Europe, Eastern Europe, Asia, Other)
- Weight (kg)
- Weight Categories (< 60 kg, ≥ 60 kg)
- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Body Mass Index (BMI) Category (kg/m^2) (BMI < 25 vs BMI ≥ 25)

RA Medical History and Characteristics

- Duration of RA Symptoms in years
- Duration of RA Diagnosis in years
- Duration of RA Diagnosis Categories (< 5 year or ≥ 5 year)
- Prior exposure to bDMARDs (failed 1 or 2 biologics with the same mechanism of action; failed ≥ 3 biologics with the same mechanism of action and/or ≥ 2 biologics with multiple mechanisms of action)
- Number of prior bDMARD received (1, 2, 3, ≥ 4)
- Concomitant csDMARDs at baseline (MTX alone, MTX and other csDMARDs, csDMARDs other than MTX)

ACR and/or DAS Components at Baseline

- Tender joint count (TJC68) defined as the number of tender joints out of 68 assessed joints
- Swollen joint count (SJC66) defined as the number of swollen joints out of 66 assessed joints
- Tender joint count (TJC28) defined as the number of tender joints out of 28 assessed joints used for DAS28 calculation
- Swollen joint count (SJC28) defined as the number of swollen joints out of 28 assessed joints used for DAS28 calculation
- Physician's global assessment of disease activity (mm on a 100-mm horizontal visual analogue scale [VAS])
- Patient's assessment of pain within last week (mm on a 100-mm horizontal (VAS))
- Patient's global assessment of disease activity within last 24 hours (mm on a 100-mm horizontal VAS)
- Health Assessment Questionnaire Disability Index of the (HAQ – DI) (range: 0 to 3)
- High sensitivity C-reactive protein (hsCRP) (mg/L)
- Erythrocyte sedimentation rate (ESR) (mm/hr)

Other Baseline RA Disease Characteristics

- Morning stiffness (severity and duration)
- Anti-cyclic citrullinated peptide (Anti-CCP) (units)
- Anti-CCP status: Positive or Negative
- Rheumatoid Factor (RF) (units)
- Rheumatoid Factor (RF) status: Positive or Negative
- Percentage of subjects on oral steroid at baseline
- Oral steroid dose (prednisone equivalent) at baseline
- DAS28 [ESR]

- DAS28 [CRP]
- DAS28 [ESR] Categories:
  a. DAS28 [ESR] > 5.1 (High Disease Activity)
  b. DAS28 [ESR] ≤ 5.1
- DAS28 [CRP] Categories:
  a. DAS28 [CRP] > 5.1 (High Disease Activity)
  b. DAS28 [CRP] ≤ 5.1
- Clinical Disease Activity Index (CDAI)
- CDAI categories:
  a. CDAI > 22 (High Disease Activity)
  b. CDAI ≤ 22
- Simplified Disease Activity Index (SDAI)
- SDAI categories:
  a. SDAI > 26 (High Disease Activity)
  b. SDAI ≤ 26

Clinical Tests at Screening

- Chest x-ray
- Tuberculin PPD skin test, QuantiFERON TB Gold test
- Serum pregnancy test

Immunization History

- BCG immunization
- Herpes Zoster immunization
- Hepatitis B immunization

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine Use (current, former, never, or unknown)
- Alcohol Use (current, former, never, or unknown)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. Medical history data will be summarized and presented for the FAS population using body systems and conditions/diagnoses as captured on the CRF. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of subjects with a condition/diagnosis will be summarized for each randomized treatment group as well as overall. Subjects reporting more than one condition/diagnosis within a body system will be counted only once for that body system. No statistical comparison will be performed for medical history reporting.

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by each randomized treatment group as well as overall for the FAS. Prior medications are those medications taken prior to the first dose of study drug. This includes medications with a start date before the first study drug administration date, regardless of the end date of these medications. Medications taken on the day of the first dose of study drug are not counted as prior medications. Concomitant medications are those medications, other than study drug, taken after the first dose of study drug and within 1 day of the last dose of study drug. This includes medications with a start date between first study drug administration and last study drug administration + 1 day, as well as, medications with a start date prior to first dose of study drug and which are ongoing after first dose of study drug. Medications taken on the day of the first dose of study drug are counted as concomitant medications.

The number and percentage of subjects who received a prior medication and the number and percentage of subjects who received a concomitant medication will be tabulated separately by the generic name assigned by the most current version of the World Health Organization (WHO) Drug Dictionary.

8.0 Efficacy Analyses

8.1 General Considerations

The primary and secondary efficacy endpoints will be analyzed based on the FAS. The primary and secondary efficacy endpoints as specified in Section 3.0 will be tested at significance level alpha = 0.1 (two-sided).

For all efficacy endpoints, the descriptive statistics will be provided by treatment group. The statistics include number of observations, mean, standard deviation, minimum, median, and maximum for continuous variables; and number and percentage of subjects for categorical variables.

Unless otherwise specified, categorical variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, stratified by prior bDMARD use. Continuous variables will be analyzed using Mixed-Effect Model Repeated Measure (MMRM) method.

Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the correct stratum to which they should have been randomized.

"Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given.

8.2 Handling of Missing Data

Missing data will be imputed using the following methods for the efficacy analyses:

Non-Responder Imputation (NRI)

An NRI approach will categorize any subject who has missing value for categorical variables at a specific visit as non-responder for that visit. In addition, subjects who prematurely discontinue from randomized study drug will be considered as non-responders for all subsequent visits after discontinuation.

As Observed (AO)

The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. Regardless of treatment switching or premature discontinuation of study drug, all observed data will be used in the analysis.

Mixed-Effect Model Repeated Measure (MMRM)

MMRM analysis will be conducted for repeated continuous measurements. Data collected after a subject discontinues study drug will be considered as missing. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factors, and the continuous fixed covariate of baseline measurement. An unstructured variance covariance matrix will be used. The parameter estimation is based on the assumption of data being missing at random and using the method of restricted maximum likelihood (REML).

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is change from baseline in DAS28 (CRP) at Week 12.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint

MMRM under assumption of data being missing at random will be used for primary efficacy analyses, unless otherwise specified.

8.3.3 Primary Efficacy Analyses

The first primary analysis is to test the superiority of ABBV-599 compared to placebo at Week 12 using MMRM (see Section 8.3.5.1). Two-sided 0.1 is to be used throughout in all efficacy analysis.

The second primary analysis is to test a pre-specified set of dose-response models among ABBV-105 dose groups and the placebo group at Week 12 using the Multiple Comparison Procedure – Modeling (MCP-Mod) method with borrowing historical placebo data (see Section 8.3.5.3).

8.3.4 Additional Analyses of the Primary Efficacy Endpoint

The additional analyses for the primary endpoint are

(a) comparison of ABBV-599 to each ABBV-105 dose group;

(b) comparison of ABBV-599 to the upadacitinib group;

(c) comparison of each ABBV-105 group to the placebo group.

For the comparisons (a) - (c) 90% CI from MMRM of the treatment difference will be used (see Section 8.3.5.1). For comparison (b) and (c) historical data borrowing is also to be considered (see Section 8.3.5.2).

More details of the planned analyses can be found in Section 8.3.5 below.

8.3.5 Details for Primary Efficacy Analysis

8.3.5.1 Mixed-Effect Model Repeated Measure (MMRM) for Pair-Wise Comparison

In the pair-wise comparison analyses without historical data borrowing, the treatment mean will be estimated from MMRM method. One MMRM model including all treatment groups up to visit Week 12 data will be used.

MMRM analysis will be conducted for primary endpoint. Data collected after a subject discontinues study drug will be considered as missing. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, prior bDMARD use, and baseline DAS28 (CRP) measurement. An unstructured variance covariance matrix will be used. The parameter estimation will be based on the assumption of data being missing at random and using the method of restricted maximum likelihood (REML).

Least-square mean (SE) and 90% confidence interval (CI) at Week 12 for each treatment group will be presented together with between treatment difference of least-square means (SE), 90% CI and p-value from t-test for the difference = 0.

8.3.5.2 Historical Data Borrowing Analysis

Historical control data will be used in the comparison of ABBV-105 vs placebo (borrowing placebo data) and of ABBV-599 vs upadacitinib (borrowing upadacitinib data).

Meta-analysis results from the historical control data will be used as prior distribution for control group mean and a non-informative prior will be used for treatment group mean. Based on posterior means and variances for treatment and control groups, the probability of (Treatment mean – Control mean < 0) given the data i.e., $P(\mu_T - \mu_C < 0|Data)$ will be calculated. The details for historical data borrowing analysis is in Appendix D.

- **ABBV-105 vs Placebo (Borrowing Placebo)**

Historical data for placebo group consists of 349 subjects from three clinical trials with similar populations, similar inclusion and exclusion criteria; and they were conducted in a similar way as the current study. A meta-analysis shows that predicted mean (SD) and 95% confidence interval for mean are –0.80 (1.5) and [–1.06, –0.54].

In the ABBV-105 vs Placebo comparisons, 20 historical placebo subjects will be borrowed, which targets control of Bayesian type I error rate/False Positive Rate < 10%

and not exceeding study placebo sample size of 20. That is: $n_h = 20, \mu_h = -0.80$ and its
SE $\sigma_h = 1.5/\sqrt{20} = 0.335$. Therefore a conjugate prior distribution for placebo would be N (–0.80, 0.335^2). For ABBV-105, a "flat" non-informative prior probability distribution will be assumed, resulting in the posterior distribution totally based on the data. For supportive analysis, the dynamic borrowing based on observed bias may be performed as detailed in Appendix D.

- **ABBV-599 vs upadacitinib (Borrowing upadacitinib)**

For the upadacitinib control group versus ABBV-599, historical data consist of 206 subjects from two upadacitinib studies (Phase 2: Study M13-550 Balance I and Phase 3: Study M13-542 Select Beyond). Active upadacitinib treatment group of 15 mg is considered. The meta-analysis results in the predicted estimated mean (SD) and 95% confidence interval are –2.32 (1.57) and [–2.54, –2.11].

For this comparison, 40 historical placebo subjects will be borrowed, which targets control of false positive rate < 10% and not exceeding the study sample size 40. That is: $n_h = 40, \mu_h = -2.32,$ and $\sigma_h = 1.57/\sqrt{40} = 0.25$. Therefore the prior distribution for upadacitinib would be N (–2.32, 0.25^2). For ABBV-599, a "flat" non-informative prior probability distribution will be assumed, resulting in the posterior distribution totally based on the data. A dynamic borrowing based on observed bias may be performed as supportive analysis.

8.3.5.3 Dose-Response Modeling for ABBV-105

The dose-response relationship among ABBV-105 dose groups and the placebo group will be characterized for the primary endpoint using the Multiple Comparison Procedure - Modeling (MCP-Mod) method [Pinheiro 2006, Bretz 2005]. The response based on the primary analysis approach above will be used, and ADDPLAN DF software will be used to perform the MCP-Mod analyses. The posterior mean (SD) for placebo group (i.e., with

historical borrowing) will be used as primary analysis. The data directly from MMRM for placebo will also be analyzed as sensitivity analysis.

A set of 6 pre-specified standardized candidate dose-response models, as described in Table 1 will be utilized to examine the dose-response relationship. A statistically significant dose-response relationship will be declared if at least one model is identified by the MCP-Mod method to be statistically significant at $\alpha = 0.1$ two-sided. The fitted dose-response curves will be presented graphically for all statistically significant models along with confidence bands. The minimum effective dose (MED) will be identified for each statistically significant model based on the pre-specified clinical meaningful target of –0.88. The weighted MED across all significant models will be calculated, with weight being inverse of model AIC.

Table 1. Candidate Models

Model	$f(d, \theta)$ d = dose, θ = Model Parameters	$f^0(d, \theta)$ Standardized Model	Initial Value(s) for Parameter(s)
Linear	$E_0 + \delta d$	d	NA
Exponential	$E_0 + E_1\left[exp\left(\frac{d}{\delta}\right) - 1\right]$	$exp\left(\frac{d}{\delta}\right) - 1$	$\delta = 20$
Logistic	$E_0 + \frac{E_{max}}{1 + exp\left(\frac{ED_{50} - d}{\delta}\right)}$	$\frac{1}{1 + exp\left(\frac{ED_{50} - d}{\delta}\right)}$	$ED_{50} = 5$, $\delta = 10$
EMax	$E_0 + \frac{E_{max} d}{ED_{50} + d}$	$\frac{d}{ED_{50} + d}$	$ED_{50} = 5$
sigEMax	$E_0 + \frac{E_{max} d^h}{{ED_{50}}^h + d^h}$	$\frac{d^h}{{ED_{50}}^h + d^h}$	$ED_{50} = 5$, $h = 2$
Quadratic	$E_0 + \beta_1 d + \beta_2 d^2$	$d + \frac{\beta_2}{\lvert\beta_1\rvert} d^2$	$\delta = -0.01$

Steps of MCPMod:

1. Choose a candidate set of ***S*** models as in Table 1.
2. Compute the optimum contrast for each model.
3. Use contrast test to find the significant ***T*** models while preserving FWER.
4. Use AIC criteria to find the most significant model from the significant ***T*** models found from Step 3.
5. Use the model found from Step 4 to fit observed data from the study and make inference (e.g., to find Minimum Effective Dose (MED) or the dose achieving certain amount of maximum effect), or use all significant models to make inference about the weighted target dose of interest.

8.4 Secondary Efficacy Analyses

Continuous efficacy variables will be analyzed using MMRM method. It will be used for primary inference purpose the same way as for primary efficacy endpoint analysis.

Categorical efficacy variables will be analyzed using the CMH test controlling for stratification variables. The risk difference and 90% confidence intervals for the differences in each comparison will be constructed. NRI imputation will be used for missing data imputation. Point estimate of the response rate for each treatment group will also be summarized with a two-sided 90% confidence interval using Wilson's score method.

In addition, as observed (AO) analyses will be used as sensitivity analyses for DAS28 (CRP) and ACR20/50/70 response rates at all visits.

8.5 Additional Efficacy Analyses

No additional efficacy analyses are planned.

8.6 Efficacy Subgroup Analyses

Subgroup analysis for the primary endpoint will be conducted by the subgroups specified in Table 2. The difference in the primary efficacy endpoint between the treatment groups in each subgroup will be assessed using MMRM method.

Table 2. Subgroups for Primary Efficacy Analysis

Subgroup Factor	Categories
Age	18 – < 40 years old 40 – < 65 years old ≥ 65 years old
Sex	Male Female
BMI	< 25 ≥ 25
Race	White Black Asian Other
Baseline Rheumatoid Factor Status	Positive Negative
Baseline Anti-CCP Antibody Status	Positive Negative
Baseline DAS28 (CRP)	> 5.1 ≤ 5.1
No. of prior anti-TNF biologics	0 1 ≥ 2
Use of non-anti-TNF biologics	YES NO
Number of prior bDMARD received	1 2 3 ≥ 4
Duration of RA diagnosis	< 6 Months 6 Months – < 1 Year 1 Year – < 3 Years ≥ 3 Years

9.0 Safety Analyses

9.1 General Considerations

Safety analyses will include reporting of adverse events, laboratory, and vital signs measurements. Safety analyses will be carried out using the Safety Analysis Set.

The following summary statistics will be presented for subjects who have both baseline and post-baseline values for laboratory parameters and vital signs: the mean value at Baseline and at each respective protocol specified visit, and the mean, standard deviation and median for changes from Baseline. Categorical data will be summarized using frequencies and percentages. The number of non-missing values will be given. Missing safety data will not be imputed.

A subject's actual treatment will be determined by the most frequent dose regimen received.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

A treatment-emergent Adverse Event (TEAE) is defined as an adverse event with an onset date that is on or after the first dose of study drug, and no more than 30 days after the last dose of study drug; or an adverse event with onset date before the first dose of study drug,

but increased in severity on or after the first dose of study drug, and no more than 30 days after the last dose of study drug.

Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study drug start date).

The number and percentage of subjects experiencing TEAEs will be summarized.

9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- All TEAEs
- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs of special interest (AESIs)
- TEAEs leading to discontinuation of study drug
- TEAE leading to death
- All deaths

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate.

For AESIs, the point estimate and 95% CI (using normal approximation) will be provided for the treatment difference in AE percentages.

9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

The number and percentage of subjects experiencing adverse events will be tabulated by SOC and MedDRA PT by "as treated" treatment groups and overall. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The following summaries of adverse events will be generated:

- All TEAEs
- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs leading to discontinuation of study drug
- TEAE leading to death
- Frequent AEs (reported in 2% of subjects or more in any treatment group)

Subjects reporting more than one adverse event for a given MedDRA preferred term will be counted only once for that term (most severe incident for the severity tables and most related incident for the relationship tables). Subjects reporting more than one type of adverse event within a SOC will be counted only once for that SOC. Subjects reporting more than one type of adverse event will be counted only once in the overall total. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported. In addition, TEAEs will be summarized by PT and sorted by decreasing frequency for the total active group.

9.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

TEAEs will be summarized by event rate per 100 subject years, defined as

100 * (Number of TEAEs)/(Total Patient Years)

where total subject years is defined as the sum of the study drug exposure of all subjects normalized by 365.25 and rounded to 1 decimal place.

9.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

All serious adverse events (SAEs), deaths, and adverse events leading to discontinuation of study drug will be listed. The number and percentage of subjects experiencing SAEs (including deaths) and adverse events leading to discontinuation of study drug will be tabulated by SOC and PT or each treatment group.

9.2.6 Adverse Events of Special Interest

The Adverse Events of Special Interest (AESI) categories will be summarized and presented by "as treated" treatment group and overall using SOC and MedDRA PT. The AESI categories will be identified by the following search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) in Appendix B.

9.2.7 Adverse Events by Maximum Severity

TEAEs will also be summarized by maximum severity by "as treated" treatment group and overall. If a subject has an AE with an unknown severity, then the subject will be counted in the severity category of unknown, even if the subject has another occurrence of the same event with a severity present. The only exception is that if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

9.2.8 Adverse Events by Maximum Relationship

TEAEs will also be summarized by maximum relationship to treatment, as assessed by the investigator, by "as treated" treatment group and overall. If a subject has a TEAE with an unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship present. The only exception is if the subject has another occurrence of the

same TEAE with a relationship assessment of "reasonable possibility." In this case, the subject will be counted under the "reasonable possibility" category.

9.2.9 Adverse Events by "Reasonably Possibly Related" Relationship

TEAEs and reasonably possibly related AEs occurring for more than 2% of the subjects in any of the treatment groups will also be summarized by MedDRA SOC and PT. If a subject has an AE with an unknown relationship, then the subject will be counted in as 'related.'

9.2.10 Frequent (> 5%) Adverse Events and Reasonably Possibly Related Adverse Events by Preferred Term in Decreasing Frequency

TEAEs and reasonably possibly related AEs occurring for more than 5% of the subjects in any of the treatment arms will be summarized by MedDRA PT in decreasing frequency separately.

9.2.11 Listing of Adverse Events

The following additional summaries of AEs will be prepared.

- Listing of Subjects with Treatment-Emergent AESIs
- Listing of Subjects with Pretreatment SAEs
- Listing of Subjects with Treatment-Emergent SAEs
- Listing of all AEs that led to discontinuation of study drug
- Listing of all deaths.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline, where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. The

clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized.

Mean change from baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups using ANOVA model.

The baseline and post-baseline laboratory observations will be categorized as Grade 1, Grade 2, Grade 3, and Grade 4 according to OMERACT criteria (Rheumatology Common Toxicity Criteria v.2.0). For creatine phosphokinase and creatinine, NCI CTC criteria will be used.

For each laboratory variable, shift tables will be generated that cross tabulate the subjects' as deemed appropriate by "as treated" treatment group:

- Category of the baseline value versus category of the final value.
- Category of the baseline value versus maximum category.
- Category of the baseline value versus minimum category.

Note that the minimum/maximum category is used, rather than the category of the minimum/maximum value. The two may be different due to variation in the reference range.

No statistical tests will be performed for this analysis.

Laboratory abnormalities will be evaluated based on Potentially Clinically Significant (PCS) criteria (Appendix C). For each laboratory PCS criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCS criteria.

According to FDA's Guidance for Industry "Drug-Induced Liver Injury: Premarketing clinical evaluation" (July 2009), when aminotransferase (AT) abnormalities indicating hepatocellular injury are accompanied by evidence of impaired hepatic function (bilirubin elevation > 2 × ULN), in the absence of evidence of biliary obstruction (i.e., significant elevation of ALP) or some other explanation of the injury (e.g., viral hepatitis, alcohol hepatitis), the combined finding (i.e., Hy's Law cases) represents a signal of a potential for the drug to cause severe DILI.

For the purpose of assessing for potential Hy's law cases, the frequencies and percentages of subjects with post baseline liver-specific function test values that meet the following criteria of potential clinical interest will be summarized by "as treated" treatment group:

- ALT ≥ 3 × ULN
- ALT ≥ 5 × ULN
- ALT ≥ 10 × ULN
- ALT ≥ 20 × ULN
- AST ≥ 3 × ULN
- AST ≥ 5 × ULN
- AST ≥ 10 × ULN
- AST ≥ 20 × ULN
- Total Bilirubin Level (TBL) ≥ 2 × ULN
- Alkaline phosphatase ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 2 × ULN

9.4 Analysis of Vital Signs

Vital sign variables include sitting systolic blood pressure, sitting diastolic blood pressure, pulse, body temperature, and weight. The criteria for potentially clinically significant vital sign findings are presented in Appendix C.

A treatment-emergent value is defined as a value with a date that is on or after the first dose of study drug, and no more than 30 days after the last dose of study drug.

The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups using ANOVA model.

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix C). For each vital sign PCS criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCS criteria.

9.5 Safety Subgroup Analyses

No planned safety subgroup analyses.

9.6 Other Safety Analyses

ECG is collected at screening visit, Week 2, Week 4, Week 8, and Week 12. ECG findings will be summarized by treatment group for each parameter and visit.

10.0 Other Analyses

No other analyses are planned.

11.0 Interim Analyses

An unblinded interim analysis to review safety will be conducted by an independent internal DMC at 3 time points and ad hoc as needed: when approximately 50% of subjects have completed Week 4 evaluation, when approximately 80% of subjects have completed Week 4 evaluation, and when approximately 80% of subjects have completed Week 12 evaluation. The sponsor, study sites and subjects will remain blinded for the duration of the study.

11.1 Data Monitoring Committee

An internal data monitoring committee (DMC) composed of persons independent of study team and with relevant expertise in their field will review unblinded safety data from the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

Since there are no efficacy analyses for early stopping, no alpha adjustment is needed.

12.0 Overall Type-I Error Control

Hypothesis testing in the primary analyses (see Section 8.3.3) will be performed in a ranked fashion as follows based on two-sided significance level of $\alpha = 0.1$.

1) The superiority of ABBV-599 compared to placebo at Week 12 using MMRM.
2) The dose-response among ABBV-105 dose groups and the placebo group at Week 12 using MCP-Mod method.

All other tests will be performed at two-sided statistical significance level of $\alpha = 0.1$ without multiple testing α adjustment.

Since there are no efficacy analyses for early stopping planned for the DMC review, no alpha spending is needed due to the DMC review.

13.0 Version History

Table 3. SAP Version History Summary

Version	Date	Summary
1.0	29 May 2018	Original version
2.0	16 Jul 2018	The changes made between the second and the first version are: • Small editorial changes for readability and to maintain consistency with the protocol • Clarification of cutoff days for laboratory and vital sign values. • An update to the DMC evaluation checkpoints. • The addition of the OMERACT Criteria, which was referenced in the safety section, to the Appendix
3.0	13 June 2019	The changes made between this and the previous version are: • An update to visits with ECG data collected in Section 9.6 to maintain consistency with the protocol • Addition of an interim efficacy analysis to maintain consistency with the protocol • Updated the AESI table to maintain consistency with the protocol • Updates to the historical borrowing methods detailed in Appendix D • Change of formatting to adhere to a new SAP template
4.0	23 Jan 2020	The change made between this and the previous version are: • Updated Potential Clinically Important Criteria for safety endpoints in Appendix C • Removed interim analysis due to protocol amendment • Changed secondary endpoints from "Proportion of subjects achieving CR (DAS28 CRP) at Week 12" and "Proportion of subjects achieving LDA (DAS28 CRP) at Week 12" to "Proportion of subjects achieving CR (DAS28 CRP) at all visits" and "Proportion of subjects achieving LDA (DAS28 CRP) at all visits," respectively, as the summaries by visits are of interest. • Added "Proportion of subjects achieving LDA (DAS28 CRP) but not CR at all visits" and "Proportion of subjects achieving LDA (CDAI) but not CR at all visits" as secondary endpoints, as these summaries are of interest. • Added AO analyses for DAS28 CRP and ACR20/50/70 response rates at all visits as sensitivity analyses per protocol. • Updated the criteria for Liver-Related Elevations to align with Immunology RA project conventions.

14.0 References

1. Ibrahim J, Chen M. Power prior distribution for regression models. Stat Sci. 2000;15:46-60.
2. Hobbs BP, Carlin BP, Mandrekar SJ, et al. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. Biometrics. 2011;67(3):1047-56.
3. Schmidli H, Gsteiger S, Roychoudhury S, et al. Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information. Biometrics. 2014;70(4):1023-32.
4. Neuenschwander B, Capkun-Niggli G, Branson M, et al. Summarizing historical information on controls in clinical trials. Clin Trials. 2010;7(1):5-18.
5. Baeten D, Baraliakos X, Braun J, et al. Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: A randomised, double-blind, placebo-controlled trial. Lancet. 2013;382(9906):1705-13.
6. Pocock SJ. The combination of randomized and historical controls in clinical trials. J Chronic Dis. 1976;29(3):175-88.
7. Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. J Biopharm Stat. 2006;16(5):639-56.
8. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics. 2005;61(3):738-48.
9. Genovese MC, Kremer J, Zamani O, et al. Baricitinib in Patients with Refractory Rheumatoid Arthritis. N Engl J Med. 2016;374(13):1243-52.
10. Burmester GR, Blanco R, Charles-Schoeman C, et al. Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: a randomised phase 3 trial. Lancet. 2013;381(9865):451-60.

11. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. Stat Med. 2007;26(9):1964-81.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix B. Definition of Adverse Events of Special Interest

Adverse Events of Special Interest (AESI) will be identified using the following search criteria:

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection	CMQ		"Opportunistic Infection"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ		"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Broad	Skin Malignant tumours (Broad SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			"Malignant tumours" SMQ removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder (Hepatic Events and Increased Hepatic Transaminases)	SMQ	Narrow	"Drug Related Hepatic Disorders"
Gastrointestinal Perforations	SMQ	Narrow	"Gastrointestinal Perforation"
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia (Veliparib Product Specific)"
Herpes Zoster	CMQ		"Herpes Zoster"
Creatine Phosphokinase (CPK) Elevation	PT		Search only for the PT of "Blood creatine phosphokinase increased"
Renal Dysfunction	SMQ	Narrow	"Acute Renal, Failure"
Tuberculosis	CMQ		"Tuberculosis"
Adjudicated Cardiovascular Events	Output from CAC		

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCS) laboratory findings are determined by OMERACT criteria in Table C-1, and the PCS criteria for vital sign findings are described in Table C-2.

Table C-1. OMERACT Criteria

Rheumatology Common Toxicity Criteria v.2.0 **Based on Woodworth TG, et al. Standardizing assessment of adverse effects in rheumatology clinical trials II. Status of OMERACT Drug Safety Working Group May 2006: OMERACT 8. Standardizing Assessment and Reporting of Adverse Effects in Rheumatology Clinical Trials: Enabling Description of Comparative Safety Profiles for Antirheumatic Therapies** (Note that for L9. CPK and L11. Creatinine, the criteria in this table is replaced by the NCI CTC grade, as the NCI CTC grade is used for analysis for these two parameters.)				
	1 – Mild Asymptomatic, or transient Short duration (< 1 week) No change in life style No medication or OTC	**2 – Moderate** Symptomatic Duration (1 – 2 weeks) Alter lifestyle occasionally Meds relieve. (may be prescription), Study drug continued	**3 – Severe** Prolonged symptoms, reversible, major functional impairment Prescription meds/partial relief May be hospitalized < 24 hr Temporary study drug discontinuation, or/and dose reduced	**4 – Includes Life Threatening** At risk of death Substantial disability, especially if permanent. Multiple meds Hospitalized > 24 hr Study drug discontinued
A. Allergic/Immunologic				
A1. Allergic reaction/ hypersensitivity (includes drug fever)	Transient rash: drug fever < 38°C: transient, asymptomatic bronchospasm	Generalised urticaria responsive to meds; or drug fever > 38°C, or reversible bronchospasm	Symptomatic bronchospasm requiring meds; symptomatic urticaria persisting with meds, allergy related oedema/angioedema	Anaphylaxis, laryngeal/pharyngeal oedema, requiring resuscitation
A2. Autoimmune reaction	Seriologic or other evidence of autoimmune reaction, but patient asymptomatic: all organ function normal and no treatment is required (e.g., vitiligo)	Evidence of autoimmune reaction involving a non-essential organ or functions, requiring treatment other than immunosuppressive drugs (e.g., hypothyroidism)	Reversible autoimmune reaction involving function of a major organ or toxicity requiring short term immunosuppressive treatment (e.g., transient colitis or anaemia)	Causes major organ dysfunction, or progressive, not reversible, or requires long-term administration of high dose immunosuppressive therapy

A3. Rhinitis (includes sneezing, nasal stuffiness, post-nasal discharge)	Transient, non-prescription meds relieve	Prescription med. required, slow	Corticosteroids or other prescription med. with persistent disabling symptoms such as impaired exercise tolerance	NA
A4. Serum sickness	Transient, non-prescription meds relieve	Symptomatic, slow response to meds (e.g., oral corticosteroids)	Prolonged; symptoms only partially relieved by meds; parenteral corticosteroids required	Major organ dysfunction, requires long-term high-dose immunosuppressive therapy
A5. Vasculitis	Localised, not requiring treatment; or rapid response to meds; cutaneous	Symptomatic, slow response to meds (e.g., oral corticosteroids)	Generalised, parenteral corticosteroids required or/and short duration hospitalisation	Prolonged, hospitalisation, ischemic changes, amputation
B. Cardiac				
B1. Arrhythmia	Transient, asymptomatic	Transient, but symptomatic or recurrent, responds to meds	Recurrent/persistent; maintenance prescription	Unstable, hospitalisation required, parenteral meds
B2. Cardiac function decreased	Asymptomatic decline in resting ejection fraction by > 10%, but < 20% of baseline value	Asymptomatic decline of resting ejection fraction ≥ 20% of baseline value	CHF responsive to treatment	Severe or refractory CHF
B3. Edema	Asymptomatic (e.g., 1 + feet/ calves), self-limited, no therapy required	Symptomatic (e.g., 2 + feet/ calves), requires therapy	Symptoms limiting function (e.g., 3 + feet/ calves, 2 + thighs), partial relief with treatment prolonged	Anasarca; no response to treatment
B4. Hypertension (new onset or worsening)	Asymptomatic, transient increase by > 20 mmHg (diastolic) or to > 150/100 if previously normal, no therapy required	Recurrent or persistent increase > 150/100 or by > 10 mmHg (diastolic), requiring and responding readily to treatment	Symptomatic increase > 150/100, > 20 mmHg, persistent, requiring multi agency therapy, difficult to control	Hypertensive crisis
B5. Hypotension (without underlying diagnosis)	Transient, intermittent, asymptomatic, orthostatic decrease in blood pressure > 20 mmHg	Symptomatic, without interference with function, recurrent or persistent > 20 mmHg decrease, responds to treatment	Syncope or symptomatic, interferes with function, requiring therapy and sustained medical attention, dose adjustment or drug discontinuation	Shock

B6. Myocardial ischaemia	Transient chest pain/ECG changes; rapid relief with nitro	Recurring chest pain, transient ECG ST-T changes; treatment relieves	Angina with infarction, no or minimal functional compromise, reduce dose or discontinue study drug	Acute myocardial infarction, arrthymia or/and CHF
B7. Pericarditis/ pericardial effusion	Rub heard, asymptomatic	Detectable effusion by echocardiogram, symptomatic NSAID required	Detectable on chest x-ray, dyspnoea; or pericardiocentesis; requires corticosteroids	Pulsus alternates with low cardiac output; requires surgery
B8. Phlebitis/ thrombosis/ Embolism (excludes injection sites)	Asymptomatic, superficial, transient, local, or no treatment required	Symptomatic, recurrent, deep vein thrombosis, no anticoagulant therapy required	Deep vein thrombosis requiring anticoagulant therapy	Pulmonary embolism
C. General (constitutional)				
C1. Fatigue/malaise (asthenia)	Increase over baseline; most usual daily functions maintained, short term	Limits daily function intermittently over time	Interferes with basic ADL, persistent	Unable to care for self, bed or wheelchair bound > 50% of day debilitating, hospitalisation
C2. Fever (pyrexia) (note: fever due to drug allergy should be coded as allergy)	Transient, few symptoms 37.7 – 38.5°C	Symptomatic, recurrent 38.6 – 39.9°C. Relieved by meds	≥ 40°C; ≤ 24 h, persistent symptoms; partial response to meds	≥ 40°C, debilitating, > 24 h, hospitalisation; no relief with meds
C3. Headache	Transient or intermittent, no meds or relieved with OTC	Persistent, recurring, non-narcotic analgesics relieve	Prolonged with limited response to narcotic medicine	Intractable, debilitating, requires parenteral meds
C4. Insomnia	Difficulty sleeping, short term, no interfering with function	Difficulty sleeping interfering with function, use of prescription med	Prolonged symptoms, with limited response to narcotic meds	Debilitating, hospitalisation; no relief with meds
C5. Rigors, chills	Asymptomatic, transient, no meds, or non-narcotic meds relieve	Symptomatic, narcotic meds relieve	Prolonged symptoms, with limited response to narcotic meds	Debilitating, hospitalisation; no relief with meds
C6. Sweating (diaphoresis)	Episodic, transient	Frequent, short term	Frequent, drenching, disabling	Dehydration, requiring IV fluids/hospitalization > 24 hrs
C7. Weight gain	5% – 9.9%	10% – 19.9%	20% – 30%	NA
C8. Weight loss	5% – 9.9%	10% – 19.9%	20% – 30%	NA

D. Dermatologic				
D1. Alopecia	Subjective, transient	Objective, fully reversible	Patchy, wig used, partly reversible	Complete, or irreversible even if patchy
D2. Bullous eruption	Localised, asymptomatic	Localised, symptomatic, requiring treatment	Generalised, responsive to treatment; reversible	Prolonged, generalised, or requiring hospitalisation for treatment
D3. Dry skin	Asymptomatic, controlled with emollients	Symptoms eventually (1 – 2 wks) controlled with emollients	Generalised, interfering with ADL > 2 wks, persistent pruritis, partially responsive to treatment	Disabling for extended period, unresponsive to ancillary therapy and requiring study drug discontinuation for relief
D4. Injection site reaction	Local erythema, pain, pruritis, < few days	Erythema, pain, oedema, may include superficial phlebitis, 1 – 2 wks	Prolonged induration, superficial ulceration; includes thrombosis	Major ulceration necrosis requiring surgery
D5. Petechiae (without vasculitis)	Few, transient asymptomatic	Dependent areas, persistent up to 2 wks	Generalised, responsive to treatment; reversible	Prolonged, irreversible, disabling
D6. Photosensitivity	Transient erythema	Painful erythema and oedema requiring topical treatment	Blistering or desquamation, requires systematic corticosteroids	Generalised exfoliation or hospitalisation
D7. Pruritis	Localised, asymptomatic, transient, local treatment	Intense, or generalised, relieved by systematic medication	Intense or generalised; poorly controlled despite treatment	Disabling, irreversible
D8. Rash (not bullous)	Erythema, scattered macular/popular eruption; pruritis transient; TOC or no meds	Diffuse macular/popular eruption or erythema with pruritus; dry desquamation; treatment required	Generalised, moist desquamation, requires systemic corticosteroids; responsive to treatment; reversible	Exfoliative or ulcerating; or requires hospitalisation; or parenteral corticosteroids
D9. Induration/fibrosis/ Thickening (not sclerodermal)	Localized, high density on palpation, reversible, no effect on ADL and not disfiguring	Local areas < 50% body surface, not disfiguring, transient interference with ADL, reversible	Generalized, disfiguring, interferes with ADL, reversible	Disabling, irreversible, systemic symptoms

E. Ear/Nose/Throat				
E1. Hearing loss	Transient, intermittent, no interference with function	Symptomatic, treatment required, reversible	Interferes with function; incomplete response to treatment	Irreversible deafness
E2. Sense of smell	Slightly altered	Markedly altered	Complete loss, reversible	Complete loss, without recovery
E3. Stomatitis	Asymptomatic	Painful, multiple, can eat	Interferes with nutrition, slowly reversible	Requires enteral support; residual dysfunction
E4. Taste disturbance (dysgeusia)	Transiently altered; metallic	Persistently altered; limited effect on eating	Disabling, effect on nutrition	NA
E5. Tinnitus	Intermittent, transient, no interference with function	Requires treatment, reversible	Disabling, or associated with hearing loss	Irreversible deafness
E6. Voice changes (includes hoarseness, loss of voice, laryngitis)	Intermittent hoarseness, able to vocalise	Persistent hoarseness, able to vocalise	Whispered speech, slow return of ability to vocalise	Unable to vocalize for extended
E7. Xerostomia (dry mouth)	Transient dryness	Relief with meds	Interferes with nutrition, slowly reversible	Extended duration interference with nutrition, requires parenteral nutrition
F. Eye/Ophthalmologic				
F1. Cataract	Asymptomatic, no change in vision, non-progressive	Symptomatic, partial visual loss, progressive	Symptoms impairing function, vision loss requiring treatment, including surgery	NA
F2. Conjunctivitis	Asymptomatic, transient, rapid response to treatment	Symptomatic, responds to treatment, changes not interfering with function	Symptoms prolonged, partial response to treatment, interferes with function	NA
F3. Lacrimation increased (tearing, watery eyes)	Symptoms not requiring treatment, transient	Symptomatic, treatment required, reversible	Unresponsive to treatment with major effect on function	NA
F4. Retinopathy	Asymptomatic, non-progressive, no treatment	Reversible change in vision; readily responsive to treatment	Disabling change in vision ophthalmological findings reversible, sight improves over time	Loss of sight

F5. Vision changes (e.g., blurred, photophobia, night blindness, vitreous floaters)	Asymptomatic, transient, no treatment required	Symptomatic, vision changes not interfering with function, reversible	Symptomatic, vision changes interfering with function	Loss of sight
F6. Xerophtalmia (dry eyes)	Mild scratchiness	Symptomatic without interfering with function, requires artificial tears	Interferes with vision/function, corneal ulceration	Loss of sight
G. Gastrointestinal				
G1. Anorexia	Adequate food intake, minimal weight loss	Symptoms requiring oral nutritional supplementation	Prolonged, requiring iv support	Requires hospitalization for nutritional support
G2. Constipation	Asymptomatic, transient, responds to stool softener, OTC laxatives	Symptomatic, requiring prescription laxatives, reversible	Obstipation requiring medical intervention	Bowel obstruction. Surgery required
G3. Diarrhea	Transient, increase of 2 – 3 stools/day over pre-treatment (no blood or mucus), OTC agents relieve	Symptomatic, increase 4 – 6 stools/day, nocturnal stools, cramping, requires treatment with prescription meds	Increase > 6 stools/day, associated with disabling symptoms, e.g., incontinence, severe cramping, partial response to treatment	Prolonged, dehydration, unresponsive to treatment, requires hospitalization
G4. Dyspepsia (heartburn)	Transient, intermittent, responds to OTC antacids, H-2 blockers	Prolonged, recurrent, requires prescription meds, relieved by meds	Persistent despite treatment, interferes with function, associated with GI bleeding	NA
G5. GI bleed (gastritis, gastric or duodenal ulcer diagnosed-define aetiology)	Asymptomatic, endoscopic finding, haemocult + stools, no transfusion, responds rapidly to treatment	Symptomatic, transfusion ≤ 2 units needed; responds to treatment	Haematemesis, transfusion 3 – 4 units, prolonged interference with function	Recurrent, transfusion > 4 units, perforation, requiring surgery, hospitalisation
G6. Haematochezia (rectal bleeding)	Haemorrhodial, asymptomatic, no transfusion	Symptomatic, transfusion ≤ 2 units, reversible	Recurrent, transfusion > 3 – 4 units	> 4 units, hypotension, requiring hospitalization

G7. Hepatitis	Laboratory abnormalities, asymptomatic, reversible	Symptomatic laboratory abnormalities, not interfering with function, slowly reversible	Laboratory abnormalities persistent > 2 wks, symptoms interfere with function	Progressive, hepato-renal, anasarca, pre-coma or coma
G8. Nausea, or nausea/vomiting (use diagnostic term)	Transient, intermittent, minimal interference with intake, rapid response to meds	Persistent, recurrent, requires prescription meds, intake maintained	Prolonged, interferes with daily function and nutritional intake, periodic iv fluids	Hypotensive, hospitalization, parenteral nutrition, unresponsive to out-patient management
G9. Pancreatitis	Anylase elevation, intermittent nausea/vomiting, transient, responds rapidly to treatment	Amylase elevation with abdominal pain, nausea, occasional vomiting, responsive to treatment	Severe, persistent abdominal pain with pancreatitic enzyme elevation, incomplete or slow response to treatment	Complicated by shock, haemorrhage (acute circulatory failure)
G10. Proctitis	Perianal pruritus, haemorrhoids (new onset), transient, or intermittent, relieved by OTC meds	Tenesmus or ulcerations, anal fissure, responsive to treatment, minimal interference with function	Unresponsive to treatment, marked interference with function	Mucosal necrosis with haemorrhage, infection, surgery required
H. Musculoskeletal				
H1. Avascular necrosis	Asymptomatic MRI changes, non-progressive	MRI changes and symptoms responsive to rest and analgesia	MRI changes, symptoms requiring surgical intervention	Wheelchair bound; surgical repair not possible
H2. Arthralgia	Intermittent transient symptoms, no meds or relieved by OTC meds	Persistent or recurrent symptoms, resolve with meds, little effect on function	Severe symptoms despite meds impairs function	Debilitating, hospitalisation required for treatment
H3. Leg cramps	Transient, intermittent, does not interfere with function	Recurrent symptoms, minimally interferes with function or sleep, responds to meds	Persistent, prolonged interference with function or sleep, partial or no response to meds	NA
H4. Myalgia	Occasional; does not interfere with function	Frequent, requires meds (non-narcotic); minor effects on function	Major change in function/lifestyle, narcotic pain meds	Debilitating, profound weakness, requires wheelchair, unresponsive to meds

I. Neuropsychiatric				
I1. Anxiety or Depression (mood alteration)	Symptomatic, does not interfere with function; no meds	Frequent symptoms, responds to meds; interferes with ADL at times	Persistent, prolonged symptoms, partial or no response to meds, limits daily function	Suicidal ideation or danger to self
I2. Cerebrovascular ischaemia	NA	Single transient ischaemic event, responsive to treatment	Recurrent transient ischaemic events	Cerebrovascular vascular accident with permanent disability
I3. Cognitive disturbance	Subjective symptoms, transient, intermittent, not interfering with function	Objective symptoms, persisting, interferes with daily function occasionally	Persistent, or worsening objective symptoms; interferes with routine daily routine	Debilitating/ disabling and permanent; toxic psychosis
I4. Depressed consciousness (somnolence)	Observed, transient, intermittent, not interfering with function	Somnolence or sedation, interfering with function	Persistent, progressive, obundation, stupor	Coma
I5. Inability to concentrate	Subjective symptoms, does not interfere with function	Objective findings, interferes with function	Persistent, prolonged objective findings or organic cause	NA
I6. Insomnia (in absence of pain)	Occasional difficulty sleeping, transient intermittent, not interfering with function	Recurrent difficulty sleeping; requires meds for relief; occasional interference with function	Persistent or worsening difficulty sleeping; severely interferes with routine daily function	NA
I7. Libido decreased	Decrease in interest	Loss of interest; influences relationship	Persistent, prolonged interfering with relationship	NA
I8. Peripheral motor neuropathy	Subjective or transient loss of deep tendon reflexes; function maintained	Objective weakness, persistent, no significant impairment of daily function	Objective weakness with substantial impairment of function	Paralysis
I9. Peripheral sensory neuropathy (sensory disturbance)	Subjective symptoms without objective findings, transient, not interfering with function	Objective sensory loss, persistent, not interfering with function	Prolonged sensory loss or paraethesias interfering with function	NA

I10. Seizure	NA	Recurrence of old seizures, controlled with adjustment of medication	Recurrence/ exacerbation with partial response to medication	Recurrence not controlled, requiring hospitalization; new seizures
I11. Vertigo (dizziness)	Subjective symptoms, transient, intermittent, no treatment	Objective findings, recurrent, meds relieve, occasionally interfering with function	Persistent, prolonged, interfering with daily function; partial response to medication	Debilitating without response to medication, hospitalization
J. Pulmonary				
J1. Asthma	Occasional wheeze, no interference with activities	Wheezing, requires oral meds, occasional interference with function	Debilitating, requires nasal O_2	Requires ventilator assistance
J2. Cough	Transient, intermittent, occasional OTC meds relieve	Persistent, requires narcotic or other prescription meds for relief	Recurrent, persistent coughing spasms without consistent relief by meds, interferes with function	Interferes with oxygenation; debilitating
J3. Dyspnea	Subjective, transient, no interference with function	Symptomatic, intermittent or recurring, interferes with exertional activities	Symptomatic during daily routine activities, interferes with function, treatment with intermittent nasal O_2 relieves	Symptomatic at rest, debilitating, requires constant nasal O_2
J4. Pleuritic pain (pleurisy)	Transient, intermittent symptoms, no treatment or OTC meds relieve	Persistent symptoms, requires prescription meds for relief	Prolonged symptoms, interferes with function, requires frequent narcotic pain relief	Debilitation, requiring hospitalisation
J5. Pneumonitis (pulmonary infiltrates)	Asymptomatic radiographic changes, transient, no treatment required	Symptomatic, persistent, requiring corticosteroids	Symptomatic, requiring treatment including O_2	Debilitating, not reversible; or requiring assisted ventilation
J6. Pulmonary function decreased (FVC or carbon monoxide diffusion capacity – DLCO)	76% – 90% of pre-treatment value	51% – 75% of pre-treatment value	26% – 50% of pre-treatment value	≤ 25% of pre-treatment value

Laboratory Data				
K. Haematology				
K1. Hgb (g/dl) decrease from pre-treatment	1.0 – 1.4	1.5 – 2.0	2.1 – 2.9, or Hgb < 8.0, > 7.0	≥ 3.0; or Hgb < 7.0
K2. Leukopenia (total WBC) × 1000	3.0 – 3.9	2.0 – 2.9	1.0 – 1.9	< 1.0
K3. Neutropenia (× 1000)	1.5 – 1.9	1.0 – 1.4	0.5 – 0.9	< 0.5
K4. Lymphopenia (× 1000)	1.5 – 1.9	1.0 – 1.4	0.5 – 0.9	< 0.5
K5. Platelets (× 1000)	75 – LLN	50 – 74.9	20 – 49.9; platelet transfusion required	< 20; recurrent platelet transfusions
L. Chemistry				
L1. Hypercalcaemia (mg/dl)	1.1 × ULN – 11.5	11.6 – 12.5	12.6 – 13.5; or symptoms present	> 13.5; or associated coma
L2. Hyperglycemia (mg/dl) Fasting	140 – 160	161 – 250	251 – 500	> 500, or associated with ketoacidosis
L3. Hyperkalaemia (mmol/l)***	5.5 – 5.9	6.0 – 6.4	6.5 – 7.0 or any ECG change	> 7.0 or any arrhythmia
L5. Hypocalcaemia (mg/dl)	0.9 × LLN – 7.8	7.7 – 7.0	6.9 – 6.5; or associated with symptoms	< 6.5 or occurrence of tetany
L6. Hypoglycemia (mg/dl)	55 – 64 (no symptoms)	40 – 54 (or symptoms present)	30 – 39 (symptoms impair function)	< 30 or coma
L7. Hyponatraemia (mmol/l)***	-	125 – 129	120 – 124	< 120
L8. Hypokalaemia (mg/dl)***	-	3.0 – 3.4	2.5 – 2.9	< 2.5
L9. CPK (also if polymyositis-disease)****	> ULN – 1.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 10.0 × ULN	> 10.0 × ULN
L10. Serum uric acid	1.2 – 1.6 × ULN	1.7 – 2.9 × ULN	3.0 – 5.0 × ULN or gout	NA
L11. Creatinine (mg/dl)****	> 1 – 1.5 × Baseline; > ULN – 1.5 × ULN	> 1.5 – 3.0 × Baseline; > 1.5 – 3.0 × ULN	> 3.0 baseline; > 3.0 – 6.0 × ULN	> 6.0 × ULN

L12. SGOT (AST)	1.2 – 1.5 × ULN	1.6 – 3.0 × ULN	3.1 – 8.0 × ULN	> 8.0 × ULN
L13. SGPT (ALT)	1.2 – 1.5 × ULN	1.6 – 3.0 × ULN	3.0 – 8.0 × ULN	> 8.0 × ULN
L14. Alkaline phosphatase	1.1 – 1.5** × ULN	1.6 – 3.0 × ULN	3.0 – 5.0 × ULN	> 5.0 × ULN
L15. T. bilirubin	1.1 – 1.4 × ULN	1.5 – 1.9 × ULN	2.0 – 3.0 × ULN	> 3.0 × ULN
L16. LDH	1.3 – 2.4 × ULN	2.5 – 5.0 × ULN	5.1 – 10 × ULN	> 10 × ULN
M. Urinalysis				
M1. Haematuria	Micro only	Gross, no clots	Clots, transfusion < 2 units	Transfusion required
M2. Proteinuria (per 24 h)	300 – 500 mg (tr/1 +)	501 – 1999 mg (2 +)	2 – 5.0 g (3 +) nephrotic syndrome	5.0 g (4 +) anasarca
M3. WBC in urine	NA	NA	Indicating acute interstitial nephritis	Associated with acute renal failure
M4. Uric acid crystals	Present without symptoms	NA	With stones or symptoms of stones (e.g., renal colic)	Causing renal outflow obstruction and hospitalization

*: in L11, 1.5 – 1.8 × ULN is changed to 1.4 – 1.8 × ULN.

**: in L14, 1.1 – 2.0 × ULN is changed to 1.1 – 1.5 × ULN.

***: in L3, L7 and L8, mg/dl is changed to mmol/l.

****: NCI CTC grade.

Table C-2. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and/or decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and/or increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and/or decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and/or increase ≥ 10 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and/or decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and/or increase ≥ 15 bpm from Baseline
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

Appendix D. Details for Historical Data Borrowing

Statistical methods for incorporating historical data in a new study include power prior [Ibrahim 2000], commensurate prior [Hobbs 2011] and robust mixture prior [Schmidli 2014]. All these methods are Bayesian approaches. We took a similar approach using meta-analytical-predictive method to summarize historical control data as a basis for informative prior for the control group and to determine the maximal number of subjects to be borrowed (effective sample size [Neuenschwander 2010]). For the treatment group, we use a noninformative prior so that information for the treatment group only depends on in-trial data [Baeten 2013]. We systematically evaluate the impact of bias (defined as the difference between historical control data and concurrent control data) and number of historical control subjects borrowed on the (Bayesian) type I error rate/false positive rate and power/true positive rate. So the number of subjects borrowed is determined to control the inflation (if any) of two types of error rates to a reasonable extent based on a reasonable range of the magnitude of bias.

D-1 Historical Data Selection and Meta-Analysis

The historical data are carefully selected based on the similarity in target population, mechanism of action of study drugs, concomitant medications [Pocock 1976] between historical trials and the trial under design. The selection of historical trials is described in the following.

- **ABBV-105 vs Placebo (Borrowing Placebo)**

For placebo control group vs ABBV-105 treatment group, historical data for placebo group are available from three clinical trials at the time of design stage of the current trial. These historical control data consist of 349 subjects from three clinical trials with similar populations (TNF and/or Bio-IR on background MTX or csDMARDs), mechanism of action (JAK inhibition), and study designs: baricitinib Phase 3 study BEACON (n = 176, mean [SD] of DAS28-CRP change = –0.8 [1.5]) [Genovese 2016]; tofacitinib Phase 3 study STEP (n = 118, mean (SD) = –0.6 (1.5)) [Burmester 2013]; and upadacitinib-Phase II (n = 55, mean (SD) = –1.1 (1.6)) (AbbVie Study M13-550). The criteria for

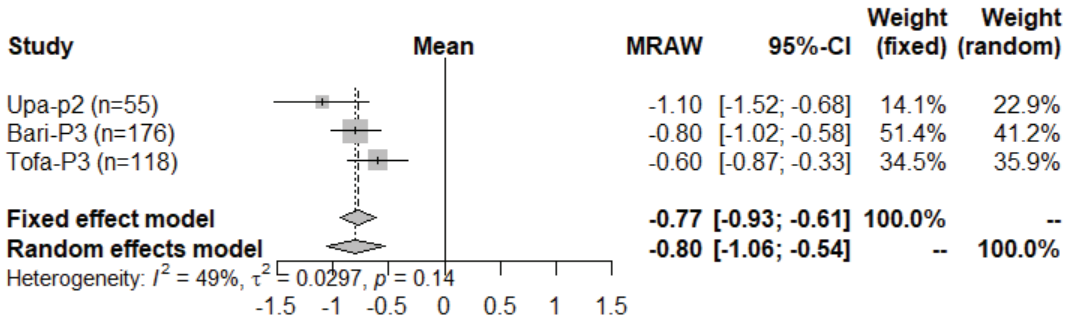
relevant historical control data selection were evaluated and met for these trials and the trial under design. These studies have similar population (TNF and Bio-IR), similar inclusion and exclusion criteria and were conducted in the similar way as the current study as shown in Table D-1.

Table D-1. Major entry criteria for historical studies and Study M16-063

Criteria	Historical Studies			Current Study
	Baricitinib (BEACON P3)	**Tofacitinib (STEP P3)**	**Upadacitinib (M13-550 P2)**	**ABBV-599 (M16-063 P2)**
Inc 1	Bio-IR	Bio-IR	Bio-IR	Bio-IR
Inc 2	18 years or older	18 years or older	18 years or older	18 years or older
Inc 3	Moderately to severely RA	Moderately to severely RA	Moderately to severely RA	Moderately to severely RA
Inc 4	csDMARDs	Methotrexate	Methotrexate	csDMARDs
Exc 1	haemoglobin < 100 g/L	haemoglobin < 90 g/L	haemoglobin < 90 g/L	haemoglobin < 90 g/L
Exc 2	WBC < 2.5 × 10^9/L	WBC < 3.0 × 10^9/L	WBC < 3.0 × 10^9/L	WBC < 3.0 × 10^9/L
Exc 3	platelet < 100 × 10^9/L	platelet < 100 × 10^9/L	platelet < 100 × 10^9/L	platelet < 100 × 10^9/L
Exc 4	eGFR < 40 mL/min/ 1.73 m^2	eGFR < 40 mL/min/ 1.73 m^2	eGFR < 40 mL/min/ 1.73 m^2	eGFR < 40 mL/min/ 1.73 m^2
Exc 5	AST/ALT > 1.5 × ULN	AST/ALT > 1.5 × ULN	AST/ALT > 1.5 × ULN	AST/ALT > 2 × ULN

The meta-analysis shows that predicted estimated mean (SD) and 95% confidence interval for mean are –0.80 (1.5) and [–1.06, –0.54] (Figure D-1) based on Empirical Bayes estimator [Sidik 2007]. The estimated heterogeneity variance $\tau^2 = 0.0297$ and historical effective sample size (the maximum number of subjects we could borrow) is 49 [Neuenschwander 2010].

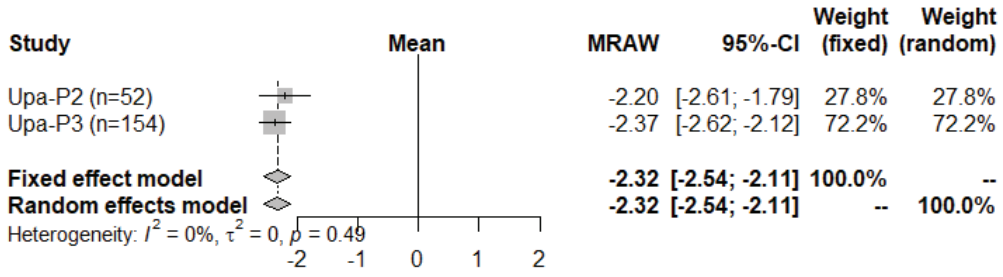
Figure D-1. Meta-Analysis for Historical Placebo Data



- ABBV-599 vs upadacitinib (Borrowing upadacitinib)

For upadacitinib control group versus ABBV-599, historical data consist of 206 subjects on 15 mg from two upadacitinib studies (Phase 2: Study M13-550 Balance I and Phase 3: Study M13-542 Select Beyond). Upadacitinib-Phase II gives n = 52, mean (SD) = –2.2 (1.5), and upadacitinib-Phase III shows n = 154, mean (SD) = –2.37 (1.6). The meta-analysis results in the predicted mean (SD) and 95% confidence interval are –2.32 (1.57) and [–2.54, –2.11] (Figure D-2) based on Empirical Bayes estimator [Sidik 2007]. Since $\tau^2 = 0$, the historical effective sample size is 206.

Figure D-2. Meta-Analysis for Historical Upadacitinib Data



In addition, upadacitinib-Phase III study data becomes available after this study design. A sensitivity analysis will be performed to include placebo data from upadacitinib-Phase III study in the meta-analysis (n = 144, mean (SD) = –1.01 (1.3)).

D-2 Borrowing size and prior distribution determination

For the primary analysis, our principle for fixed borrowing historical sample size is not to exceed the sample size in the current study. For placebo control borrowing, the fixed historical placebo borrowing sample size will not exceed 20. The upadacitinib fixed historical borrowing will not exceed 40.

- ABBV-105 vs Placebo (Borrowing Placebo)

Assume the mean for ABBV-105 is –1.65 and for placebo is –0.80 (from meta-analysis) so the absolute effect size is $\delta = 0.85$. Let r define the relative bias from historical data, i.e., $|\mu_C - \mu_h| = r\delta$ where μ_C is the true but unknown mean for control group (placebo) and μ_h is the historical data mean calculated from meta-analysis. To assess the bias impact on historical data borrowing, a reasonable limit for bias $r\delta$ would be the half confidence interval width from meta-analysis i.e., $r\delta \leq 0.26$. So

$$r \leq 0.26/0.85 = 0.31.$$

Let $a_0 = \frac{n_h}{n_h+n_c}$ be the proportion of borrowed number of historical patients (n_h) among all control patients ($n_h + n_C$ where n_C is the number of control group subjects in current study). Let μ_T and n_T be the treatment group mean and the number of subjects. Given r, and a_0, we can obtain Bayesian type I error rate/false positive rate (i.e., the probability achieving $P(\mu_T - \mu_C < 0|Data) > 1 - \alpha$ under the null hypothesis $\mu_T = \mu_C$):

$$\Phi\left[\frac{\pm\sqrt{2}ra_0\left(\Phi^{-1}(1-\alpha)+\Phi^{-1}(1-\beta)\right)-\Phi^{-1}(1-\alpha)\sqrt{1+k(1-a_0)}}{\sqrt{1+k(1-a_0)^2}}\right] \qquad (1)^*$$

and Bayesian power/true positive rate (i.e., the probability achieving $P(\mu_T - \mu_C < 0|Data) > 1-\alpha$ under the null hypothesis $\mu_C - \mu_T = \delta$):

$$\Phi\left[\frac{\sqrt{2}(1\pm r a_0)\left(\Phi^{-1}(1-\alpha)+\Phi^{-1}(1-\beta)\right)-\Phi^{-1}(1-\alpha)\sqrt{1+k(1-a_0)}}{\sqrt{1+k(1-a_0)^2}}\right]. \qquad (2)^*$$

The ± in above formula corresponds to two directions of the bias, α and $1-\beta$ are type I error and power for sample size calculation without borrowing, and $k = n_T/n_C$ is the randomization ratio in current study with reduced control group due to borrowing.

* (1) and (2) could be derived as follows:

Probability of trial success = $P(P(\mu_T - \mu_C < 0|Data) > 1-\alpha)$

$$= \Phi\left(\frac{\mu_C - \mu_T}{\sigma\sqrt{\frac{1}{n_T}+\frac{1-a_0}{n_h+n_C}}} - \frac{a_0(\mu_h-\mu_C)}{\sigma\sqrt{\frac{1}{n_T}+\frac{1-a_0}{n_h+n_C}}} - \Phi^{-1}(1-\alpha)\frac{\sqrt{\frac{1}{n_T}+\frac{1}{n_h+n_C}}}{\sqrt{\frac{1}{n_T}+\frac{1-a_0}{n_h+n_C}}}\right)$$

The treatment group sample size $n_T = \frac{2\left(\Phi^{-1}(1-\alpha)+\Phi^{-1}(1-\beta)\right)^2\sigma^2}{\delta^2}$ which is the balanced sample size without borrowing. With $k = n_T/n_C$, the power without borrowing equals

$$\Phi\left[\sqrt{\frac{2}{1+k}}\left(\Phi^{-1}(1-\alpha)+\Phi^{-1}(1-\beta)\right)-\Phi^{-1}(1-\alpha)\right].$$

With $k = n_T/n_C$, $|\mu_C - \mu_h| = r\delta$ and $a_0 = \frac{n_h}{n_h+n_c}$, (1) could be derived for $\mu_T - \mu_C = 0$ and (2) could be obtained from $\mu_C - \mu_T = \delta$.

Table D-2. Bayesian type I error/false positive rate (Error I) and power/true positive rate for ABBV-105 vs Placebo

Borrowing # of Subjects	$r = 0.1$				$r = 0.2$				$r = 0.3$			
	Error I		Power		Error I		Power		Error I		Power	
	–	+	–	+	–	+	–	+	–	+	–	+
0	0.05	0.05	0.65	0.65	0.05	0.05	0.65	0.65	0.05	0.05	0.65	0.65
10	0.028	0.041	0.741	0.793	0.023	0.049	0.712	0.816	0.019	0.058	0.682	0.838
20	0.021	0.040	0.796	0.868	0.014	0.053	0.753	0.896	0.010	0.071	0.706	0.920
30	0.018	0.041	0.831	0.907	0.011	0.060	0.780	0.934	0.007	0.085	0.722	0.955
40	0.016	0.044	0.853	0.929	0.010	0.067	0.798	0.954	0.005	0.099	0.734	0.971
50	0.016	0.047	0.868	0.943	0.009	0.074	0.811	0.965	0.005	0.113	0.742	0.980
Borrowing # of Subjects	$r = 0.4$				$r = 0.5$				$r = 0.6$			
	Error I		Power		Error I		Power		Error I		Power	
	–	+	–	+	–	+	–	+	–	+	–	+
0	0.05	0.05	0.65	0.65	0.05	0.05	0.65	0.65	0.05	0.05	0.65	0.65
10	0.015	0.069	0.651	0.858	0.012	0.080	0.619	0.876	0.010	0.094	0.587	0.893
20	**0.007**	**0.093**	**0.655**	**0.939**	0.004	0.119	0.600	0.954	0.003	0.150	0.544	0.967
30	0.004	0.118	0.657	0.970	0.002	0.158	0.588	0.980	0.001	0.206	0.515	0.988
40	0.003	0.142	0.660	0.982	0.001	0.195	0.579	0.990	0.001	0.258	0.495	0.994
50	0.002	0.163	0.662	0.989	0.001	0.227	0.573	0.994	0.000	0.303	0.481	0.997

* For sample size = 40:20 (k = 2), α = 0.05, 1 – β = 80%, δ = 0.85.

Table D-2 shows the Bayesian type I error rate/false positive rate and power/true positive rate for different number of borrowing and different relative bias (r = 0.1 - 0.6) given sample size = 40:20 (k = 2), α = 0.05, 1 – β = 80%, δ = 0.85. For $r \leq 0.4$, with the sample size = 40:20 (k = 40/20 = 2) for ABBV-105 and placebo respectively, we could borrow up to 20 historical placebo subjects while controlling type I error rate/false positive rate < 10% and not exceeding in study placebo sample size. For example, for $r = 0.3$, borrowing 20 will control false positive rate at 7%.

Therefore, for this analysis with historical data borrowing, we would borrow historical data: $n_h = 20, \mu_h = -0.80$, and its SE $\sigma_h = 1.5/\sqrt{20} = 0.335$. That is, the prior distribution for placebo would be $N(-0.80, 0.335^2)$. For ABBV-105, a "flat" non-

informative prior probability distribution will be assumed, that is the posterior distribution is totally from the data.

A supportive analysis will also be performed as dynamic borrowing based on the observed *r* and Table D-2. For example, if *r* = 0.2, we will borrow up to 50 historical sample size. On the other hand, if *r* = 0.5 in positive (+) direction, we just borrow 10; if *r* = 0.5 in negative (–) direction, we will not borrow (borrow 0). Overall, if borrowing provides a true positive rate above 65% and false positive rate less than 10%, we conduct dynamic historical borrowing by picking up the appropriate borrowing sample size to achieve maximum true positive rate.

- **ABBV-599 vs Upadacitinib**

Assume the mean for ABBV-599 is –2.65 and for upadacitinib is –2.32 so the absolute effect size is δ = 0.33. To control the bias with historical data borrowing, a reasonable limit for bias $r\delta$ would be the half confidence interval width from meta-analysis i.e., $r\delta \leq 0.215 \; or \; r \leq 0.65$. With the same formula as for ABBV-105 vs placebo, we obtained the Bayesian type I error/false positive rate and power/true positive rate in Table D-3 for *r* = 0.6/0.65/0.7, respectively.

Table D-3. Bayesian type I error/false positive rate (Error I) and power/true positive rate for ABBV-599 vs Upa

Borrowing # of Subjects	$r = 0.6$				$r = 0.65$				$r = 0.7$			
	Error I		Power		Error I		Power		Error I		Power	
	–	+	–	+	–	+	–	+	–	+	–	+
0	0.05	0.05	0.34	0.34	0.05	0.05	0.34	0.34	0.05	0.05	0.34	0.34
10	0.028	0.057	0.298	0.422	0.027	0.059	0.294	0.427	0.026	0.061	0.289	0.432
20	0.018	0.066	0.273	0.497	0.017	0.070	0.264	0.507	0.016	0.073	0.256	0.517
30	0.012	0.076	0.254	0.561	0.011	0.081	0.243	0.574	0.010	0.087	0.233	0.587
40	0.009	0.087	0.241	0.612	**0.008**	**0.093**	**0.228**	**0.628**	0.008	0.100	0.216	0.643
50	0.007	0.097	0.231	0.654	0.007	0.105	0.217	0.671	0.006	0.114	0.204	0.688

* For sample size = 60:40 (k = 1.5), α = 0.05, 1 – β = 39%, δ = 0.33.

Table D-3 shows the Bayesian type I error rate/false positive rate and power/true positive rate for different number of borrowing and different relative bias given sample size = 60:40 (k = 1.5), $\alpha = 0.05$, $1 - \beta = 39\%$, $\delta = 0.33$. For $r \leq 0.65$ in positive direction, sample size = 60:40 for ABBV-599 and upadacitinib respectively, we could borrow 40 historical placebo subjects while controlling type I error rate < 10% and no more than in study sample size 40. Therefore for historical data borrowing analysis, we have $n_h = 40, \mu_h = -2.32,$ and its SE $\sigma_h = 1.57/\sqrt{40} = 0.25.$ That is, the prior distribution for upadacitinib would be $N(-2.32, 0.25^2)$. For ABBV-599, a "flat" non-informative prior probability distribution will be assumed. That is the posterior distribution is totally based on the observed data.

In addition, a supportive analysis using dynamic borrowing will also be performed based on the observed *r* and extended Table D-3 (not presented). For example, if bias is positive and $r = 0.6$, we could borrow 50 historical sample size to control false positive rate within 10%. On the other hand, if bias is negative, we will not borrow (borrow 0). Overall, if borrowing provides a true positive rate above 34% and the false positive rate less than 10%, we conduct dynamic historical borrowing by picking up the appropriate borrowing sample size to achieve maximum true positive rate.

D-3 Analysis Methods

The prior distributions discussed in Section D-2 will be combined with the observed data in the trial to obtain posterior distributions for the treatment and control groups. The probability of Treatment mean – Control mean < 0 given the observed data i.e., $P(\mu_T - \mu_C < 0|Data)$ will be calculated.

Let $D = (x_1, \cdots, x_n)$ be the data and $x_i \sim N(\mu, \sigma^2)$. Then $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$. If σ^2 is a constant, we have the likelihood for the data

$$L(\mu|D) = p(D|\mu) \propto N(\bar{x}, \frac{\sigma^2}{n}).$$

The conjugate prior for μ is supposed to be

$$p(\mu) \propto N(\mu_0, \sigma_0^2).$$

Then the posterior is given by

$$p(\mu|D) = L(\mu|D)p(\mu) \propto N(\mu_n, \sigma_n^2) \text{ where}$$

$$\mu_n = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} \text{ and}$$

$$\sigma_n^2 = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}.$$

If $P(\mu_T - \mu_C < 0|Data) > 0.95$, the treatment group will be claimed as significantly better than the control group. This is equivalent to the hypothesis test of $\mu_T - \mu_C = 0$ at one-sided $\alpha = 0.05$ significant level.

Notice that (μ_0, σ_0^2) are prior parameters determined from historical data borrowing and meta-analysis (see Section D-2). $\left(\bar{x}, \frac{\sigma^2}{n}\right)$ are LSMean and SE from MMRM based on data in the study. For treatment group, a "flat" non-informative prior probability distribution will be assumed, that is the posterior distribution is totally based on the observed data.