Protocol B1801381

A MULTICENTER OPEN-LABEL STUDY OF ETANERCEPT WITHDRAWAL AND RETREATMENT IN SUBJECTS WITH NON-RADIOGRAPHIC AXIAL SPONDYLOARTHRITIS WHO ACHIEVED ADEQUATE 24 WEEK RESPONSE

Statistical Analysis Plan
(SAP)

Version: 2

Date: 6-Dec-2017

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B1801381 is based on the protocol dated 24JUN2015.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Location and Change	Rationale
1	Not Applicable	
2	Throughout Typo corrections	Clarification
	Section 3.1 Rearrange ASDAS section for clarity	Clarification
	Section 5.2.4 Modify time of relapse to be next scheduled visit for relapses at unscheduled visits – comparator study B1801031 did not have any option to identify flares at unscheduled visits	Clarification based on BDR 1 review
	Section 5.3 Added detail to define use of unscheduled visits in LOCF summaries	Clarification based on BDR 1 review
	Section 6.2.1 Corrected method of imputation, removed reference to logistic regression (all time to flare modelling will use Cox regression) and added specificity to results reporting.	Time to event specified LOCF, but no imputation is used. Changed univariate summaries of subgroup analysis of time to flare to use Cox models to use one consistent approach for modeling flare.
	Section 6.4 Changed method from logistic regression to Cox regression, added specificity to results reporting	Cox models to use one consistent approach for modeling flare.
	Appendix 1.1 Added NSAID classification for NSIADs not on original list	Added NSAIDS not in original classification list based on BDR 1 findings

	Appendix 2 Clarified specification of endpoints to be presented by period 1 sustained ASDAS category (Y/N)	Clarification needed after review of BDR 1 tables.
	Appendix 3 Added. Visit windows defined to include unscheduled visits in tabulations	Requested by programming to document interactive decisions.

2. INTRODUCTION

The purpose of this document is to provide further details about the statistical analysis methods specified in study protocol B1801381.

2.1. Study Objectives

2.1.1. Primary Objective

- To estimate the proportion of subjects who flare within 40 weeks following withdrawal of ETN in subjects who have achieved ASDAS CRP less than 1.3 (inactive disease).

2.1.2. Secondary Objectives

- To estimate time to flare after withdrawal of ETN, and to compare it to that in patients from B1801031 who continued ETN therapy.
- To estimate the efficacy of 12 weeks of retreatment in subjects who experience a flare after withdrawal of ETN.
- To estimate the efficacy of ETN over 24 weeks of initial treatment.
- To estimate the safety and tolerability of ETN in this population.

2.2. Study Design

This multicenter, open-label, three period study will evaluate withdrawal and retreatment of ETN in subjects with nr-ax SpA who achieved an adequate response, as measured by ASDAS CRP less than 1.3 (inactive disease) following 24 weeks of treatment. The study is expected to randomize approximately 200 subjects in Period 1, in order to have approximately 96 subjects qualify for Period 2.

2.2.1. Period 1

This is an open-label, 24-week period in which all eligible subjects with nr-ax SpA will be enrolled and treated with ETN 50 mg once weekly plus a stable background NSAID at the optimal tolerated anti-inflammatory dosage as determined by the investigator. The target for this period is therapeutic response defined as achieving Ankylosing Spondylitis Disease Activity Scale C-reactive protein (ASDAS CRP) less than 1.3 at Week 24. Subjects who

qualify at Week 24 will enter Period 2 after all Period 1 procedures are completed. Subjects who do not achieve ASDAS CRP less than 1.3 will not enter Period 2 and will complete the study following the Week 28 day follow-up phone call/visit.

2.2.2. Period 2

The Week 24 visit ends Period 1 and marks the beginning of Period 2. This is a 40-week withdrawal period where subjects will discontinue ETN following the Week 24 dose, yet maintain the background NSAID. If the subject has not flared by the Week 64 visit, their participation in the study is complete.

2.2.3. Period 3

Subjects who flare (defined as an ASDAS ESR greater than or equal to 2.1) during Period 2 will enter a retreatment period and receive approximately 12 weekly doses of open-label ETN. Subjects experiencing increased disease activity can come into the office any time for an evaluation; if flare criteria are met, open-label ETN is started at this unscheduled visit. If the subject has experienced increased disease activity during Period 2, but does not meet the protocol defined criteria for flare, it is the investigator's judgment to determine if the subject should remain in the study or discontinue the study to pursue alternative treatment.

2.2.4. Follow-up

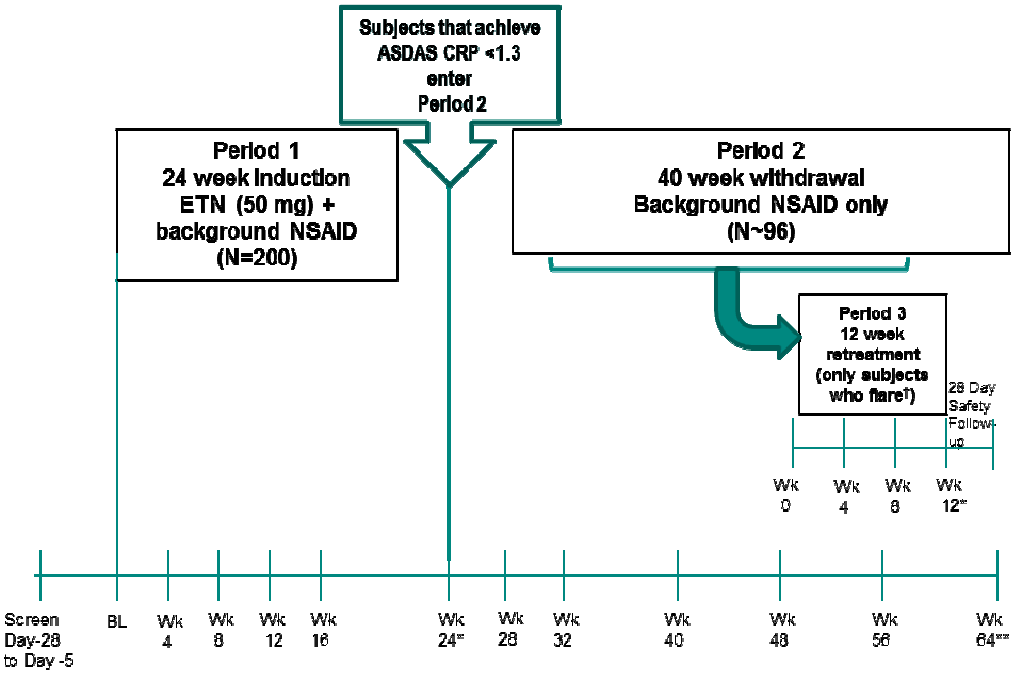
A safety follow-up phone call/visit will only be completed for subjects who:

- Receive investigational product (IP) in Period 1, but who do not enter Period 2, or,
- Enter Period 2 and discontinue from the study before the Week 28 visit or,
- Enter Period 3 (Retreatment period).

The purpose of this visit is to assess any new and/or ongoing adverse events and will be performed approximately 28 days after the last dose of IP. This follow-up visit will be done by telephone, unless local regulations require a visit to the study site.

A safety follow-up visit does not need to be completed for subjects who discontinue from the study after the Week 28 visit and who do not enter Period 3 for retreatment.

Figure 1. Study Design



†Flare is defined as ASDAS ESR ≥2.1
*The follow-up visit is only required 28 days following the last dose of IP. This will occur either at the end of Period 1 (for subjects that don't qualify for Period 2) or at the end of Period 3 (for subjects who flare and require retreatment).
**Subjects who flare at Week 64 will receive 12 weeks of OL ETN and will have their final study visit at Week 76 and their follow-up visit 28 days later.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

The primary endpoint for this study is occurrence of flare within 40 weeks (defined as an ASDAS ESR greater than or equal to 2.1) following withdrawal of ETN. The key secondary endpoint is the time to flare after withdrawal of ETN.

ASDAS is calculated from the following parameters:

1. Total back pain (BASDAI question 2) (units CM, range 0-10).
2. Subject global assessment of disease activity (units CM, range 0-10).
3. Peripheral pain/swelling (BASDAI question 3) (units CM, range 0-10).
4. Duration of morning stiffness (BASDAI question 6) (units CM, range 0-10).
5. C-reactive protein (CRP) in mg/litre or ESR in mm/h.

The ASDAS score is then calculated as follows:

ASDAS ESR = (0.08 x total back pain) + (0.11 x subject global) + (0.09 x peripheral pain/swelling) + (0.07 x duration of morning stiffness) + (0.29 x √(ESR).

ASDAS CRP = (0.121 x total back pain) + (0.110 x subject global) + (0.073 x peripheral pain/swelling) + (0.058 x duration of morning stiffness) + (0.579 x Ln(CRP+1)).

Note: for ASDAS CRP, hsCRP observed values of <2 mg/L will be converted to the value of 2 in the calculation of ASDAS CRP. Also, for values ≥2 mg/L that are below the limit of detection (conventional CRP), as indicated by "<" in the field (eg, <4.1), a value of 2 should be used in the calculation of ASDAS CRP.

3.2. Secondary Endpoint(s)

The following key secondary endpoint will be estimated:

- The time to flare following withdrawal of ETN (as measured from treatment withdrawal until ASDAS ESR greater than or equal to 2.1).

The following secondary endpoints and outcome measures will be estimated within 40 weeks following withdrawal of ETN and during the 12 week re-treatment period (if applicable):

- Occurrence of ASDAS CRP inactive disease.

ASDAS CRP inactive Disease is defined as ASDAS CRP less than 1.3:

- Occurrence of ASAS 20 and ASAS 40.

ASAS 40 responders are defined as subjects who satisfy the following criteria:

1. An improvement of at least 40% and absolute improvement of at least 2 units on a 0 to 10 cm scale (converted from 0 to 100 mm) or an improvement of 100% for those domains that have a baseline score <2 in at least 3 of the following 4 domains:

    - Subject assessment of disease activity,
    - Mean of subject assessment of total back pain,
    - Function represented by the BASFI score,
    - Inflammation represented by the mean of the two morning stiffness-related BASDAI scores.

2. No worsening at all in any of the domains.

ASAS 20 responders are defined as subjects who satisfy the following criteria:

1. An improvement of at least 20% and absolute improvement of at least 1 unit on a 0 to 10 cm scale (converted from 0 to 100 mm) in at least 3 of the following 4 domains:

    - Subject global assessment of disease activity,
    - Mean of subject assessment of total back pain,
    - Function represented by the BASFI score,
    - Inflammation represented by the mean of the two morning stiffness-related BASDAI scores.

2. Absence of deterioration (of at least 20% and absolute change of at least 1 unit) in the potential remaining domain.

    - Occurrence of ASAS partial remission.

    ASAS partial remission is based on the same 4 domains as all of the other ASAS endpoints. Partial remission is defined as a score of 2 or less (on a scale of 0-10 cm) for each of the 4 domains.

ASDAS:

ASDAS ESR and ASDAS CRP continuous endpoints, as defined in Section 6.1.1.

- Occurrence of ASDAS major improvement and clinically important improvement.

    ASDAS major improvement is defined as an improvement ≥2.0 from baseline.

    ASDAS clinically important improvement is defined as an improvement ≥1.1 from baseline.

- Nocturnal and total back pain;

    Subjects will assess their nocturnal back pain and total back pain over the last 48 hours by placing a mark that corresponds to the magnitude of their pain on a 100 mm pain scale that ranges from 0 mm (none) to 100 mm (severe). Change from baseline at each visit in each of these endpoints, as well as average back pain (calculated as the mean of the nocturnal and total), are the endpoints of interest. The reported values will be converted to cm for analysis purposes.

- Bath Ankylosing Spondylitis Functional Index (BASFI) and its components;

The BASFI consists of 10 questions related to the subject's ability to function. Each is scored by the subject on a 100 mm scale ranging from 0 (Easy) to 100 (Impossible). The BASFI score is the mean of the scores for these 10 questions. If 1 or 2 answers are missing, the BASFI score should be based on the available scores. If more than 2 are missing, then the BASFI score should be considered missing. Changes from baseline in the BASFI score, as well as the scores for each individual question, are endpoints of interest. The reported values will be converted to cm for analysis purposes.

- Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and its components;

The BASDAI consists of 6 questions related to disease activity. Each of the first 5 questions is scored by the subject on a 100 mm scale ranging from 0 (None) to 100 (Very severe). The sixth question, related to duration of morning stiffness, is on a scale for 0 (0 hours) to 100 (2 hours). The reported values will be converted to cm for analysis purposes.

The BASDAI score is obtained by computing the mean score for the 2 questions related to morning stiffness (questions 5 and 6) and then adding that value to the sum of the scores for the first 4 questions and then dividing the total by 5. This can be written as

BASDAI=(Q1+Q2+Q3+Q4+(Q5+Q6)/2)/5

The total score will range from 0 to 10.

If the answer to either question 5 or 6 is missing, then the score for the one available question can be used. If both are missing, then the BASDAI is considered missing.

Regarding the other 4 questions, if more than 1 of those is missing, then the BASDAI is not calculated.

Changes from baseline in the BASDAI score, as well as the scores for each individual question, are endpoints of interest.

- Occurrence of BASDAI 50;

The BASDAI 50 endpoint is defined as the proportion of subjects who achieved a 50% improvement (decrease) from baseline in their BASDAI score.

- High sensitivity C Reactive Protein (hsCRP);

The lab value of hsCRP will be evaluated as both a continuous endpoint as proportion normal at each visit.

Health Outcomes Assessments using the following instruments:

- EuroQoL-5D Health Questionnaire (EQ-5D).

There are 5 questions in the EQ-5D, one for each dimension of health (mobility, self care, usual activities, pain/discomfort and anxiety/depression). Each question has three possible responses (no problem in the relevant health dimension to severe problems). If a patient reports no problem in any health dimension, then his EQ-5D score is 1. Otherwise, the EQ-5D index is calculated following the table below, with the amount shown for each response subtracted from a starting (perfect health) score of 1.

EuroQol dimension	Response = 2	Response = 3
Mobility	0.069	0.314
Self-care	0.104	0.214
Usual activity	0.036	0.094
Pain/discomfort	0.123	0.386
Anxiety/depression	0.071	0.236
Any response level	0.081*	0.269
* Subtracted if any score of 2 or 3 is present – if a score of 3 is present, both 0.081 and 0.269 are subtracted.		

For example, if a subject has EQ-5D assessment scores of 1, 1, 2, 2 and 3 for the 5 dimensions, the utility weight would be:

Full Health =1.000
Constant term (for any dysfunctional state) -0.081
Mobility (level 1) -0
Self-Care (level 1) -0
Usual Activities (level 2) -0.036
Pain or discomfort (level 2) -0.123
Anxiety or depression (level 3) -0.236
Any dimension at level 3 -0.269

Therefore, the estimated value for 11223 is 1-0.081-0-0-0.036-0.123-0.236-0.269 = 0.255

EQ-5D score will be defined as missing if one or more items are missing.

In addition to the 5 questions above, there is a separate question about overall health state measured on a 0 to 100 mm scale where 0 = worst imaginable health state and 100 = best imaginable health state.

In addition to the continuous scores, the following dichotomous endpoints will be summarized:

Proportion with EQ-5D VAS score of >82 (population norm)
Proportion with EQ-5D index improvement from baseline of ≥0.05
Proportion of each EQ-5D item score (5 total) reporting ‘no problem’

- SF-36

The SF-36 is a widely used generic quality of life instrument that assesses the subject's general health and functional status. It consists of 36 questions that are grouped into 8 domains (physical functioning, vitality, social functioning, mental health, role physical, bodily pain, role emotional, and general health). Domain scores range from 0-100, with greater scores reflecting better health status. In addition, SF-36 has 2 summary scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS). The summary scores are constructed as T scores with a mean ± SD of 50 ±10 where 50 is the average score in the general population. Higher scores indicate better health status.

In addition to the continuous scores, the following dichotomous endpoints will be summarized:

- Proportion with improvement from baseline in Mental Component Score of >=2.5;
- Proportion with improvement from baseline in Mental Component Score of >=5;
- Proportion with improvement from baseline in Physical Component Score of >=2.5;
- Proportion with improvement from baseline in Physical Component Score of >=5;
- Work Productivity and Activity Impairment (WPAI).

The WPAI assesses work productivity and impairment. It is a 6-item questionnaire used to assess the degree to which a specified health problem affected work productivity and regular activities over the past 7 days. The questions are as follows:

Q1 = currently employed

Q2 = hours missed due to health problems

Q3 = hours missed other reasons

Q4 = hours actually worked

Q5 = degree health affected productivity while working (0-10 scale)

Q6 = degree health affected regular activities (0-10 scale)

Subscale scores that are calculated from these questions are:

Percent work time missed due to health problem: Q2/(Q2+Q4)

Percent impairment while working due to health problem: Q5/10

Percent overall work impairment due to health problem: Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4))*(Q5/10)]

Percent activity impairment due to health problem: Q6/10.

Each subscale score is expressed as an impairment percentage (0-100) where higher numbers indicate greater impairment and less productivity.

If a response is missing, any subscale score that requires that response will be considered missing.

- MRI SIJ/spine as measured by Spondyloarthritis Research Consortium of Canada (SPARCC).

Scoring of sacroiliac joints will also use the SPARCC method. Scoring will be based on 6 consecutive coronal slices from posterior to anterior.

Each joint will be divided into 4 quadrants. Each quadrant will be assigned a score of 0 = no lesion or 1 = increased signal. This part of the scoring allows for a total score ranging from 0-8 for the 2 joints of one coronal slice. For each slice, the score is increased by 1 for each joint that exhibits an intense signal in any quadrant (thereby allowing for an increase of up to 2 points in the total score for each slice). Also, for each slice, an additional score of 1 will be given for each joint that includes a lesion demonstrating continuous increased signal of a depth ≥1 cm from the articular surface (thereby allowing for an additional increase of up to 2 points for each slice). Therefore, the maximal score for each coronal slice is 12 and the maximum score for 6 slices is 72.

The following secondary endpoint will be estimated over 12 weeks following re-treatment of patients who flare:

- Time to ASDAS inactive disease after re-treatment.

Subjects are required to have flared to enter the retreatment period. This endpoint is the time from first dose of retreatment until the first observed event of ASDAS inactive disease. Subjects who do not achieve ASDAS inactive disease will be censored at the time of the last ASDAS evaluation in the interval.

Other endpoints:

- Subject Assessment of Disease Activity (SADA).

Change from baseline in Subject Assessment of Disease Activity - VAS at each visit. Subjects will assess the overall disease activity over the last 48 hours by placing a mark that corresponds to the magnitude of their pain on a 100 mm pain scale that ranges from 0 mm (none) to 100 mm (severe). The reported values will be converted to cm for analysis purposes.

- Physician Global Assessment (PGA).

Change from baseline in VAS Physician Global Assessment at each visit. The investigator will estimate the subject's overall disease activity over the last 48 hours by placing a mark on a 100 mm scale that ranges from 0 mm (none) to 100 mm (severe). The reported values will be converted to cm for analysis purposes.

- Bath Ankylosing Spondylitis Patient Global Assessment Score (BAS-G).

The BAS-G consists of 2 questions:

How have you been over the last week?

How have you been over the last six months?

Each is scored by the subject on a 100 mm scale ranging from 0 (Very Good) to 100 (Very Bad). The two values are averaged to obtain the BAS-G score. (If either question is unanswered, then no BAS-G score will be calculated and it will be considered missing. The reported values will be converted to cm for analysis purposes.

In addition to the average BAS-G score, the first question itself will be analyzed separately at each visit.

- Tender and swollen joint counts (44 count).

There are 44 joints to be evaluated for pain/tenderness (the specific joints involved are identified in the protocol). The tender/painful joint count is the number of joints that have pain/tenderness marked as present. The count is considered missing if more than 20% (8 of the joints have missing scores. If the number of missing joints is ≤8, then the count is (44/n) * the count based on the non-missing joints, where n is the number of non-missing joints.

The same 44 joints are to be evaluated for swelling. The swollen joint count is the number of joints that have swelling marked as present. The count is considered missing if more than 20% (8) of the joints have missing scores. If the number of missing joints is ≤8, then the count is (44/n) multiplied by the count based on the non-missing joints, where n is the number of non-missing joints.

Changes from baseline in the number of painful/tender joints and the number of swollen joints are of interest.

- Dactylitis and enthesitis score (Maastricht Ankylosing Spondylitis Enthesitis Score [MASES]).

Dactylitis: Each of the 10 fingers and 10 toes is evaluated for dactylitis. A score of 0, 1, 2 or 3 (where 0 = none, 1= mild, 2 = moderate, 3 = severe) is assigned to each. A total score which can range from 0 to 60 is obtained by adding the scores for the 20 digits. The endpoints of interest based on these data will be the proportion of subjects with 1 or more digits affected, the number of digits affected and the dactylitis total score.

MASES: There are 13 sites evaluated for enthesis: 1st costochondral joint (left/right), 7th costochondral joint (l/r), posterior superior iliac spine (l/r), posterior anterior iliac spine (l/r), iliac crest (l/r), proximal insertion of Achilles tendon (l/r) and 5th lumbar spinous process. Each site is scored as 0 or 1 depending on whether enthesis is present or absent. The MASES score is the sum of all site scores and can range from 0 to 13.

If the number of sites with missing values is 1 or 2, then the MASES score is calculated as (13/n) * the count based on the non-missing joints, where n is the number of non-missing joints. If more than 2 are missing, then the MASES score is missing.

3.3. Other Endpoints

3.3.1. Non-steroidal Anti-Inflammatory Drug (NSAID) Usage

Any NSAIDs taken as background medications during the study will be identified along with start and stop dates, average daily dose (mg) and treatment frequency (days per week). Details regarding the calculation can be found in Appendix 1.1. This intake score will be calculated for each study period.

3.4. Baseline Variables

Period 1 baseline values will be defined as the last available value of the screening and baseline visits (eg, if the baseline visit value is missing and screening is available, the screening value will be used as baseline). Period 2 baseline is defined as the last value before treatment withdrawal. Period 3 baseline is defined as the last period 2 value before re-treatment starts. P-values for change from baseline will be calculated using the paired t-test.

For response endpoints defined using change from baseline (eg, ASAS40) period 1 baseline will always be used as the reference point.

3.5. Safety Endpoints

3.5.1. Adverse Events

Adverse events (AEs) recorded throughout the study will be coded using the MedDRA dictionary. The number and percentage of patients experiencing at least one adverse event will be summarized overall, and by body system and preferred term. In any given category, subjects will only be counted once. Adverse events will further be categorized and summarized by severity and study drug relationship, as well as those that cause withdrawal, result in a dose reduction or in concomitant therapy. Adverse events that occur up to 28- days after the last dose test article administration will be attributed to the treatment period.

3.5.2. Laboratory Evaluations

Laboratory evaluations include blood chemistry and hematology, urinalysis. Plasma and whole blood biobank may also be collected. Baseline for laboratory evaluations will be obtained from the baseline visit.

3.5.3. Vital Signs

Vital signs will include sitting systolic and diastolic blood pressure, sitting pulse, weight and height. Baseline for vital sign parameters will be obtained from the baseline visit.

3.5.4. Inflammatory Bowel Disease (IBD), Psoriasis, and Acute Anterior Uveitis

3.5.4.1. Inflammatory Bowel Disease

The proportion of subjects diagnosed with IBD during each period will be determined for each treatment group. The denominator will include only those subjects who did not have a history of IBD.

The proportion of subjects who had a flare of IBD during each period will also be calculated by treatment group. The denominator will include only those subjects with a previous history of IBD.

3.5.4.2. Psoriasis

The proportion of subjects with psoriasis or with a flare of psoriasis will be calculated in a manner similar to that described for IBD in the previous section. That is, there will be separate calculations done depending on whether or not subjects had psoriasis reported as part of their medical history.

In subjects who had a flare, the proportion of subjects who had change in psoriasis morphology will also be determined.

3.5.4.3. Acute Anterior Uveitis

The proportion of subjects with acute anterior uveitis or with a flare of such will be calculated in a manner similar to that described for IBD and psoriasis. That is, there will be separate calculations done depending on whether or not subjects had uveitis reported as part of their medical history.

In addition, for those subjects newly diagnosed and for those subjects who had a flare, the proportion of subjects who received treatment and the proportion of subjects who received certain types of treatment will also be summarized. For those subjects who had a flare, the mean number of flares will be calculated.

3.5.5. Nonstudy Medications

Prior and concomitant medications will be coded using the WHO Drug dictionary. The numbers of subjects using concomitant medications will be categorized by drug category and preferred term, and presented for each treatment group.

Nonstudy medications will be summarized separately for each period.

4. ANALYSIS SETS

4.1. Full Analysis Set

4.1.1. Period 1

The full analysis set for period 1 will include all subjects who took study medication and had one evaluation after baseline.

4.1.2. Period 2

The full analysis set for period 2 will include all subjects who had at least one evaluation during period 2.

4.1.3. Period 3

The full analysis set for period 3 will include all subjects who took study retreatment medication and had at least one evaluation after restarting active therapy.

4.2. 'Per Protocol' Analysis Set

A per-protocol population may be defined for sensitivity analysis of the primary and key secondary efficacy endpoints (ASDAS inactive disease, ASAS 40). This will be defined separately for the withdrawal and retreatment periods. The per-protocol populations will be defined as subsets of each FAS population excluding subjects who have a protocol violation that is thought to potentially affect efficacy, with focus on potential violations of inclusion criteria in each period. There will be no per protocol population defined for period 1 of the study.

4.3. Safety Analysis Set

The safety analysis set for each period will include all subjects who entered the period; for periods 1 and 3, subjects must also take one dose of investigational drug.

4.4. Other Analysis Sets

N.A.

5. GENERAL METHODOLOGY AND CONVENTIONS

The final primary analysis will be performed when all subjects have completed the study using the final study database.

5.1. Hypotheses and Decision Rules

This study is an open-label study with one treatment group. Inferences will be based on point estimates and confidence intervals, rather than on hypothesis tests.

Except where stated, all confidence intervals will be 2-sided 95% confidence intervals.

5.2. General Methods

This is an open-label study so the statistical focus will be on estimation rather than hypothesis testing. Efficacy and outcomes endpoint analyses will use the full analysis set (FAS), safety analyses will use the safety analysis set.

As a general guideline, the format of tables produced will follow the specifications outlined in this analysis plan. Descriptive statistics for categorical variables (nominal or ordinal) will include the number (frequency), and percentage of subjects at each level of response. Continuous variables will be summarized by sample size, mean, mean changes, standard deviation, 95% CIs, minimum, maximum, median and 25^{th} and 75^{th} percentile values.

5.2.1. Analyses for Binary Data

Dichotomous clinical and subject rated endpoints measured during Periods 1, 2 and 3 will be summarized as number evaluated, percent with response of interest and 95% confidence intervals. For response endpoints defined using change from baseline (eg, ASAS40) period 1 baseline will always be used as the reference point.

5.2.2. Analyses for Continuous Data

For continuous clinical and subject rated endpoints measured during Periods 1, 2 and 3, both raw and change values will be summarized with the following descriptive statistics: n, mean, standard deviation, min and max, median and the 25^{th} and 75^{th} percentiles (quartiles).

Summaries will be produced by visit. For periods 2 and 3 changes from both the original (period 1) baseline and the specific period baseline will be produced. Both observed cases (time point completers) and last observation carried forward LOCF summaries will be produced. For the LOCF summaries, no baseline values will be carried forward. Period 1 baseline values will be defined as the last available value of the screening and baseline visits (eg, if the baseline visit value is missing and screening is available, the screening value will be used as baseline). Period 2 baseline is defined as the last value before treatment withdrawal. Period 3 baseline is defined as the last period 2 value before re-treatment starts. P-values for change from baseline will be calculated using the paired t-test.

5.2.3. Analyses for Categorical Data

N.A.

5.2.4. Analyses for Time to Event Data

Kaplan-Meier (KM) estimates, including plots, will be used to illustrate the time course of flare after treatment withdrawal. In addition to summarizing the data from the current study, the KM plots will show subjects from protocol B1801031 who met the entry requirements for the current trial withdrawal period. Those subjects did not have treatment withdrawn and will serve as a reference for evaluating the impact of treatment withdrawal. As a sensitivity analysis a life table approach will be used to compare the time to flare between the 2 studies, with time intervals based on visit windows.

In addition to the KM summaries, Cox proportional hazards models will be used to estimate the hazard ratio of continuing (B1801031) to withdrawn (B1801381) subjects.

5.3. Methods to Manage Missing Data

Unless otherwise specified, missing items for observations on the individual items of the derived efficacy and health outcome scales will be handled as follows. If more than 20% of the items are missing, the total scores for these scales will be considered missing. If 20% or fewer items are missing, the average of the available items will be multiplied by the total number of items to get a derived total score. Similarly, if more than 20% of the items from a subscale are missing the subscale will not be used in the analysis. If 20% or fewer items are missing then the average of the available items for that subscale will be multiplied by the total number of items in the subscale to get the derived subscale score.

Where an observed case analysis is being employed, the data will be analyzed without imputation for missing visit data. Where the last observation carried forward (LOCF) approach is being employed, efficacy assessments will be handled in the following manner. If the scheduled assessment is available it will be used. If the scheduled assessment is missing, the immediately preceding assessment, scheduled or unscheduled, will be used (eg, if assessments are to be done at weeks 4, 8 and 12 and data are available at weeks 4 and 12 but week 8 is missing, then the week 4 value will be carried forward to week 8). Baseline data will not be carried forward unless otherwise noted. For composite endpoints, missing components will not be carried forward to complete a composite endpoint with partial data available; only the complete composite endpoint value will be carried forward.

Missing values for safety endpoints will not be imputed.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. ASDAS flare

6.1.1.1. Primary Analysis

Endpoint: ASDAS flare (ASDAS ESR greater than or equal to 2.1).

- Analysis time points**:** Period 2 (over entire period).
- Analysis population**:** Period 2 FAS.
- method of imputation for missing data: LOCF.
- Primary analysis methodology**:** Descriptive summary only of proportion flared at each period 2 visit.

Reporting results:

- Descriptive: proportions (n/N) and 95% Confidence interval) at each period 2 visit.

- Supporting objective and Decision rule: Primary Objective.

Additional analyses for primary endpoint

- Method of imputation for missing data: None .
- Analysis methodology: Kaplan Meier estimate of time to flare and Cox proportional hazards model for evaluation of covariates and estimation of hazard compared to prior study B1801031 which had no withdrawal of treatment.
- Supporting objective and Decision rule: Primary Objective.
- Covariates used (Cox model): Covariates for the Cox regression model of time to flare will include, in addition to study (B1801031 vs B1801381), age, sex, geographic region, baseline CRP (normal vs. high), baseline MRI status (negative vs. Positive), disease duration, prior NSAID response (good vs. not) plus ASDAS values at weeks 12 and 24 or, in a separate model, sustained ASDAS inactive disease in period 1, defined as ASDAS ESR<1.3 at Week 12 or Week 16 (where ASDAS may not be greater than 2.1 at either time point) and at Week 24. Binary variables with small frequencies in either study will not be included in the Cox regression model.

Reporting results:

- Estimation:

  Kaplan-Meier estimates of 25%, median and 75% survival and 95% confidence intervals. Cox model based hazard ratio estimates for covariates and study.

- Tables: Present table with just HR and CI for "study" for 4 models:

  1. study only
  2. study plus baseline variables
  3. study plus baseline variables plus ASDAS continuous at 2 times
  4. study plus baseline variables plus sustained ASDAS remission

- Figures:

  Kaplan Meir plots showing both studies flare events.

6.2. Secondary Endpoint(s)

6.2.1. Binary endpoints

Endpoints: ASAS 40, ASAS 20, ASAS 5/6, , BASDAI50, the proportion of subjects with ASAS partial remission, the proportions of subjects with ASDAS inactive disease, major improvement and clinically important improvement, the proportion of subjects with at least 1 dactylitic digit, the proportions of subjects with any or persistent peripheral articular involvement EQ-5D index improvement ≥0.5,EQ-5D VAS score >82, each EQ-5D item proportion reporting "no problem", WPAI proportion employed, SF-36 PCS and MCS improvements of ≥2.5 and ≥5.0.

- Analysis time points: All Period 1 , 2 and 3 scheduled visits.
- Analysis population (method of imputation for missing data): FAS (LOCF and observed cases), where FAS populations are defined in each period in Section 4.1.
- Analysis methodology: Proportions will be summarized using the methods specified in Section 5.2.1.
- Supporting objective and Decision rule: Secondary Objectives.

Reporting results:

- Descriptive: The number evaluated, percent with response of interest and 95% confidence intervals. For response endpoints defined using change from baseline (eg, ASAS40) period 1 baseline will always be used as the reference point.

6.2.2. Continuous endpoints

Endpoints: ASDAS ESR and ASDAS CRP, BASDAI total and components, BASFI total and components, BAS-G total, SAGA, PGA, tender and swollen joint counts, dactylitis (total and number of affected digits) and enthesitis (MASES) scores, ESR, CRP, EQ-5D index and EQ-5D VAS, SF36 PCS and MCS, WPAI (4 derived subscales totals), SPARCC (SIJ and 6DVU).

- Analysis time points: All Period 1, 2 and 3 scheduled visits.
- Analysis population (method of imputation for missing data): FAS (LOCF and observed cases), where FAS populations are defined in each period in Section 4.1.
- Analysis methodology: Change from baseline will be summarized using the methods specified in Section 5.2.2.
- Supporting objective and Decision rule: Secondary Objectives.

Reporting results:

- Raw data: The sample size, mean, standard deviation, 95% CIs, minimum, maximum, median and 25th and 75th percentile values will be presented.

- Change from baseline: The sample size, mean, standard deviation, 95% CIs, minimum, maximum, median and 25th and 75th percentile values will be presented. Changes from period 1 baseline will be presented for all periods. For period 2 and 3, changes from the respective period baselines will also be presented.

6.3. Baseline and Other Summaries and Analyses

6.3.1. Baseline Summaries

Descriptive summaries of baseline demographic and disease characteristics will be provided. For binary/categorical characteristics sample size and the number and percent with each level will be and presented. For continuous characteristics the sample size, mean, standard deviation, 95% CIs, minimum, maximum, median and 25th and 75th percentile values will be presented.

6.3.2. Study Conduct and Subject Disposition

The number entering, completing and discontinuing from each treatment period will be tabulated for each treatment period. Discontinuations will be further summarized by reasons for discontinuation.

6.3.3. Concomitant Medications and Non-Drug Treatments

Prior and concomitant medications will be coded using the WHO Drug dictionary. The numbers of subjects using concomitant medications will be categorized by drug category and preferred term, and presented for each treatment group.

Concomitant medications and non-study treatments will be summarized separately for each period.

6.4. Subset Analyses

In addition to summarization of the overall population, the ASDAS flare endpoint will be summarized for key demographic and baseline disease characteristics for the B1801381 population. For each demographic and disease characteristic subgroup, the response in each subgroup category will be presented (responder count, group size, and response percent). In addition Cox regression models with factors for study, subgroup and study by subgroup interaction will be fit, separately for each subgroup. Results will be summarized with hazard ratios and 95% confidence intervals. The subgroups to be included are:

- Sex (Male, Female)

- Race (White, Non-White)

- Age category (<40 years, >=40 years)

- Weight (<70kg, ≥70 kg)
- Baseline hs-CRP status (High, Normal)
- HLA-B27 status (Positive, Negative)
- Concomitant use of DMARDs at Baseline (Yes, No)
- MRI sacroiliitis classification at Screening (Positive, Negative)
- Both positive MRI and elevated baseline hs-CRP (Yes/No)
- Baseline/screening MRI SIJ SPARCC score >2 (yes, no)
- History of IBD at Screening (Yes, No)
- History of uveitis at Screening (Yes, No)
- Disease duration <2 years (Yes, No).

Before the final model is determined, data patterns will be examined for feasibility issues, including highly correlated predictors and small cell size in subgroups. Adjustments will be made as necessary to address any such issues.

To support the interpretation of the primary analysis the following analysis will be performed:

- Sustained period one responders vs. responders at week 24 only.

An additional summary evaluating flare rates in those subjects who met the entry criteria only at week 24 and those who were sustained responders [subjects who, in addition to ASDAS ESR <1.3 at week 24, also have ASDAS ESR <1.3 at Week 12 or Week 16 (where ASDAS may not be greater than 2.1 at either time point)] will be provided. Populations and methodology will be applied as above.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Descriptive summaries of baseline demographic and disease characteristics will be provided. For binary/categorical characteristics sample size and the number and percent with each level will be and presented. For continuous characteristics the sample size, mean, standard deviation, 95% CIs, minimum, maximum, median and 25th and 75th percentile values will be presented.

6.5.2. Study Conduct and Subject Disposition

The number entering, completing and discontinuing from each treatment period will be tabulated for each treatment period. Discontinuations will be further summarized by reasons for discontinuation.

6.5.3. Study Treatment Exposure

Descriptive statistics for compliance will be provided. The proportion of compliant subjects, defined as those having taken at least 80% of the expected doses over the entire treatment interval, will be tabulated.

6.5.4. Concomitant Medications and Non-Drug Treatments

Prior and concomitant medications will be coded using the WHO Drug dictionary. The numbers of subjects using concomitant medications will be categorized by drug category and preferred term, and presented for each treatment group.

Concomitant medications and non-study treatments will be summarized separately for each period.

6.6. Safety Summaries and Analyses

Safety data (demographics, vital signs, laboratory data, adverse events, demographics, primary diagnosis and duration, medical history, concomitant medications, treatment durations and discontinuations) will be analyzed in accordance with Pfizer Data Standards (PDS).[redacted]

Safety analyses will be based on the Safety Analysis Population and will be tabulated separately for each study period and over all periods for etanercept. Specific tables to be included will be summarized in a separate document.

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation.

6.6.1. Adverse Events

The number and percentage of patients experiencing at least one adverse event will be summarized overall, and by body system and preferred term for each study period. In any given category, subjects will only be counted once. Adverse events will further be categorized and summarized by severity and study drug relationship, as well as those that cause withdrawal, result in a dose reduction or in concomitant therapy.

6.6.2. Laboratory Data

Continuous and binary laboratory data will be summarized following the approaches defined for those data types in Section 6.2.

6.6.3. Vital Signs

Continuous and binary vital signs data will be summarized following the approaches defined for those data types in Section 6.2.

7. INTERIM ANALYSES

7.1. Introduction

No interim analysis with the aim of stopping or modifying the trial in any way is planned. The study will not have a data monitoring committee.

7.2. Interim Analyses and Summaries

Descriptive analyses of Period 1 will be performed when all subject data for Period 1 is complete.

8. REFERENCES

N.A.

9. APPENDICES

Appendix 1. DATA DERIVATION DETAILS

Appendix 1.1. Calculation of NSAID Intake Total Score

The general formula for calculation of the NSAID intake total score is:
(equivalent NSAID score) × (days of intake during period of interest) × (days per week) / (period of interest in days).

The equivalent NSAID score in this formula can be obtained from the table below. The table shows the doses of various commonly used NSAIDs that are considered equivalent to a 150 mg dose of diclofenac. A 150 mg daily would be used to achieve optimal symptomatic control of a patient with axial SpA. Diclofenac 150 mg is assigned a score of 100. The equivalent score for other NSAIDs can be determined based on their dose. A 20 mg dose of piroxicam, since it is equivalent to a 150 mg dose of diclofenac is also assigned a score of 100. A 1200 mg dose of ibuprofen would have a score of 50 (since it is half of the equivalent dose).

NSAID	Dose (mg) comparable to 150 mg of diclofenac
Diclofenac	
Naproxen	1000
Aceclofenac	400
Celecoxib	200
Etodolac	600
Etoricoxib	90
Flurbiprofen	200
Ibuprofen	2400
Indometacin	150
Ketoprofen	200
Meloxicam	15
Nimesulide	200
Phenylbutazone	400
Piroxicam	20
Tenoxicam	20

As an example, suppose a 12 week period from baseline through week 12 (84 days) is of interest. During this period, a subject took piroxicam 4 times a week at a dose of 20 mg for the first 8 weeks and then switched to ibuprofen 2 times a week at a dose of 1200 mg for the last 4 weeks. Based on the table, the equivalent score for 20 mg of piroxicam is
[100 x 56 x (4/7)]/84 = 38.1,
and the score for ibuprofen is
[50 x 28 x (2/7)]/84 = 4.8.

The NSAID intake score for this subject for this 12 week period is 38.1 + 4.8 = 42.9.

Notes:

Follow up based on NSAIDS identified in the study but not included in the table above:

Lornoxicam – assigned the same dose equivalence as Piroxicam in the table

Alternative names for known medications were mapped as follows (unknown/known):

Olfen Uno/diclofenac
Orudis Retard/ketoprofen
Orudis depot/ketoprofen injection
Orudis Retard/ketoprofen
Salsalate/disalcid
Surgam/tiaprofenic acid
Trosicam/meloxicam
Xefo Rapid/lornoxicam
Xefo rapid/loroxicam
diclofenac/dicolfenac
diclofenacum natricum/diclofenac
naproxen and esomeprazole/naproxyn

Appendix 1.2. Criteria for Potentially Clinically Important (PCI) Laboratory Values

Laboratory test results that were Grade 3 or Grade 4 according to the National Cancer Institute Common Terminology Criteria for Adverse Events) will be considered Potentially Clinically Important (PCI).

Laboratory Test	Grade	Criteria
Neutrophils	1	1.5 <= value < 2.0
	2	1.0 <= value < 1.5
	3	0.5 <= value < 1.0
	4	value < 0.5
Hemoglobin	1	100 <= value < Lower range
	2	80 <= value < 100
	3	65 <= value < 80
	4	value < 65
Lymphocytes	0	value >= 1.5
	1	1.5 > value >= 1.0
	2	1.0 > value >= 0.5
	3	0.5 > value
Platelets	0	value >= Lower range
	1	Lower range > value >= 75.0
	2	75.0 > value >= 50.0
	3	50.0 > value >= 25.0
	4	25.0 > value
White Blood Count	0	value >= 4.0
	1	4.0 > value >= 3.0
	2	3.0 > value >= 2.0
	3	2.0 > value >= 1.0
	4	1.0 > value
Alkaline Phosphatase	0	value <= 1.0*Upper range
	1	1.0*Upper range < value <= 2.5*Upper range
	2	2.5*Upper range < value <= 5.0*Upper range
	3	5.0*Upper range < value <= 20.0*Upper range
	4	20.0*Upper range < value
SGOT/AST	0	value <= 1.0*Upper range
	1	1.0*Upper range < value <= 3.0*Upper range
	2	3.0*Upper range < value <= 5.0*Upper range
	3	5.0*Upper range < value <= 20.0*Upper range
	4	20.0*Upper range < value
SGPT/ALT	0	value <= 1.0*Upper range
	1	1.0*Upper range < value <= 3.0*Upper range
	2	3.0*Upper range < value <= 5.0*Upper range
	3	5.0*Upper range < value <= 20.0*Upper range

Laboratory Test	Grade	Criteria
	4	20.0*Upper range < value
Total bilirubin	1	value <= 1.0*Upper range
	2	1.0*Upper range < value <= 1.5*Upper range
	3	1.5*Upper range < value <= 3.0*Upper range
	4	3.0*Upper range < value
Creatinine	0	value <= 1.0*Upper range
	1	1.0*Upper range < value <= 1.5*Upper range
	2	1.5*Upper range < value <= 3.0*Upper range
	3	3.0*Upper range < value <= 6.0*Upper range
	4	6.0*Upper range < value
Albumin	0	value > 35.0 (=LLN)
	1	35.0 >= value >= 30
	2	30.0 > value >= 20.0
	3	**20.0 > value**
	4	**(no values given)**
Urea	0	value <= 1.5*Upper range
	1	1.5*Upper range < value <= 2.0*Upper range
	2	2.0*Upper range < value <= 3.0*Upper range
	3	3.0*Upper range < value <= 4.0*Upper range
	4	4.0*Upper range < value
Potassium - high and low	0	3.4 (=LLN) < value < 5.0 (=ULN)
	1	5.0 <= value <= 5.5 or <3.4 >= value >= 3.0
	2	5.5 < value <= 6.0 or <3.4 >= value >= 3.0
	3	6.0 < value <= 7.0 or 3.0>value >= 2.5
	4	value > 7.0 or value < 2.5
Sodium - high & low	0	134 (=LLN)< value < 146 (=ULN)
	1	146 <= value <= 150 or 140 >= value >= 130
	2	150 < value <= 155 or (**no values given**)
	3	155 < value <= 160 or 130 > value >= 120
	4	value > 160 or value < 120

Appendix 1.3. Criteria for Potentially Clinically Important (pci) Vital Signs Values

Test	**Definition**
Sitting Pulse Rate	Increase from baseline >= 15 beats/min and >= 120 beats/min
	Decrease from Baseline >= 15 beats/min and <= 50 beats/min
Sitting Systolic Blood Pressure	Increase from baseline >= 20 mm Hg and >= 160 mm Hg
	Decrease from baseline >= 20 mm Hg and <= 90 mm Hg
Sitting Diastolic Blood Pressure	Increase from baseline >= 15 mm Hg and >= 90 mm Hg
	Decrease from baseline >= 15 mm Hg and <= 60 mm Hg
Weight	Increase from baseline >= 15% and >= 25 lbs (11.36 Kg)
	Decrease from baseline >= 15% and <= 25 lbs (11.36 Kg)

Appendix 2. STATISTICAL METHODOLOGY DETAILS

Appendix 2.1. Flare using ASDAS CRP

Subjects will be entered into the retreatment period based on occurrence of flare, defined as an ASDAS ESR value ≥2.1 during the withdrawal period. A secondary set of efficacy endpoint analyses for the retreatment period will include only those subjects who also had an ASDAS CRP value ≥2.1.

Appendix 2.2. Sustained Remission

For the withdrawal and retreatment periods, sensitivity analyses for the relapse, ASDAS inactive disease and ASAS40 endpoints by period 1 sustained inactive disease status (sustained or not, as defined under "Covariates") will be provided. Tables presenting these endpoints by period 1 inactive disease status (Yes/No) will be provided.

Appendix 2.3. Analysis Intervals For EfficacyAnd Other Endpoints

Induction (Baseline – Week 24)/ Withdrawal (Week 28-64) / Retreatment (R0-R12) treatment period intervals

	Target week	Target day	BASDA,BASFI, SADA, VAS PGA, WPAI	Joint Assessment, BAS-, EQ-5D, SF-36	Hematology/ Blood Chemistry	Urine	Height/Weight
Baseline		1	<=1	<=1		<=1	<=1
Week 4	4	28	2-42				
Week 8	8	56	43-70				
Week 12	12	84	71-98	2-98	2-98		
Week 16	16	112	99-140				
Week 24	24	168	141-182 (or EOS1) *	99-182 (or EOS1) *	99-182 (or EOS1) *		2-182 (or EOS1) *
Week 28	28	196	183-210				
Week 32	32	224	211-252	183-252			
Week 40	40	280	253-308				
Week 48	48	336	309-364	253-364			
Week 56	56	392	365-420				
Week 64	64	448	421-EOS2	365-EOS2	183-EOS2	2-EOS 2	183-EOS2
Week R0	1***	1	<=1	<=1	<=1		
Week R4	4	28	2-42				
Week R8	8	56	43-70				
Week R12	12	84	71-EOS3	>=2	>=2		>=1

*For subjects not entering the withdrawal period, no data should be assigned to periods 2 or 3.

*** Day 1 is relative to the first day in Retreatment Period.