NOVARTIS

Clinical Development

Secukinumab (AIN457)

CAIN457F2308 / NCT02896127

A randomized, double-blind, placebo-controlled, phase III multicenter study of subcutaneous secukinumab in prefilled syringes, to compare efficacy at 16 weeks with placebo and to assess safety and tolerability up to 52 weeks in subjects with active Ankylosing Spondylitis

Statistical Analysis Plan (SAP) Amendment 3

Author: [redacted], Trial Statistician; [redacted], Trial Statistician

Document type: SAP Documentation

Document status: Final

Release date: August 16, 2018

Number of pages: 49

Document History – Changes compared to previous version of SAP.

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
19-Mar-2018	Prior to DBL	[redacted]	[redacted]	[redacted]
		To provide separate summaries for SAEs and significant AEs by site in China	Amendment 1	13.1
		To align with the changes in the protocol to include a primary analysis after all patients have completed the Week 16 visit.	Amendment 1	14.0
22-Jul-2018	Prior to DBL	To clarify significant AE	Amendment 2	13.1
22-Jul-2018	Prior to DBL	To follow convention in seed specification in MI	Amendment 2	16.2.3.1
16-Aug-2018	Prior to DBL	Added sensitivity analyses for hierarchy variables using patient's global assessment for disease activity (VAS)	Amendment 3	11.5

Table of contents

List of abbreviations

AE	Adverse Event
ALT	Alanine Aminotransferase/ Glutamic Pyruvic Transaminase/ GPT
AS	Ankylosing Spondylitis
ASAS	Assessment of Spondyloarthritis International Society criteria
[redacted]	[redacted]
AST	Aspartate Aminotransferase/ Glutamic Oxaloacetic Transaminase/ GOT
ASQoL	Ankylosing Spondylitis Quality of Life Questionnaire
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
[redacted]	[redacted]
BSL	Baseline
CRP (hsCRP)	(high sensitivity) C-Reactive Protein
CRP+	hsCRP > ULN (as defined by the central lab)
DMARD	Disease Modifying Antirheumatic Drug
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	electronic Case Report/Record Form
[redacted]	[redacted]
[redacted]	[redacted]
FAS	Full Analysis Set
GGT	Gamma Glutamyl Transferase
HLA	Human Leukocyte Antigen
IRT	Interactive Response Technology
IVRS / IWRS	Interactive Voice Response System / Interactive Web Response System
MAR	Missing At Random
[redacted]	[redacted]
MCS	Mental Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model Repeated Measures
MTX	Methotrexate

NRI	Non-Responder Imputation
NSAID	Non-Steroidal Anti-Inflammatory Drug
PCS	Physical Component Summary
PD	Pharmacodynamic
PG	Pharmacogenetic
PK	Pharmacokinetic
PRN	According to need, as required
PRO	Patient Reported Outcome
QoL	Quality of Life
RA	Rheumatoid Arthritis
SAE	Serious Adverse Event
s.c.	Subcutaneous(ly)
SCR	Screening
SF-36	Medical Outcome Short Form (36) Health Survey
TNF	Tumor Necrosis Factor
TNF-IR	TNFα Inhibitor Incomplete Responders
ULN	Upper Limit of Normal
VAS	Visual Analog Scale
[redacted]	[redacted]

1 Introduction

Data will be analyzed by Novartis according to the data analysis section 9 of the study protocol. The statistical methodology is described below and any deviations from the protocol are documented. Additional detailed information regarding the analysis methodology is contained in the Appendix section.

2 Study objectives

The primary objective is to demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo in subjects with active AS based on the proportion of subjects achieving an ASAS20 (Assessment of SpondyloArthritis International Society criteria) response.

The following are the secondary objectives:

1. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the proportion of subjects achieving an ASAS40 response
2. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the change from baseline of high sensitivity C-Reactive Protein (hsCRP)
3. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the proportion of subjects meeting the ASAS 5/6 response criteria
4. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the change from baseline in total Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)
5. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the change from baseline in Short Form 36 physical component score (SF-36 PCS)
6. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the change from baseline in Ankylosing Spondylitis Quality of Life score (ASQoL)
7. To demonstrate that the efficacy of secukinumab 150 mg s.c. at Week 16 is superior to placebo based on the proportion of subjects achieving an ASAS partial remission
8. Overall safety and tolerability of secukinumab compared to placebo as assessed by vital signs, clinical laboratory values and adverse event monitoring

3 Data presentation

To support China AS submission, tables and figures will be generated for Chinese population and overall separately. Listings will be generated for overall and will be sorted by region (China and non-China), except for the listing of randomization allocation to treatment which will be sorted by the randomization number.

Summary statistics for continuous variables will include N, mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum. Summary statistics for discrete variables will be presented in the number and percent of subjects in each category.

If not otherwise specified, p-values will be presented as two-sided p-values and two-sided confidence intervals will be displayed.

The default level of significance will be set to 5% (two-sided, family-wise type-I-error).

Data analyses will be presented by treatment regimen. Efficacy and safety data for the placebo-controlled period (or the entire treatment period as appropriate) will be presented by the following 2 treatment groups. Subjects may be included in more than one treatment group for some analyses (e.g. exposure-adjusted adverse events over the entire treatment period). The 2 treatment groups represent the regimens subjects will be eligible to be randomized to for the first 16 weeks of the study.

- Secukinumab regimen: secukinumab 150 mg (1 mL, 150 mg/mL) s.c. PFS at BSL, Weeks 1, 2, and 3, followed by administration every four weeks starting at Week 4
- Placebo regimen: placebo (1 mL) s.c. PFS at BSL, Weeks 1, 2, 3, 4, 8, and 12, followed by secukinumab 150 mg (1 mL, 150 mg/mL) administration every four weeks starting at Week 16

Note that the treatment groups above for a subject may differ depending on the time period of the analysis and whether one assesses the subject for efficacy or safety.

Furthermore, additional aspects of efficacy, safety and tolerability of secukinumab will be investigated.

Comparative efficacy data

Comparative efficacy analyses (i.e. inferential efficacy comparisons with placebo) will focus on the time period when both active drug and the placebo are given in a manner suitable for making comparisons (e.g. double-blind). For AIN457F2308 this is the first 16-weeks of treatment. Comparative efficacy will be performed based on the FAS population using the randomized treatment. Data collected after Week 16 will generally be summarized descriptively on the FAS population using treatment sequence.

Efficacy data following treatment switch

Data will also be presented after Week 16, by a combination of the ‘original’ and ‘switch’ treatment groups and will be referred to as treatment sequence. These treatment sequences represent the treatment combinations the subjects experience over the course of the entire trial in case of switch (e.g., placebo patients who are reassigned to 150 mg secukinumab at Week 16).

4 All listings will be presented by treatment sequence.Subjects and treatments

4.1 Analysis sets

The following analysis sets will be used in this trial:

Randomized set: The randomized set will be defined as all subjects who were randomized. Unless otherwise specified, mis-randomized subjects (mis-randomized in IRT) will be excluded from the randomized set.

Mis-randomized subjects are defined as those subjects who were mistakenly randomized into the IVR prior to the site confirming all eligibility criteria had been met and to whom no study medication was given. Mis-randomized patients are treated as screen failures.

Full analysis set (FAS): The FAS will be comprised of all subjects from the randomized set to whom study treatment has been assigned. Following the intent-to-treat principle, subjects will be evaluated according to the treatment assigned to at randomization, but actual stratum, if stratified randomization is used.

Safety set: The safety set includes all subjects who took at least one dose of study treatment during the treatment period. Subjects will be evaluated according to treatment received.

4.2 Treatment groups

The summaries by treatment will primarily be performed by the randomized treatment but also by combination of randomized treatment and switch treatment. For some safety summaries (e.g. exposure-adjusted) the 'switch' treatment may be summarized separately.

- Randomized treatment:
  - AIN457 150 mg
  - Placebo
- Treatment sequence:
  - AIN457 150 mg
  - Placebo – AIN457 150 mg
- Switch treatments (for patients who cross-over):
  - AIN457 150 mg

5 Subgroup definitions

The primary and secondary endpoints will be analyzed by region (China and non-China). Analyses by previous use of TNF-α inhibitor will also be provided for total population and Chinese population.

6 Assessment windows, baseline and post baseline definitions, missing data handling

Baseline and post-baseline definitions

In general, a *baseline* value refers to the last measurement made prior to administration of the first dose of study treatment. A post-baseline value refers to a measurement taken after the first dose of study treatment.

Analysis visit windows

Analysis visit windows will be used for the data that is summarized by visit; they are based on the study evaluation schedule and comprise a set of days around the nominal visit day. For any assessment, there are protocol defined scheduled visits around which analysis visit windows were created to cover the complete range of days within the study. The analysis visit windows and rules for dealing with multiple measurements within the windows are described in Appendix 16.1.

7 Subject disposition, background and demographic characteristics

7.1 Subject disposition

The number of subjects screened will be presented. In addition, the reasons for screen failures will be provided. The number and percentage of subjects in the randomized set who completed the study periods and who discontinued the study prematurely (including the reason for discontinuation) will be presented at the end of each treatment period (Week 16, Week 52, Week 60), if appropriate, for each treatment group and all subjects.

For each protocol deviation (PD), the number and percentage of subjects for whom the PD applies will be tabulated.

7.2 Background and demographic characteristics

The following common background and demographic variables will be summarized:

Continuous variables:

- Age (which is derived from date of birth and the date of informed consent)
- Height
- Weight
- Body mass index (BMI) = (body weight in kilograms) / (height in meters)^2

For BMI, height and body weight, the last value prior to randomization is used. If there is no weight recorded prior to taking of study drug, BMI will be missing.

Categorical variables:

- Age categories (<65 years, 65 years and older, 75 years and older)
- Gender
- Race
- Ethnicity
- Smoking status at baseline

Baseline disease characteristics will also be summarized for the following variables:

- Patient's global assessment of disease activity and other ASAS components, hsCRP (mg/L), [redacted] prior use (yes/no) of TNF-alpha inhibitors, use (yes/no) and separate dose of methotrexate (mg/week), sulfasalazine (g/day) and systemic corticosteroids (mg/day) at randomization, time since first diagnosis of AS (years), modified New York criteria for AS,

HLA-B27, [redacted] total back pain (VAS), nocturnal back pain (VAS), total BASDAI score, spinal pain (BASDAI question #2) and BASMI components (all seven in original units).

Unless otherwise specified, summary statistics will be presented for continuous variables for each treatment group and for all subjects (total) in the randomized set. The number and percentage of subjects in each category will be presented for categorical variables for each treatment group and all subjects (total) in the randomized set.

Unless otherwise specified, analyses will be based on the randomized set.

8 Medical history

Any condition entered on the *Relevant medical history / current medical conditions* CRF will be coded using the MedDRA dictionary. They will be summarized by system organ class (SOC) and preferred term (PT) of the MedDRA dictionary. Summaries for cardiovascular medical history and ankylosing spondylitis medical history will also be provided.

Smoking history will be summarized by treatment group.

Unless otherwise specified, analyses will be based on the randomized set.

9 Study treatment

The analysis of study treatment data will be based on the safety set. The number of active and placebo injections will be summarized by treatment group. The duration of exposure to study treatment will also be summarized by treatment group. In addition, the number and percentage of subjects with cumulative exposure levels (e.g. any exposure, ≥ 1 week, ≥ 2 weeks, ≥ 3 weeks, ≥ 4 weeks, ≥ 8 weeks, etc.) will be presented.

Duration of exposure will be defined as the time from first dose of study treatment to the time of treatment switch (for subjects who switch treatment) or minimum of (last dose of the treatment + 84 days) and (last visit date). For subjects who discontinue, this will be the subject's last visit in the corresponding treatment period. Patients who switch treatment during the study (e.g. from placebo to active treatment) will have exposure to both medications using the appropriate start and stop dates.

Duration of exposure (years) = duration of exposure (days) / 365.25

Duration of exposure (100 subject years) = duration of exposure (years) / 100

The analyses of duration of exposure described above will be done for the entire study treatment period.

10 Concomitant medication

Prior and concomitant medications will be summarized in separate tables by treatment group.

Prior medications are defined as treatments taken and stopped prior to first dose of study treatment. Any medication given at least once between the day of first dose of study treatment

and within 84 days after last dose will be a concomitant medication, including those which were started pre-baseline and continued into the period where study treatment is administered.

Medications will be presented in alphabetical order, by Anatomical Therapeutic Classification (ATC) codes and grouped by anatomical main group. Tables will show the overall number and percentage of subjects receiving at least one treatment of a particular ATC code and at least one treatment in a particular anatomical main group.

Significant prior and concomitant surgeries and procedures will be summarized by primary system organ class and MedDRA preferred term.

Prior surgeries and procedures are defined as surgeries and procedures done prior to first dose of study treatment. Any surgeries and procedures done between the day of first dose of study treatment and within 84 days after last dose will be a concomitant surgeries and procedures, including those which were started pre-baseline and continued into the period where study treatment is administered.

The number and percentage of subjects receiving prior ankylosing spondylitis therapy will be presented by randomized treatment group as well as the reasons for stopping their therapies (primary lack of efficacy, secondary lack of efficacy, lack of tolerability, other) and the total duration of exposure to ankylosing spondylitis therapies previously.

Prior or concomitant medication will be identified by comparing recorded or imputed start and end dates of medication taken to the reference start date.

Presentation by dose

Methotrexate, sulfasalazine and systemic corticosteroid intake at randomization will be presented based on the amount taken per time unit, e.g. mg/day.

Since different steroids have different strengths the following multiplication factors will be used to convert a dose in mg into a prednisone equivalent dose in mg:

Cortisone (0.20), hydrocortisone (0.25), prednisolone (1.0), triamcinolone (1.25), methylprednisolone (1.25), dexamethasone (6.67), betamethasone (8.33) (corticosteroid converter website).

The reported dose and frequency of intake will be converted into the desired units. If the frequency is missing or specified as ‘per needed’, ‘unknown’, ‘once’, ‘other’ or if the dose or dose unit is missing, then the medication will not be part of the presentation.

11 Efficacy evaluation

11.1 Description of efficacy variables

Assessment of Spondyloarthritis International Society criteria (ASAS) response criteria

The ASAS response measures consist of the following assessment domains (Sieper 2009)

Main ASAS domains:

1. Patient's global assessment of disease activity measured on a VAS scale
2. Patient's assessment of back pain, represented by either total or nocturnal pain scores, both measured on a VAS scale. For ASAS response analyses, the total back pain will be used.
3. Function represented by BASFI average of 10 questions regarding ability to perform specific tasks as measured by VAS scale
4. Inflammation represented by mean duration and severity of morning stiffness, represented by the average of the last 2 questions on the 6-question BASDAI as measured by VAS scale (at least one question is needed)

Additional assessment domains:

5. Spinal lateral flexion, represented by one of the components of BASMI
6. C-reactive protein (acute phase reactant)

ASAS components

Patient's global assessment of disease activity (VAS)

The patient's global assessment of disease activity will be performed using a 100 mm visual analog scale (VAS) ranging from not severe to very severe, after the question *"How active was your disease on average during the last week?*".

Patient's assessment of total back pain and nocturnal back pain intensity (VAS)

The patient's assessment of back pain will be performed using a 100 mm VAS ranging from no pain to unbearable pain, after the question "*Based on your assessment, please indicate what is the amount of back pain at any time that you experienced during the last week?*" and *"Based on your assessment, please indicate what is the amount of back pain at night that you experienced during the last week?"*.

Bath Ankylosing Spondylitis Functional Index (BASFI)

The BASFI is a set of 10 questions designed to determine the degree of functional limitation in those subjects with AS. The ten questions were chosen with a major input from subjects with AS. The first 8 questions consider activities related to functional anatomy. The final 2 questions assess the subjects' ability to cope with everyday life. A 100 mm VAS is used to answer the questions. The mean of the ten questions gives the BASFI score – a value between 0 and 10. In the case that some of the BASFI questions are missing, then the average of the non-missing items is used (Braun 2009, van Tubergen 2001).

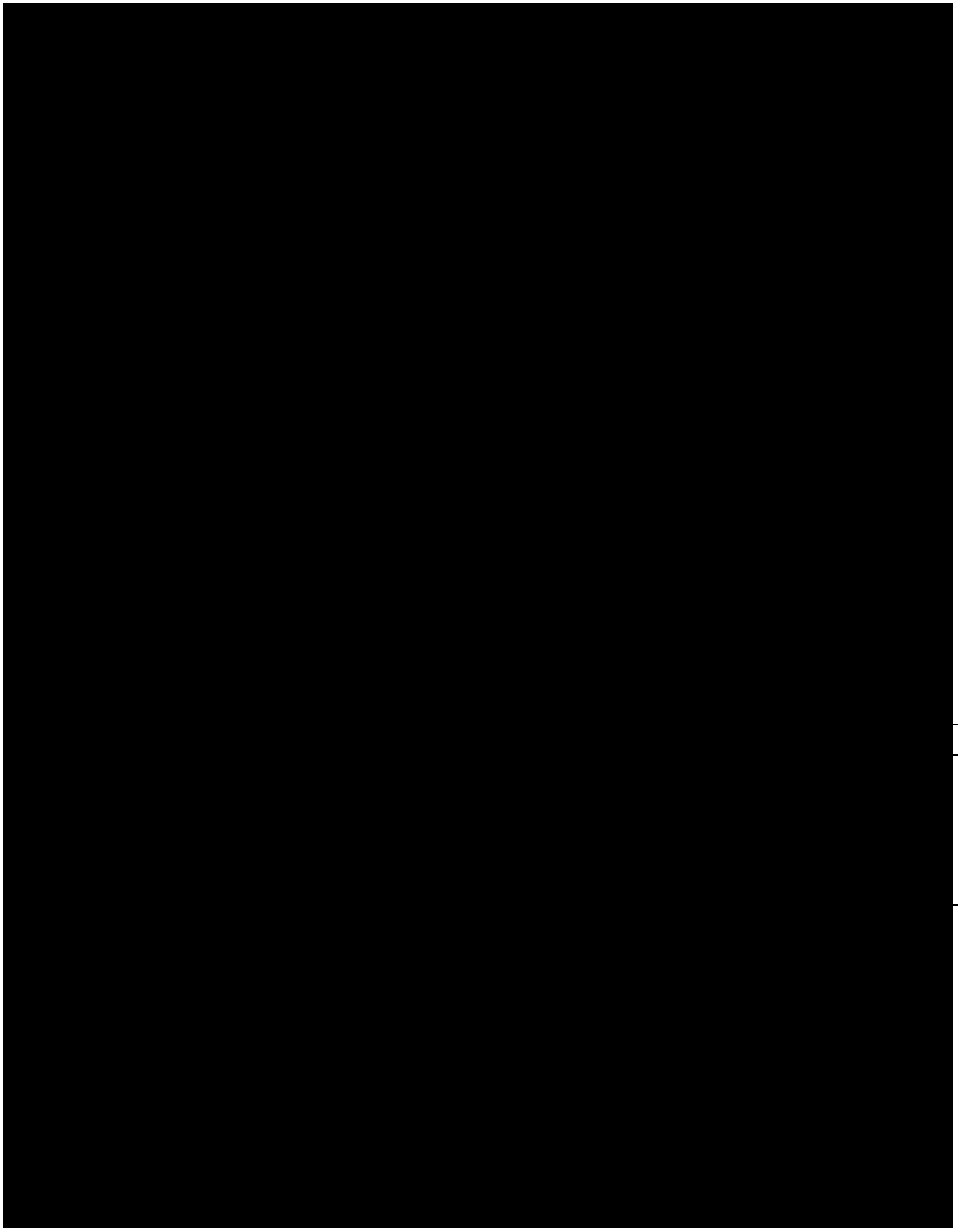
Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

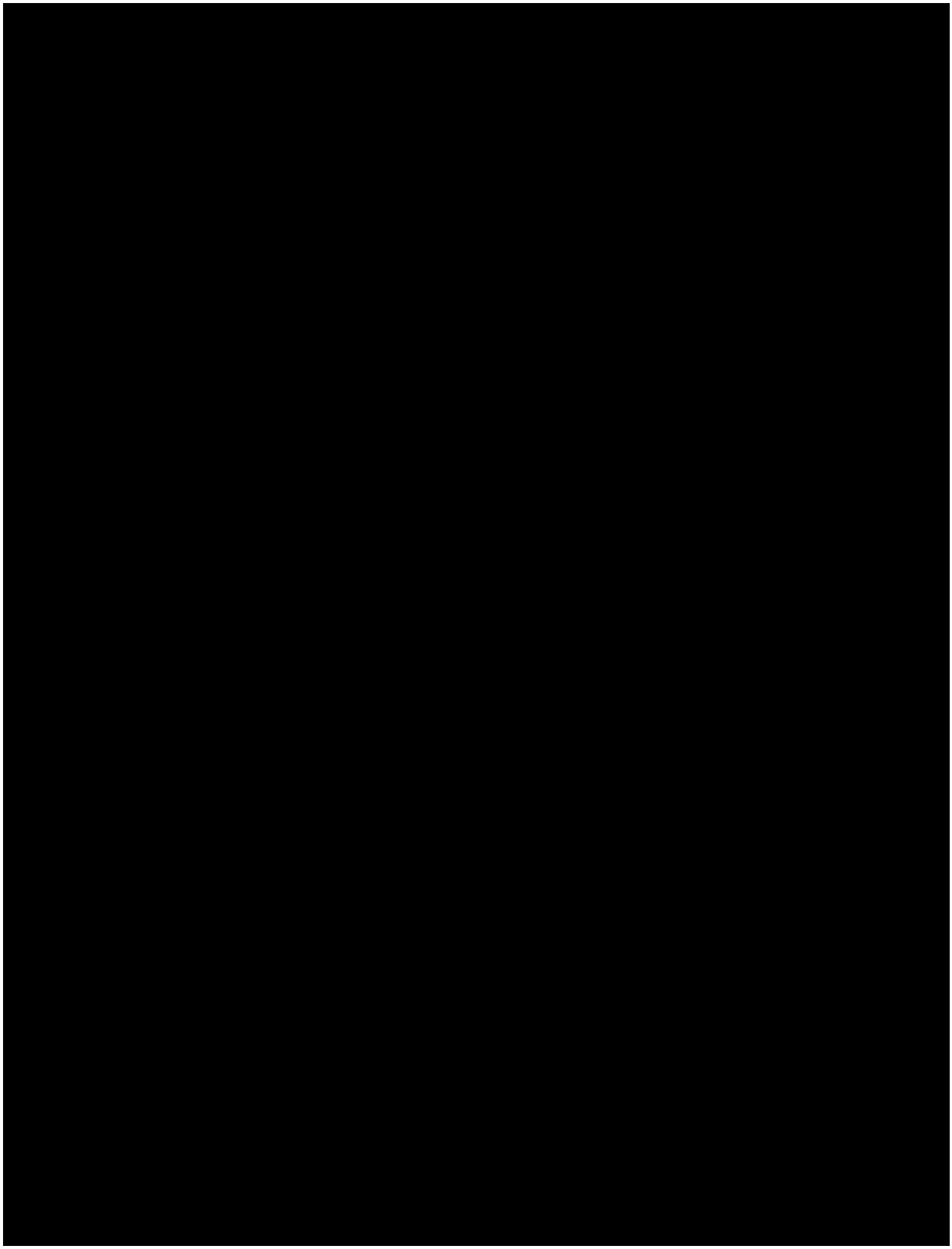
The gold standard for measuring and evaluating disease activity in AS is the BASDAI. The BASDAI consists of a 0 through 10 scale (0 being no problem and 10 being the worst problem), which is used to answer 6 questions pertaining to the 5 major symptoms of AS:

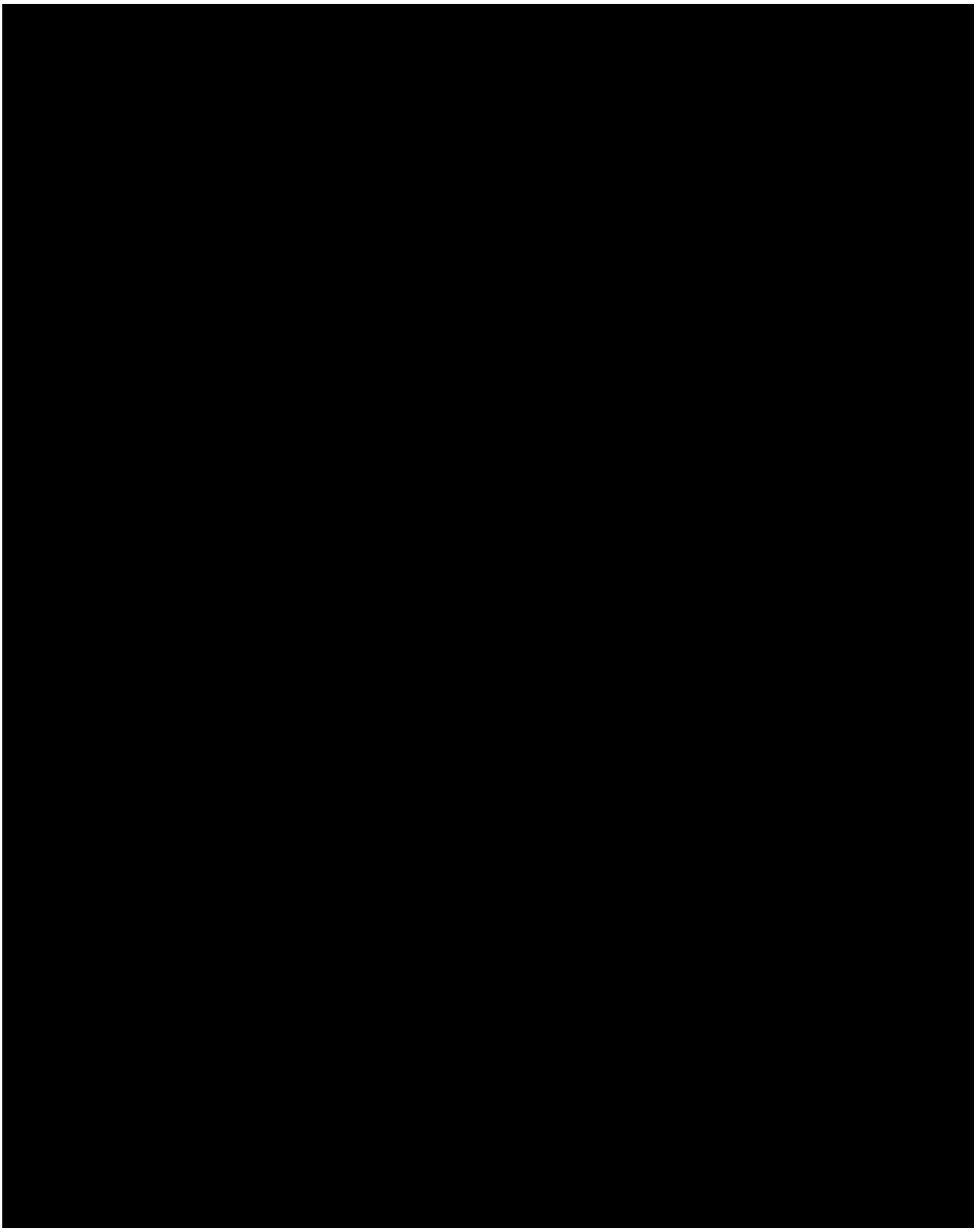
1. Fatigue
2. Spinal pain
3. Joint pain / swelling
4. Areas of localized tenderness
5. Morning stiffness severity
6. Morning stiffness duration

To give each symptom equal weighting, the mean (average) of the two scores relating to morning stiffness is taken and is then added to the sum of the first 4 questions. The resulting 0 to 50 score is divided by 5 to give a final 0 – 10 BASDAI score. Scores of 4 or greater suggest suboptimal control of disease, and subjects with scores of 4 or greater are usually good candidates for either a change in their medical therapy or for enrollment in clinical trials evaluating new drug therapies directed at AS. BASDAI is a quick and simple index taking between 30 secs and 2 mins to complete. At least 4 questions should be non-missing to calculate the BASDAI score. Otherwise, BASDAI score will be missing (Haywood 2002). If both Q5 and Q6 are missing or one of Q1 to Q4 is missing, then the total sum should be divided by 4

instead of 5. If two of Q1 to Q4 is missing and both Q5 and Q6 are not missing, then the sum should be divided by 3.







11.2 Description of health-related quality of life variables

Ankylosing Spondylitis Quality of Life (ASQoL)

The ASQoL is a self-administered questionnaire designed to assess health-related quality of life in adult patients with Ankylosing Spondylitis. The ASQoL contains 18 items with a dichotomous yes/no response option. A single point is assigned for each "yes" response and no points for each "no" response resulting in overall scores that range from 0 (least severity) to 18 (highest severity). As such, lower score indicate better quality of life. Items include an assessment of mobility/energy, self-care and mood/emotion. The recall period is "at the moment," and the measure requires approximately 6 minutes to complete.

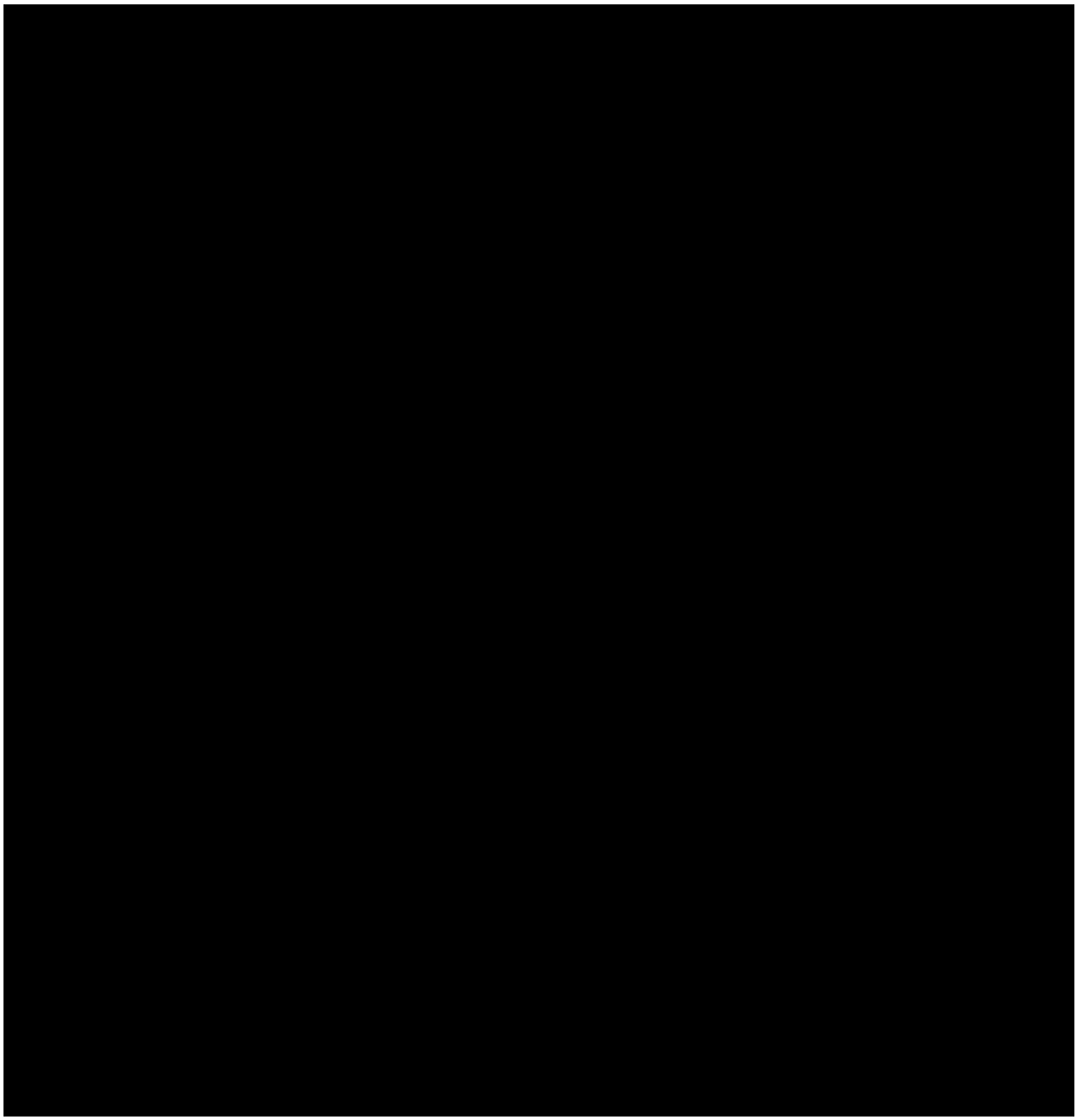
At least 15 answered questions are required to be able to calculate ASQoL using mean imputation, (sum of answered)/(number answered)*18 (Doward 2003)

SF-36

The Short Form Health Survey (SF-36) is a widely used and extensively studied instrument to measure health-related quality of life among healthy subjects and patients with acute and

chronic conditions. It consists of eight subscales (domains) that can be scored individually: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role- Emotional, and Mental Health. Two overall summary scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS) also can be computed. The SF-36 has proven useful in monitoring general and specific populations, comparing the relative burden of different disease, differentiating the health benefits produced by different treatments, and in screening individual patients. The eight domains are based on a scale from 0-100 while PCS and MCS are norm-based scores with a mean of 50 and a standard deviation of 10.

Quality metric uses weighted maximum likelihood estimation, a modified version of item response theory (IRT) to estimate scale scores when a respondent is missing multiple items. The PCS summary score measure requires scores for seven scales, one of which must be the PF scale and the MCS score also requires scores for seven scales, one of which must be the MH scale. Only one item is needed for each of the multi-item domains.



11.3 Handling of missing data

Missing data for ASAS20/40 response and other binary efficacy variables (e.g., ASAS5/6, etc.) for data up to Week 16 will be handled as follows:

1. Patients who drop out of the trial for any reason will be considered as non-responders from the time they drop out through Week 16.
2. Patients who do not have the required data to compute responses (e.g., ASAS components) at baseline and at the specific timepoint will be classified as non-responders at the specific timepoint.

Patients who are unblinded prior to the scheduled timepoint will be considered non-responders from the time of unblinding up to Week 16. The primary analysis will use non-responder imputation.

Continuous variables (e.g., ASAS components) will be analyzed using a mixed-effects model repeated measures (MMRM) which is valid under the missing at random (MAR) assumption. As such, single-point imputation of missing data will not be performed (e.g., LOCF). For analyses of these parameters, if all post-baseline values are missing then these missing values will not be imputed and this patient will be removed from the analysis of the corresponding variable, i.e., it might be that the number of patients providing data to an analysis is smaller than the number of patients in the FAS. Model will be based on randomized treatment (Week 1 to Week 16).

Data collected after Week 16 will generally be presented as 'observed case'; i.e. all available data for each time point will be included in the analyses.

11.4 Analysis of primary variable

The primary efficacy variable is response to treatment according to the ASAS20 criteria at Week 16. The analysis of the primary variable will be based on the FAS patients.

Description

The ASAS Response Criteria (ASAS20) is defined as an improvement of ≥20% and ≥1 unit on a scale of 10 in at least three of the four main domains and no worsening of ≥20% and ≥1 unit on a scale of 10 in the remaining domain.

Statistical analyses

The statistical hypothesis for ASAS20 being tested is that there is no difference in the proportion of subjects fulfilling the ASAS20 criteria at Week 16 in secukinumab regimen versus placebo regimen.

Let p_0 denote the proportion of ASAS20 responders at Week 16 for Placebo regimen and p_1 denote the proportion of ASAS20 responders at Week 16 for Secukinumab 150 mg regimen,

In statistical terms, H_1: $p_1 = p_0$, H_{A1}: $p_1 \neq p_0$, i.e.

H_1: Secukinumab 150 mg regimen is not different to placebo regimen with respect to ASAS20 response at Week 16

The primary analysis will be conducted via logistic regression with treatment, randomization stratum (region) as factors and baseline weight as a covariate. Odds ratios and 95% CI will be presented comparing secukinumab 150 mg regimen to placebo.

Supportive analyses

Sensitivity analyses and supportive analyses will be conducted in order to provide evidence that the results seen from the primary analysis are robust.

In order to determine the robustness of the logistic regression model used for the primary analysis, ASAS20 response at Week 16 will also be evaluated using a non-parametric

ANCOVA model (Koch 1998) with the same independent variables as the logistic regression model.

Interactions between treatment and selected baseline demographics and disease characteristics will be explored.

The impact of missing data on the analysis results of ASAS20 response will be assessed as well by repeating the logistic regression model using different ways to handle missing data. These may include, but are not limited to:

- Multiple imputation
- Observed data analysis

11.5 Analysis of secondary variables

The secondary efficacy variables include,

- response to treatment at Week 16 according to the ASAS40 criteria
- change from baseline in hsCRP at Week 16
- response to treatment at Week 16 according to the ASAS5/6 criteria
- change from baseline in total BASDAI score at Week 16
- change in SF-36 PCS from baseline at Week 16
- change in ASQoL from baseline at Week 16
- response to treatment at Week 16 according to the ASAS partial remission

All analysis will be done in the FAS population.

Testing strategy to control type I error

The following null hypotheses will be included in the testing strategy, and type-I-error will be set such that a family-wise type-I-error of 5% is kept:

Primary objectives:

H_1: secukinumab 150 mg regimen is not different to placebo regimen with respect to signs and symptoms (ASAS20 response) at Week 16

Secondary objectives:

H_2: secukinumab 150 mg regimen is not different to placebo regimen with respect to signs and symptoms (ASAS40 response) at Week 16

H_3: secukinumab 150 mg regimen is not different to placebo regimen with respect to change from baseline in hsCRP at Week 16

H_4: secukinumab 150 mg regimen is not different to placebo regimen with respect to ASAS 5/6 response at Week 16

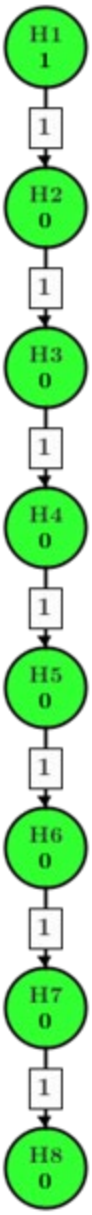
H_5: secukinumab 150 mg regimen is not different to placebo regimen with respect to change from baseline in total BASDAI at Week 16

H_6: secukinumab 150 mg regimen is not different to placebo regimen with respect to change from baseline in SF-36 PCS at Week 16

H_7: secukinumab 150 mg regimen is not different to placebo regimen with respect to change from baseline in ASQoL at Week 16

H_8: secukinumab 150 mg regimen is not different to placebo regimen with respect to ASAS partial remission criteria at Week 16

Figure 11-1 Testing strategy



The family-wise error will be set to α=5% and it will be controlled with the proposed closed testing procedure.

The secondary efficacy variables and the method for adjusting for multiplicity are described below. Secondary efficacy variables will be analyzed using the FAS population.

ASAS40 at Week 16

The ASAS 40 response is defined as an improvement of ≥40% and ≥2 units on a scale of 10 in at least three of the four main domains and no worsening at all in the remaining domain.

Response at Week 16 to ASAS40 in the FAS will be evaluated using a logistic regression model with treatment, randomization stratum (region) as factors and baseline weight as a covariate. Odds ratios and 95% CI will be presented for comparison performed between secukinumab and placebo at Week 16.

hsCRP at Week 16

For the hsCRP, since evidence from the literature would suggest that the data is not normally distributed (Huffman 2006), analysis will be performed on the $\log_e$ ratio of the treatment value vs. baseline value (calculated by dividing the post-baseline value by the baseline value and then applying the $\log_e$ transformation) to normalize the distribution of hsCRP at each analysis visit. Between-treatment differences in the hsCRP relative to baseline will be evaluated using a mixed-effect model repeated measures (MMRM) with treatment group, analysis visit, and randomization stratum (region) as factors and $\log_e$ baseline hsCRP and baseline weight as continuous covariates. Treatment by analysis visit and $\log_e$ baseline hsCRP by analysis visit will be included as interaction terms in the model. An unstructured covariance structure will be assumed for the model. The significance of the secukinumab treatment effect will be determined from comparison performed between secukinumab regimen and placebo at Week 16. The estimate and the 2-sided 95% confidence intervals obtained from the model will be back transformed to the original scale.

ASAS 5/6 at Week 16

The ASAS 5/6 improvement criteria is an improvement of ≥20% in at least five domains.

The proportion of subjects meeting the response criteria at Week 16 will be evaluated using a logistic regression model with treatment group, randomization stratum (region) as factors and baseline weight as a covariate. Odds ratios and 95% CI will be presented for comparison performed between secukinumab regimen and placebo at Week 16.

BASDAI at Week 16

Between-treatment differences in the change from baseline in BASDAI will be evaluated using a mixed-effect model repeated measures (MMRM) with treatment group, analysis visit, and randomization stratum (region) as factors, baseline BASDAI score and baseline weight as continuous covariates. Treatment by analysis visit and baseline BASDAI score by analysis visit will be included as interaction terms in the model. An unstructured covariance structure will be assumed for the model. The significance of the secukinumab treatment effect at Week 16 will be determined from comparison performed between secukinumab regimen and placebo at Week 16.

SF-36 PCS and ASQoL at Week 16

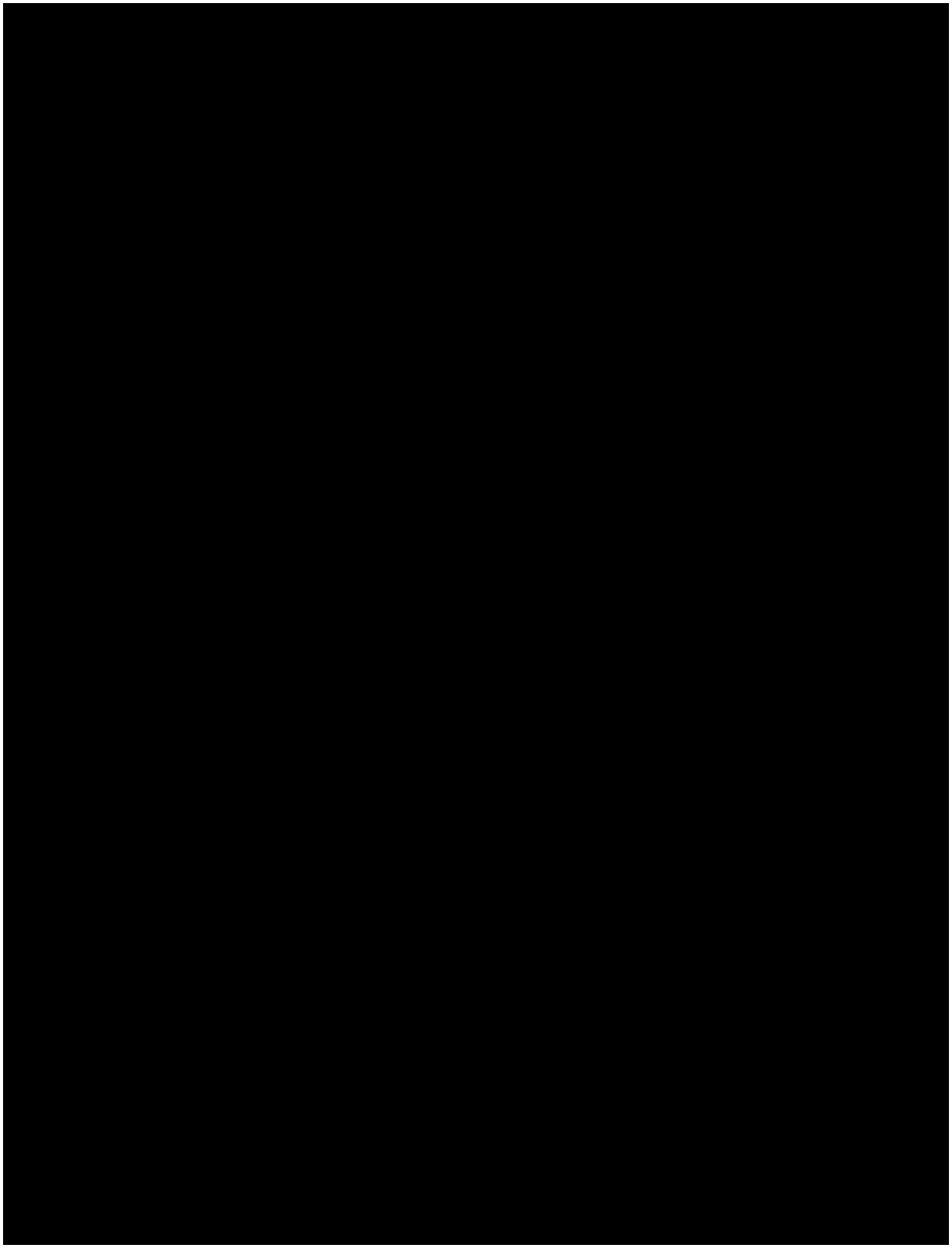
See Section 11.7.

ASAS partial remission at Week 16

Response at Week 16 to ASAS partial remission criteria will be evaluated using a logistic regression model with treatment, randomization stratum (region) as factors and baseline weight as a covariate. Odds ratios and 95% CI will be presented for comparison performed between secukinumab regimen and placebo at Week 16.

Sensitivity analysis for hierarchy variables using patient's global assessment of disease activity (VAS)

During the data cleaning review process prior to database lock, a patient was found to have had a baseline value for "patient's global assessment of disease activity (VAS)" completed out of sequence during a post-baseline visit, as documented in the vendor's audit trail. Since this variable is a component of ASAS20, ASAS40 and ASAS 5/6 that require the calculation of change from baseline, a sensitivity analysis using the same logistic regression analysis for these hierarchy variables excluding this baseline value will be done.



11.7 Analysis of health-related quality of life variables

Ankylosing Spondylitis Quality of Life (ASQoL)

For the change in ASQoL scores, between-treatment differences in the change in ASQoL scores will be evaluated using a mixed effect repeated measures model (MMRM). Treatment group, analysis visit, randomization stratum (region) will be used as categorical factors and baseline ASQoL score and baseline weight as continuous covariates. Treatment by analysis visit and baseline ASQoL score by analysis visit will be included as interaction terms in the model. An unstructured covariance structure will be assumed for the model. The significance of the secukinumab treatment effect at different analysis visits will be determined from comparison performed between secukinumab regimen and placebo at the appropriate analysis visits.

The proportion of subjects with improvements from baseline in ASQoL score meeting or exceeding minimal clinically important difference (MCID) will also be analyzed using a logistic regression model with treatment and randomization stratum (region) as factors, baseline ASQoL score and baseline weight as covariates up to Week 16. Odds ratios and 95% CI will be presented for treatment comparison. Post Week 16 the proportion of responders will be descriptively summarized along with its 95% confidence intervals for each treatment sequence.

SF-36

The following variables will be evaluated:

- SF-36 domain scores (based on a scale of 0-100).
- SF-36 PCS and MCS scores (norm-based scores).
- SF-36 PCS responder (improvement of $\geq$ 2.5 points, Lubeck 2004)

For the change in SF-36 summary scores (PCS and MCS), between-treatment differences will be evaluated using a mixed effect repeated measures model (MMRM). Treatment group, analysis visit and randomization stratum (region) will be included as categorical factors and baseline SF-36 score (PCS or MCS) and baseline weight as continuous covariates. Treatment by analysis visit and baseline SF-36 score (PCS or MCS) by analysis visit will be included as interaction terms in the model. An unstructured covariance structure will be assumed for the model. The significance of the secukinumab treatment effect at different analysis visits will be determined from comparison performed between secukinumab regimen and placebo at the appropriate analysis visits.

In the responder analyses, treatment groups will be compared with respect to response to treatment using a logistic regression model with treatment and randomization stratum (region)

as factors, baseline SF-36 PCS score and baseline weight as covariates up to Week 16. Odds ratios and 95% CI will be presented for treatment comparison. Post Week 16, the proportion of responders will be descriptively summarized along with its 95% confidence intervals for each treatment sequence.

The SF-36 domain scores will be summarized.

[redacted]

12 [redacted] evaluations (change / add PD, [redacted]/PD, Biomarkers, as needed)

[redacted]

12.2 Pharmacogenetics

Not applicable.

12.3 Biomarkers

Not applicable.

12.4 PK/PD

Not applicable.

13 Safety evaluation

Summaries may performed separately for initial (Week 1-16) and entire treatment periods. The analyses of the follow-up period will be limited to summaries for treatment-emergent adverse events, serious adverse events and risks based on adverse events.

Safety analyses will be performed on treatment received or actual treatment as described below:

The actual treatment or treatment received for summaries of safety data will differ to the treatment assigned at randomization only if a subject received the wrong treatment during the entire study.

For those subjects who received not the treatment randomized, i.e. who received erroneously the wrong treatment at least once, an additional AE listing will be prepared displaying which events occurred after the treatment errors.

13.1 Adverse events

The crude incidence of treatment emergent adverse events (i.e. events started after the first dose of study treatment or events present prior to the first dose of study treatment but increased in severity based on preferred term and on or before last dose date + 84 days) will be summarized

by primary system organ class and preferred term. Confidence intervals for the crude rate will be derived as described in Section 16.2.4.1. In addition, exposure time-adjusted incidence rates including 95% confidence intervals will be provided for the entire treatment period (see Section 16.2.5.1). A graphical display of the crude incidence rates and exposure-adjusted rates will be presented for all AEs and serious AEs by system organ class.

Adverse events reported will be presented in descending frequency according to its incidence in the secukinumab group starting from the most common event. Summaries (crude incidences only) will also be presented for AEs by severity and for study treatment related AEs. If a particular AE 'severity' is missing, this variable will be listed as missing and treated as missing in summaries. If a subject reported more than one adverse event with the same preferred term, the adverse event with the greatest severity will be presented. If a subject reported more than one adverse event within the same primary system organ class, the subject will be counted only once with the greatest severity at the system organ class level, where applicable.

Separate summaries will be provided for adverse events suspected to be related to study drug, deaths, serious adverse events, and adverse events leading to discontinuation and adverse events requiring concomitant medication.
Separate summaries will be provided for SAEs and significant AEs that led to treatment discontinuation or dose adjustment or interruption by site in China.

Adverse events will also be reported separately by SMQ according to MedDRA, using a narrow search. The MedDRA version used for reporting the study will be described in a footnote.

A listing of non-treatment emergent adverse events will be provided. These adverse events occurred before the first dose of the study treatment. The crude incidence rate will be provided without treatment information.

For SAEs occurred during screening a listing will be prepared for all subjects screened including screening failures.

When adjudication is required of major cardiovascular events, a summary of those types of events as reported by the investigator and confirmed by adjudication will be provided.

An overview of the safety analyses which will be performed for treatment emergent AEs, labs, ECG and vital signs for each analysis period is described in Table 13-1.

Table 13-1 Overview of analyses on some safety endpoints

Analysis period	AEs & SAEs	AEs by severity	Study drug related AEs	AEs-SMQ	Risk	Notables for (vitals/ ECG), lab criteria
Day 1 – Week 16	• crude incidence	• crude incidence	• crude incidence	• crude incidence	• crude incidence	• crude incidence
Entire Treatment	• crude incidence • exposure time adjusted incidence	• crude incidence	• crude incidence	• exposure time adjusted incidence	• crude incidence • exposure time adjusted incidence	• crude incidence

Exposure-adjusted incidence rates will be done for the following:
- at the PSOC for AE and SAE and Level 1 for Risks and SMQ analyses

Analysis period	AEs & SAEs	AEs by severity	Study drug related AEs	AEs-SMQ	Risk	Notables for (vitals/ ECG), lab criteria

- at the PT level for common AEs, which is defined as at least 2% of the patients in the AIN457 group during the initial treatment period or events that had an incidence rate of at least 5.0 cases per 100 subject-years in the AIN457 group during the entire treatment period

If adjudication is performed, the adjudication events (myocardial infarction, stroke, and cardiovascular death) will be listed.

Algorithms for date imputations will be provided in Programming Datasets Specifications.

13.2 Laboratory data

The summary of lab data will only include treatment emergent data, which are defined as those lab assessments after the first dose of study treatment and on or before last dose + 84 days.

Reported laboratory assessments with either a less than or greater than sign ("<" or ">") will be used for analysis after removal of the sign and conversion to standard unit. These laboratory data will be displayed in listings using the standard unit with the reported sign ("<" or ">").

The summary of laboratory evaluations will be presented for three groups of laboratory tests (hematology, chemistry and urinalysis). In addition to the individual laboratory parameters the ratios "total cholesterol / HDL" and "apolipoprotein B / apolipoprotein A1" will be derived and summarized.

For urinalysis, frequency tables will be presented.

Descriptive summary statistics for the change from baseline to each study visit will be presented. These descriptive summaries will be presented by laboratory test and treatment group. Change from baseline will only be summarized for subjects with both baseline and post baseline values and will be calculated as:

change from baseline = post baseline value – baseline value

For each parameter, the maximum change (maximum decrease and maximum increase) from baseline, if appropriate for each study phase, will be analyzed analogously.

In addition, shift tables will be provided for all parameters to compare a subject's baseline laboratory evaluation relative to the visit's observed value. For the shift tables, the normal laboratory ranges will be used to evaluate whether a particular laboratory test value was normal, low, or high for each visit value relative to whether or not the baseline value was normal, low, or high. If appropriate, the shifts to the most extreme laboratory test value within a treatment phase (either initial or entire) will be presented as well (including category "high and low"). These summaries will be presented by laboratory test and treatment group.

The following laboratory parameters will be analyzed with respect to numerical Common Terminology Criteria for Adverse Events (CTCAE) grades, given in Table 13-2: hemoglobin, platelets, white blood cell count, neutrophils, lymphocytes, creatinine, total bilirubin (TBL), gamma-glutamyl transferase (GGT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), glucose, cholesterol, triglycerides (TG).

These summaries will be split into hematology and chemistry for study level reports and the pooled summary of clinical safety.

Table 13-2 CTCAE grades for laboratory parameters to be analyzed

CTCAE v4.0 Term	Grade 1	Grade 2	Grade 3	Grade 4
HGB decreased (Anemia)	<LLN – 100 g/L	<100 – 80 g/L	<80 g/L	
Platelet count decreased	<LLN – 75.0 x10e9 /L	<75.0 - 50.0 x10e9 /L	<50.0 – 25.0 x10e9 /L	<25.0 x 10e9 /L
White blood cell decreased	<LLN - 3.0 x 10e9 /L	<3.0 - 2.0 x 10e9 /L	<2.0 - 1.0 x 10e9 /L	<1.0 x 10e9 /L
Neutrophil count decreased	<LLN - 1.5 x 10e9 /L	<1.5 - 1.0 x 10e9 /L	<1.0 - 0.5 x 10e9 /L	<0.5 x 10e9 /L
Lymphocyte count decreased	<LLN - 0.8 x 10e9/L	<0.8 - 0.5 x 10e9 /L	<0.5 - 0.2 x 10e9 /L	<0.2 x 10e9 /L
Creatinine increased*	>1 - 1.5 x baseline; >ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 baseline; >3.0 - 6.0 x ULN	>6.0 x ULN
TBL increased	>ULN - 1.5 x ULN	>1.5 - 3.0 x ULN	>3.0 - 10.0 x ULN	>10.0 x ULN
GGT increased	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALT increased	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
AST increased	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALP increased	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
Glucose increased (Hyperglycemia)	; >ULN - 8.9 mmol/L	>8.9 - 13.9 mmol/L	>13.9 - 27.8 mmol/L	>27.8 mmol/L
Glucose decreased (Hypoglycemia)	<LLN - 3.0 mmol/L	<3.0 - 2.2 mmol/L	<2.2 - 1.7 mmol/L	<1.7 mmol/L
Cholesterol high	>ULN - 7.75 mmol/L	>7.75 - 10.34 mmol/L	>10.34 - 12.92 mmol/L	>12.92 mmol/L
Hypertriglyceridemia	1.71 - 3.42 mmol/L	>3.42 - 5.7mmol/L	>5.7 - 11.4 mmol/L	>11.4 mmol/L

*Note: for “creatinine increased” the baseline criteria do not apply.
Note: Grade 4 Hemoglobin events were defined as fatal anemia.

Shift tables will be presented comparing baseline laboratory result (CTCAE grade) with the worst results (expressed in CTCAE grade) during the treatment phase (either initial or entire) analyzed. Of note, baseline will be defined as last assessment prior to first dosing in initial treatment phase. Subjects with abnormal laboratory values will be listed and values outside the normal ranges will be flagged.

Summaries for newly occurring or worsening clinically notable lipid abnormalities will also be provided cumulatively for each of the following parameters and categories:

- HDL:
  - <=LLN
  - <0.8 x LLN
- LDL, cholesterol, triglycerides:
  - >=ULN
  - >1.5 x ULN
  - >2.5 x ULN

Newly occurring or worsening liver enzyme abnormalities will also be summarized based on the event criteria given in Table 13-3:

Table 13-3 Liver-related events

Parameter	Criterion
ALT	>3xULN; >5xULN; >8xULN;>10xULN, >20xULN
AST	>3xULN; >5xULN; >8xULN >10xULN; >20xULN
ALT or AST	>3xULN; >5xULN; >8xULN >10xULN; >20xULN
TBL	>1.5xULN, >2xULN, >3xULN,
ALP	>2xULN, >3xULN. >5xULN
ALT or AST & TBL	ALT or AST>3xULN & TBL >2xULN; ALT or AST >5xULN & TBL >2xULN; ALT or AST >8xULN & TBL >2xULN; ALT or AST >10xULN & TBL >2xULN
ALP & TBL	ALP >3xULN & TBL >2xULN ALP >5xULN & TBL >2xULN
ALT or AST & TBL & ALP	ALT or AST>3xULN & TBL >2xULN & ALP <2xULN **(Hy's Law**) Note: elevated ALP may suggest obstruction as a consequence of gall bladder or bile duct disease; ALP may also be increased in malignancy. FDA therefore terms Hy's Law cases as indicators of *pure hepatocellular injury*. This does not mean that cases of ALT or AST >3xULN & TBL >2xULN & ALP ≥2xULN may not result in severe DILI.

Notes:
In studies which enroll subjects with pre-existing liver disease, baseline LFT may be increased above ULN; in such a case it is meaningful to add the condition "and worse than baseline" to the abnormality criteria

For a combined criterion to be fulfilled, all conditions have to be fulfilled on the same visit. The criteria are not mutually exclusive, e.g. a subject with ALT = 6.42xULN is counted for ALT >3xULN and ALT>5x ULN.

Individual subject data listings will be provided for subjects with abnormal laboratory data. Data of subjects with newly occurring or worsening liver enzyme abnormalities will be listed in an additional listing.

Boxplots over time will be presented for selected laboratory parameters (neutrophils, liver and lipid parameters).

13.3 Vital signs

The summary of vital signs will only include treatment emergent data, which are defined as those vital sign measurements after the first dose of study treatment and on or before last dose + 84 days.

Analysis in vital sign measurement using descriptive summary statistics for the change from baseline for each post-baseline visit will be performed. These descriptive summaries will be presented by vital sign and treatment group. Change from baseline will only be summarized for subjects with both baseline and post-baseline values and will be calculated as:

change from baseline = post-baseline value – baseline value

The number and percentage of subjects with newly occurring notable vital signs will be presented. Criteria for notable vital sign abnormalities are provided in Table 13-4:

Table 13-4 Criteria for notable vital sign abnormalities

Vital sign (unit)	Notable abnormalities
Systolic blood pressure (mmHg)	>= 140 mmHg or < 90 mmHg
Diastolic blood pressure (mmHg)	>=90 mmHg or <60 mmHg
Pulse (bpm)	> 100 bpm or <60 bpm

13.4 Electrocardiogram (ECG)

The summary of ECG will only include treatment emergent data, which are defined as those ECG measurements after the first dose of study treatment and on or before last dose + 84 days.

The following quantitative variables will be summarized: ventricular rate, RR interval, PR interval, QRS duration, QT interval, and corrected QT interval (QTc). Both Bazett (QTcB) and Fridericia (QTcF) corrections will be presented for QTc.

QTc will be summarized by computing the number and percentage of subjects (including 95% confidence intervals for pooled analyses, e.g. DMC or SCS) with:

- QTc > 500 msec
- QTc > 480 msec
- QTc > 450 msec
- QTc changes from baseline > 30 msec
- QTc changes from baseline > 60 msec
- PR > 250 msec

Summary statistics will be presented for ECG variables by visit and treatment group.

In addition, shift tables comparing baseline ECG interpretation (normal, abnormal, not available, total) with the worst on-study interpretation (normal, abnormal, not available, total) will be provided.

A listing of all newly occurring or worsening abnormalities will be provided, as well as a by-subject listing of all quantitative ECG parameters.

[redacted]

13.6 Compound specific safety evaluation

Safety topics of interest, such as risks defined in the Safety Profiling Plan (SPP), Risk Management Plan or topics of interest regarding signal detection or routine analysis are defined in the Program Case Retrieval Sheet that is stored in CREDI at the path Cabinets/CREDI Projects/A/AIN457A/Integrated Medical Safety.

The crude incidence and exposure-adjusted incidence rates for SPP risks will be summarized. In addition, listings will be provided presenting which subjects experienced which risk.

Important note: For the evaluation of SPP risks primary and secondary system organ classes of the MedDRA dictionary will be considered.

14 Interim analyses

The primary analysis will be performed after all subjects have completed the Week 16 visit in order to support regulatory filing. As the primary analysis is scheduled at Week 16, no adjustments will be made to the testing strategy to control the family-wise type I error. The investigators, site personnel and monitors will continue to remain blinded to the treatment each subject received at randomization all patients have completed the study (Week 60 Follow up) and the final database lock has occurred. Additional analyses may be performed to support interactions with health authorities, as necessary.

15 Determination of sample size

The sample size calculation is driven by the patient exposure requirements in the study population. An overall type I error (2-sided) 5% will be used to control type I error. Secukinumab 150 mg regimen will be tested versus placebo with respect to the primary endpoint (ASAS20 response at Week 16). A sample size of 300 patients in secukinumab 150 mg group and 150 patients in placebo group (randomization ratio = 2:1) is chosen in order to expose 300 patients to secukinumab for safety evaluation and also to achieve adequate power for the primary and secondary endpoints for this study.

Analysis of an unpublished phase III study showed a placebo response rate of about 28.4% and active drug response rate of 61.1% after 16 weeks for ASAS20. Using these assumptions, overall, the power for the ASAS20 endpoint should be about 99% with 300 patients in secukinumab 150 mg group and 150 patients in placebo group based on Fisher's exact test (nQuery Advisor 7.0).

15.1 Power for analysis of secondary variables

A summary of the assumptions and power for the primary and secondary efficacy parameters using the same unpublished study is shown in Table 15-1 for binary endpoints and Table 15-2 for continuous endpoints.

Table 15-1 Summary of power for binary endpoints

Endpoint	Response Rate		Power
	Secukinumab 150 mg (N=300)	Placebo (N=150)	
ASAS20	61.1%	28.4%	99%
ASAS40	36.1%	10.8%	99%
ASAS5/6	43.1%	8.1%	99%
ASAS partial remission	13.9%	4.1%	92%

Table 15-2 Summary of power for continuous endpoints

Endpoint	Mean change from baseline		Common standard deviation	Power
	Secukinumab 150 mg (N=300)	Placebo (N=150)		
hsCRP	-0.60	0.12	0.84	99%
Total BASDAI	-2.19	-0.85	2.10	99%

Endpoint	Mean change from baseline		Common standard deviation	Power
	Secukinumab 150 mg (N=300)	Placebo (N=150)		
SF-36 PCS	6.06	1.92	6.65	99%
ASQoL	-4.00	-1.37	4.45	99%

16 Appendix

16.1 Visit Windows

When visit windows are used, all visits will be re-aligned, i.e., they will be mapped into one of the visit windows. E.g., if the *Week 4* visit of a subject is delayed and occurs on Day 46 instead of on Day 29, it will be re-aligned to visit window *Week 8*. In the case of major deviations from the visit schedule, or due to unscheduled visits, several assessments of a subject may fall in a particular visit window (either scheduled or unscheduled). Statistical approaches to handle multiple assessments in a given visit window are specified below.

For lab/ECG/vital signs, follow-up (F/U) visit is excluded from analysis visit mapping window. Only assessments that come as F/U nominal visit will be directly assigned as analysis F/U visit. Other assessments that are beyond the last on-treatment visit window (W60) or after nominal F/U visit date won't be mapped to any analysis visit. F/U visit will not be included in the summary tables by visit.

Of note, subjects are allowed to have gaps in visits. All data collected will be displayed in listings.

Table 16-1 Analysis visit windows

Analysis Visit	Target Day	Analysis Visit Window	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Baseline	1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
Week 1	8	2-11	2-11					
Week 2	15	12-18	12-18	2-22				
Week 3	22	19-25	19-25					
Week 4	29	26-43	26-43	23-43	2-71			2-57
Week 8	57	44-71	44-71	44-71				
Week 12	85	72-99	72-99	72-99				58-99
Week 16	113	100-127	100-127	100-127	72-141	2-141	2-239	100-141
Week 20	141	128-155	128-155	128-155				
Week 24	169	156-183	156-197	156-197	142-267	142-267		142-197
Week 28	197	184-211						
Week 32	225	212-239	198-253	198-253				198-253
Week 36	253	240-267						
Week 40	281	268-295	254-323	254-323				254-323
Week 44	309	296-323						
Week 48	337	324-351						
Week 52	365	352-379	324-379	324-379	268-379	268-379	240-379	352-379
Week 56	393	380-407						
Week 60	421	408-435	380-435	380-435	380-435	380-435	380-435	380-435

Group1: Patient's global assessment of disease activity (VAS), Patient's assessment of back pain intensity (VAS), BASFI, BASDAI, Vital signs, hematology, blood chemistry, urinalysis

Group2: [redacted] hsCRP

Group3: SF-36, [redacted], [redacted], ASQoL, [redacted] Lipids

Analysis Visit	Target Day	Analysis Visit Window	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Group4: [redacted], Weight, [redacted] Cardiovascular panel								
Group5: ECG								
Group6: Urine pregnancy test								
* The first day of administration of randomized study treatment (first dose) is defined as Day 1.								

The following rules are used to determine the window for an applicable visit post baseline: "Lower limit" = "upper limit of prior applicable visit" + 1. "Upper limit" = "target day of current visit" + integer part of ("target day of next applicable visit" – "target day of current visit")/2. Lower limit of the first applicable visit is always Day 2.

The mapping described above applies to all visits (not just scheduled visits). Repeat and/or unscheduled visits (which will be numbered in the database according to new NCDS standards) will be mapped for analysis purposes in the same way as scheduled visits. This leaves the possibility, then, for multiple measurements within an analysis window. The following conventions will be used to determine the appropriate measurement to be summarized in the event of multiple measurements within a visit window.

Table 16-2 Rules for flagging variables

Timing of measurement	Type of data	Rule
Baseline	All data	The last measurement made prior to administration of the first dose of study treatment – note this may include measurements taken on the day of randomization (e.g. lab). Baseline assessments scheduled for and captured on Day 1 will be considered baseline measurements regardless of the time of assessment. If a patient did not receive any dose of study treatment then the randomization date will be used. Only the date part will be considered if there is only one assessment on Day 1 but if there are multiple assessments on Day 1, then the following rules will apply: 1. If time of assessment exist, • select the last available measurement prior to the reference start date/time considering time • if no measurement prior to the reference start date/time then considering time select the earliest measurement post reference start date/time 2. If time of assessment does not exist the measurement from the lowest CRF visit number will be used.
Post-baseline efficacy	All data	The measurement closest to the target day will be used. In the event two measurements are taken equally apart (e.g. 1 day before target date and 1 day after) the first one will be used. If the patient switches from placebo to AIN within the window the following rules apply: 1. If the analysis visit window is ≤ week 16, then • If available, the closest measurement to the target date which is on or before the switch date will be used • If there are no data on or before the switch then the closest measurement after the switch to target will be used 2. If the analysis visit window is > week 16, then • If available, the closest measurement to the target date which is after the switch date will be used • If there are no data after the switch then the closest to target before or on the switch date will be use Cases where the same parameter is recorded more than once on the same date will be handled as follows: • If time of completion exists the earliest measurement will be use • If time does not exist the measurement from the lowest CRF visit number will be used
Post-baseline safety	Summary visit information (e.g. lab, ECG, etc.)	The measurement closest to the target day will be used. In the event two measurements are taken equally apart (e.g. 1 day before target date and 1 day after) the first one will be used. If the patient switches from placebo to AIN within the window the following rules apply: 1. If the analysis visit window is ≤ week 16, then

Timing of measurement	Type of data	Rule
		• If available, the closest measurement to the target date which is on or before the switch date will be used • If there are no data on or before the switch then the closest measurement after the switch to target will be used 2. If the analysis visit window is > week 16, then • If available, the closest measurement to the target date which is after the switch date will be used • If there are no data after the switch then the closest to target before or on the switch date will be use Cases where the same parameter is recorded more than once on the same date will be handled as follows: • If time of completion exists the earliest measurement will be used • If time does not exist the measurement from the lowest CRF visit number will be used • If CRF visit number is the same the average value will be used
Post-baseline safety	Notable abnormalities (e.g. lab)	The most extreme measurement in the window will be used. Note this means a patient can have a notably high and notably low measurement within a window

16.2 Statistical methodology and assumptions

16.2.1 Analysis of continuous data

16.2.1.1 Summary statistics for continuous data

Summary statistics (including N, mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum) will be provided for continuous data by visit and treatment group.

16.2.1.2 Analysis of covariance

Univariate model

An analysis of covariance (ANCOVA) model will be used to analyze some endpoints. The model will include factors and covariates as specified for respective analysis. The SAS code below outlines a template for the analysis where covariates can be added or removed as required:

ods output diffs=*lsm_diff* lsmeans=*lsm_est*;

proc mixed data=;

by *visit*;

class *treatment strata*;

model *outcome = treatment strata baseline weight*;

lsmeans *treatment* / diff cl;

run;

Least-square-mean (LSM) estimates for each treatment group and LSM difference, confidence intervals and p-value for the difference between secukinumab and placebo can be obtained.

Repeated measures analysis

Some endpoints will be analyzed using a longitudinal model that comprises several visits. The model used will be mixed model repeated measures (MMRM) with factors, covariates, interactions and covariance structure as specified for respective analysis. The SAS code below outlines a template for the analysis where covariates and interaction terms can be added or removed as required:

ods output diffs=*lsm_diff* lsmeans=*lsm_est*;

proc mixed data=;

class *treatment strata visit*;

model *outcome = treatment strata visit baseline weight treatment*visit baseline*visit* / ddfm=kr;

repeated *visit* / type=un subject=;

lsmeans *treatment*visit* / diff cl;

run;

Least-square-mean (LSM) estimates for each treatment group and LSM difference, confidence intervals and p-value for the difference between secukinumab and placebo will be calculated at appropriate analysis visits.

In case the MMRM model does not converge the following sequential steps will be used:

1. change to type=cs. If still no convergence, perform step 2.

2. remove covariates in the following order until convergence: *weight*, *baseline*visit*, *strata*.

16.2.1.3 Non-parametric analysis of covariance

A non-parametric ANCOVA model (Koch 1998) will be used as sensitivity analysis for certain binary endpoints and also for continuous endpoints that risk having a large deviation from a normal distribution.

The macro NParCov3 will be used (Zink 2012). Input dataset to the macro should only contain the two treatments to be compared and only data from one visit.

The macro call will follow the below templates where covariates (in numeric format only) may be added or removed as required (max one strata variable can be specified).

For continuous variables

%nparcov3(outcomes=*outcome*, covars=*cov1 cov2*, c=*1*, hypoth=*alt*, strata=*strata*, trtgrps=*treatment*, transform=*none*, combine=*first*, dsnin=, dsnout=*out*);

Data sets "*_out_deptest*" and "*_out_ci*" will be automatically created by the macro and contains the following:

- "*_out_deptest*" provides estimate and p-value for the treatment difference, and
- "*_out_ci*" provides a 95% confidence interval for the treatment difference.

For binary variables

%nparcov3(outcomes=*outcome*, covars=*cov1*, c=*1*, hypoth=*alt*, strata=*strata*, trtgrps=*treatment*, transform=*logistic*, combine=*first*, dsnin=, dsnout=*out*);

The odds ratio and confidence interval can be obtained from "*_out_ratioci*".

16.2.2 Analysis of binary and categorical data

16.2.2.1 Summary statistics for binary and categorical data

Summary statistics for discrete variables will be presented in contingency tables and will include count and frequency in each category. If applicable, confidence intervals will be derived as well based on the score method including continuity correction (Newcombe 1998):

With z as (1-α/2)-quantile of the standard normal distribution (SAS: z=probit(1-α/2), n as total number of subjects (i.e. number of subjects in the denominator), p as estimated crude incidence (number of subjects with event / n) and q = 1-p

Then the lower limit is

$$L = 100 \times \max\left(0, \frac{2np + z^2 - 1 - z\sqrt{z^2 - 2 - 1/n + 4p(nq+1)}}{2(n+z^2)}\right)$$

and the upper limit is

$$U = 100 \times \min\left(1, \frac{2np + z^2 + 1 + z\sqrt{z^2 + 2 - 1/n + 4p(nq-1)}}{2(n+z^2)}\right).$$

In addition, if L > p then L = p and if U < p then U = p.

For binary response variables the placebo-adjusted response rates including 95% confidence interval will be derived.

SAS code for risk difference:

proc freq data=;

tables response*trt/ riskdiff;

run;

Note the response value should be sorted with '1' ahead of '0'.

16.2.2.2 Logistic regression

Certain binary outcome variables, e.g. response outcomes, will be evaluated using a logistic regression model. The model will include factors and covariates as specified for respective analysis. The SAS code below outlines a template for the analysis where covariates can be added or removed as required:

ods output diffs=*lsm_diff* convergencestatus=*conv_status*;

proc logistic data=;

by *visit*;

class *treatment strata* / param=glm;

model *response = treatment strata weight*;

lsmeans *treatment* / diff cl exp;

run;

In cases where separation is a concern, e.g. 0% or 100% response for some treatment and covariate level combination, an exact logistic regression model may be applied to all visits. To help with convergence, this model will not include any continuous covariates. The SAS code below outlines a template for the analysis where covariates can be added or removed as required:

ods output exactoddsratio=*exact_or*;

proc logistic data=*data* exactonly;

by *visit*;

class *treatment strata* / param=ref;

model *response = treatment strata*;

exact *treatment* / estimate=odds;

run;

When exact logistic regression is unable to be implemented (due to computational complexity as the procedure can lead to extremely long run times), then Fisher's exact test will be applied. In this case, only a p-value for a test of equal response in the two groups can be obtained (no odds ratios or confidence intervals can be estimated.)

ods output fishersexact=*fisher*;

proc freq data=;

by *visit*;

table *treatment***response* / fisher;

run;

Input dataset should only contain data from the two treatment groups to be compared.

16.2.3 Imputation methods

16.2.3.1 Multiple Imputation

A linear regression model will be used to perform multiple imputation (MI) under a missing-at-random (MAR) assumption. To help preserve the relationship between outcome and covariates within each treatment a separate model will be run for each treatment. This will also

help ensure that the imputation model does not make stronger assumptions on data relations than the analysis model.

The SAS code below outlines a template for the analysis where covariates and visits can be added or removed as required. To ensure that results can be replicated the data should be sorted by subject number before running the model (the data should be in horizontal format with one subject per dataset row).

proc mi data= seed=4572308 nimpute=100 out=*mi_out*;

by *treatment*;

class *strata*;

fcs reg (/details);

var *strata weight value1 value2 value3*;

run;

Where in the template code the continuous variable to be imputed is *value* (e.g. *value1* could be the baseline value and *value2* the first post-treatment measurement of the variable to be imputed.) Normally, all data collection visits during the analysis period of interest would be included in the model. Including variables using a CLASS statement instead of a BY statement should help facilitate model convergence also when the number of non-missing data points are low for some specific covariate level and visit combination. The FCS option is used to ensure that also non-monotone missing data can be handled in an appropriate way.

If convergence is not obtained the following sequential steps will be used:

1. remove *weight*. If still no convergence, perform step 2

2. remove *status*.

For a situation where several variables need to be imputed using separate models (e.g. using independent models to impute each component needed to derive a response variable *V*) a step-wise process needs to be implemented as outlined below:

1. Run the SAS code as described above for the first variable to be imputed
2. Run the SAS code as described above for the next variable to be imputed (but with the following changes: “data=*mi_out*”, “out=*mi_out2*”, “by *treatment _imputation_*”, “nimpute=1”)
3. Repeat step 2, but with input dataset equal to the output dataset from the prior step, until all *j* variables have been imputed resulting in a dataset named mi_out_j
4. Derive the variable *V* from within mi_out_j

The required analysis (e.g. ANCOVA) is then performed separately within each imputation dataset (as identified by variable *_imputation_*). To obtain the final result of the imputation process the analysis result from each imputation dataset needs to be combined according to Rubin’s rules as outlined below:

ods output parameterestimates=*mi_result*;

proc mianalyze data=;

modeleffects *estimate*;

stderr *estimate_se*;

run;

The *estimate* and *estimate_se* parameters come from the analysis model used to analyze the imputed variable within each imputation dataset (e.g. from the lsmean estimate of the treatment difference and its standard error obtained from PROC LOGISTIC or PROC MIXED.)

To obtain binary response rates and confidence intervals for individual treatment groups the following process should be followed (exemplified for one visit):

ods output binomialprop=*bin_est*;

proc freq data=;

by *treatment _imputation_*;

table *response* / binomial (cl=wilson correct);

run;

Then apply a logit transformation on the saved proportions and derive its standard error:

data *bin_est*; set *bin_est*;

estimate=log(*_bin_*/(1-*_bin_*));

estimate_se=*e_bin*/(*_bin_**(1-*_bin_*));

run;

The transformed binomial proportion estimates and its standard errors are then combined by applying Rubin's rules as described above using PROC MIANALYZE. Before presenting the combined data it needs to be transformed back as follows:

data *mi_result*; set *mi_result*;

prop_est=1/(1+exp(-*estimate*));

prop_lower=1/(1+exp(-*lclmean*));

prop_upper=1/(1+exp(-*uclmean*));

run;

If all responses are imputed as 0 (or 1) for all imputation datasets for a specific treatment group then the between-imputation-variation will be zero. The combined final response rate would be presented as seen in any of the imputed datasets but the 95% CI will be undefined.

If after imputation all responses are either 0 or 1 for a combination of treatment group and imputation dataset it will not be possible to perform a logit transformation and the response rate (0% or 100%) will be presented without 95% CI.

16.2.4 Crude incidence and related risk estimates

16.2.4.1 Crude incidence and 100*(1-α)% confidence interval

For n subjects, each at risk to experience a certain event with probability π, the crude incidence is estimated as $p=x/n$, where x is the number of subjects with the event.

Absolute and relative frequencies will be displayed as well as 95% confidence interval for the relative frequency based on the score method including continuity correction (Newcombe 1998).

With z as $(1-\alpha/2)$-quantile of the standard normal distribution (SAS: z=PROBIT(1-α/2), n as total number of subjects (i.e. number of subjects in the denominator), and p as estimated crude incidence (number of subjects with event / n) it is q=1-p.

Then the lower limit is

$$L = \max\left(0, \frac{2np + z^2 - 1 - z\sqrt{z^2 - 2 - 1/n + 4p(nq+1)}}{2(n+z^2)}\right)$$

and the upper limit is

$$U = \min\left(1, \frac{2np + z^2 + 1 + z\sqrt{z^2 + 2 - 1/n + 4p(nq-1)}}{2(n+z^2)}\right).$$

In addition, if L > p then L = p and if U < p then U = p.

If appropriate, an exact 100*(1-α)% confidence interval (Clopper-Pearson 1934) will be obtained by using the SAS procedure PROC FREQ with the EXACT BINOMIAL statement. However, the confidence interval derived via the score method including continuity correction will be the default in safety analyses.

16.2.4.2 Odds ratio and 100*(1-α)% confidence interval

For an investigational drug group with n_1 subjects at risk, independent from the control group (e.g. placebo or comparator) with n_0 subjects at risk, of whom x_1 and x_0 experience a certain event with probability π_1 and π_0 respectively, the odds ratio is estimated as

$\frac{p_1/(1-p_1)}{p_0/(1-p_0)}$ with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$. A conditional exact 100*(1-α)% confidence interval can be obtained by using the SAS procedure PROC FREQ with statement EXACT OR. However, to be able to adjust for covariates odds ratios will primarily be obtained from PROC LOGISTIC.

16.2.4.3 Risk difference and 100*(1-α)% confidence interval

For an investigational drug group with n_1 subjects at risk, independent from the control group (e.g. placebo or comparator) with n_0 subjects at risk, of whom x_1 and x_0 experience a certain event, the risk difference is estimated as p_1-p_0 with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$.

Exact unconditional confidence limits for the risk difference can be obtained with SAS procedure PROC FREQ and option RISKDIFF in the TABLES statement, specifying the RISKDIFF option also in the EXACT statement.

16.2.5 Exposure adjusted incidence rate and related risk estimates

16.2.5.1 Exposure adjusted incidence rate and 100*(1-α)% confidence interval

It will be assumed that for each of n subjects in a clinical trial the time t_j (j=1,...,n) to the first occurrence of a certain event is observed, or if the event was not experienced, the (censored) time to the end of the observation period. The sequence of first occurrences of an event will be modeled to follow approximately a Poisson process with constant intensity θ. The rate parameter θ will be estimated as $\lambda = D/T$, where $T = \sum_{j=1}^{n} t_j$ and D is the number of subjects with at least one event. Conditionally on T, an exact 100*(1-α)% confidence interval for a Poisson variable with parameter θT and observed value D can be obtained based on (Garwood 1936), from which an exact 100*(1-α)% confidence interval for D/T will be derived as follows (Sahai 1993; Ulm 1990):

Lower confidence limit $L = \frac{0.5 c_{\alpha/2,2D}}{T}$ for D>0, 0 otherwise,

Upper confidence limit $U = \frac{0.5 c_{1-\alpha/2,2D+2}}{T}$

Where $c_{\alpha,k}$ is the αth quantile of the Chi-square distribution with k degrees of freedom. The example below shows how this should be handled for cases where subjects switch treatment. In particular for summarizing 'Any AIN' as a group, one should take into consideration the sequence of treatments while calculating exposure time for subjects.

Table 16-3 Examples for calculating exposure time for incidence rates (IR)

1st treatment / total exposure time	2nd treatment / total exposure time	AE event onset (in days from study start)	Exposure for IR
Placebo / 100 days	AIN457 150 mg / 200 days	Day 50 (during 1st treatment) Day 110 (10 days into 2nd treatment)	Placebo: 50 days AIN457 150 mg: 10 days Any AIN: 10 days

16.3 Additional sample size considerations

There are no additional sample size considerations for this study.

17 References

[redacted]

Bretz F, Maurer W, Brannath W, et al. (2009) A graphical approach to sequentially rejective multiple test procedures. Stat Med; 28(4):586-604.

Cella DF, Tulsky DS, Gray G et al. (1993) The Functional Assessment of Cancer Therapy (FACT) scale: Development and validation of the general measure. Journal of Clinical Oncology; 11(3):570-579.

Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26; 404–413.

Corticosteroid converter (data retrieved December 12, 2013): http://www.globalrph.com/steroid.cgi

Davis JC, Revicki D, van der Heijde D, et al (2007) Health-related quality of life outcomes in patients with active ankylosing spondylitis treated with adalimumab: results from a randomized controlled study. Arthritis & Rheumatism; 57(6):1050-1057.

[redacted]

Doward LC, Spoorenberg A, Cook SA, et al. (2003) Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis; 62(1):20-26.

Garwood F (1936). Fiducial limits for the Poisson distribution. Biometrika, 46; 441–453.

[redacted]

Heinze G (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. Statistics in Medicine; 25:4216–4226

Huffman K, Samsa G, Slentz C, et al. (2006) Response of high-sensitivity C-reactive protein to exercise training in an at-risk population. Am Heart J; 152(4):793-800.

Inman RD, David JC Jr, et al. (2008) Efficacy and safety of golimumab in patients with ankylosing spondylitis: results of a randomized, double-blind, placebo-controlled, phase III trial. Arthritis Rheum; 58(11):3402-12.

Koch GG, Tangen CM, Jung JW, et al. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine; 17:1863-92.

Lubeck, DP. (2004) Patient-reported outcomes and their role in the assessment of rheumatoid arthritis. Pharmacoeconomics 22(1): 27-38.

[redacted]

[redacted]

Newcombe, RG. (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine; 17: 857-872.

Sahai H, Khurshid Anwer (1993). Confidence intervals for the mean of a poisson distribution: a review. Biom J, 35 (7); 857-867

Sieper J, Rudwaleit M, Baraliakos X. (2009)The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis.; 68 Suppl 2:ii1-44.

Ulm K (1990). A simple method to calculate the confidence interval of a standard mortality ratio. American Journal of Epidemiology, 131(2); 373-375

van der Heijde D, Dijkmans B, et al. (2005) Efficacy and safety of infliximab in patients with ankylosing spondylitis: Results of a randomized, placebo-controlled trial (ASSERT). Arthritis Rheum; 52(2):582-591.

van der Heijde D, Kivitz A, Schiff MH, et al. (2006) Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum; 54(7):2136-46.

van der Heijde D, Landewe R, Feldtkeller E. (2008) Proposal of a linear definition of the Bath Ankylosing Spondylitis Metrology Index (BASMI) and comparison with the 2-step and 10-step definitions. Ann Rheum Dis; 67: 489-93.

van Tubergen A, Landewe R, van der Heijde D, et al. (2001) Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. *Arthritis Rheum*; 45: 430–38