NOVARTIS

Clinical Development

AIN457/Secukinumab/Cosentyx®

CAIN457K2340 / NCT03259074

A randomized, partially-blinded, active-controlled multicenter study of secukinumab to demonstrate reduction of radiographic progression versus GP2017 (adalimumab biosimilar) at 104 weeks and to assess the long term safety, tolerability and efficacy up to 2 years in patients with active ankylosing spondylitis

Statistical Analysis Plan (SAP) Amendment 3

Author:	Trial Statistician, [redacted]

Document type:	SAP Documentation
Document status:	Final
Release date:	10-May 2022
Number of pages:	49

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
27-Jun-2018	Prior to DBL	In general the SAP was amended to align with the changes in the protocol (via amendment 1).	Amendment 1	
		Updated the primary and secondary objectives		1.2.1, 1.2.2
		Defined the syndesmophyte subset and MRI subset		2.2
		Rewrote the estimand attributes to align with the changes in the protocol		2.5.1, 2.5.2, 2.6.1., 2.6.2
		Updated the handling of missing data for the primary and secondary objectives		2.5.3, 2.6.3
		Updated the testing hierarchy and power		2.6.2, 3
		[redacted]		[redacted]
		Added additional details for some of the efficacy variables' description		5
		Added references		6
03-Feb-2021	Prior to DBL	In general the SAP was amended to align with the changes in the protocol (via amendment 2)	Amendment 2	
		Added COVID-19, ECG, and TEAE in the list of abbreviations		2.5.1, 2.5.2, 2.6.2
		Updated the analysis from using odds ratio to difference in proportion		

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
		Updated the supportive analyses		2.5.4, 2.6.4
		Added the subsection "Electrocardiogram (ECG)" and the relative details		2.8.3.2
		Added the section "COVID-19 related supportive [redacted]" and the relative details		2.15, 2.15.1, 2.15.2
		[redacted]		[redacted]
		Updated the details for logistic regression		5.3.2.2
		Added a tipping point analysis for continuous endpoints.		5.3.3.1
		Added a reference		6
05-May-2022	Prior to DBL	In general the SAP was amended to add template programming codes and specify the data completion for the primary analysis	Amendment 3	
		Deleted a paragraph on prior and concomitant ankylosing spondylitis therapy as these data are not collected		2.4.2
		Corrected the notation in the model		2.5.3
		Added the unused retrieved dropout based model as possible sensitivity analysis for the primary endpoint		2.5.4
		Specified how to complete missing post baseline scores		5.1

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
		Added a template SAS code for the ANCOVA model		5.4.1.2
		Added a template programming code and description for the MCMC based multiple imputation		5.4.3.2

Table of contents

List of abbreviations

AE	adverse event
ALT/SGPT	alanine aminotransferase/ serum glutamic pyruvic transaminase
ANCOVA	analysis of covariance
AS	ankylosing spondylitis
ASAS	Ankylosing SpondyloArthritis International Society
ASDAS	Ankylosing Spondylitis Disease Activity Score
[redacted]	[redacted]
ASspiMRI-a	Ankylosing Spondylitis Spine Magnetic Resonance Imaging - activity
AST/SGOT	aspartate aminotransferase/ serum glutamic oxaloacetic transaminase
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BMI	body mass index
BSL	baseline
COVID-19	Coronavirus disease of 2019
CRF	case report/record form (paper or electronic)
[redacted]	[redacted]
CTC	Common Toxicity Criteria
ECG	Electrocardiogram
EMA	European Medical Agency
ENRADAS	Effects of Non-Steroidal Anti-Inflammatory Drugs on RAdiographic Damage in Ankylosing Spondylitis
[redacted]	[redacted]
FAS	full analysis set
FDA	Food and Drug Administration
GP2017	GP2017 (adalimumab biosimilar)
IL	interleukin
IN	Investigator Notification
IRT	interactive response technology
LLN	lower limit of normal
MedDRA	Medical dictionary for regulatory activities
[redacted]	[redacted]
MMRM	mixed-effects model repeated measures
MRI	magnetic resonance imaging/image
mSASSS	modified Stoke Ankylosing Spondylitis Spinal Score
MTX	methotrexate
OASIS	Outcome in Ankylosing Spondylitis International Study
NSAID	non-steroidal anti-inflammatory drug
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
PRN	pro re nata

[redacted]	[redacted]
[redacted]	[redacted]
SAE	serious adverse event
s.c.	subcutaneous(ly)
SCR	screening
SI	sacroiliac
TEAE	treatment emergent adverse events
TNF	Tumor Necrosis Factor
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 clinical study protocol. That 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

1.1 Study design

This multicenter study uses a randomized, partially-blinded, active-controlled, parallel-group design in subjects with AS. A screening period (SCR) of up to 10 weeks before randomization will be used to assess eligibility followed by 104 weeks of treatment. Two follow-up visits at Weeks 112 and 120 will occur thereafter.

At baseline (BSL), approximately 837 subjects whose eligibility is confirmed will be randomized to one of three treatment groups (1:1:1).

- Group 1: secukinumab 150 mg [1 x 1.0 mL s.c. plus placebo (1 x 1.0 mL s.c.)] at BSL, Weeks 1, 2 and 3, followed by administration every four weeks starting at Week 4
- Group 2: secukinumab 300 mg (2 x 1.0 mL s.c.) at BSL, Weeks 1, 2 and 3, followed by administration every four weeks starting at Week 4
- Group 3: GP2017 (adalimumab biosimilar) 40 mg (1 x 0.8 mL s.c.) at BSL followed by administration every two weeks starting at Week 2

Although study treatment (secukinumab vs GP2017) will be provided in an open-label fashion to the subjects, secukinumab study treatment will be blinded to the dose. Subjects in groups 1 and 2 will know that they are on secukinumab, but will be blinded to which treatment group and will not know whether they are receiving secukinumab 150 mg or 300 mg. Subjects in the GP2017 treatment group will know that they are receiving GP2017 (adalimumab biosimilar) 40 mg. Investigators and study site staff will be blinded in the same manner.

The primary endpoint analysis may be performed after all subjects complete the assessments associated with the primary endpoint visit (Week 104) in order to support regulatory filing. The final analysis will be conducted after all subjects complete the study at Week F120.

1.2 Study objectives and endpoints

1.2.1 Primary objective

The primary objective is to demonstrate the proportion of subjects on secukinumab (150 mg s.c. or 300 mg s.c.) with no radiographic progression as measured by mSASSS at Week 104 is superior to subjects on GP2017 (adalimumab biosimilar 40 mg s.c.).

1.2.2 Secondary objectives

1. To demonstrate the change from baseline in mSASSS in subjects on secukinumab (150 mg s.c. or 300 mg s.c.) is superior to GP2017 (adalimumab biosimilar 40 mg s.c.) at Week 104.

2. To demonstrate the proportion of subjects with a syndesmophyte at baseline with no new syndesmophytes at Week 104 on secukinumab (150 mg s.c. or 300 mg s.c.) is superior to GP2017 (adalimumab biosimilar 40 mg s.c.).
3. To evaluate the Berlin SI joint edema score in subjects on secukinumab (150 mg s.c. or 300 mg s.c.) at Week 104 versus GP2017 (adalimumab biosimilar 40 mg s.c.) (in a subset of subjects at selected sites).
4. To evaluate the ASspiMRI-a Berlin modification score in subjects on secukinumab (150 mg s.c. or 300 mg s.c.) at Week 104 versus GP2017 (adalimumab biosimilar 40 mg s.c.) (in a subset of subjects at selected sites).
5. To evaluate ASAS 20 response, ASAS 40 response, ASAS partial remission and ASDAS inactive disease in subjects on secukinumab 150 mg s.c. compared to secukinumab 300 mg s.c. at Week 104.
6. Overall safety and tolerability of secukinumab.

2 Statistical methods

2.1 Data analysis general information

Summary statistics for continuous variables will include N, mean, standard deviation, minimum, lower quartile, median, upper quartile, and 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 will be presented by the following 3 treatment groups.

- Secukinumab 150 mg
- Secukinumab 300 mg
- GP2017 40 mg

2.2 Analysis sets

The following analysis sets will be used in this study:

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 IRT prior to the site confirming all eligibility criteria had been met and to whom no study medication was given. Mis-randomized subjects 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 at randomization.

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.

Syndesmophyte Subset: The syndesmophyte subset will include all subjects in the FAS who have syndesmophyte at baseline. A subject with at least one individual vertebral score ≥ 2 for any of the interpretable locations will be classified as having a syndesmophyte.

MRI Subset: The MRI subset will include all subjects in the FAS who have an MRI performed at the selected centers.

2.2.1 Subgroup of interest

The primary endpoint(s) and secondary endpoints will be evaluated by gender and by smoking status. Subjects with missing data in a subgroup category will be omitted for the corresponding subgroup analysis.

2.3 Subject disposition, demographics and other baseline characteristics

2.3.1 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 the treatment period and follow-up period, 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 by deviation category. If not reported separately, this includes the PDs related to COVID-19.

2.3.2 Background and demographic characteristics

The following common background and demographic variables will be summarized:

- Gender, age, race, ethnicity, weight, height, and BMI.

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

- Subject's global assessment of disease activity and other ASAS components, hsCRP, use (yes/no) and separate dose of MTX (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, mSASSS, number of syndesmophytes, HLA-B27, total back pain (VAS), nocturnal back pain (VAS), total BASDAI score and spinal pain (BASDAI question #2).

2.3.3 Medical history

A history of AS with focus on previous extra-articular involvement and past therapies for AS will be obtained. Any other significant prior or active medical condition at the time of signing informed consent will be recorded and coded using the MedDRA dictionary. These medical conditions will be summarized by primary system organ class and preferred term.

To establish a baseline level of cardiovascular risk, the number and percentage of subjects with pre-solicited cardiovascular risk factors will be summarized by treatment group. The number of cardiovascular risk factors that each subject has will also be summarized by treatment group. If it is unknown whether or not a subject currently or previously experienced a specific cardiovascular risk factor, it will be assumed that cardiovascular risk factor did not occur for that subject.

Smoking history will be summarized by treatment group.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

The analysis of study treatment data will be based on the safety set. The number of visits with injections received will be presented 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, ≥ 2 weeks, ≥ 4 weeks, ≥ 8 weeks, etc.) will be presented.

Duration of exposure of a treatment will be defined as the time from first dose of the treatment to the minimum of (last dose of the treatment + 84 days) and (last visit date).

Duration of exposure of a treatment for the pre- and during COVID-19 pandemic periods will be defined as the duration of exposure of a treatment that fell into the pre- and during COVID-19 pandemic periods, respectively. These periods are defined in Section 2.15.1.

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.

2.4.2 Prior, concomitant and post therapies

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 randomized study treatment and within 84 days after the 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 non-drug therapies 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 the date of the last study visit will be concomitant surgeries and procedures, including those which were started pre-baseline and continued into the period where study treatment is administered.

NSAID use will be summarized.

2.5 Analysis of the primary objective

2.5.1 Primary endpoint

The primary efficacy objective is to demonstrate that the proportion of subjects on secukinumab (150 mg or 300 mg) with no radiographic progression is superior to GP2017 40 mg at Week 104. The analysis of the primary efficacy objective will be based on the following estimand:

- Population – defined through appropriate inclusion/exclusion criteria to reflect the targeted AS population
- Variable – binary response variable indicating no radiographic progression defined as a change from baseline in mSASSS at Week 104 of ≤ 0.5
- Intercurrent event–regardless of adherence to randomized treatment
- Population-level summary – difference in proportions of responders between the secukinumab and GP2017 arms

2.5.2 Statistical hypothesis, model, and method of analysis

The statistical hypothesis for no radiographic progression being tested is that there is no difference in the proportion of subjects with no radiographic progression at Week 104 in the secukinumab 150 mg or secukinumab 300 mg versus the GP2017 40 mg regimen.

Let p_j denote the proportion of subjects with no radiographic progression at Week 104 for treatment regimens j, j=0, 1 or 2 where

- 0 corresponds to GP2017 40 mg regimen,
- 1 corresponds to secukinumab 150 mg
- 2 corresponds to secukinumab 300 mg

In statistical terms, H_j: $p_j = p_0$, H_{Aj}: $p_j \neq p_0$,

H_1: secukinumab 150 mg is not different to GP2017 regimen with respect to proportion of subjects with no radiographic progression at Week 104

H_2: secukinumab 300 mg is not different to GP2017 regimen with respect to proportion of subjects with no radiographic progression at Week 104

The primary analysis will be conducted via logistic regression with treatment as a factor and includes baseline mSASSS as a covariate. Difference in marginal response proportions with p-value and respective 95% confidence interval will be estimated from the logistic regression model.

2.5.3 Handling of missing values/censoring/discontinuations

Missing Week 104 data for a given subject who discontinued in the study will be imputed using an imputation model based on the observed data of subjects in the same randomized arm who also discontinued study treatment but for whom Week 104 X-ray data are available ('retrieved dropout approach', EMA 2010) and based on the observed data of the given subject.

A logistic regression model will be fitted to the Week 104 X-ray data for the retrieved dropout patients, while adjusting for treatment, baseline mSASSS and post-baseline slope in mSASSS.

The post-baseline slope is defined as the last available post-baseline change in mSASSS score divided by study day on which the X-ray was taken. Recent literature suggests that in AS patients treated with TNF inhibitors the spinal radiographic progression followed a linear course during the first 4 years of follow-up (Maas, et. al. 2017). Thus, the linearity assumption is expected to be reasonable. The covariates of baseline mSASSS and slope will be standardized to have a mean of 0 and standard deviation of 1 before model fitting.

PROC MCMC will be used for fitting the logistic regression model and multiple imputation of the Week 104 value, with weakly informative priors and model code as shown below,

prior beta1 ~ normal(0,sd=2);

prior beta2 ~ normal(0,sd=2);

prior beta3 ~ normal(0,sd=2);

prior beta4 ~ normal(0,sd=3);

prior beta5 ~ normal(0,sd=3);

pi = logistic(beta1*trt_GP2017 + beta2*trt_SEC150 + beta3*trt_SEC300 + beta4*sqrt(base_stand) + beta5*slope_stand);

model noprog ~ binary(p = pi);

where trt_GP2017 – binary indicator for GP2017 40 mg

trt_SEC150 – secukinumab 150 mg

trt_SEC300 – secukinumab 300 mg

sqrt(base_stand) – square root of the baseline mSASSS (standardized)

slope_stand – the last available post-baseline change in mSASSS score divided by study day on which the X-ray was taken (standardized)

noprog – no radiographic progression at Week 104

If all post-baseline values are missing for a subject, the subject will be imputed based on a similar model as above but just based on treatment and baseline value as covariates (without the post-baseline slope variable). Subjects with a missing baseline will be excluded from the analysis.

Missing Week 104 data due to an incomplete number of interpretable locations for a given subject will be imputed using the same imputation model as described above but based on the observed data of all subjects who provided Week 104 data and based on the observed data of the given subject.

In this way multiple complete data sets are created and for each of these completed datasets the response rate will be calculated and then combined using Rubin's rule.

The potential sparsity of the retrieved data may preclude fitting the proposed model to impute the subjects with missing X-ray data at Week 104. In this case, a fallback approach will be utilized.

In case of low sample size of the retrieved dropout data (< 8 for any treatment group), the imputation model can lead to unstable imputations. In this case imputations will be performed based on a model without treatment as a covariate.

pi = logistic(beta1+ beta4*sqrt(base_stand) + beta5*slope_stand)

In case there are less than 8 retrieved dropout patients overall, the same imputation model as above will be used (without treatment as a covariate) but based on all patients (not only retrieved dropout patients).

2.5.4 Supportive analyses

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

Sensitivity analyses will be conducted to determine the impact of missing data handling on the primary estimand.

- Tipping point analysis
- In case an imbalance in the distribution of any baseline demographics and disease characteristics impacting response (e.g., gender and smoking status) is seen, additional analyses may be performed adjusting for these variables in the model
- The unused models described in Section 2.5.3 may be used for sensitivity analysis, if appropriate (e.g., if they converge).

Additional supplementary analyses may be conducted to examine different estimands which are also deemed clinically relevant, such as the estimand based on the hypothetical strategy which would examine the treatment effect in outcomes where biologic rescue was not available to subjects.

The effect of the COVID-19 pandemic on the study will be monitored and assessed. Additional supplementary analyses to study the potential effect of the pandemic on the primary estimands may be performed, and are described in Section 2.15.

2.6 Analysis of secondary efficacy objectives

2.6.1 Secondary endpoints

The secondary efficacy variables are listed below.

1. Change from baseline in mSASSS at Week 104
2. No new syndesmophyte at Week 104
3. Change from baseline in ASspiMRI-a Berlin modification score at Week 104
4. Change from baseline in Berlin SI joint edema score at Week 104
5. ASAS20 at Week 104
6. ASAS40 at Week 104
7. ASAS partial remission at Week 104
8. ASDAS inactive disease at Week 104

2.6.2 Statistical hypothesis, model, and method of analysis

The secondary efficacy variables and the method for adjusting for multiplicity are described below. All analysis will be based on the FAS population unless otherwise specified.Modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) at Week 104

Change from baseline in mSASSS will be evaluated using ANCOVA model with treatment as a factor and baseline mSASSS as a covariate.

Estimand definition:

- Population – defined through appropriate inclusion/exclusion criteria to reflect the targeted AS population
- Variable – mSASSS change from baseline at Week 104
- Intercurrent event –regardless of adherence to randomized treatment
- Summary measure – difference in means

Absence of new syndesmophytes at Week 104

The proportion of subjects with no new syndesmophyte will be evaluated using a logistic regression model with treatment group as a factor and baseline mSASSS as a covariate.

Difference in marginal response proportions with p-value and respective 95% confidence interval will be estimated from the logistic regression model.

Estimand definition:

- Population – defined through subjects with syndesmophyte at baseline, within the targeted AS population defined by the inclusion/exclusion criteria
- Variable – binary response variable indicating absence of new syndesmophyte at Week 104. Absence of new syndesmophyte is defined as having individual vertebral scores < 2 for each interpretable location which had no syndesmophyte at baseline.
- Intercurrent event – regardless of adherence to randomized treatment
- Population-level summary – difference in proportion of absence of new syndesmophte between the secukinumab and GP2017 arms

MRI of spine and SI joints

MRI analysis will be based on a subgroup of subjects who have MRI performed on spine and SI joints at selected sites.

The change from baseline to Week 104 in ASspiMRI-a Berlin modification score and Berlin SI joint edema score will be evaluated using a non-parametric ANCOVA model with treatment as a factor and baseline score as a covariate. Pairwise comparison will be performed for secukinumab (150 mg s.c. or 300 mg s.c.) versus GP2017. These will be assessed and analyzed outside of the testing strategy.

If values are missing for the baseline or post-baseline visit, the subject will be excluded from the analysis.

Summary statistics of observed data by visit and change from baseline will be provided for each treatment regimen. Summary statistics include mean, standard deviation, minimum, lower quartile, median, upper quartile and maximum.

ASAS 20, ASAS 40, ASAS partial remission and ASDAS inactive disease at Week 104

Response at Week 104 to ASAS 20, ASAS 40, ASAS partial remission and ASDAS inactive disease will be evaluated using a logistic regression model with treatment group as a factor and baseline score (if appropriate) as a covariate. These will be assessed and analyzed outside of the testing strategy.

Testing strategy

The following hypotheses will be included in the testing strategy, which controls the family-wise two-sided type I error at 5% level. The two primary hypotheses (H1 and H2) form the first family, which will be tested first at 5% level using the Hochberg (1988) procedure. The second family consists of the secondary hypotheses (H3 through H6), which will be tested only if both primary hypotheses have been rejected, and will be tested at 5% level using the graphical approach for sequentially rejective procedures (Bretz 2009) as illustrated in Figure 2-1.

Primary hypotheses:

H1: secukinumab 150 mg is not different from GP2017 40 mg with respect to proportion of subjects with no radiographic progression at Week 104

H2: secukinumab 300 mg is not different from GP2017 40 mg with respect to proportion of subjects with no radiographic progression at Week 104

Secondary hypotheses:

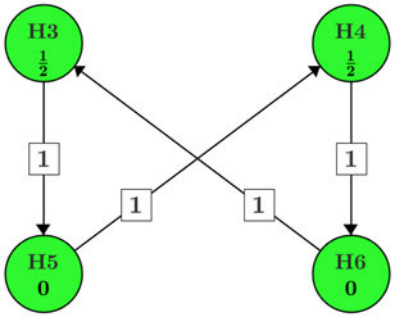
H3: secukinumab 150 mg is not different from GP2017 40 mg with respect to change from baseline in modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) at Week 104

H4: secukinumab 300 mg is not different from GP2017 40 mg with respect to change from baseline in modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) at Week 104

H5: secukinumab 150 mg is not different from GP2017 40 mg with respect to proportion of subjects with a syndesmophyte at baseline with no new syndesmophytes at Week 104

H6: secukinumab 300 mg is not different from GP2017 40 mg with respect to proportion of subjects with a syndesmophyte at baseline with no new syndesmophytes at Week 104

Figure 2-1 Testing strategy for the second family after both H1 and H2 have been rejected



The family-wise error rate will be controlled strongly at α=5% two-sided level with the proposed hierarchical testing strategy. In the first family, the Hochberg procedure will reject H1 and H2, if both of them are significant at level α simultaneously. Otherwise, it will reject H1 if it is significant at level α/2, or reject H2 if it is significant at level α/2. Because H1 and H2 represent two doses of secukinumab versus GP2017, the correlation between the two test statistics will be positive. In this case, the type I error is controlled for the Simes test (Sarkar and Chang 1997) and, thus, the family-wise error rate is controlled strongly for the Hochberg procedure. Once both primary hypotheses within the first family are rejected, then the α will be passed to the second family starting with each of the hypotheses H3 and H4 tested simultaneously at α/2. Then based on the rejection of one or both of these hypotheses (H3 and H4), the next endpoint will be tested hierarchically for each dose (through H5 and/or H6), respectively. This procedure will continue (pending rejection of the null hypotheses) until H5 and/or H6 are/is rejected respectively; then the respective α/2 can be passed on to the other secukinumab regimen's hierarchy of hypotheses in the second family (i.e., H3 or H4), if they are not already rejected at α/2. Of note, in the description above, rejection of a hypothesis refers to rejection of the two-sided hypothesis; however, the level of a rejected hypothesis is only passed on according to the graphical procedure for the test of another hypothesis if the treatment effect is in favor of secukinumab.

2.6.3 Handling of missing values/censoring/discontinuations

Missing Week 104 change from baseline values in mSASSS for a given subject will be imputed using an imputation model based on the observed data of subjects in the same randomized arm who also discontinued study treatment but for whom Week 104 X-ray data are available ('retrieved dropout approach', EMA 2010). Recent literature suggests that in AS patients treated with TNF inhibitors, the spinal radiographic progression followed a linear course during the first 4 years of follow-up (Maas, et. al. 2017). Therefore, a parametric model assuming a normal distribution and a linear time trend for the mean scores will be used to impute missing Week 104 values. The model will be fitted to the average of the three readers' change from baseline in mSASSS.

We will consider a mixed-effects model for the change in mSASSS functional scale measurement. Let y_{ij} be the j^{th} change from baseline measurement of subject i, at time t_{ij}:

$$y_{ij} = \alpha_i t_{ij} + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim N(0, \sigma^2).$$

The subject specific intercept α_i will be modelled using treatment and the subject's baseline mSASSS value as fixed effects and assuming a subject-specific normally distributed random effect.

PROC MCMC will be used for fitting the mixed-effects model and multiple imputation of the Week 104 value, with weakly informative priors and model code as shown below,

random gamma ~ normal(0, sd=sd_re) subject=usubjid;

prior beta1 ~ normal(0,sd=100);

prior beta2 ~ normal(0,sd=100);

prior beta3 ~ normal(0,sd=100);

prior beta4 ~ normal(0,sd=100);

prior beta5 ~ normal(0,sd=100);

prior sd_re ~ uniform(0,20);

prior sd ~ uniform(0,20);

mu = (beta1*trt_GP2017 + beta2*trt_SEC150 + beta3*trt_SEC300 + beta4*base + beta5*sq_base_stand + gamma)*ady/365.25;

model chg ~ normal(mu, sd=sd);

where trt_GP2017 – binary indicator for GP2017 40 mg

trt_SEC150 – binary indicator for secukinumab 150 mg

trt_SEC300 – binary indicator for secukinumab 300 mg

base – baseline mSASSS

sq_base_stand – square root of baseline mSASSS (standardized)

gamma – random effect

ady – number of days when X-ray assessment was done after randomization

chg – change from baseline in mSASSS at Week 104

For absence of new syndesmophytes at Week 104, the handling of missing data will be the same as the primary endpoint.

For MRI endpoints, if values are missing for the baseline or post-baseline visit, the subject will be excluded from the analysis.

Multiple imputation (MI) approach under missing-at-random (MAR) assumptions will be applied to handle missing data for ASAS20, ASAS40, ASAS partial remission and ASDAS inactive disease.

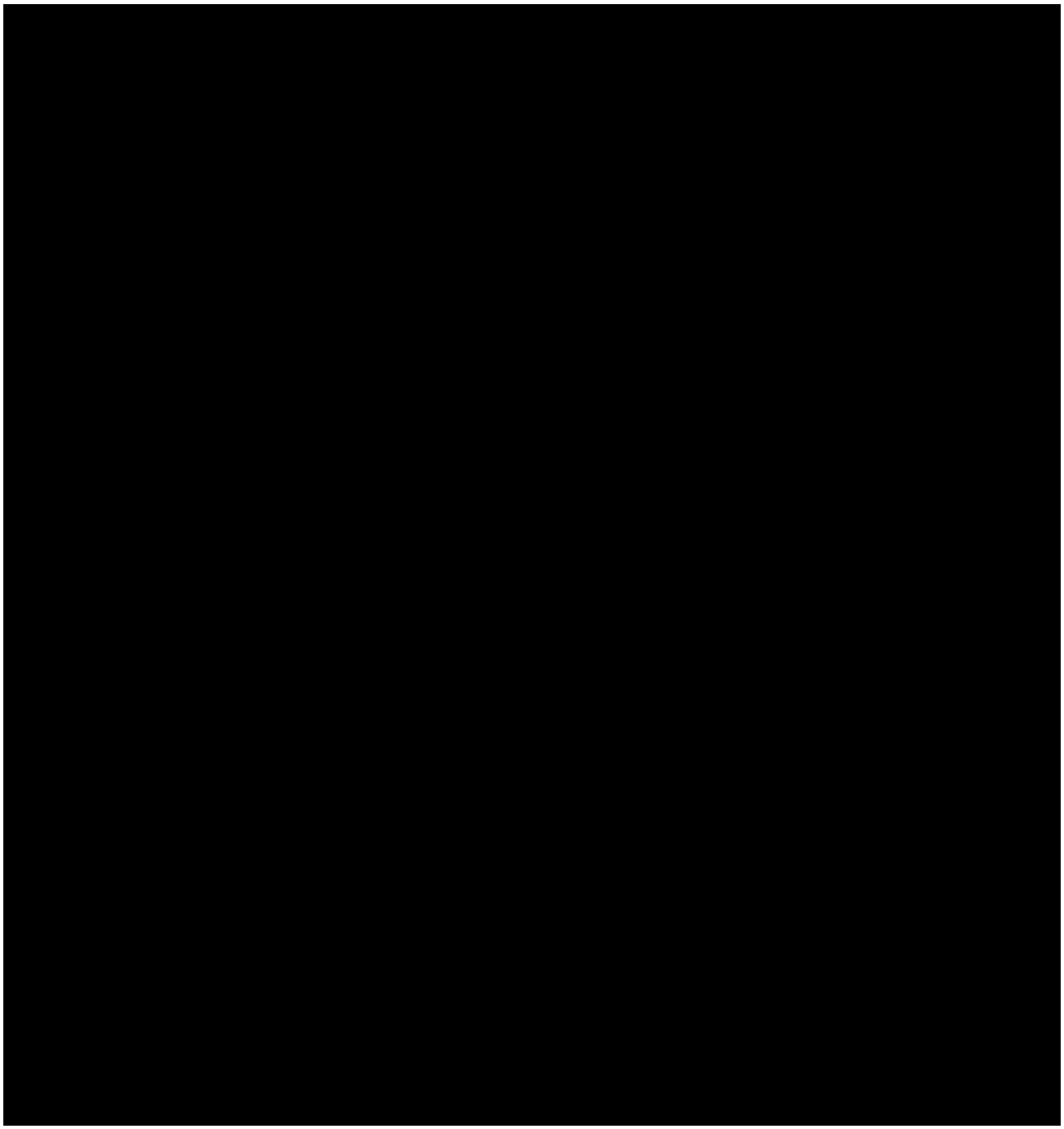
2.6.4 Supportive analyses

Sensitivity analyses and supplementary analyses will be conducted in order to provide evidence that the results seen from the key secondary objectives are robust.

For the change from baseline in mSASSS at Week 104 a sensitivity analyses will be conducted to determine the impact of missing data handling on the primary estimand by changing the covariates in the imputation model (e.g., using different ways of adjusting for baseline mSASSS, adding additional covariates). Additional supplementary analyses may be conducted to examine different estimands which are also deemed clinically relevant, such as the estimand based on the hypothetical strategy which would examine the treatment effect in outcomes where biologic rescue was not available to subjects. In this case the data after the occurence of the intercurrent event would be imputed under the MAR assumption.

For the proportion of subjects with absence of new syndesmophyte at Week 104 the same sensitivity and supplementary analyses used in the primary objective will be done.

The effect of the COVID-19 pandemic on the study will be monitored and assessed. An additional supportive analysis related to its effect may be performed as described in Section 2.15.



2.8 Safety analyses

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 did not receive the treatment to which they were randomized (i.e., who erroneously received the wrong treatment at least once), an additional AE listing will be prepared displaying which events occurred after the treatment errors.

In addition for subjects who discontinue study treatment but continue with study participation, an additional AE listing will be prepared displaying which events occurred after the study treatment discontinuation.

2.8.1 Adverse events (AEs)

The crude incidence of treatment emergent adverse events (TEAEs) (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. In addition, exposure-adjusted incidence rates including 95% CI will be provided for the entire treatment period. A graphical display of the crude incidence rates and exposure-adjusted incidence 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 their incidence in total secukinumab group (combining all secukinumab treatment groups), 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, adverse events leading to discontinuation, and adverse events leading to temporary dose interruption.

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.

Non-treatment emergent adverse events will be listed.

For SAEs that 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, on-treatment laboratory tests and vital signs for each analysis period is described below.

Table 2-1 Overview of analyses on some safety endpoints

Analysis period	AEs & SAEs	AEs by severity	Study drug related AEs	AEs-SMQ	Risk	Notables for vital signs and lab criteria
Entire Treatment	crude incidence exposure-adjusted incidence	crude incidence	crude incidence	crude incidence	crude incidence exposure-adjusted incidence	crude incidence

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

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

2.8.2 Laboratory data

The summary of laboratory data will only include on-treatment data, which are defined as those lab assessments after the first dose of study treatment and on or before the 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 units. These laboratory data will be displayed in listings using the standard units 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, 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: 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 2-2 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	See note below
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: Grade 4 Hemoglobin events are defined as life-threatening anemia events and will not be displayed in the table, as a numerical range is not provided in the CTCAE.

* For “creatinine increased” the baseline criteria do not apply.

Shift tables will be presented comparing baseline laboratory result (CTCAE grade) with the worst results (expressed in CTCAE grade) during the treatment phase (either the initial phase up to Week 16 or the entire treatment phase) analyzed. Of note, baseline will be defined as the last assessment prior to first dosing in the 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 below:

Table 2-3 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 laboratory criteria**) 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.

Note:
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).

2.8.3 Other safety data

2.8.3.1 Vital signs

The summary of vital signs will only include on-treatment 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 below:

Table 2-4 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

2.8.3.2 Electrocardiogram (ECG)

Listings of ECG abnormalities at baseline will be provided. The following will be considered as notable values:

- QT > 500 msec;
- QTcF > 450 msec (males), QTcF > 460 msec (females);
- QTcF change from baseline > 30 msec, >60 msec;
- PR > 250 msec.

[redacted]

2.8.3.4 Compound specific safety evaluation

Safety topics of interest, such as risks defined in the Safety Profiling Plan (SPP), Risk Management Plan (RMP) 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 and preferred terms of the MedDRA dictionary will be considered, as defined in the Program Case Retrieval Sheet and RMP.

[redacted]

[redacted]

[redacted]

2.15 COVID-19 related supportive and exploratory analyses

The effect of the COVID-19 pandemic on the study will be monitored and assessed. For instance, tables comparing TEAEs, PDs, and other relevant safety endpoints by COVID-19 period, and week or visit as applicable, will be shown.

Supportive analyses considering hypothetical estimands in case patients have been affected by COVID-19, e.g., missed doses, missed assessment, etc., due to COVID-19 (to be determined via the COVID-19 PDs) will be considered, as needed. In such a case the data after the occurence of a COVID-19 related intercurrent event would be omitted from the analysis and imputed under the MAR assumption.

2.15.1 COVID-19 periods

The table below defines the start dates of the pandemic by regions.

Table 2-5 Start of the COVID-19 pandemic by region

Region/Country	Start Date	End Date
China	01-Jan-2020	End date has not yet been defined
South Korea	20-Feb-2020	End date has not yet been defined
Japan	21-Feb-2020	End date has not yet been defined
Rest of the World	01-Mar-2020	End date has not yet been defined

2.15.2 Treatment emergent adverse events

The COVID-19 related cases in the treatment emergent adverse events listing will be included to be able to assess the potential impact of COVID-19.

3 Sample size calculation

An overall type I error (2-sided) of 5% will be used to control type I error. A total of 279 subjects per group is deemed appropriate to achieve adequate power for the primary and secondary endpoints for this study.

Based on the modeling of data from an unpublished non-interventional study that compared a blinded re-read of radiographic x-ray images from the secukinumab phase III AS study CAIN457F2305 (MEASURE 1) (Braun 2016) and a historical cohort of AS subjects treated with NSAIDs only (ENRADAS), a no progression rate of 66% for secukinumab-treated subjects and 51% for NSAIDs-treated subjects at Week 104 were calculated. Published data for adalimumab-treated AS subjects compared to a cohort of AS subjects treated only with NSAIDs (OASIS) showed no difference in change from baseline in mSASSS at Week 104 (van der Heijde 2009). Therefore, the 51% no progression rate calculated from the subjects who took NSAIDs will be used for GP2017 (adalimumab biosimilar). Assuming 80% of subjects will complete the 2-year study treatment period and provide both baseline and Week 104 evaluable radiographic films with a 66% no progression rate at Week 104 and assuming subjects who discontinue study treatment will have a 51% no progression rate at Week 104, a new no progression rate of 63% for secukinumab was calculated using the definition of the estimand. Therefore, a no progression rate of 63% for secukinumab and 51% for GP2017 were used as assumptions to calculate the sample size using a two-group Chi-square test of equal proportion (nQuery Advisor 7.0) with approximately 73% power at α=0.025. And the power to reject at least one hypotheses using the Hochberg procedure is 88%.

Analysis of an unpublished Phase III study showed a change from baseline in mSASSS at Week 104 with a mean of 0.52 and standard deviation of 2.5 for secukinumab subjects. With a sample size of 279 subjects in each treatment group, a treatment difference of 0.58 (i.e., assuming GP2017 has a mean of 1.1 and the same standard deviation) yields approximately 68% power at $\alpha = 0.025$.

Analysis of an unpublished Phase III study showed that 68% of randomized AS subjects had one or more syndesmophytes at baseline, and for this subset, the proportion of subjects with no new syndesmophyte at Week 104 was 70% for secukinumab. With a sample size of 190 subjects in each treatment group (about 68% of the planned randomized 279 subjects), a treatment difference of 15% (i.e., assuming a GP2017 response rate of 55%) yields approximately 78% power at $\alpha = 0.025$.

4 Change to protocol specified analyses

No change from protocol specified analysis was made.

5 Appendix

5.1 Description of efficacy variables

X-ray

X-rays will be obtained as defined in the schedule of assessments in the protocol and according to the imaging acquisition guidelines provided by the central imaging vendor. They will be analyzed according to the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS) [redacted].

The maximum mSASSS score for vertebral sites in the cervical vertebra is 36, and in the lumbar verterba is 36, resulting in a total maximum mSASSS of 72.

[redacted]

If a joint or bone is not visible (e.g. poor film quality, missing imaging, severe misalignment, flexion deformity, dislocation) at the timepoint, the individual joint or bone will be coded as 'Not Visible' (N). If radiographs at the timepoint show a joint or bone with surgical fusion, replacement (prosthesis), or amputation, then the joint or bone will be scored 'Surgically Modified' (S). For joints and bones with end stage disease, scores 'N' and 'S' should not be used. Any 'N' or 'S' will be considered null in the calculation of the total score.

Only scores of radiographs with ≤ 3 missing vertebral unit (VU) per segment (either cervical or lumbar) will be used. There are 12 VUs per segment. [redacted]

Missing post-baseline value

1. A missing value for a post-baseline VU will be replaced with the nearest value available. For Week 52: if both baseline VU and Week 104 VU are available, the baseline VU value will be used.
2. The mean spinal segment's progression score (either cervical or lumbar) per subject will be calculated.
3. The mean spinal segment's progression score will be added to the imputed post-baseline value in #1.

Missing baseline value

1. A missing baseline score of a VU will be replaced with the value of the post-baseline score.
2. The mean spinal segment's progression score (either cervical or lumbar) per subject will be calculated.
3. The mean spinal segment's progression score will be subtracted to the imputed baseline value #1.

Missing both baseline and post-baseline value

1. If both baseline and post-baseline value are missing for the VU, then the average of the other available VUs from this spinal segment at baseline will be used to replace the missing VUs at baseline.
2. Then follow the steps for imputing missing post-baseline value.

Status scores refer to the score in each of the available time points (at baseline and post-baseline). Progression scores are calculated as the difference between the status scores of two time points. Two-year progression scores refer to the progression occurring within 2 years, that is, status score at Week 104 minus the status score at baseline.

The readings of the X-rays and the scoring will be performed centrally. Three central independent radiograph readers, all blinded to subject's identity, treatment arm, and radiograph sequence, will analyze the digitized images. The statistical analysis will use the average score from the three readers. No adjudication will be performed.

No radiographic progression in mSASSS

No radiographic progression in mSASSS is defined as a change from baseline in mSASSS at Week 104 of ≤ 0.5. While the primary analyses will be based on this definition, additional analyses using ≤ 0 and < 2 will be performed.

Syndesmophyte

A syndesmophyte will be defined as a score of ≥ 2 for any individual vertebral edge within evaluable vertebral units. A new syndesmophyte is defined as an individual vertebral edge with a score of 0 or 1 at baseline that changes to a score of 2 or 3 at Week 104.

MRI

MRI will be performed at selected sites in a subgroup of approximately 30% of all randomized subjects. MRI will be performed to assess SI and spinal inflammation using a scoring system for quantification of AS-related pathologies, to investigate whether these changes are affected by treatment with secukinumab or GP2017. The images will be obtained as defined in the schedule of assessments in the protocol and according to the imaging acquisition guidelines provided by the central imaging vendor.

For the Berlin SI joint edema score, the maximum score is 24. For the ASspiMRi-a Berlin modification score, the maximum score is 69.

For the ASspiMRI-a Berlin modification score the maximum score for VUs in the cervical spine is 18, in the thoracic spine is 36, and in the lumbar spine is 15. Therefore, the total maximum Berlin score for all scored VUs is 69.

If a joint or bone is not visible (e.g., poor image quality, missing imaging, etc.) at a timepoint, the individual joint or bone will be coded as 'Not Visible' (N). If MRI scans at a timepoint show a joint or bone with surgical alteration then the joint or bone will be scored 'Surgically Modified' (S). For joints and bones with end stage disease, scores 'N' and 'S' should not be used. Any 'N' or 'S' will be considered null in the calculation of the total score.

Assessment of SpondyloArthritis International Society criteria (ASAS)

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

ASAS domains:

1. Subject's global assessment of disease activity measured on a VAS
2. Subject's assessment of back pain, represented by either total or nocturnal pain scores, both measured on a VAS scale
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

ASAS Response Criteria 20% (ASAS 20)

ASAS 20 response is defined as an improvement of ≥ 20% and ≥ 1 unit on a scale of 10 in at least three of the four domains and no worsening of ≥ 20% and ≥ 1 unit on a scale of 10 in the remaining domain.

ASAS Response Criteria 40% (ASAS 40)

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 domains and no worsening at all in the remaining domain.

ASAS partial remission criteria

The ASAS partial remission criteria are defined as a value not above 2 units in each of the four domains on a scale of 10.

Subject's global assessment of disease activity (VAS)

The subject'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?*".

Subject's assessment of back pain intensity (VAS)

The subject'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 10 questions were chosen with 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 0 through 10 scale (captured as a continuous

VAS) is used to answer the questions. The mean of the ten scales gives the BASFI score – a value between 0 and 10. If there are missing questions the average of the non-missing items is used (Braun 2009, van Tubergen 2001).

Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

The BASDAI consists of a 0 through 10 scale (0 being no problem and 10 being the worst problem, captured as a continuous VAS), 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 (called enthesitis, or inflammation of tendons and ligaments)
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. The mean of questions 5 and 6 is added to the scores from questions 1-4. 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 studies evaluating new drug therapies directed at AS. BASDAI is a quick and simple index taking between 30 seconds and 2 minutes 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 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 the sum should be divided by 3.

High sensitivity C-reactive protein (hsCRP)

This assessment will be performed in order to identify the presence of inflammation, to determine its severity, and to monitor response to treatment.

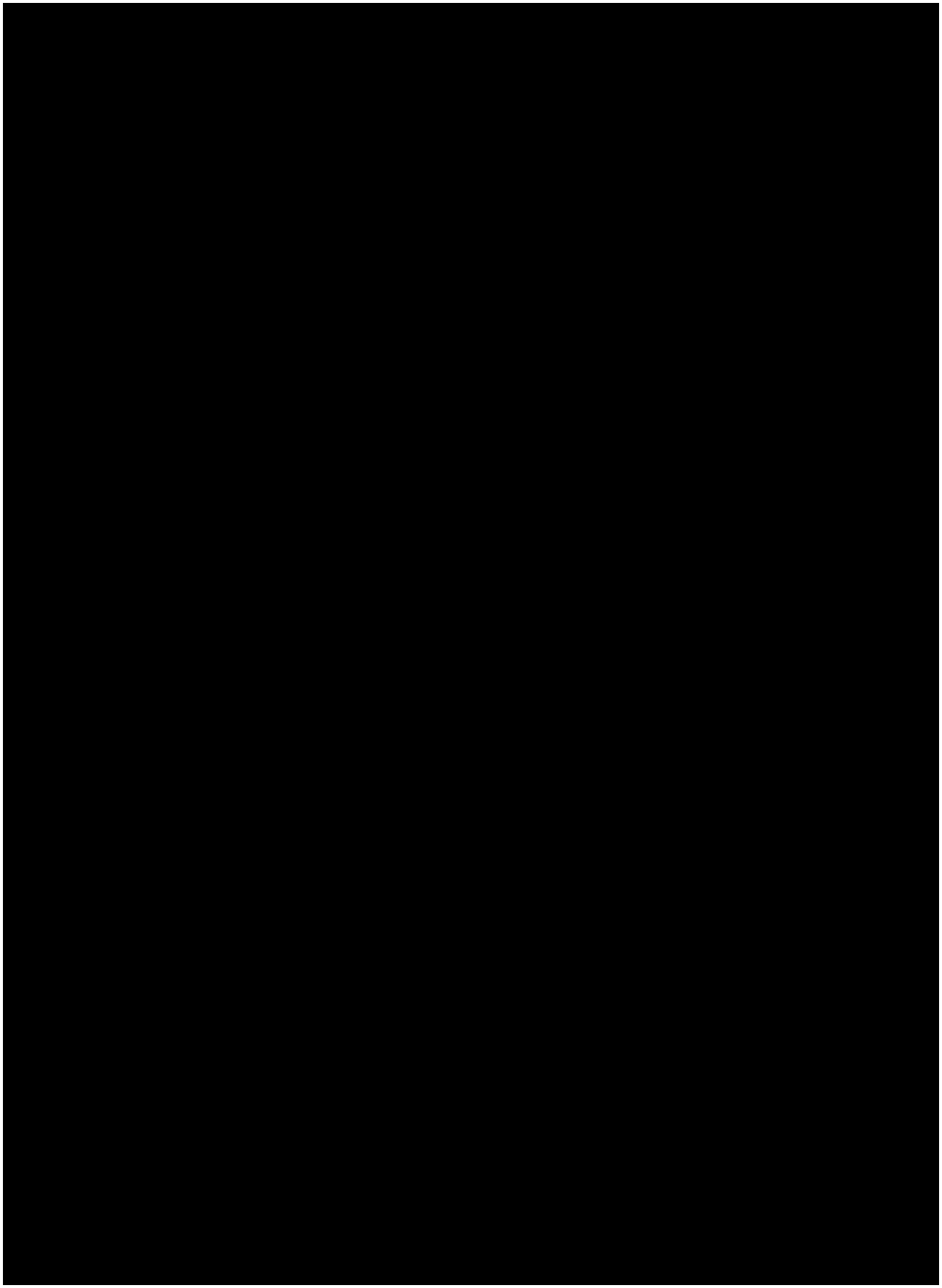
ASDAS-CRP and ASDAS response categories

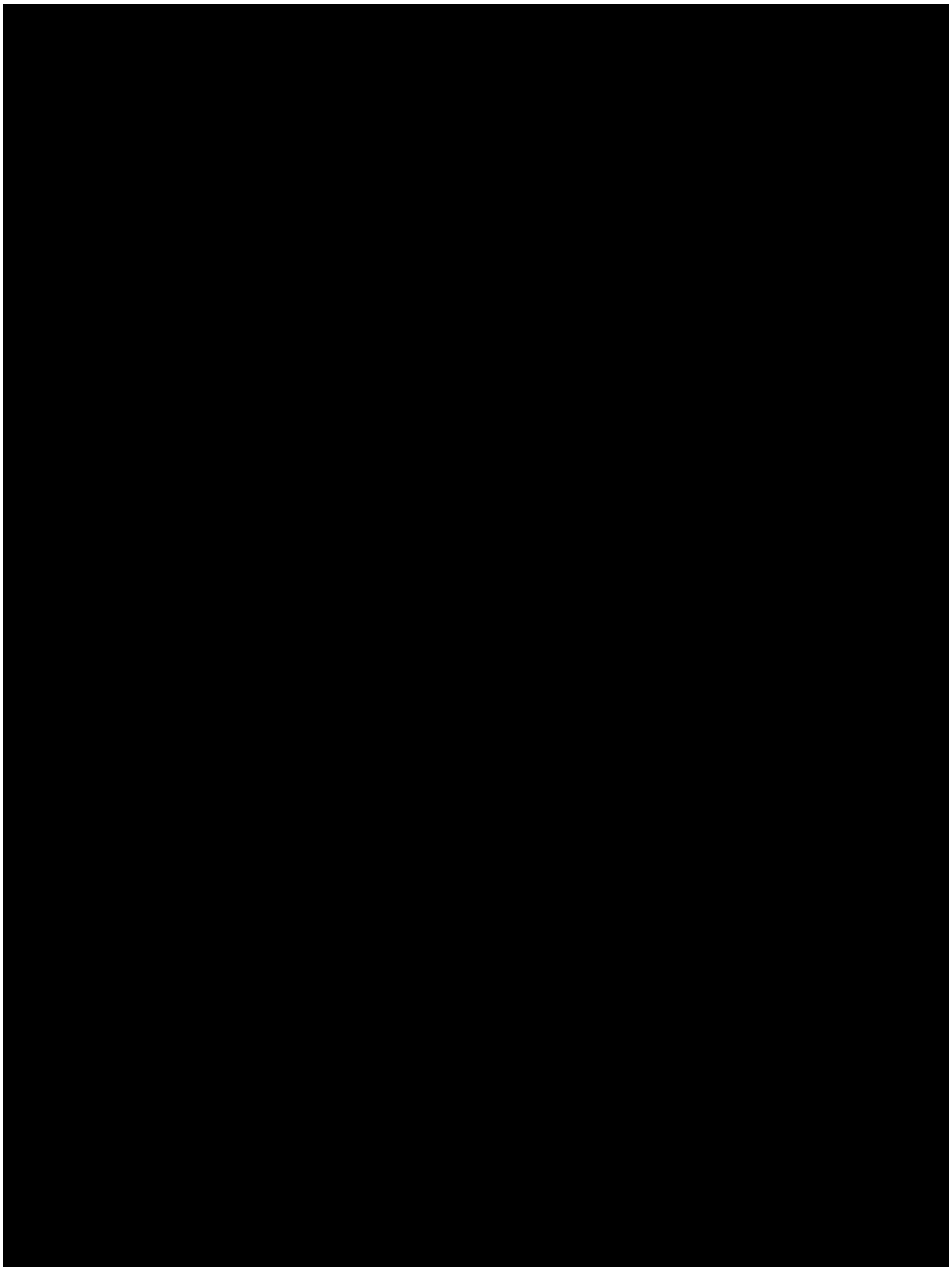
The Ankylosing Spondylitis Disease Activity Score (ASDAS) is a composite index to assess disease activity in AS.

The ASDAS-CRP (Ankylosing Spondylitis Disease Activity Score) will be utilized to assess the disease activity status. Parameters used for the ASDAS include spinal pain (BASDAI question 2), the subject's global assessment of disease activity, peripheral pain/swelling (BASDAI question 3), duration of morning stiffness (BASDAI question 6) and C-reactive protein (CRP) in mg/L (Sieper 2009, Lukas 2009).

Disease activity states are inactive disease, moderate disease activity, high disease activity, and very high disease activity. The 3 values selected to separate these states were < 1.3 between inactive disease and moderate disease activity, < 2.1 between moderate disease activity and high disease activity, and > 3.5 between high disease activity and very high disease activity.

Selected cutoffs for improvement scores were a change ≥ 1.1 unit for "minimal clinically important improvement" and a change ≥ 2.0 units for "major improvement" (Machado 2011).





5.3 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 laboratory and 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 (W104) 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 5-3 Analysis visit windows

Analysis Visit	Target Day	Analysis Visit Window	Group 1	Group 2	Group 3	Group 4	Group 5
Baseline	1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 30	≤ 7
Week 4	29	2-57	2-57				
Week 8	57						
Week 12	85	58-99	58-99				
Week 16	113	100-141	100-141	2-239	2-239		54-173
Week 20	141						
Week 24	169	142-225	142-225				
Week 28	197						
Week 32	225						
Week 36	253						
Week 40	281	226-323	226-323				
Week 44	309						
Week 48	337						
Week 52	365	324-407	324-407	240-547	240-407	276-455	276-455
Week 56	393						

Analysis Visit	Target Day	Analysis Visit Window	Group 1	Group 2	Group 3	Group 4	Group 5
Week 60	421						
Week 64	449	408-491	408-491		408-491		
Week 68	477						
Week 72	505						
Week 76	533	492-589	492-589		492-589		
Week 80	561						
Week 84	589						
Week 88	617						
Week 92	645	590-687	590-687		590-687		
Week 96	673						
Week 100	701						
Week 104	729	688-785	688-785	548-785	688-785	640-849	640-819

Group 1: Subject's global assessment of disease activity (VAS), Subject's assessment of back pain intensity (VAS), BASFI, BASDAI, [redacted], hsCRP, vital signs, hematology, blood chemistry, urinalysis

[redacted]

Group 3: Weight
Group 4: Lipids, X-ray
Group 5: MRI

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 5-4 Rules for flagging variables

Timing of measurement	Type of data	Rule
Baseline	All data	Baseline is defined as the last available measurement recorded on or before the reference start date (only date part)/Day 1. If there are multiple assessments on Day 1, the following rules apply: 1. If assessment time exists: • select the last available measurement prior to reference start date/time considering time • if no measurement prior to reference start date/time considering time, select the earliest measurement post reference start date/time considering time 2. If assessment time does not exist: select the available measurement from the lowest CRF visit number
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 subject switches treatment (e.g., from placebo to AIN) within the window the following rules apply: • If available, the closest measurement to the target date which is on or before the switch date will be used

Timing of measurement	Type of data	Rule
		• If there are no data on or before the switch date then the closest measurement after the switch to target will be used 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
Post-baseline safety	Summary visit information (e.g., lab, 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 subject switches treatment (e.g., from placebo to AIN) within the window the following rules apply: • 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 date then the closest measurement after the switch to target will be used 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
Post-baseline safety	Notable abnormalities (e.g., lab)	The most extreme measurement in the window will be used. Note this means a subject can have a notably high and notably low measurement within a window

5.4 Statistical methodology and assumptions

5.4.1 Analysis of continuous data

5.4.1.1 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.

5.4.1.2 Non-parametric analysis of covariance

Certain continuous endpoints will be analyzed with an ANCOVA model. The model will include factors and covariates as specified for the respective analyses. The SAS code below outlines a template for the analysis where factors and covariates can be added or removed as required.

SAS code example:

```
proc mixed data=mydata;
        class TRT01PN;
        model CHG = TRT01PN BASE / s ddfm=kr;
        lsmeans TRT01PN / diff cl;
run;
```

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.

5.4.2 Analysis of binary and categorical data

5.4.2.1 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

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

Note: If L > 100 xp then L = 100 x p and if U < 100 x p then U = 100 x p.

5.4.2.2 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 marginal standardization method will be used to calculate the mean response rate in each treatment group as well as their difference. This method uses the same fitted logistic model, but involves using the model to predict, for each patient in the study, the mean outcome assuming assignment to each particular treatment group in turn, assuming each patient's

observed values for the other baseline covariates (i.e., baseline score). Averaging these predictions for each treatment group provides the estimate of the mean response rate for each treatment group. Then the difference will be derived based on the estimated mean response rates comparing secukinumab and GP2017 arms. The SAS code below outlines a template for the analysis where covariates can be added or removed as required:

The macro Margins (Predictive margins and average marginal effects) will be used.

SAS code example as the following,

```
%Margins(data = mydata,
        class = treatment,
        response = response,
        roptions = event='1',
        model = treatment baseline_score,
        dist = binomial,
        margins = treatment,
        options = cl diff)
```

For cases where the convergence status indicates that the model did not reach appropriate convergence (conv_status is not 0), no risk difference or p-value will be presented from that model.

However, if the issue relates to the primary timepoint of Week 104 then the following steps will be followed:

1. Remove *baseline_score* from the model. If there are still issues perform step 2
2. Use Fisher's exact test as described below.

When Fisher's exact test is applied, only a p-value for a test of equal response in the two groups can be obtained (no risk difference or confidence intervals can be estimated.)

```
ods output fishersexact=fisher;
proc freq data=mydata;
    by visit;
    table treatment*response / fisher;
run;
```

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

5.4.3 Imputation methods

5.4.3.1 Tipping point

The goal of the tipping point analysis is to identify assumptions about the missing data and presumed non-response under which the conclusions change, i.e., under which there is no longer evidence of a treatment effect. Then, the plausibility of those assumptions can be discussed.

For binary variables

Defining all cases with missing data at a specific visit as uncertain cases (i.e., the subject could be either a responder or a non-responder) the following notations are made for a comparison of two treatment regimens:

R_i: Number of observed responders from subjects randomized to regimen *i*

NR_i: Number of observed non-responders from subjects randomized to regimen *i*

M_i: Number of uncertain response cases from subjects randomized to regimen *i*

Where *i*=1, 2 denotes the two regimens to be compared.

A Chi-square test can now be performed comparing the two regimens for each possible combinations of uncertain response. Table 16-3 shows an outline of all the possibilities for the comparisons, where J_i takes values from 0 to M_i for regimen *i*.

Table 5-5 Counts in tipping point analysis

Response	Regimen 1 (N=N_1)	Regimen 2 (N=N_2)
Yes (responder)	$R_1 + J_1$	$R_2 + J_2$
No (non-responder)	$NR_1 + (M_1-J_1)$	$NR_2 + (M_2-J_2)$

SAS code for Chi-square test of a specific assignment of uncertain cases:

ods output chisq=*chisquare*;

proc freq data=;

table *treatment**response* / chisq;

weight *n*;

run;

The input dataset should only include the two regimens to be compared. Number of subjects included in each of the four cells of Table 16-3 (e.g., $R_1 + J_1$) is denoted by *n*.

In order to also take covariates into consideration a logistic regression model may also be performed for selected combinations of response distributions (e.g. worst case scenario where J_1=0 and J_2=M_2).

For continuous variables

Tipping point to compare each AIN group with the GP2017 group will be performed as follows:

1) For any patient i randomized to the AIN group with an imputed mSASSS value resulting in a change from baseline value in mSASSS at Week 104 $CB_{i,mSASSS}$, let $CB_{i,mSASSS}^{tipping} = CB_{i,mSASSS} + d$
2) For any patient j randomized to the GP2017 group with an imputed mSASSS value resulting in a change from baseline value in mSASSS at Week 104 $CB_{j,mSASSS}$, let $CB_{j,mSASSS}^{tipping} = CB_{j,mSASSS} + b$
3) Let d = -1, -2/3, -1/3, …, 2 and b = -1, -2/3, -1/3, …, 2

4) For each combination, repeat the same ANCOVA analysis with the same model used in the primary analysis and the newly computed changes from baseline
5) Identify the values of d and b that reverts significance
6) A different range of values for d and b can be used if no tipping point was identified or if it helps the clinical interpretation

5.4.3.2 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=4572340 nimpute=100 out=*mi_out*;

by *treatment*;

fcs reg (/details);

var *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.

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, logistic regression with Margins macro) 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.

For the datasets generated with the PROC MCMC routine, the same analysis as described in this section will be performed. PROC MCMC will be run twice with 100000 iterations, a thinning of 1000, and burn-in of 10000 in order to obtain 200 partially completed datasets. 100 datasets will have imputed data for subjects who discontinued the study, and the other 100 datasets will have imputed data for subjects with missing data for other reasons (e.g., due to an incomplete number of interpretable locations). The two sets of 100 partially completed datasets will be patched together matching their iteration number. That is, the datasets corresponding to the i-th iteration of the first 100 datasets will be merged with the datasets corresponding to the i-th iteration of the second 100 datasets. The SAS code below outlines a template for the procedure:

```
PROC MCMC data= seed=4572340 nbi=10000 nmc=100000 thin=1000 missing=CCMODELY;
    parms beta1 beta2 beta3 beta4 beta5;
    prior beta1 beta2 beta3 ~ normal(0,var=2);
    prior beta4 beta5 ~ normal(0,var=3);
    pi = logistic(beta1 * trt_GP2017 + beta2 * trt_SEC150 + beta3 * trt_SEC300 + beta4 * sqrt(base_stand) + beta5 * slope_stand);
    model y ~ binary(pi);
    preddist outpred=;
    run;
```

Analogous codes will be used to impute the missing data due to other reasons (other than study discontinuation) as well as the change from baseline in mSASSS at Week 104 (secondary endpoint).

5.4.4 Crude incidence and related risk estimates

5.4.4.1 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

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

Note: 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.

6 Reference

Braun J, Pham T, Sieper J, Davis J, van der Linden Sj, Dougados M, van der Heijde D (2003) International ASAS consensus statement for the use of anti-tumour necrosis factor agents in patients with ankylosing spondylitis. Ann Rheum Dis.; 62: 817-824

Braun J, McHugh N, Singh A, Wajdula JS, Sato R (2009) Improvement in patient-reported outcomes for patients with ankylosing spondylitis treated with etanercept 50 mg once-weekly and 25 mg twice-weekly. *Rheumatology (Oxford)*; 46: 999-1004

Braun J, Baraliakos X, Deodhar A, et al (2016) Effect of secukinumab on clinical and radiographic outcomes in ankylosing spondylitis: 2-year results from the randomised phase III MEASURE 1 study. Ann Rheum Dis; 0:1–8. doi:10.1136/annrheumdis-2016-209730.

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

[redacted]

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

[redacted]

[European Medicines Agency, Committee for Medicinal Products for Human Use (2010)] CHMP Guideline on Missing Data in Confirmatory Clinical Trials. EMA/CPMP/EWP/1776/99 Rev1.

Haywood KL, et al. (2002) Disease-specific, patient-assessed measures of health outcome in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology; 41:1295-1302.

Hochberg Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika; 75(4):800-802.

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.

[redacted]

Lukas C, Landewé R, Sieper J (2009) Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis; 68(1):18-24.

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

Maas F (2017) Reduction in Spinal Radiographic Progression in Ankylosing Spondylitis Subjects Receiving Prolonged Treatment With Tumor Necrosis Factor Inhibitors. Arthritis Care & Research; 69(7):. 1011–1019.

Machado PM, Landewé RB, van der Heijde DM (2011) Endorsement of definitions of disease activity states and improvement scores for the Ankylosing Spondylitis Disease Activity Score: results from OMERACT 10. J Rheum; 38(7):1502-06.

[redacted]

Predictive margins and average marginal effects: http://support.sas.com/kb/63/038.html

Ramiro et al. (2013) Scoring radiographic progression in ankylosing spondylitis: should we use the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) or the Radiographic Ankylosing Spondylitis Spinal Score (RASSS)? Arthritis Research & Therapy ; 15:R14

[redacted]

Sarkar SK & Chang CK. (1997) The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics. Journal of the American Statistical Association; 92(440): 1601-1608.

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

van der Heijde D, Salonen D, Weissman BN, et al (2009) Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. Arthritis Res & Therapy; 11:R127.

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

Zink RC and Koch GG (2012). NParCov3: A SAS/IML Macro for Nonparametric Randomization-Based Analysis of Covariance. Journal of Statistical Software; 50(3): 1-17.