NCT05018806

STATISTICAL ANALYSIS PLAN

Protocol title:	**A Phase 2, randomized, double-blind, placebo-controlled, multicenter proof-of-concept study evaluating efficacy and safety of rilzabrutinib in adult patients with moderate-to-severe atopic dermatitis who are inadequate responders or intolerant to topical corticosteroids**
Protocol number:	**ACT17207**
Compound number (INN/Trademark):	**SAR444671** **Rilzabrutinib/Trademark not available**
Study phase:	**Phase 2**
Short Title:	**Proof of concept study of rilzabrutinib in adult patients with moderate-to-severe atopic dermatitis**
Statistician:	[redacted]
Statistical project leader:	[redacted]
Date of issue:	**27-Oct-2022**
Regulatory agency identifier number(s):	
IND:	134595
EudraCT:	2021-001704-15
NCT:	NCT05018806
WHO:	U1111-1261-7565
EUDAMED:	Not applicable
Other:	Not applicable

Date: 27-Oct-2022 Total number of pages: 71

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

This Statistical Analysis Plan (SAP) for study ACT17207 is based on Protocol Amendment 03 dated 24-Aug-2022. This section summarizes major changes to the statistical analysis features in the SAP.

The first participant was randomized on 06-Oct-2021. This SAP is approved before the analyses for BID (twice a day) cohort is conducted.

Table 1 - Major changes in statistical analysis plan

SAP Version	Approval Date	Changes	Rationale
1	Current version	The use of rescue medications considered as intercurrent events were changed from high potency TCS or systemic medication defined in protocol to any potency TCS (including low to medium potency TCS and high potency TCS), Crisaborole, Calcineurin inhibitors or systemic medication or phototherapy in the statistical analysis	The use of other rescue medications could also confound efficacy assessments.

1 INTRODUCTION

1.1 STUDY DESIGN

This is a global, multicenter, Phase 2, double-blind, 2-arm, parallel treatment, placebo-controlled randomized proof-of-concept (PoC) study with 2 staggered cohorts (2 arms in each cohort) to evaluate the efficacy and safety of rilzabrutinib in adult participants (aged at least 18 years) with moderate-to-severe AD and intolerance or inadequate response to topical corticosteroids (TCS).

Participants will be enrolled into 2 staggered dose regimen cohorts, twice a day (BID) cohort and three times a day (TID) cohort, starting with the BID cohort. These 2 dose regimen cohorts will be evaluated with the same double-blind fashion and planned assessments in the study.

After a screening phase of up to 4 weeks, participants in the main study will be centrally randomized (using permuted block randomization schedule) via Interactive Response Technology (IRT) in a 3:2 ratio to 1 of the 2 intervention groups (rilzabrutinib and matching placebo) and treated double-blind for approximately 16 weeks. Randomization will be stratified by immunoglobulin E (IgE) levels (< or ≥ 300 UI/mL) at screening. Approximately 120 participants in total will be randomized in a 3:2 ratio to receive either rilzabrutinib or matching placebo on Day 1 (Week 0, Visit 2) in the study. In the BID cohort, approximately 50 participants will receive either rilzabrutinib (n = 30) or matching placebo (n = 20) twice a day, and in the TID cohort, approximately 70 participants will receive either rilzabrutinib (n = 42) or the matching placebo (n = 28) three times a day.

[redacted]

[redacted]

The study duration consists of the following periods:

- Screening: up to 4 weeks
- Double-blind investigational medicinal product (IMP) treatment period: 16 weeks ± 3 days
- Post-treatment follow-up: 1 week ± 3 days.

1.2 OBJECTIVES AND ENDPOINTS

Table 2 - Objectives and endpoints

Objectives	Endpoints
Primary	
• Assess the efficacy of rilzabrutinib in participants with atopic dermatitis (AD)	• Percentage change from baseline to Week 16 in Eczema Area and Severity Index (EASI) score.
Secondary	
• Assess the efficacy of rilzabrutinib at different time points	• Proportion of participants with Investigator's Global Assessment (IGA) of 0 or 1 (disease free or almost disease free) compared to placebo at Week 16. • Proportion of participants achieving EASI-75 (reduction of EASI score by ≥ 75% from baseline) at Week 16. • Proportion of participants with reduction of weekly average of daily peak pruritus Numerical Rating Scale (PP-NRS) of ≥ 4 points from baseline at Week 16. • Time to onset of effect on pruritus (≥ 4 points reduction of weekly average of daily PP-NRS from baseline during the 16-week treatment period). • Absolute change from baseline to Week 16 in EASI score • Proportion of participants achieving EASI-50/90 (reduction of EASI score by ≥ 50% or ≥ 90% from baseline) at Week 16. • Change from baseline to Week 16 in percent body surface area (BSA) of EASI. • Change (absolute and percent change) on weekly average of daily PP-NRS based on daily participant assessments documented in their electronic diary (e-diary) from baseline to Week 16. • Proportion of participants achieving IGA*BSA-50/75/90 (reduction of IGA*BSA by ≥ 50% or 75% or 90% from baseline) at Week 16.
• Assess the safety of rilzabrutinib	• Incidence of treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), and adverse events of special interest (AESIs) from baseline to Week 16. • Incidence of study investigational medicinal product (IMP) discontinuation and withdrawals due to TEAEs from baseline to Week16.

[redacted]

Objectives	Endpoints
[redacted]	

1.2.1 Estimands

The primary estimands defined for main endpoints are summarized in Table 3 below. More details are provided in Section 3.

The comparison of interest is SAR444671 individual dose vs. matching placebo.

Table 3 - Summary of primary estimand for main endpoints

Endpoint Category	Estimands			
	Endpoint[a]	Population	Intercurrent event(s) handling strategy	Population-level summary (Analysis and missing data handling)
Primary objective: Assess the efficacy of rilzabrutinib in participants with AD				
Primary endpoint	Percent change from baseline to Week 16 in EASI score	ITT	The intercurrent events will be handled as follows: • Discontinuation of study intervention: all data collected after discontinuation will be included in the analysis (treatment policy strategy). • Receiving selected rescue medications[b] before Week 16: data collected post the use of rescue treatments will be set to missing values and the worst postbaseline value on or before the time of the rescue treatments usage will be used to impute the missing Week 16 value (for participants whose postbaseline value are all missing, the baseline will be used to impute) (hypothetical strategy).	Mean percent change difference between rilzabrutinib and placebo from ANCOVA. The missing data are handled as follows: • After discontinuation of the study intervention due to lack of efficacy prior to Week 16: "worst observation carried forward (WOCF)" approach will be used to impute missing data if needed. • After discontinuation of the study intervention due to reasons other than lack of efficacy prior to Week 16: multiple imputation (MI) approach will be used to impute missing endpoint value, and this MI will use all participants excluding participants who have been rescued with selected rescue medications[b] before Week 16 and excluding participants who discontinue due to lack of efficacy prior to Week 16. Each of the imputed complete data will be analyzed by fitting the ANCOVA model with percent change in EASI score from baseline to Week 16 as response variable, study intervention (rilzabrutinib and placebo) and randomization stratification of screening IgE level (< or ≥300 UI/mL) as fixed effects, and baseline EASI score as covariates. Statistical inference obtained from all imputed data will be combined using Rubin's rule. Descriptive statistics

Endpoint Category	Estimands			
	Endpoint[a]	**Population**	**Intercurrent event(s) handling strategy**	**Population-level summary (Analysis and missing data handling)**
				including the number of subjects, mean, standard error, and LS means will be provided. In addition, difference in LS means and the corresponding 90% CI will be provided along with the p-values.
Secondary objective: Assess the efficacy of rilzabrutinib at different time points				
Secondary endpoint	Proportion of participants with Investigator's global Assessment (IGA) of 0 or 1 (disease free or almost disease free) compared to placebo at Week 16.	ITT	The intercurrent events will be handled as follows: • Discontinuation of study intervention before Week 16: Off-study intervention data up to Week 16 will be included in the analysis (treatment policy strategy). • Receiving selected rescue medications[b] prior to Week 16: Participants will be considered as non-responders (composite strategy).	CMH test adjusted by randomization stratification of screening IgE level (< or ≥300 UI/mL). The missing data at Week 16 will be handled as follows: Participants will be considered as non-responders.
	Time to onsite of effect on pruritus (≥4 points reduction of weekly average of daily PP-NRS from baseline during the 16-week treatment period).	ITT	The intercurrent events will be handled as follows: • Discontinuation of study intervention before Week 16: Off-study intervention data up to Week 16 will be included in the analysis (treatment policy strategy). • Receiving the selected rescue medications[b] prior to Week 16: Analyses will be censored at Week 16 (composite strategy).	Hazard ratio (HR) between rilzabrutinib and placebo, estimated using Cox proportional hazard model stratified by screening IgE level (< or ≥300 UI/mL). The missing data will be handled as follows: Analyses will be censored at the time of last PP-NRS assessment if participants discontinue the study follow-up before Week 16.

a Additional secondary objective/endpoints are not included in this table but would be handled with a similar strategy as the endpoint type (ie. continuous, proportion, time-to-event).

b Selected rescue medications and/or prohibited procedures are listed in Table 5

2 ANALYSIS POPULATIONS

The following populations for analyses are defined.

Table 4 - Populations for analyses

Population	Description
Screened	All participants who signed the ICF.
Randomized	All participants from screened population who have been allocated to a randomized intervention by IRT regardless of whether the intervention was received or not.
Intent-to-treat (ITT)	All randomized participants. Participants will be analyzed according to the intervention allocated by randomization.
Safety	All randomized participants who have taken at least 1 dose of study intervention, regardless of the amount of treatment administered. Participants will be analyzed according to the intervention they actually received.
Pharmacokinetic (PK)	All participants from the safety population with at least one post-baseline PK result with adequate documentation of dosing and sampling dates and times. Participants having received only placebo will not be part of the PK population. Participants will be analyzed according to the intervention they actually received.

Participants exposed to study intervention before or without being randomized will not be considered randomized and will not be included in any analysis population. The safety experience of these participants will be reported separately.

Randomized participants for whom it is unclear whether they took the study intervention will be considered as exposed and will be included in the safety population as randomized.

For any participant randomized more than once, only the data associated with the first randomization (except if the first randomization is done by error) will be used in any analysis population. The safety experience associated with any later randomization will be reported separately.

For participants receiving more than one study intervention during the study, the intervention group for as-treated analyses will be the intervention group as randomized if the participant received at least one administration as randomized.

3 STATISTICAL ANALYSES

3.1 GENERAL CONSIDERATIONS

In general, continuous data will be summarized using the number of observations available, mean, standard deviation (SD), median, Q1, Q3, minimum, and maximum. Categorical and ordinal data will be summarized using the count and percentage of participants.

The baseline value is defined as the last available value prior to the first dose of IMP if the participant is treated, or the last available value up to randomization if the participant is not exposed to IMP unless below eDiary data (PP-NRS) or otherwise specified.

The baseline for weekly average PP-NRS score is defined as the average of daily non-missing scores obtained during the 7 days prior to randomization.

Unless otherwise specified, analyses will be performed by intervention group (and overall for baseline and demographics characteristics).

For the main study, two database locks will be performed.

- A partial database lock will be performed to evaluate the BID cohort before the final database lock when all randomized participants in the BID cohort have completed their 16-week study intervention period, including early discontinuation. Only the key data will be analyzed at the time of the partial database lock.
- A final database lock will be performed for the main study when all randomized participants in TID cohort have completed their 16-week study intervention period, including early discontinuation.

For the Japan substudy, the database lock will be performed when all randomized participants in the Japan substudy have completed their 16-week study intervention period, including early discontinuation.

The analyses for the main study will be conducted after last participant's last study visit in the BID and TID cohort, respectively. A pooled analysis of the BID and TID cohort may also be performed as appropriate. The analyses for the Japan sub-study will be performed separately with descriptive analyses after the last participant's last study visit in the Japan substudy.

Observation period

The observation period will be divided into 2 segments:

- The **pre-treatment period** is defined as the period up to first investigational medicinal product (IMP) administration.

- The **treatment-emergent (TE) period** is defined as the period from the first IMP administration to the follow-up visit. The treatment-emergent period includes the following 2 periods:
  - The on-treatment period is defined as the period from the first IMP administration to the last administration of the IMP + 1 day
  - The post-treatment period is defined as the period from the end of the on-treatment period to the end of the treatment-emergent period.

The on-study period is defined as the time from randomization until the end of the study defined as the last scheduled visit for those who completed the study and the end-of-study date collected on electronic case report form (e-CRF) page "Completion of End of Study" for those who did not complete the study. If death is the end-of-study reason, date of death will be used.

3.2 PRIMARY ENDPOINT(S) ANALYSIS

3.2.1 Definition of endpoint(s)

The primary endpoint is the percent change from baseline to week 16 in EASI score.

Eczema Area and Severity Index (EASI)

The EASI is a composite index with scores ranging from 0 to 72. Four AD disease characteristics (erythema, thickness [induration, papulation, edema], scratching [excoriation], and lichenification) will each be assessed for severity by the Investigator or designee. The involvement will be assessed as a percentage by body area of head, trunk, upper limbs, and lower limbs, and converted to a score of 0 to 6. In each body region, the area is expected as 0, 1 (1% to 9%), 2 (10% to 29%), 3 (30% to 49%), 4 (50% to 69%), 5 (70% to 89%), or 6 (90% to 100%).

For each body region (head, trunk, upper limbs, and lower limbs), EASI score = (sum of the four AD disease characteristics scores) multiply (converted area score of percent of body surface involvement in the region). The total EASI score is the weighted total of the body region EASI scores using the weights 10% = head, 20% = upper limbs, 30% = trunk, and 40% = lower limbs. The minimum possible EASI score is 0 and the maximum possible EASI score is 72 where a higher score indicates increased extent and severity of AD.

3.2.2 Main analytical approach

The primary endpoint will be analyzed with the main estimand as defined in Table 3. The analysis population for the efficacy endpoints will be the ITT population.

The primary endpoint will be analyzed using an analysis of covariance (ANCOVA) model with intervention group, randomization stratification of screening IgE levels (level (< or ≥ 300 UI/mL) as fixed effects, and baseline EASI score as covariates, with intercurrent event strategy and missing data handling as defined in Table 3. Specifically, data after participants taking the rescue medications and/or phototherapy in Table 5 will be set to missing and the worst postbaseline

value on or before the time of the medication usage will be used to impute the missing Week 16 value (for participants whose postbaseline values are all missing, the baseline will be used to impute).

Participants who discontinue the study intervention prematurely are encouraged to follow the planned visits and in those participants who did not take the rescue medications and/or phototherapy in Table 5, all data collected after study intervention discontinuation will be used in the analysis. For these participants, missing data may still happen despite all efforts have been tried to collect the data after study intervention discontinuation. For participants who discontinue due to lack of efficacy, all data collected after discontinuation will be used in the analysis, and a WOCF approach will be used to impute missing data if needed.

For participants who discontinued not due to lack of efficacy, or had missing value at Week 16 due to any other reasons including COVID-19, a multiple imputation (MI) approach will be used to impute missing endpoint value, and this MI method will use all participants excluding participants who have taken the rescue medications and/or phototherapy prior to Week 16 and excluding participants who discontinue the study intervention due to lack of efficacy prior to Week 16.

Each of the imputed complete data will be analyzed by fitting an ANCOVA model as described above. The imputation number will be about 40. Statistical inference obtained from all imputed data will be combined using Rubin's rule. Descriptive statistics including the number of subjects mean, standard error, and LS means will be provided. In addition, difference in LS means and the corresponding 90% CI will be provided along with the p-values. See Section 5.5 for the sample SAS code for the imputation and the analysis.

Handling of the use of rescue medications

Blinded review of the rescue/prohibited treatment (medication or procedure) by the clinical team based on Table 5 will be implemented before database locks by considering the type of medication or procedure, indication, timing, frequency and the potential impact of the use of the prohibited medications/procedures and rescue medications.

For the purpose of statistical analysis of efficacy, only the use of any potency TCS (Low to medium potency TCS and High potency TCS), Crisaborole, Calcineurin inhibitors, Systemic medications (Systemic corticosteroids and Non-steroidal systemic immunosuppressive drugs) and phototherapy or regular use (more than 2 visits per week) of a tanning booth/parlor in Table 5 will be considered as intercurrent events and therefore be handled in the estimands for endpoints defined in Table 3.

Table 5 - Rescue medications and prohibited medications/procedures

Medication/procedure	Comment	Selection criteria[a]
Rescue medications		
Low to medium potency TCS	No IMP discontinuation[c]	CDG10429[b] (Low to medium potency TCS study level)
Crisaborole (cAMP-specific phosphodiesterase -4 inhibitor)	No IMP discontinuation[c]	CDG10366[b] (CRISABOROLE mono and multi-ingredients)
Calcineurin inhibitors (for problem areas only, eg, face, neck, intertriginous and genital areas).	No IMP discontinuation[c]	CDGsn00418[b] (Calcineurin inhibitors Narrow)
High potency TCS	IMP to be discontinued	CDG10430[b] High potency TCS study level
Systemic corticosteroids	IMP to be discontinued	CDGsn00010[b] Corticosteroids NARROW
Non-steroidal systemic immunosuppressive drugs	IMP to be discontinued	CDG10446[b] Nonsteroidal anti-inflammatory drugs (NSAIDs) Narrow
Prohibited medications/procedures		
Immunosuppressive or immunomodulating drugs	No IMP discontinuation	CDG10443[b] (Immunosuppressive or immunomodulating agent and prohibited medication study level)
Live vaccines	No IMP discontinuation	CDG20004[b] (Live/attenuated vaccines)
Phototherapy or regular use (more than 2 visits per week) of a tanning booth/parlor	No IMP discontinuation	CMQ10547[b] HLT Phototherapies
Proton Pump inhibitors	No IMP discontinuation	CDG00593[b] (PROTON PUMP INHIBITORS)
Known systemic strong-to-moderate inducers or inhibitors of CYP3A[d]	No IMP discontinuation	CDGsn00225[b] Moderate CYP3A inducers Narrow CDGsn00234[b] Moderate CYP3A inhibitors Narrow CDGsn00253[b] Strong CYP3A inhibitors Narrow CDGsn00265[b] Strong CYP3A inducers Narrow

a The estimand for the intercurrent event handling strategy will be as follows: hypothetical for continuous endpoints, and composite for responder and time-to-event endpoints.
b CDG=company drug groupings, CMQ=company MedDRA query.
c IMP to be discontinued if reason for treatment is “Rescue Therapy” and the route is not “Topical” or “Cutaneous”.
d Exclude if route is ophtalmic, conjunctival, cutaneous, topical, or inhalation

The detailed information of rescue medications and/or prohibited medications/procedures including duration of use and incidence of use during the planned study intervention period will be summarized by intervention group and by categories of medications (Low to medium potency TCS, Crisaborole, Calcineurin inhibitors, high potency TCS or systemic medications, any rescue medications that lead to IMP discontinuation, prohibited medications/procedures).

In addition, the time (week) of first rescue medication (including any potency TCS, Crisaborole, Calcineurin inhibitors or systemic rescue medication or phototherapy) taken will be analyzed using Kaplan-Meier method. The time (week) of first high potency TCS or systemic rescue medication and the time (week) of first rescue medication that leads to IMP discontinuation will also be provided.

The time (week) of first prohibited medication/procedure will be analyzed using Kaplan-Meier method as appropriate.

Listing of rescue medication use will be presented as appropriate.

3.2.3 Sensitivity analysis

No sensitivity analysis is planned for primary analysis.

3.2.4 Supplementary analyses

As-observed analysis (including all data after taking rescue medications/procedures)

The data collected after taking all the rescue medications and/or phototherapy in Table 5 will be included in the supplementary analysis to evaluate the robustness of the primary analysis results with respect to the method of handling data while taking the rescue medications (eg, treatment policy strategy). In addition, for participants discontinuing the study intervention before Week 16, their off-study intervention values measured up to Week 16 will be included in the analysis. For participants with missing data at Week 16 regardless of reason (s), a multiple imputation approach will be used to impute the missing endpoint value.

Additional rescue analysis (WOCF to impute all data after taking the corresponding rescue medications)

Two additional rescue analyses will be performed: 1) the data collected after taking high potency TCS or systemic rescue medications in Table 5 will be set to missing and the WOCF on or before the time of the medication usage will be used to impute the missing Week 16 value in this supplementary analysis (hypothetical strategy); 2) the same intercurrent event handling strategy will be applied to the use of any rescue medications in Table 5 that lead to IMP discontinuation.

In addition, for both of the additional rescue analyses, for participants discontinued the study intervention before Week 16, their off-study intervention values measured up to Week 16 will be included in the analysis. For participants with missing data at Week 16, the missing data handling approach applied in the primary analysis will be used.

Placebo data pooling analysis

An additional supplementary analysis will also be conducted for the TID cohort only using the rilzabrutinib data from TID Cohort and placebo data from both BID and TID cohort. The same intercurrent events handling strategy and missing data handling strategy used in the primary analysis will be applied

Historical placebo data borrowing analysis

In order to improve the quantitative decision making and provide supportive data to demonstrate efficacy of rilzabrutinib, the historical placebo data will be borrowed. Data from historical study which has similar target population and study design will be used as prior information supplementing the placebo arm, using a Bayesian approach. The placebo data collected in a most recently conducted external study with similar design and population - Upadacitinib (Rinvoq M18-891) study (1), will be used as the historical placebo data. The historical placebo data will be used in both BID and TID cohort, respectively. For TID cohort, an additional supplementary analysis will also be conducted by borrowing placebo data from BID cohort. In terms of the estimand, the same intercurrent events handling strategy and missing data handling strategy applied in the primary analysis will be used.

The external study is considered an acceptable source of historical data to some extent based on Pocock´s (2) "acceptability" criteria:

- In the Rinvoq study, participants in the placebo arm received the same treatment as in the placebo arm of the current ACT17207 study
- The external Rinvoq study was conducted from 2018 to 2020 with the similar requirements for participant eligibility
- EASI will be assessed in the same conditions in the external study and the current study
- The distribution of important participant characteristics in the external study should be comparable with those in the current study, due to the similarity of eligibility criteria
- The treatment evaluations are standardized and objective, and the standard of care are the same in the previous and current study, so similar results are expected regardless of different sponsors of the study
- To our knowledge, there is no other indications leading one to expect different results between the external study and current study

Based on the above considerations, the information of placebo arm from the historical study will be borrowed for the supplemental analysis. However, in order to take into account potential unexpected difference between historical placebo data and current placebo data, a dynamic borrowing approach will be used in the analysis. The amount of historical placebo data that will be borrowed will depend on the degree of similarity between historical placebo data and current placebo data regarding the mean percent change in EASI score from baseline to Week 16. Further details about dynamic borrowing are provided in Section 5.6.

3.3 SECONDARY ENDPOINT(S) ANALYSIS

The secondary endpoints detailed in this section are secondary efficacy endpoints. Safety endpoints analyses are defined in Section 3.6.

3.3.1 Key secondary endpoint(s)

3.3.1.1 Definition of endpoint(s)

The key secondary endpoints are:

- Proportion of participants with IGA of 0 or 1 (disease free or almost disease free) compared to placebo at Week 16
- Proportion of participants achieving EASI-75 (reduction of EASI score by ≥ 75% from baseline) at Week 16
- Proportion of participants with reduction of weekly average of daily PP-NRS of ≥ 4 points from baseline at Week 16

Investigator Global Assessment (IGA) of disease activity

Investigator Global Assessment (IGA) is a static 5-point measure of disease severity based on an overall assessment of the skin lesions on a 5-point scale (0 = clear, 1 = almost clear, 2 = mild disease, 3 = moderate disease, 4 = severe disease).

Peak Pruritus Numeric Rating Scale (PP-NRS)

The Pruritus NRS is a simple assessment tool that patients will use to report the intensity of their pruritus (itch) during a daily recall period that has been used in clinical studies of AD. Participants will be asked the following 2 questions:

- For average itch intensity: “On a scale of 0 to 10, with 0 being 'no itch' and 10 being the 'worst itch imaginable', how would you rate your itch overall (on average) during the previous 24 hours?”
- For maximum itch intensity: “On a scale of 0 to 10, with 0 being 'no itch' and 10 being the 'worst itch imaginable', how would you rate your itch at the worst moment during the previous 24 hours?”

Participants will be asked to provide answers to these two questions from 7 days prior to the randomization visit to the end of treatment. The baseline pruritus NRS average score for maximum itch intensity (i.e., Peak Pruritus NRS) will be determined based on the average of daily NRS scores for maximum itch intensity (the daily score ranges from 0 to 10) during the 7 days immediately preceding randomization. A minimum of 4 daily scores out of the 7 days is required to calculate the baseline average score. For participants who do not have at least 4 daily scores reported during the 7 days immediately preceding the planned randomization date, randomization should be postponed until this requirement is met, but without exceeding the 28-day maximum duration for Screening. The post-baseline pruritus NRS average score for maximum itch intensity will be determined based on the average of daily NRS score for maximum itch intensity within a week (7 days).

3.3.1.2 Analytical approach

The key secondary endpoints will be analyzed with the main estimand, ie, intercurrent event handling strategy and missing data handling method as defined in Table 3 and the analysis will be conducted by using CMH test adjusted by randomization stratification of screening IgE levels (< or ≥ 300 UI/mL). Comparisons of the response rates between rilzabrutinib and placebo will be derived. In addition, odds ratio and response rate difference as well as the corresponding 90% confidence intervals (CI) will be provided along with the p-values.

As defined in Table 3 for participants who discontinue the study intervention before Week 16, their off-study intervention values measured up to Week 16 will be included in the analysis. Participants taking the rescue medications and/or phototherapy (see Table 5) prior to Week 16 or having missing data at Week 16 will be considered as non-responders.

3.3.2 Supportive secondary endpoint(s)

The other secondary endpoints are as follows:

- Time to onsite of effect on pruritus (≥ 4 points reduction of weekly average of daily PP-NRS from baseline during the 16-week treatment period)
- Absolute change from baseline to Week 16 in EASI score
- Proportion of participants achieving EASI-50/90 (reduction of EASI score by ≥ 50% or ≥ 90% from baseline) at Week 16
- Change from baseline to Week 16 in percent BSA of EASI
- Absolute and percent change on weekly average of daily PP-NRS based on daily participant assessments documented in their e-diary from baseline to Week 16
- Proportion of participants achieving IGA*BSA-50/75/90 (reduction of IGA*BSA score by ≥ 50% or ≥ 75% or ≥ 90% from baseline) at Week 16

Body surface area (BSA) of EASI

Body surface area (BSA) affected by AD will be assessed for each section of the body and will be reported as a percentage of all major body sections combined.

IGA*BSA score

IGA*BSA score is calculated by multiplying the IGA score by the BSA, giving a score range of 0 to 400 (maximum IGA = 4 and maximum BSA = 100).

Continuous secondary efficacy endpoints

Continuous secondary efficacy endpoints that measure continuous responses will be analyzed in the same fashion as the primary endpoint using primary statistical model.

Binary secondary efficacy endpoints

Binary secondary efficacy endpoints will be analyzed in the same fashion as the key secondary endpoints.

Time-to-event secondary efficacy endpoint

The primary estimand for time-to-event secondary efficacy endpoints is defined in Table 3 and will be analyzed using Cox proportional hazard model stratified by screening IgE level (< or ≥ 300 UI/mL). The missing data will be handled as follows: analyses will be censored at the time of last PP-NRS assessment if participants discontinue the study follow-up before Week 16. The hazards ratio, its 90% confidence interval and p-value will be reported. Kaplan-Meier curves will be also provided.

For time to first PP-NRS response (PP-NRS by ≥ 4 from baseline), participants who receive the rescue medications and/or phototherapy (see Table 5), data prior to start of the medications will be used, but after medication start, the participant's data will not be used and they will be censored at Week 16 (ie, Day 113). For other participants, all available data up to Week 16 (ie, Day 113) including those collected during the off-treatment period will be used. Participants without events will be censored at Day 113 or their last PP-NRS assessment date if discontinued from the study, whichever is earlier.

3.4 TERTIARY/EXPLORATORY ENDPOINT(S) ANALYSIS

3.4.1 Analytical approach

[redacted]

[redacted]

3.5 MULTIPLICITY ISSUES

No adjustments for multiplicity are planned for this Phase 2 study. More specifically, no adjustments will be made in comparing the treatment groups based on the secondary efficacy endpoints and no adjustments will be made for the subgroup analyses. Further, no adjustments will be made for the comparisons between dose levels either. Reported p-values for the secondary endpoints will be nominal p-values.

3.6 SAFETY ANALYSES

All safety analyses will be performed on the safety population as defined in Section 2, unless otherwise specified, using the following common rules:

- The analysis of the safety variables will be essentially descriptive, and no testing is planned.
- Safety data in participants who do not belong to the safety population (eg, exposed but not randomized) will be provided.

3.6.1 Extent of exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure and compliance and summarized within the safety population.

Duration of IMP exposure

Duration of IMP exposure is defined as last IMP administration date – first IMP administration date + 1 day, regardless of unplanned intermittent discontinuations. Duration of IMP exposure will be summarized quantitatively using number of participants, means, SD, min, median and max. In addition, duration of IMP exposure will also be summarized categorically by number and percentage for each of the following categories and cumulatively according to these categories as well:

- > 0 to ≤ 2 weeks
- > 2 to ≤ 4 weeks
- > 4 to ≤ 8 weeks
- > 8 to ≤ 12 weeks
- > 12 to ≤ 16 weeks
- > 16 weeks

Additionally, the cumulative duration of treatment exposure (expressed in participant-years) will be provided.

Treatment compliance

A given administration will be considered noncompliant if the participant did not receive the planned dose as required by the protocol. No imputation will be made for participants with missing or incomplete data.

Percentage of treatment compliance for a participant will be defined as the number of administrations that the participant was compliant divided by the total number of administrations that the participant was planned to take from the first administration of IMP up to the actual last administration of IMP.

Treatment compliance will be summarized quantitatively (number, mean, SD, median, minimum and maximum) and categorically: < 80%, ≥ 80%.

Cases of overdose (defined as a single ingestion of a dose equal to or greater than 1200 mg (≥ 3 times the indicated single dose of 400 mg), provided that it has been taken at once or over 2-hour period) will be listed.

3.6.2 Adverse events

General common rules for adverse events

All adverse events (AEs) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the Medical Dictionary for Regulatory Activities (MedDRA) version currently in effect at Sanofi at the time of database lock.

The AEs will be analyzed in the following 2 categories:

- Pre-treatment AEs: AEs that developed, worsened or became serious during the pre-treatment period.
- Treatment-emergent adverse events (TEAE)s: AEs that developed, worsened or became serious during the treatment-emergent period

There is no post-treatment AEs as treatment-emergent period covers the entire study duration. Similarly, the deaths will be analyzed in the pre-treatment, and treatment-emergent periods. The primary AE analyses will be on TEAEs. Pre-treatment AEs will be described separately.

An AE with incomplete or missing date/time of onset (occurrence, worsening, or becoming serious) will be classified as a TEAE unless there is definitive information to determine it is a pre-treatment AE.

If the assessment of the relationship to IMP is missing for an AE, this AE will be assumed as related to IMP. If the severity is missing for 1 of the treatment-emergent occurrences of an AE,

the severity will be imputed with the maximal severality of the other occurrences. If the severity is missing for all the occurrences, the severity will be left as missing.

Multiple occurrences of the same event in the same participant will be counted only once in the tables within a treatment phase.

The AE tables will be sorted as indicated in Table 6.

Table 6 - Sorting of AE tables

AE presentation	Sorting rules
SOC, HLGT, HLT and PT	By the internationally agreed SOC order and by alphabetic order of HLGTs, HLTs and PTs.
SOC, HLT and PT	By the internationally agreed SOC order and by alphabetic order of HLTs and PTs.
SOC and PT	By the internationally agreed SOC order and decreasing frequency of PTs[a,b]
PT	By decreasing frequency of PTs[a]

a Sorting will be based on the rilzabrutinib group
b The table of all TEAEs presented by primary SOC and PT will define the presentation order for all other tables (eg, treatment-emergent SAE) presented by SOC and PT, unless otherwise specified.

Analysis of all adverse events

The overview of TEAE with the details below will be generated:

- Any TEAE
- Any Severe TEAE
- Any treatment-emergent SAE
- TEAE leading to death
- Any TEAE leading to permanent intervention discontinuation
- Any treatment-emergent AESI
- Any TEAE related to IMP

The AE summaries of Table 7 will be generated with number (%) of participants experiencing at least one event.

Table 7 - Analyses of adverse events

Type of AE	MedDRA levels
All TEAE	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
	Primary SOC
	PT
	Primary and secondary SOC, HGLT, HLT and PT
Common TEAE (≥2 participants in any group)	Primary SOC and PT
TEAE related to IMP as per Investigator's judgment	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
TEAE by maximal intensity	Primary SOC and PT
Treatment emergent SAE	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
Treatment emergent SAE related to IMP as per Investigator's judgment	Primary SOC, HGLT, HLT and PT
TEAE leading to permanent intervention discontinuation	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
TEAE leading to death[b]	Primary SOC, HGLT, HLT and PT
Pretreatment AE	Overview[a]
	Primary SOC and PT

a Will include the following AE categories: any AEs, any serious AEs, any AEs leading to death, any AEs leading to permanent intervention discontinuation

b Death as an outcome of the AE as reported by the Investigator in the AE page

Analysis of deaths

In addition to the analyses of deaths included in Table 6 the number (%) of participants in the following categories will be provided if needed:

- Deaths during the study period or treatment-emergent period
- Deaths in non-randomized participants or randomized but not treated participants

Analysis of adverse events of special interest (AESIs)

Adverse events of special interest (AESIs) be selected for analyses as indicated in Table 8. Number (%) of participants experiencing at least one event will be provided for each event of interest, by PT if applicable. Tables will be sorted as indicated in Table 6.

Table 8 - Selections for AESIs

AESIs	Selection
Pregnancy of a female participant entered in a study as well as pregnancy occurring in a female partner of a male participant entered in a study with IMP	"Pregnancy" or "Partner Pregnancy" checked on the Pregnancy eCRF page as reported by the investigator
Symptomatic overdose (serious or nonserious) with IMP	Symptomatic Overdose is answered Yes, with Overdose of IMP answered Yes on AE eCRF.
Increase in ALT	ALT > 3 x ULN
Severe infections including opportunistic infection	Primary SOC = 'Infections and infestations' and with severe status
Tuberculosis or initiation of medications for suspected tuberculosis	"Tuberculosis" checked on the Infection Event Form (Type of Organisms) as reported by investigator or medications (CDG00737) initiated after screening for suspected tuberculosis is recorded by the Investigator in the Prior and Concomitant Medications eCRF
Diagnosed and biologically proven SARS-CoV-2 infection	CMQsn00237 COVID-19 Narrow.
Major hemorrhagic events[b]	CMQsn00038 Haemorrhages Narrow search[a] and "moderate" or "severe" ticked in Adverse Events eCRF page.
Serious adverse events of cytopenia	CMQsn00027 Haematopoietic cytopenias Narrow[a] and "serious" answered Yes on AE eCRF or "severe" answered Yes on AE eCRF for non-serious AEs.
Atrial fibrillation	CMQ10050 based on the single PT = "Atrial fibrillation"[a].

a The list of terms may be adjusted according to MedDRA version changes
b Include symptomatic bleeding in a critical area or organ such as the central nervous system (CNS), or intraocular bleeding, resulting in an SAE

Analysis of the gastrointestinal events:

For the gastrointestinal (GI) events, the following summaries will be provided:

- Number (%) of participants with any TEAE
- Number (%) of participants with any treatment emergent SAE
- Number (%) of participants with any AE leading to death
- Number (%) of participants with any TEAE leading to permanent study intervention discontinuation
- Number (%) of participants with any TEAE related to IMP reported by investigator
- Number (%) of participants with any TEAE by maximum intensity, corrective treatment and final outcome
- Time to onset of first TEAE and cumulative incidence at specified time points (Kaplan-Meier estimates at Week 16)
- The cumulative incidence and incidence by interval

- Average duration in days of all events
- Cumulative duration in days across all events

Kalan-Meier curves will also be provided, when appropriate, for the time to IMP discontinuation due to adverse event with any GI events.

The number (%) of participants with any GI events by primary SOC and PT will also be provided.

The frequency of TEAEs over time will be provided, when appropriate, for GI events (eg. nausea, vomiting and diarrhea) using weekly time intervals up to Week 16, ie, [0-1] week, (1,2] weeks, (2,3] weeks, (3,4] weeks, etc. In each time interval, the numerator in the calculation of percentages will be the number of participants with at least 1 TEAE occurring in this time interval. Two types of analyses will be included:

- Only the first event will be counted for each participant and all recurrent events will not be included, and the denominator for the calculation of percentages will be the number of participants at risk at the beginning of the time interval who did not experience a first event in the preceding intervals; graphical presentations will also be provided for the percentages of participants with nausea, vomiting and diarrhea as appropriate by week; and
- The recurrent events in subsequent intervals will be counted once for each participant in the numerator of the corresponding interval, and the denominator for the calculation of percentages will be the number of participants at risk at the beginning of the time interval.

3.6.3 Additional safety assessments

3.6.3.1 Laboratory variables, vital signs and electrocardiograms (ECGs)

The following laboratory variables, vital signs and electrocardiogram (ECG) variables will be analyzed. They will be converted into standard international units.

- Hematology:
  - Red blood cells and platelets and coagulation: hemoglobin, hematocrit, red blood cell count, platelet count, Prothrombin time (expressed as international normalized ratio), activated partial thromboplastin time
  - White blood cells: white blood cell count, neutrophils, lymphocytes, monocytes, basophils, eosinophils
- Clinical chemistry:
  - Metabolism: C-reactive protein, creatine phosphokinase.
  - Renal function: creatinine, creatinine clearance, blood urea nitrogen. Creatinine clearance will be derived with the equation of Cockcroft and Gault using weight assessed at the same visit as creatinine.

- Liver function: alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, lactate dehydrogenase, total, direct and indirect bilirubin
- Pregnancy test: Serum β-human chorionic gonadotropin (all female participants) at screening and urine pregnancy tests at visits except V3.
- Hepatitis screen: hepatitis B surface antigen (HBs Ag), hepatitis B core antibody (HBc Ab) (total and IgM), hepatitis C virus antibodies (HCV Ab) will be tested at screening (V1) if these tests have not been performed within 3 months prior to the screening visit. In case of results showing HBc Ab (positive), an hepatitis B virus (HBV) deoxyribonucleic acid (DNA) testing will be performed and should be confirmed negative prior to randomization. In case of results showing HCV Ab (positive), an HCV ribonucleic acid (RNA) testing will be performed and should be confirmed negative prior to randomization
- TB test (performed for participants without documented negative screening for TB via a negative QFT within 3 months prior to screening (and if required per local standard of care, a chest X-ray)).
- HIV antibody at screening visit if not been performed within 3 months prior to the screening visit.

- Urinalysis:
  - Urinalysis: pH, protein, glucose, ketones, bilirubin, blood, nitrites, urobilinogen, leukocytes by dipstick and specific gravity. Microscopic examination will be performed if dipstick analysis of blood or protein is abnormal.
- Vital signs: pulse rate (beats per minute), systolic and diastolic blood pressure (mmHg) in a semi-supine position after 5 minutes, and body temperature (degrees Celsius).
- ECG variables: heart rate, PR, QRS, QT, and QTc intervals/ECG assessments will be described as normal or abnormal.

Data below the lower limit of quantitation/detection limit (LLOQ) will be replaced by half of the LLOQ, data above the upper limit of quantification will be replaced by ULOQ value.

Quantitative analyses

When relevant, for laboratory variables, and vital signs above, descriptive statistics for results and changes from baseline will be provided for each analysis window, the last value and the worst value (minimum and/or maximum value depending on the parameter) during the on-treatment period. These analyses will be performed using central measurements only (when available), for laboratory variables.

For all parameters, mean changes from baseline with the corresponding standard error will be plotted over time.

Analyses according to PCSA

Potentially clinically significant abnormality (PCSA) analyses will be performed based on the PCSA list currently in effect at Sanofi at the time of the database lock. For parameters for which no PCSA criteria are defined, similar analyses will be done using the normal range, if applicable.

Analyses according to PCSA will be performed based on the worst value during the treatment-emergent period, using all measurements (either local or central, either scheduled, nonscheduled or repeated).

For laboratory variables, vital signs and ECG variables above, the incidence of participants with at least one PCSA during the treatment-emergent period will be summarized regardless of the baseline level and according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

Additional analyses for drug-induced liver injury

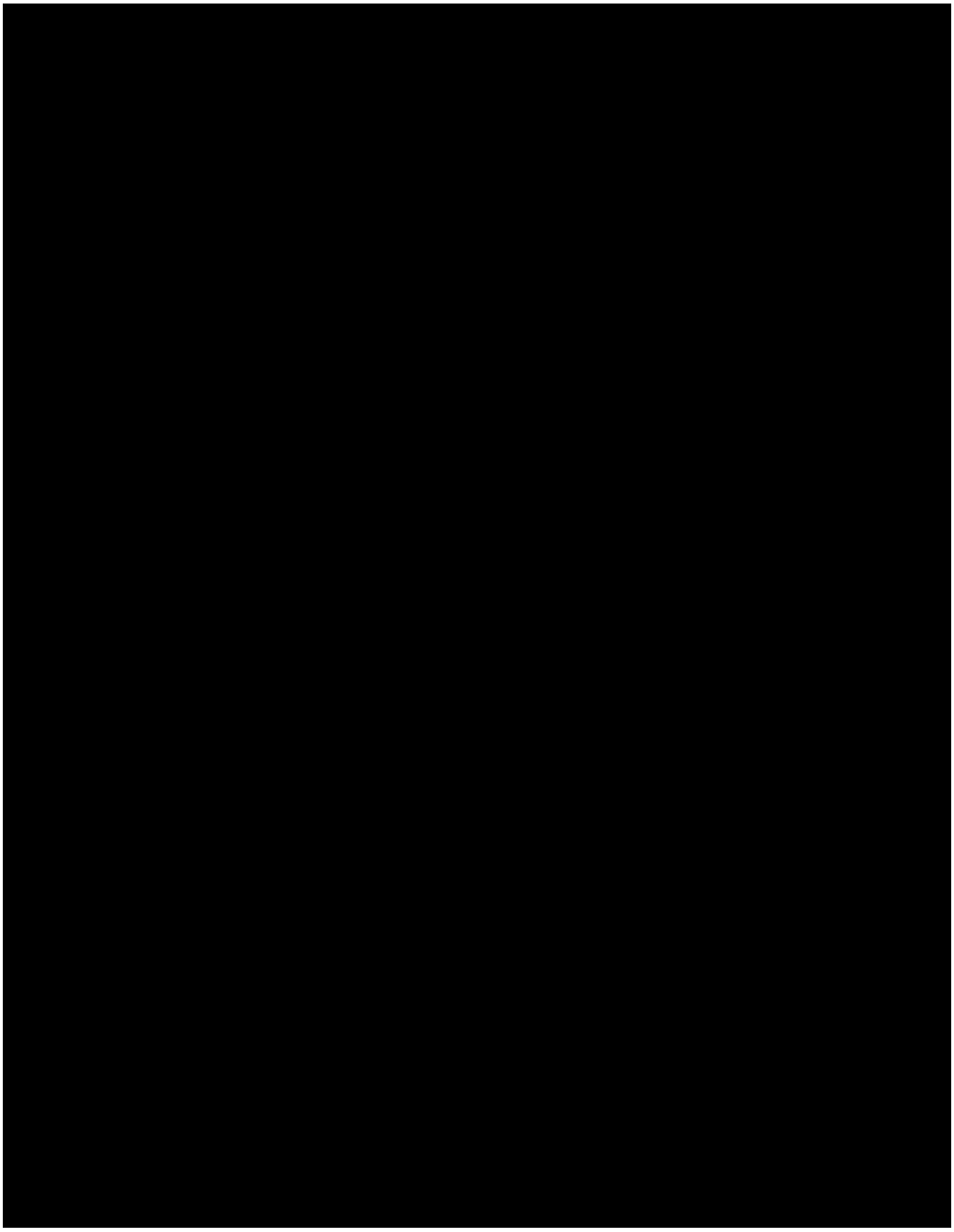
The following additional analyses will be performed for drug-induced liver injury:

- Time to onset of the initial alanine aminotransferase (ALT) or aspartate aminotransferase (AST) elevation (> 3 x ULN) during the treatment-emergent period will be analyzed using Kaplan-Meier method.
- Time to onset of the initial total bilirubin elevation (> 2 x ULN) during the treatment-emergent period will be analyzed using Kaplan-Meier method.
- A graph of the distribution of peak values of ALT versus peak values of total bilirubin during the treatment-emergent period will be provided.
- For each liver function test (eg, ALT), participants having experienced a PCSA (eg, ALT > 3 ULN) will be summarized using the following categories: Returned to baseline PCSA status (or returned to value ≤ ULN in case of missing baseline) before last IMP dose, returned to baseline PCSA status after last IMP dose, never returned to baseline PCSA status, No assessment after elevation. This summary will be performed by categories of elevation (ALT > 3, > 5, > 10, > 20 ULN).

3.7 OTHER ANALYSES

3.7.1 Other variables and/or parameters

[redacted]



[redacted]

3.7.2 Subgroup analyses

To assess the homogeneity of the treatment effect across different subgroup levels, subgroup analyses will be performed on the primary endpoint with respect to the following subgroups (categories with fewer than 5 participants may be combined with other categories):

- Age group (≥ 18-< 40, ≥ 40 - < 65, ≥ 65 years; < 40, ≥ 40 years)
- Gender (Male, Female)
- Ethnicity (Not Hispanic or Latino, Hispanic or Latino)
- Duration of AD (< 26, ≥ 26 years)
- Baseline weight (< 60, ≥ 60- < 90, ≥ 90 kg)
- Baseline BMI (< 25, ≥ 25- < 30, ≥ 30 kg/m^2)
- Baseline IGA moderate versus severe (3 versus 4)
- Baseline EASI category (< median, ≥ median)
- Baseline EASI category (< 16, ≥ 16) (for TID cohort only)
- Randomization stratification of IgE levels (< or ≥ 300 UI/mL)
- Region (North America, Other)

For the primary endpoint, the subgroup analyses will be performed based on the imputed datasets from the primary analysis. To test the interaction between study intervention and subgroup factor, an ANCOVA model incorporating subgroup-by-study intervention interaction will be built for each subgroup factor. The model will include all the covariates (except the randomization strata in the main statistical model) plus the subgroup variable (if not one of the covariates adjusted in the main model already) and the subgroup-by-study intervention interaction. Statistical inference obtained from all imputed data will be combined using Rubin's rule. A p-value for the test of interaction will be provided based on the combined inference.

In each subgroup, the primary endpoint will be analyzed using the primary approach for the primary endpoint, but on the specific subgroup of the imputed primary analysis population. Descriptive statistics including number of patients, mean, standard error, and LS means for each

subgroup will be provided. In addition, difference in LS means and the corresponding 90% CI will be provided for each subgroup. Forests plots will be provided.

For the key secondary endpoints, the subgroup analyses will be performed based on the imputed datasets from the primary analysis of the corresponding endpoint. To test the interaction between study intervention and subgroup factor, a logistic regression model incorporating subgroup-by-study intervention interaction will be built for each subgroup factor. The model will include all the covariates (except the randomization strata in the statistical model) plus the subgroup variable and the subgroup-by-study intervention interaction. A p-value for the test of interaction will be provided.

In each subgroup, the treatment effects for each key secondary endpoint will be provided, as well as the corresponding 90% CI, using the same method as applied to the analytical approach to each key secondary endpoint. Forest plots will be provided.

3.8 INTERIM ANALYSES

No interim analysis is planned.

3.9 CHANGES TO PROTOCOL PLANNED ANALYSES

This section summarizes major statistical changes in the protocol amendment(s).

Table 9 - Statistical changes in protocol amendment(s)

Amendment Number	Approval Date	Changes	Rationale
02	03-Jan-2022	Section 9.1 Sample size: added additional TID cohort (42 participants receiving rilzabrutinib /28 participants receiving placebo) which will begin enrolling after participants enrolled in BID cohort. Adjusted BID cohort sample size from 42 in rilzabrutinib arm and 28 in placebo arm to 30 in rilzabrutinib and 20 in placebo.	Dose regimen reconsidered to ensure sufficient inhibition in mast cells and basophils activation that play an important role in Type 2 predominant inflammatory diseases such as atopic dermatitis.
03	24-Aug-2022	Added the following secondary endpoint: Proportion of participants achieving IGA*BSA-50/75/90 (reduction of IGA*BSA by ≥50% or 75% or 90% from baseline) at Week 16	To explore the assessment of AD by using IGA*BSA as an alternative measure of the disease severity and improvement overtime in participants with AD.
03	24-Aug-2022	[redacted]	[redacted]

4 SAMPLE SIZE DETERMINATION

Approximately 120 participants will be randomized in a 3:2 ratio: to rilzabrutinib versus matching placebo split into approximately 50 and 70 participants in the BID and TID cohorts, respectively.

Sample size calculations were performed to ensure reasonable accuracy for the estimation of the treatment difference in percent change in EASI score from baseline to Week 16 between rilzabrutinib and placebo. For the BID cohort, a sample size of 50 participants enrolled with a ratio of 3:2 between rilzabrutinib and placebo and a common SD of 40 will provide a half-width for the 2-sided 90% CI of less than or equal to 19.0%. For the TID cohort, a sample size of 70 participants enrolled with a ratio of 3:2 between rilzabrutinib and placebo and a common SD of 40 will provide a half-width for the 2-sided 90% CI of less than or equal to 16.1%.

[redacted]

5 SUPPORTING DOCUMENTATION

5.1 APPENDIX 1 LIST OF ABBREVIATIONS

AD:	atopic dermatitis
AE:	adverse event
AESIs:	adverse events of special interest
ALT:	alanine aminotransferase
ANCOVA:	analysis of covariance
ATC:	anatomic category
BID:	twice a day
BSA:	body surface area
BTK:	Bruton's tyrosine kinase
DNA:	deoxyribonucleic acid
EASI:	eczema area and severity index
EASI-75:	reduction of EASI score by ≥ 75% from baseline
ECG:	electrocardiogram
e-CRF:	electronic case report form
e-diary:	electronic diary
ET:	early termination
GI:	gastrointestinal
HBc Ab:	hepatitis B core antibody
HBs Ag:	hepatitis B surface antigen
HBV:	hepatitis B virus
HCV Ab:	hepatitis C virus antibodies
HGLT:	high level group term
HLT:	high level term
IGA:	investigator's global assessment
IgE:	immunoglobulin E
IgG:	immunoglobulin G
IgM:	immunoglobulin M
IMP:	investigational medicinal product
IRT:	interactive response technology
ITT:	intent-to-treat
LLT:	lower-level term
LOCF:	last observation carried forward
MedDRA:	medical dictionary for regulatory activities
PBMCs:	peripheral blood mononuclear cells
PCSA:	potentially clinically significant abnormality
PK:	pharmacokinetics
PoC:	proof-of-concept
PP-NRS:	peak prutitus Numerical Rating Scale
PT:	preferred term
RNA:	ribonucleic acid

SAEs:	serious adverse events
SD:	standard deviation
SOC:	system organ class
TARC:	thymus and activation-regulated chemokine
TCS:	topical corticosteroids
TE:	treatment-emergent
TEAE:	treatment-emergent adverse event
TID:	three times a day
WHO-DD:	World Health Organization-drug dictionary
WOCF:	worst observation carried forward

5.2 APPENDIX 2 PARTICIPANT DISPOSITIONS

The number (%) of participants included in each of the analysis populations listed in Table 4 will be summarized. Reasons for exclusion from the population without trial impact (disruption) due to COVID-19 will be summarized.

Screen failures are defined as participants who consent to participate in the study but are not subsequently randomized. The number (%) of screen failures and reasons for screen failures will be provided in the screened population.

The number (%) of participants in the following categories will be provided:

- Randomized participants
- Randomized but not exposed participants
- Randomized and exposed participants
- Participants who completed the study intervention period as per protocol
- Participants who did not complete the study intervention period as per protocol and main reason for permanent intervention discontinuation.
- Participants who completed the study period as per protocol.
- Participants who did not complete the study period as per protocol and main reason for study discontinuation.
- Status at last study contact

Reasons for permanent study intervention and study discontinuation “adverse event” and “other reasons” will be split as related versus not related to COVID-19, if applicable.

The number (%) of exposed and not randomized participants and nonrandomized but exposed participants (if applicable) will also be summarized.

In addition, the number (%) of participants screened, randomized, with permanent intervention discontinuation and with early study discontinuation will be provided by country and site.

If needed, number (%) of participants with any reasons for exclusion from the population without trial impact (disruption) due to COVID-19 in the randomized population will be provided.

Protocol deviations

Critical and major protocol deviations (automatic or manual) will be summarized in the randomized population as well as displayed separately as related versus not related to COVID-19 if applicable. In addition, major or critical deviations potentially impacting the primary endpoint analysis may be summarized.

5.3 APPENDIX 3 DEMOGRAPHICS AND BASELINE CHARACTERISTICS, PRIOR OR CONCOMITANT MEDICATIONS

Demographics, baseline characteristics, medical surgical history

The following demographics and baseline characteristics, medical and surgical history and disease characteristics at baseline will be summarized using descriptive statistics in the randomized population.

Demographic and baseline characteristics

- Age in years as quantitative variable and in categories ($\geq$ 18 to < 40, $\geq$ 40 to < 65, $\geq$ 65)
- Gender (Male, Female)
- Race (White, Black or African American, Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, Multiple, Unknown, Not reported)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino, Not reported, Unknown)
- Country
- Region (North America, Other)
- Weight in Kg (quantitative and qualitative variable: (< 60, $\geq$ 60-< 90, $\geq$ 90 kg)
- BMI in kg/m^2 (quantitative and qualitative variable: (< 25, $\geq$ 25-< 30, $\geq$ 30 kg/m^2)

Baseline safety and efficacy parameters (apart from those listed above) will be presented along with the safety and efficacy summaries.

Medical (or surgical) history includes all the relevant medical (or surgical) history during the lifetime of the participant. Medical and surgical history will be coded to a LLT, PT, HLT, HLGT, and associated primary SOC using the MedDRA version currently in effect at Sanofi at the time of database lock.

Participant or family history of allergy (aeroallergens, dust mites, cates, and mold) will also be summarized.

Disease characteristics at baseline

The following baseline disease characteristics will be summarized by intervention group:

- Duration of AD and grouping (years; < 26 and ≥ 26) to be derived as
  - (Year of randomization – Year of first diagnosis of AD) + (month of randomization – month of first diagnosis of AD)/12
- Age of onset of AD (years)
- Randomization stratification of IgE levels (< or ≥ 300 UI/mL)
- Baseline IGA score (summary and frequency by each score level)
- Baseline EASI score
- Baseline EASI score by category
  - Moderate (< 21)
  - Severe EASI (≥ 21 - < 50)
  - Very severe EASI (≥ 50)
- Baseline EASI score (< median, ≥ median)
- Baseline BSA of EASI
- Baseline BSA of EASI by category (≥ 10 to < 30, ≥ 30 to < 50, ≥ 50 %)
- Baseline BSA*IGA
- Baseline BSA*IGA by category (≥ 0 to ≤ 30, > 30 to ≤ 130, > 130 to ≤ 400)
- Baseline PP-NRS
- Baseline total IgE in serum (UI/mL)
- Baseline total IgE in serum by category (< 300 UI/mL, ≥ 300 UI/mL)
- Baseline total IgG in serum (UI/mL)
- Baseline total IgG in serum by category (< 210 UI/mL, ≥ 210 UI/mL)
- Baseline total IgM in serum (UI/mL)
- Baseline total IgM in serum by category (< 300 UI/mL, ≥ 300 UI/mL)
- Baseline TARC in serum (pg/mL)
- Baseline BHRA status (BHRA-/BHRA+)
- Frequency of alcohol drinking in the last 12 months (never, occasional, at least monthly, at least weekly, at least daily) and number of drinks on a typical day (1 or 2, > 2)

Prior or concomitant medications

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version currently in effect at Sanofi at the time of database lock.

- Prior medications/procedures are those the participant received prior to first IMP intake. Prior medications/procedures can be discontinued before first administration or can be ongoing during treatment period.
- Concomitant medications/procedures are any medications received by the participant concomitantly to the IMP, from the first administration of IMP to the last IMP intake + 1 day.
- Post-treatment medications/procedures are those the participant received in the period running from the end of the concomitant medications period up to the end of the study.
- A given medication/procedure can be classified as a prior medication/procedure and/or as a concomitant medication/procedure and or as post-treatment medication/procedure. If it cannot be determined whether a given medication/procedure was taken prior or concomitantly or post, it will be considered as prior, and concomitant, and post-treatment medication/procedure.

The prior and concomitant medications/procedure will be summarized for the randomized, by anatomic and therapeutic level. Participants will be counted once in each ATC category (anatomic or therapeutic) linked to the medication.

Medications will be summarized by intervention group according to the WHO-DD dictionary, considering the first digit of the anatomic category (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and participants will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore, participants may be counted several times for the same medication.

The table for prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across intervention groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

The tables for concomitant medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence in the rilzabrutinib group. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

Procedures will be summarized by intervention group by primary SOC (sorted by internationally agreed order) and PT (sorted in alphabetical order), sorting is based on the overall incidence across intervention groups.

Background intervention

Participants will be required to apply background topical treatment (ie. moisturizers/emollients) in the e-diary everyday from screening visit up to the EOT.

The compliance of moisturizers (emollients) used from screening visit up to EOT, is defined as the (number of days moisturizers used during the period) / (number of days within the period) x 100%, will be summarized as appropriate by intervention group.

5.4 APPENDIX 4 DATA HANDLING CONVENTIONS

Demographic formulas

Age of onset of AD is calculated as:

Year of AD diagnosis – Year of birth

BMI is calculated as:

Weight in kg / (height^2 in meters)

Renal function formulas

For adults, creatinine clearance (CLcr) value will be derived using the equation of Cockroft and Gault:

$$\text{CLcr (ml/min)} = (140 - \text{age}) \times \text{weight (kg)} \times (1 - 0.15 \times \text{sex (0-M, 1-F)}) / (0.814 \times \text{creatinine (}\mu\text{mol/l)})$$

CLcr will be calculated using the last weight measurement on or before the visit of the creatinine measurement and age at the lab sampling day. Here age is calculated as following:

Age = age collected at screening + integer part of (lab sampling analysis day/365.25)

Analysis windows for time points

The following analysis windows will decide how the scheduled and/or unscheduled visits will be used in the by-visit analyses of efficacy, safety, PK and biomarker variables.

A measurement (scheduled or unscheduled) will be used if it is available and measurement date is within the analysis window.

In general, the following order will be used to select the record for analysis at given visit:

- Scheduled visit
- Early termination (ET) or end of study (EOS), whichever comes first if scheduled visit not available

- Unscheduled visit if both scheduled visit and ET/EOT/EOS are not available.

After applying these time windows, if multiple assessments are associated to the same time point, the closest from the targeted study day will be used. If the difference is a tie, the value after the targeted study day will be used. If multiple valid values exist within a same day, then the first value of the day will be selected.

If there is no measurement for a given parameter in an analysis window, data will be considered missing for the corresponding visit.

For the eDiary data (PP-NRS), all available values of daily measurements will be assigned to each week window according to Table 10, and then weekly average score will be calculated.

Table 10 - Time window for eDiary efficacy variable

Time	Targeted study day	Day range for calculating weekly score
Baseline (Week 0)	1	-7 to -1
Week 1	8	1 to 8
Week 2	15	9 to 15
Week 3	22	16 to 22
Week 4	29	23 to 29
Week 5	36	30 to 36
Week 6	43	37 to 43
Week 7	50	44 to 50
Week 8	57	51 to 57
Week 9	64	58 to 64
Week 10	71	65 to 71
Week 11	78	72 to 78
Week 12	85	79 to 85
Week 13	92	86 to 92
Week 14	99	93 to 99
Week 15	106	100 to 106
Week 16	113	107 to 113

Study days are calculated considering Day 1 as the day of first administration of intervention (or the day of randomization for participant not exposed).

For other efficacy variables, all available values of scheduled measurements will be assigned to the appropriate visit window according to Table 11. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same number of days away from the target day, the latest one will be used.

Table 11 - Time window for efficacy variables

		Time windows for		
Visit	Target study Day	EASI	BSA of EASI	IGA
Visit 1 (Week -4 to Day -1)	<1	<-14	<-14	
Visit 2 (Week 0)	1	-14-1^{-}	-14-1^{-}	1^{-}
Visit 3 (Week 2)	15	1^{+}-22	1^{+}-22	1^{+}-22
Visit 4 (Week 4)	29	23-43	23-43	23-43
Visit 5 (Week 8)	57	44-71	44-71	44-71
Visit 6 (Week 12)	85	72-99	72-99	72-99
Visit 7 (Week 16)	113	>99	>99	>99

Study days are calculated considering Day 1 as the day of first administration of intervention (or the day of randomization for participant not exposed).
1^{-}: before 1st dose date/time (or before randomization date/time if participant not exposed); 1^{+}: after 1st dose date/time (or after randomization date/time for participant not exposed)

Safety assessment

For the safety assessment, the reference date for the derivation of relative days of events or findings will be the date of first IMP administration. Selected safety variables will be summarized by the analysis window define in Table 12 for the by visit descriptive analysis. All available values from central lab will be assigned to the appropriate visit window. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same number of days away from the target day, the latest one will be used.

Table 12 - Time window for safety endpoints

Visit	Target study Day	Time windows for							
		Vital signs	Hematology, Chemistry[a] and Coalgulation	Urinalysis	COVID-19 test/Hep B and C, HIV testing and QuantiFERON (TB)	Serum Pregnancy test	Urine Pregnancy test	Physical examination[b]	ECG
Visit 1 (Week -4 to Day -1)	<1	<-14	<-14	1^-	1^-	<-14		<-14	1^-
Visit 2 (Week 0)	1	-14-1^-	-14-1^-				-14-1^-	-14-1^-	
Visit 3 (Week 2)	15	1^+-22	1^+-22					1^+-22	
Visit 4 (Week 4)	29	23-43	23-43				1^+-43	23-43	
Visit 5 (Week 8)	57	44-71	44-71				44-71	44-71	
Visit 6 (Week 12)	85	72-99	72-99				72-99	72-99	
Visit 7 (Week 16)	113	100-116	100-116				100-116	100-116	1^+
Visit 8 (Week 17)	120	>116	>116				>116	>116	

Study days are calculated considering Day 1 as the day of first administration of intervention (or the day of randomization for participant not exposed).

1^-: up to 1^{st} dose date/time; 1^+: after 1^{st} dose date/time;

a For LDH, creatinine, blook urea nitrogen and hs-CRP, the samples will be collected at baseline, Week 8 and Week 16 visit; For CPK, the samples will be collected at baseline and Week 16 visit.

b Complete physical examination including height and weight at screening V1 and abbreviated physical examination including weight after V1.

Pharmacokinetics assessment

For the pharmacokinetics variables summary, the reference date for the derivation of relative days of measurements will be the date of first IMP administration of the participant is treated with study intervention, or the randomization date if the participant is not treated. Pharmacokinetics variables will be summarized by the analysis window defined in Table 13 for the by visit descriptive analyses. All available values of measurements will be assigned to the appropriate visit window. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same distance to the target day, the latest one will be used.

[redacted]

Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will be used for computation of baseline, the last on-treatment value, analysis according to PCSAs and the shift summaries for safety. They will also be included in the by-visit summaries if they are re-allocated to scheduled visits.

Missing data handling for safety analysis

For categorical variables, participants with missing data are not included in calculations of percentages unless otherwise specified. When relevant, the number of participants with missing data is presented.

Handling of computation of study intervention duration if investigational medicinal product end of study intervention date is missing:

For the calculation of the study intervention duration, the date of the last dose of IMP is equal to the date of last administration reported on the eDiary or in the eCRF, whichever is the latest. If this date is missing, the exposure duration should be left as missing.

The last dose administration should be clearly identified in the eDiary or the eCRF and should not be approximated by the last returned package date.

Handling of adverse events with missing or partial date/time of onset:

Adverse event of missing or partial onset date/time will be classified as treatment-emergent when the partial adverse event onset date/time information does not indicate that the adverse event started prior to study intervention or after the treatment-emergent adverse event period. The handling rules are for categorization purpose only and will not be used in listings.

Missing or partial start date/time of AE will be imputed using the following rules:

- If it has missing minute or missing hour/minute or missing day/hour/minute, and if the remaining non-missing date/time parts are the same as the counterparts in date/time of the first dose of IMP, then impute missing parts with the same parts from date/time of the first dose of IMP. If the date/time is later than the AE end date/time, then use AE end date/time instead. Otherwise impute with 00 for hour/minute or 01 for day. If the date is before informed consent, use the informed consent date instead.
- If only year is not missing, and if it's less than the first dose year, then use informed consent day/month, or if it's equal to first dose year, then use the first dose day/month, or if it's after the first dose year, then use January 01. If the date is after AE end date, use AE end date instead.
- After the above process, if the AE with imputed date/time is not the first record of an AE, and the imputed date/time is earlier than the start date/time of the previous record, then use the start date/time of previous record instead.
- If AE start year is missing, no imputation can be done.

Missing or partial date/time of end date/time of adverse event and adverse event resolution will not be handled.

Handling of adverse events when date and time of first investigational medicinal product administration is missing:

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the adverse event considered as such in the frequency tables of possibly related adverse events, but no imputation should be done at the data level.

Handling of missing severity of adverse events:

If the severity is missing for 1 of the treatment-emergent occurrences of an adverse event, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, a "missing" category will be added in the summary table.

Handling of medication missing/partial dates:

To determine whether a medication is prior medication or concomitant medication or both, the missing medication start date is estimated as early as possible, and the missing medication end date is estimated as late as possible. If the medication start date is missing, the onset day will not be calculated in medication listings.

Prior medication start date:

If start day is missing, and start month and year are not missing: Impute the start day using the first day of the month. Imputation flag is 'D';

If start month is missing, and start year is not missing: Impute the day and month using 01 January. Imputation flag is 'M'.

If start year is missing: Impute start date using 6 months before inform consent date. Imputation flag is 'Y'.

A special note: for start date with year missing, the general principle is not to impute. However, in order to simplify the programming flow, the imputation is proposed to align with the protocol which specifies to collect up to 6 months prior medication. Since the start date of prior medication will not be used in any analysis, the rule will not impact the analysis result.

Prior medication end date:

If end day is missing, and end month and year are not missing: Impute end date using the last day of the month. If this leads to a date on or after first dose intake date, use first dose intake date -1 instead. Imputation flag is 'D'.

If end month is missing, and end year is not missing: Impute end date using 31 December as the day and month. If this leads to a date on or after first dose intake date, use first dose intake date - 1 instead. Imputation flag is 'M'

If end year is missing: Impute end date using the first dose intake date -1. Imputation flag is 'Y'.

Concomitant medication start date:

The imputation rule for concomitant medication start date is the same as AE start date. Concomitant medication end date If end day is missing, and end month and year are not missing: Impute end date using the last day of the month. If this leads to a date after end of study follow up date, use the last visit study date instead. Imputation flag is 'D'.

If end month is missing, and end year is not missing: Impute end date using 31 December as the day and month. If this leads to a date after end of study follow up date, use the last study visit date instead. Imputation flag is 'M'.

If end year is missing: Impute date using the end of last study visit date. Imputation flag is 'Y'.

Medication coding:

Medications whose ATC level 4 cannot be coded will be summarized by setting ATC4=ATC2 in the table programs. However, these uncoded ATC level 4 records still need to be confirmed with study DM and study medical manager.

Handling of potentially clinically significant abnormalities:

If a participant has a missing baseline, he/she will be grouped in the category "normal/missing at baseline."

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing, eg, for eosinophils the PCSA is > 0.5 GIGA/L or > ULN if ULN ≥ 0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

5.5 APPENDIX 5 SAMPLES SAS CODE

The multiple imputation and analysis model for the primary analysis of percent change from baseline to Week 16 in EASI score will be built with the following sample SAS code.

1. 40 datasets with a monotone missing pattern will be obtained, induced by Markov Chain Monte Carlo (MCMC) method on participants who have not taken the rescue medications or have discontinued study intervention due to lack of efficacy prior to Week 16.

```
proc mi data=dat_etd seed=17207 nimpute=40 out=dat_mc;
    mcmc impute=monotone;
    var strata trt01p easibl W2_easi_pchg … W16_easi_pchg;
run;
```

2. For each of the imputed dataset with monotone missing pattern in step 1, the remaining missing data will be imputed using the regression method for the monotone pattern with adjustment for randomization strata of screening IgE levels (< or ≥ 300 UI/mL) and baseline value of the response variable.

```
proc mi data=dat_mc nimpute=1 seed=70271 out=dat_mi;
    by _imputation_;
    class strata trt01p;
    monotone method=reg;
    var strata trt01p easibl W2_easi_pchg … W16_easi_pchg;
run;
```

3. Each of the 40 imputed datasets will be merged with the one dataset imputed by WOCF approach, and then be analyzed using the main statistical model. These 40 imputed datasets will be saved.

```
%macro w1;
      %do i=1%to 40;
      data wocf&i.;
      set wocf;
      _imputation_=&i.;
      run;
      %end;
      data wocf_all;
      set %do j=1 %to 40; wocf&j. %end;;
      run;
%mend w1;

%w1

data dat_imp;
      set dat_mi wocf_all;
Run;

proc sort data=dat_imp;
      by _imputation_;
run;

proc glm data= dat_imp;
      by _imputation_;
      class strata trt01p;
      model W16_easi_pchg = strata trt01p easibl;
      lsmeans trt01p / stderr;
      estimate 'Diff Rilzabrutinib vs Placebo' trt01p -1 1;
      ods output LSMeans=implsmeans Estimates=implsmeandiff;
run;
```

4. Applying Rubin's rule to combine analysis results (point estimates and standard errors) from 40 imputations using PROC MIANALYZE for the LS means and difference in LS means between rilzabrutinib and placebo. Sample code:

```
proc sort data= implsMeans; by trt01pn _imputation; run;
proc mianalyze data= implsmeans;
      by trt01pn;
```

```
      modeleffects lsmean;
      stderr stderr;
      ods output ParameterEstimates=lsmeans;
run;

proc mianalyze data= implsmeandiff;
      modeleffects estimate;
      stderr stderr;
      ods output ParameterEstimates=lsmeandiff;
run;
```

5.6 METHODS AND SIMULATIONS FOR BAYESIAN DATA BORROWING APPROACH

5.6.1 Methods

For the test of the primary endpoint, a Bayesian version of the t-test will be used. In its frequentist version, the t-test is based on the test statistic:

$$T = \frac{\hat{\delta}}{\hat{\sigma} * \sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

where: $\hat{\delta}$ is the difference (rilzabrutinib – placebo) on percent change in EASI score from baseline to Week 16, $\hat{\sigma}$ is the observed standard deviation, n_1 is the number of participants in the rilzabrutinib arm and n_2 is the number of participants in the placebo arm.

Therefore, the alternative hypothesis of the two-sided t test is T≠0. It can also be noted that:

$$T \neq 0 \Leftrightarrow \frac{\hat{\delta}}{\hat{\sigma}} \neq 0$$

Note that efficacy of rilzabrutinib will be based on the posterior probability that the mean percent change in EASI score from baseline to Week 16 of rilzabrutinib group is greater than that of the placebo group and a negative difference (rilzabrutinib – placebo) would be indicative of the superiority of rilzabrutinib. Thus, a Bayesian version of the t-test can be based on the posterior probability:

$$\Pr[\frac{\delta}{\sigma} < 0 \mid data]$$

A posterior probability of at least 95% will be considered significant.

Simulations confirmed the validity of the proposed Bayesian test. In particular, these simulations showed that when a non-informative prior distribution is used for all parameters of the model, the proposed Bayesian test is very similar to the Frequentist t-test. See
Section 5.6.2.3 for details.

Historical placebo data on percent change in EASI score from baseline to Week 16

Mean and standard deviation of the percent change in EASI score from baseline to Week 16 in the placebo arm of the external historical study are described in Table 14.

Table 14 - Percent change in EASI score from baseline to Week 16 in participants from the placebo arm of historical study

Number of participants	Mean	Standard Deviation
n0=142	μ0=-34.51%	σ0=30.9%

Dynamic borrowing of historical placebo data

In the supplemental analysis, an informative prior distribution will be used for parameter μ_2 (the mean percent change in EASI score from baseline to Week 16 in participants from the placebo arm of the current study). The prior distribution will be based on data collected from 142 participants treated with placebo in the historical study. The simulations are based on the placebo data from the historical study.

The amount of information borrowed from historical study will be estimated from the degree of similarity between placebo data from the historical study and the current study. A parameter denoted by a_0 (with values from 0 to 1) will control the amount of information borrowed from the historical placebo data. The prior distribution of μ_0 (the mean percent change in EASI score from baseline to Week 16 in participants from the placebo arm of the historical study) will be raised to the power a_0 and restandardized to a proper distribution. This is equivalent to multiplying the variance of the prior distribution by $1/a_0$. The extreme case of $a_0 = 1$ corresponds to full borrowing, whereas the other extreme case of $a_0 = 0$ corresponds to no borrowing.

Following Psioda and Ibrahim (3), the parameter a_0 will be set to:

$$\widehat{a0} = \left(\frac{\mathcal{L}(\widehat{\mu_0}|data)}{\mathcal{L}(\widehat{\mu_2}|data)}\right)^{s0}$$

$$= \exp\left(-\frac{n_2}{2*\sigma^2}(\mu_0 - \mu_2)^2\right)^{s0}$$

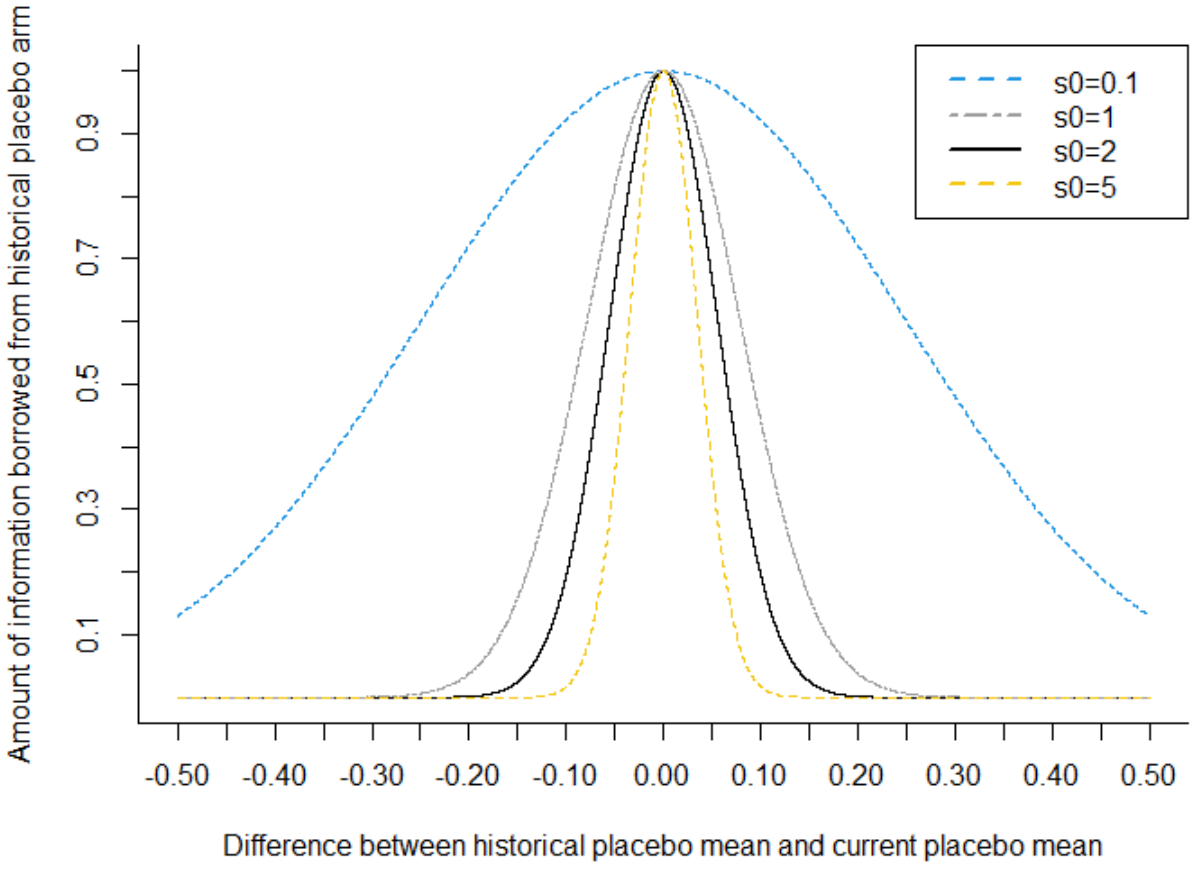
Where $\mathcal{L}(.|data)$ is the likelihood of the data from the current placebo arm, μ_0 is the historical placebo mean, μ_2 is the current placebo mean, n_2 is the number of participants in current placebo arm, σ^2 is the common deviation of the current data and $s_0 > 0$ is a calibration parameter. For s_0 close to 0, virtually all the information will be borrowed regardless of the magnitude of the difference between $\widehat{\mu_0}$ and $\widehat{\mu_2}$. For large value of s_0, the historical placebo data will be discounted greatly for even modest difference.

As shown in Figure 1 the amount of information borrowed from historical placebo data will be maximal if there is no difference between the historical placebo data and the current placebo data with regards to the mean percent change in EASI score from baseline to Week 16 in the placebo

arm. The amount of information borrowed will decrease with increasing difference and will depend on the calibration parameter s_0.

Simulations were performed to evaluate the power and Type I error control while borrowing historical placebo data. In order to not let the historical placebo data dominate the results, the information borrowed was restricted so that the effective historical sample size (EHSS) will not exceed 10 for the BID cohort (30: 20 between rilzabrutinib and placebo) and 14 for the TID cohort (42: 28 between rilzabrutinib and placebo). The simulations in this SAP are based on BID cohort. Simulations showed that a value $s_0 = 2$ provides relative higher power, while preserving the Type I error rate. Further details about simulations are provided in Section 5.6.2.

Figure 1 - Amount of information borrowed from historical placebo data as a function of the difference between historical and current placebo data



Prior distribution of parameters of the statistical model

For μ_2 (the mean percent change in EASI score from baseline to Week 16 in participants from the placebo arm), the prior distribution will be a normal distribution with

mean = $\widehat{\mu_0}$ and variance = $\frac{\widehat{\sigma_0^2}}{\widehat{a0}n0}$.

Simulations were conducted to evaluate the probability to declare superiority of rilzabrutinib. Different scenarios regarding the actual treatment effect on percent change in EASI score from baseline to Week 16 were considered:

- A scenario where rilzabrutinib is superior to placebo, with an effect size higher than expected
- A scenario where rilzabrutinib is superior to placebo, with an effect size as expected
- A scenario where rilzabrutinib is superior to placebo, with an effect size lower than expected
- A scenario where rilzabrutinib is not superior to placebo.

These scenarios are summarized in Table 15.

Table 15 - Simulated scenarios for the primary endpoint

Endpoint	Simulated scenarios			
	Mean in placebo arm	Mean difference between rilzabrutinib arm and placebo arm	Common SD	Effect size (ES)
Percent change in EASI score from baseline to Week 16	-25%	-30%	35%	0.857
		-25%	35%	0.714
		-20%	35%	0.571
		0%	35%	0
	-30%	-30%	35%	0.857
		-25%	35%	0.714
		-20%	35%	0.571
		0%	35%	0
	-35%	-30%	35%	0.857
		-25%	35%	0.714
		-20%	35%	0.571
		0%	35%	0
	-40%	-30%	35%	0.857
		-25%	35%	0.714

Endpoint	Simulated scenarios			
	Mean in placebo arm	Mean difference between rilzabrutinib arm and placebo arm	Common SD	Effect size (ES)
		-20%	35%	0.571
		0%	35%	0
	-45%	-30%	35%	0.857
		-25%	35%	0.714
		-20%	35%	0.571
		0%	35%	0

All combinations of scenarios regarding the percent change in EASI from baseline to Week 16 were assessed. A total of 1, 000 simulated datasets were generated for each of the 20 scenarios. In each simulated dataset, a total of 50 participants from the BID cohort of the current study (30 participants on rilzabrutinib and 20 participants on placebo) with available data on EASI were generated. Percent change in EASI score for each participant were randomly generated from a normal distribution with the mean specific to each scenario and the common SD is assumed to be 35% from the observed data of current study.

Each simulated dataset was then analyzed with the following methods:

- Bayesian approach with dynamic borrowing of historical placebo data (as described in Section 3.2.4)
- Frequentist approach with 2-sided t test, only using data from current study.

For the simulated datasets, superiority of rilzabrutinib was declared if the following condition was met:

- Bayesian approach: $\Pr[\frac{\delta}{\sigma} < 0 \mid data]$ >0.95
- Frequentist approach: one-sided p-value was < 0.05 (two-sided p-value <0.10).

5.6.2 Simulations

5.6.2.1 Probability to declare superiority of rilzabrutinib

Probability to declare superiority of rilzabrutinib is presented in Table 16.

- For all scenarios where a difference between groups exists, the Bayesian approach with dynamic borrowing increases the power according to the similarity between the historical placebo data and the current placebo data. When the observed placebo effect of current placebo arm is lower than that of the historical placebo arm, the borrowed data will favor the current placebo arm and vice versa.

- If the mean percent change in EASI score in the current placebo arm has a lower effect than that in the historical placebo arm, the amount of the increase is generally less than or around 5%.
- If the mean percent change in EASI score in the current placebo arm is similar with that in the historical placebo arm, the amount of the increase is generally about or more than 10%.
- If the mean percent change in EASI score in the current placebo arm has a higher effect than that in the historical placebo arm, the amount of the increase is around 10% but the increase is also discounted due to the increase of the difference between current placebo mean and historical placebo mean.

- For scenario where no between-group difference exists and mean percent change in EASI score in the current placebo arm is very different from that in the historical placebo arm, the impact on the Type I error is minimal.

Table 16 - Probability to declare superiority of rilzabrutinib under different scenarios and with different statistical approaches

Scenario		Probability to declare superiority of rilzabrutinib			
Assumed current placebo mean	Difference (ES)	Bayesian approach with dynamic borrowing		Frequentist approach not borrowing historical placebo data	
		Power/Type I error	EHSS	30:20	30:30
-25%	-30% (0.857)	90.9%	8	90%	94%
	-25% (0.714)	79.4%		78%	86%
	-20% (0.571)	59.6%		62%	70%
	0% (0)	3.8%		5%	
-30%	-30% (0.857)	94.6%	10	90%	94%
	-25% (0.714)	83.7%		78%	86%
	-20% (0.571)	67.7%		62%	70%
	0% (0)	3.4%		5%	
-35%	-30% (0.857)	96%	10	90%	94%
	-25% (0.714)	88.0%		78%	86%
	-20% (0.571)	74.8%		62%	70%
	0% (0)	3.9%		5%	
-40%	-30% (0.857)	94.5%	9	90%	94%
	-25% (0.714)	88.9%		78%	86%
	-20% (0.571)	76.8%		62%	70%
	0% (0)	4.9%		5%	

Scenario		Probability to declare superiority of rilzabrutinib			
Assumed current placebo mean	Difference (ES)	Bayesian approach with dynamic borrowing		Frequentist approach not borrowing historical placebo data	
		Power/Type I error	EHSS	30:20	30:30
-45%	-30% (0.857)	92%	7	90%	94%
	-25% (0.714)	85.4%		78%	86%
	-20% (0.571)	75%		62%	70%
	0% (0)	7.3%		5%	

$EHSS = n_0 * a_0$, where a_0 is the information borrowed from historical placebo arm and n_0 is the sample size of historical placebo arm. The EHSS is the average value of the 1,000 simulations.
Probabilities based on 1,000 simulated trials for each scenario. Each simulated trial included 50 participants with available data on EASI score (30 participants from Rilzabrutinib and 20 participants from placebo group).
Data were simulated assuming a SD for percent change in EASI score from baseline to Week 16 of 35%.
For the Bayesian approach with dynamic borrowing, superiority of rilzabrutinib was declared if $Pr\left[\frac{\delta}{\sigma} < 0 \mid data\right] > 0.95$
For the frequentist approach, superiority of rilzabrutinib was declared if the 1-sided t-test was significant at the α=0.05 (2-sided α = 0.10).

5.6.2.2 Assessment of the calibration parameter

As discussed in Section 3.2.4, the calibration parameter s_0 controls how historical placebo data will be discounted in case of difference between historical placebo data and current placebo data. For s_0 close to 0, virtually all the information will be borrowed regardless of the magnitude of the difference. For large value of s_0, the historical placebo data will be discounted greatly even for modest difference.

The impact of different values of s_0 on the power and Type I error was assessed and is presented in Table 17. Type I error rate was defined as the probability to declare superiority of rilzabrutinib under the hypothesis of no effect of rilzabrutinib on EASI score. Such a scenario would imply that the mean percent change in EASI score from baseline to Week 16 in the current placebo arm would greatly differ from the one observed in the historical placebo arm.

Small values of s_0 such as 0.1 or 1 would maximize the increase in power but would considerably impact the Type I error when the observed placebo has relatively higher treatment effect. On the other hand, large values such as 5 would reduce the power and would limit the interest of borrowing historical placebo arm data. An intermediate value of $s_0 = 2$ appeared to increase the power compared to a Frequentist approach not borrowing data from historical placebo arm (88% compared to 78%) when the difference is -25% between rilzabrutinib and placebo, whereas limiting the impact on the Type I error (3.9% at 1-side under Bayesian approach compared to 5% at 1-side under frequentist approach) when there is no difference between the historical placebo data and the current placebo data.

As discussed in Section 3.2.4, historical study and current study are expected to be similar in terms of population, treatment and assessment of the primary endpoint. It is believed that historical placebo data and current placebo data would not largely differ. The amount of information borrowed from the historical placebo arm has been capped to not exceed 10 in order

not to let the historical placebo data dominate the results. For this reason, the dynamic borrowing approach with $s_0 = 2$ is considered reasonable.

Table 17 - Power and Type I error for different values of the calibration parameter (S0) under different assumed true placebo means

Assumed current placebo mean		Bayesian approach with dynamic borrowing and calibration parameter (s_0)								Frequentist approach (no borrowing)	
		0.1	EHSS	1	EHSS	2	EHSS	5	EHSS	30:20	30:30
-25%	Power	78.8%		78.6%		79.4%		79.9%		78%	86%
	Type I error	0.6%	10	2.7%	10	3.8%	8	4.1%	6	5%	
-30%	Power	83.6%		83.4%		83.7%		83.0%		78%	86%
	Type I error	1.4%	10	2.1%	10	3.4%	10	4.1%	8	5%	
-35%	Power	88.9%		88.6%		88.0%		83.5%		78%	86%
	Type I error	2.9%	10	3.2%	10	3.9%	10	4.5%	8	5%	
-40%	Power	93.6%		92.5%		88.9%		83.6%		78%	86%
	Type I error	4.6%	10	4.5%	10	4.9%	9	4.9%	7	5%	
-45%	Power	96.4%		91.6%		85.4%		79.9%		78%	86%
	Type I error	8.0%	10	8.0%	9	7.3%	7	6.9%	5	5%	

$EHSS = n_0 * a_0$, where a_0 is the information borrowed from historical placebo arm and n_0 is the sample size of historical placebo arm. The EHSS is the average value of the 1,000 simulations.
Probabilities based on 1,000 simulated trials for each scenario. Each simulated trial included 50 participants with available data on EASI score (30 participants from Rilzabrutinib and 20 participants from placebo group).
Data were simulated assuming a SD for percent change in EASI score from baseline to Week 16 of 35%.
Power was defined as the probability to declare superiority of rilzabrutinib under the scenario of effect size 25% for percent change in EASI score from baseline to Week 16. Type I error rate was defined as the probability to declare superiority of rilzabrutinib under the scenario of effect size 0 for percent change in EASI score from baseline to Week 16.
For the Bayesian approach with dynamic borrowing, superiority of rilzabrutinib was declared if $Pr\left[\frac{\delta}{\sigma} < 0 \mid data\right] > 0.95$
For the frequentist approach, superiority of rilzabrutinib was declared if the 1-sided t-test was significant at the α=0.05 (2-sided alpha = 0.10).

5.6.2.3 Comparison of Frequentist approach vs. Bayesian approach with non-informative prior distribution

The Bayesian version of the t test has been developed for the purpose of this study. In order to better understand its characteristics, additional simulations have been run using a non-informative prior distribution for all the parameters of the model, including μ1 (the mean percent change in EASI score from baseline to Week 16 in the rilzabrutinib group). It is expected that using a non-informative prior distribution, the Bayesian approach should yield similar results compared to the Frequentist approach.

Similarity was assessed by comparing the one-sided p-value from the t test performed in a Frequentist framework to the posterior probability $Pr\left[\frac{\delta}{\sigma} < 0 \mid data\right]$ from the Bayesian approach with non-informative prior distribution.

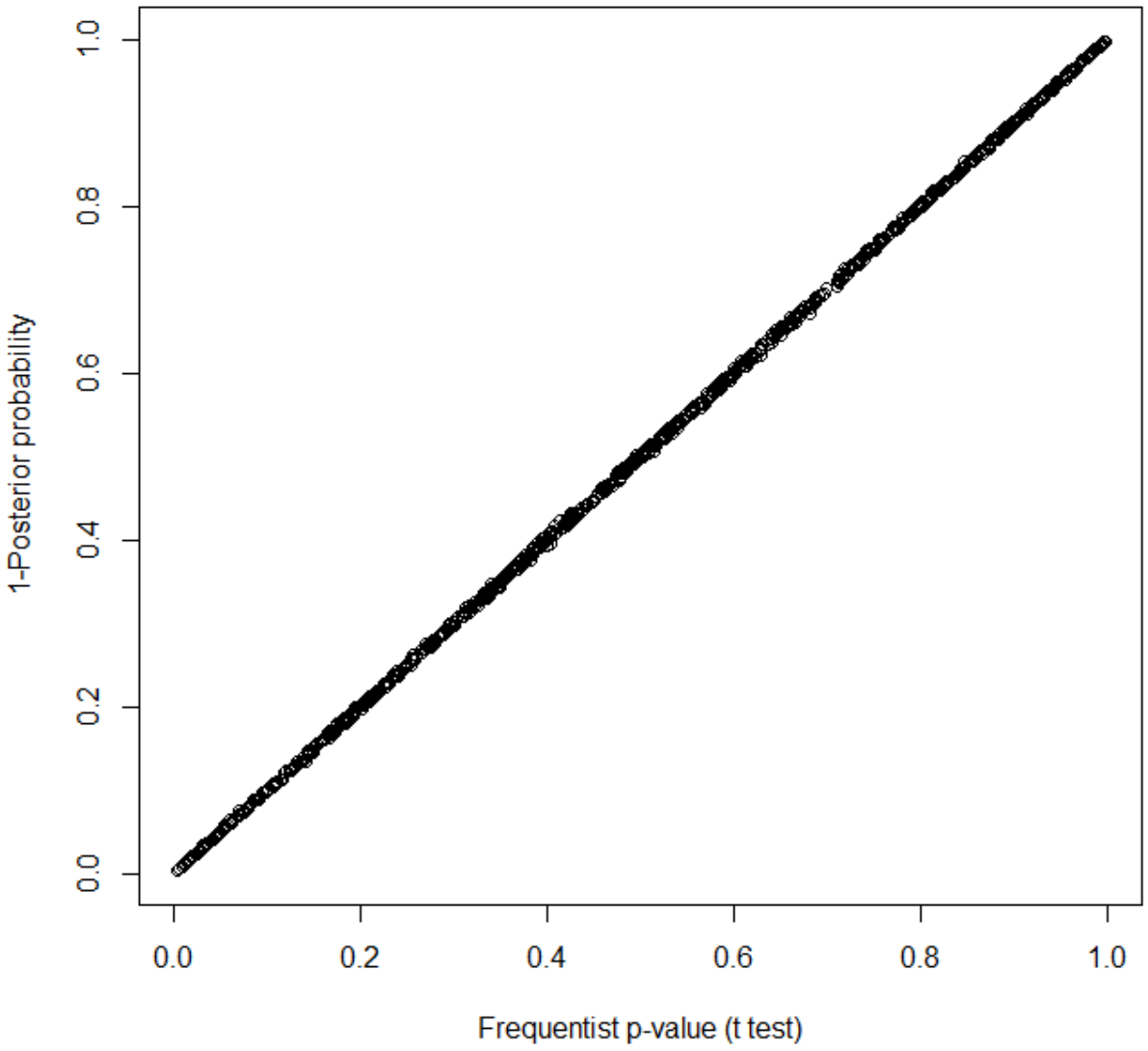
Data from 1,000 simulated trials generated under the null hypothesis were analyzed using these two approaches. For each simulated trial, the one-sided p-value from the Frequentist t test was plotted against $1 - Pr\left[\frac{\delta}{\sigma} < 0 \mid data\right]$ from the Bayesian analysis with non-informative prior.

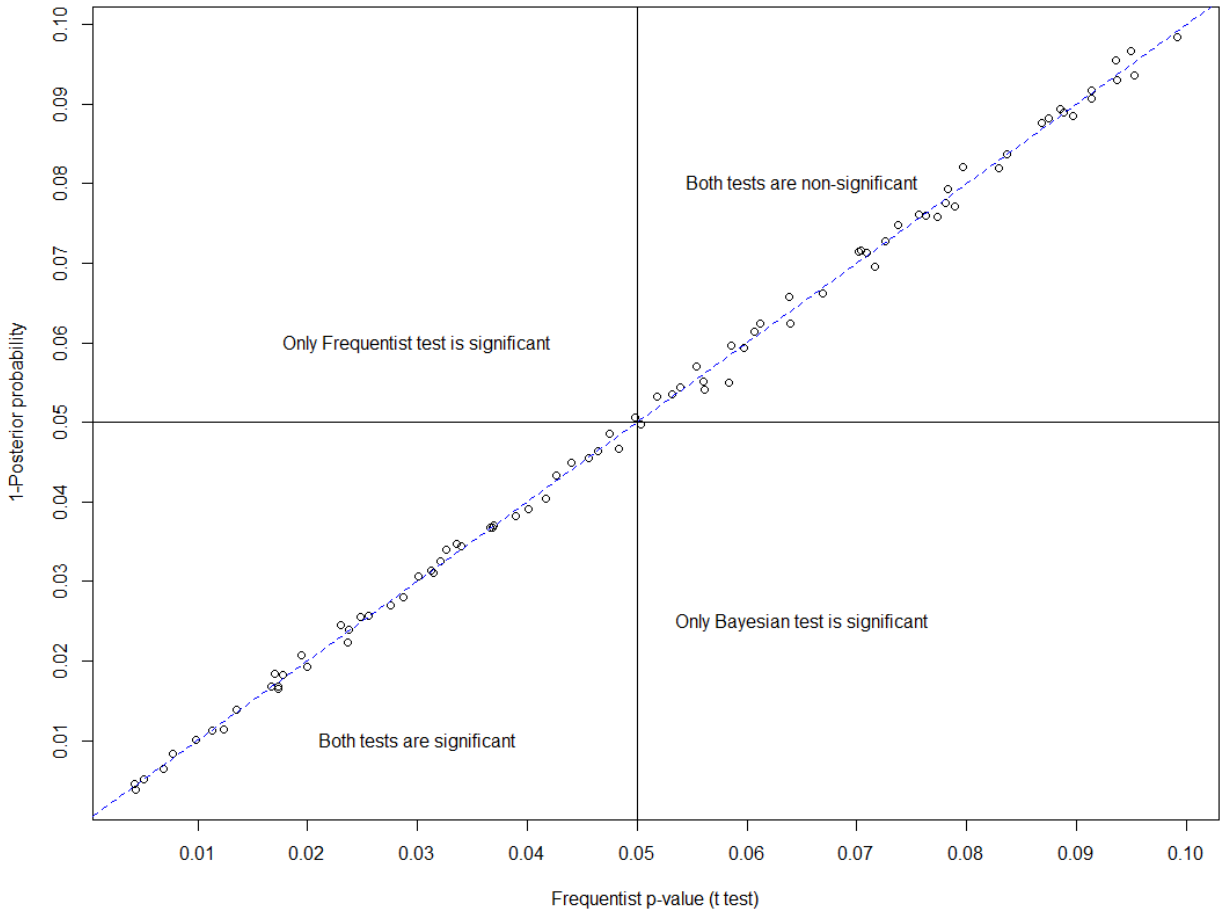
Figure 2 shows that all points were very close to the identity line, showing very strong agreement between the Frequentist approach and the Bayesian approach when a non-informative prior distribution is used. In addition, the second part of this figure (zoom in the zone from 0 to 0.1) shows that in the vast majority of simulated cases, the Frequentist and Bayesian approach led to the same conclusion regarding the significance of the treatment effect. Consistent conclusions were found in 99.8% of simulated trials. When inconsistent conclusions were reached (in 0.2% of simulated cases), they were similarly distributed between cases where only the Frequentist approach was significant (0.1%) and cases where only the Bayesian approach was significant (0.1%).

These findings confirm the validity of the proposed Bayesian version of the t test.

Of note, these simulations were generated under the null hypothesis, where p-value are uniformly distributed in the interval [0,1]. Similar results were observed in other scenarios.

Figure 2 - Comparison of Frequentist p-value vs. Bayesian posterior probability in simulated trials





Based on 1,000 simulated trials.
Each dot represents the pair of one-sided p-value (alpha=0.05) from t test versus $1 - Pr\left[\frac{\delta}{\sigma} < 0 \mid data\right]$ from a simulated trial.
Each simulated trial included 30 participants from rilzabrutinib and 20 participants from placebo group as planned.
Data were simulated under the null hypothesis (no difference between groups with regards to percent change in EASI score from baseline to week 16) and assuming a SD of 35%.

5.6.2.4 Amount of information borrowed from historical placebo data

The amount of information borrowed from historical placebo data will depend on the degree of similarity between historical placebo data and current placebo data regarding the mean percent change in EASI score from baseline to Week 16. Simulations were performed to evaluate what would be the expected amount of information borrowed, under different scenarios regarding the mean percent change in EASI score from baseline to Week 16 in the placebo arm of current study:

- Mean percent change of -25% (lower effect and different to that of historical placebo data)
- Mean percent change of -30% (similar but slightly lower effect than of historical placebo data)
- Mean percent change of -35% (same as that of historical placebo data)

- Mean percent change of -40% (similar but slightly higher effect than that of historical placebo data)
- Mean percent change of -45% (higher effect and different to that of historical placebo data)

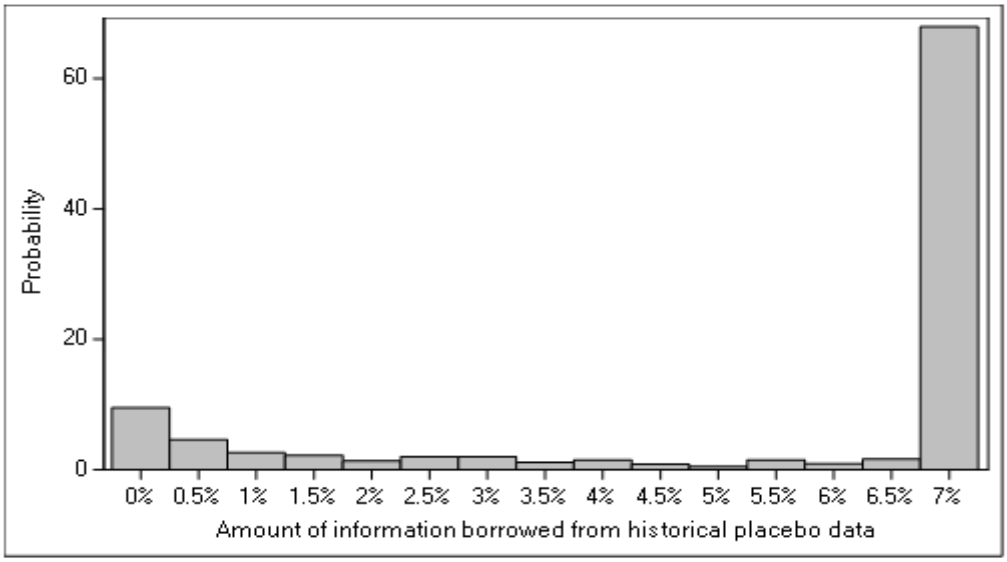
The amount of information borrowed from historical placebo data was calculated in 1,000 simulated trials for each scenario, and the median of these 1,000 simulated trials is presented in Table 18. The complete distribution under the 5 scenarios is presented in Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7.

Table 18 - Amount of information borrowed from historical placebo data under different scenarios

Assumed mean percent change in EASI score from baseline to Week 16 in the placebo arm of current study	Amount of information borrowed from historical placebo arm Median (Q1 ; Q3) from 1,000 simulated trials
-25%	7.04% (3.58%; 7.04%)
-30%	7.04% (7.04%; 7.04%)
-35%	7.04% (7.04%; 7.04%)
-40%	7.04% (7.04%; 7.04%)
-45%	7.04% (1.63%; 7.04%)

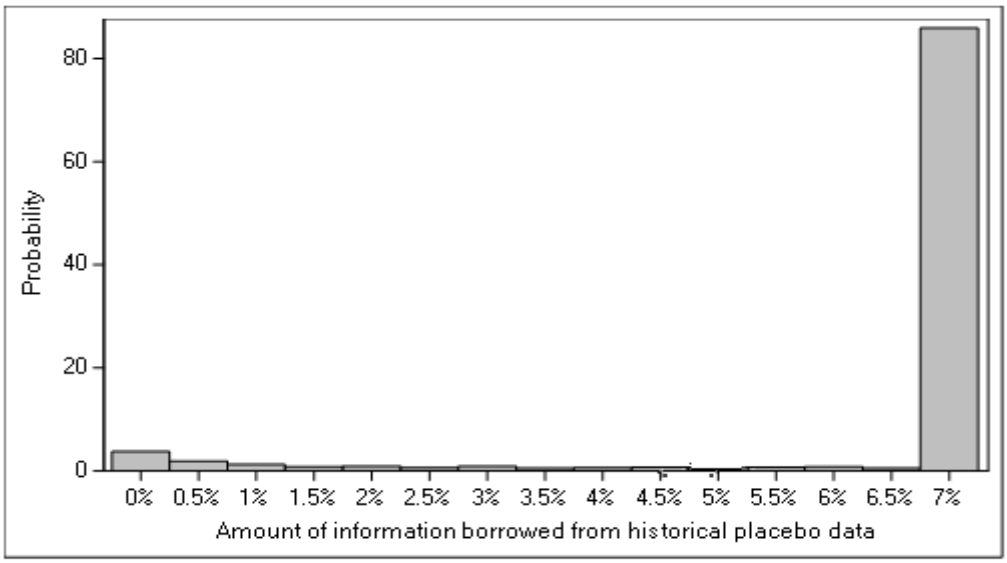
Based on 1,000 simulated trials for each scenario.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

Figure 3 - Amount of information borrowed from the historical study, assuming percent change of -25% in the placebo arm of the current study



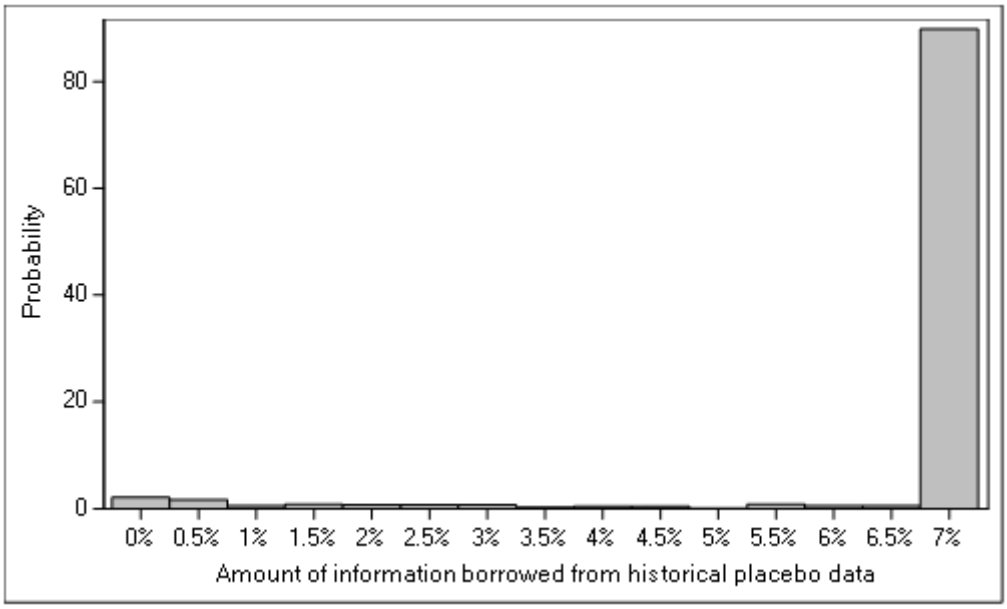
Based on 1,000 simulations.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

Figure 4 - Amount of information borrowed from the historical study, assuming percent change of -30% in the placebo arm of the current study



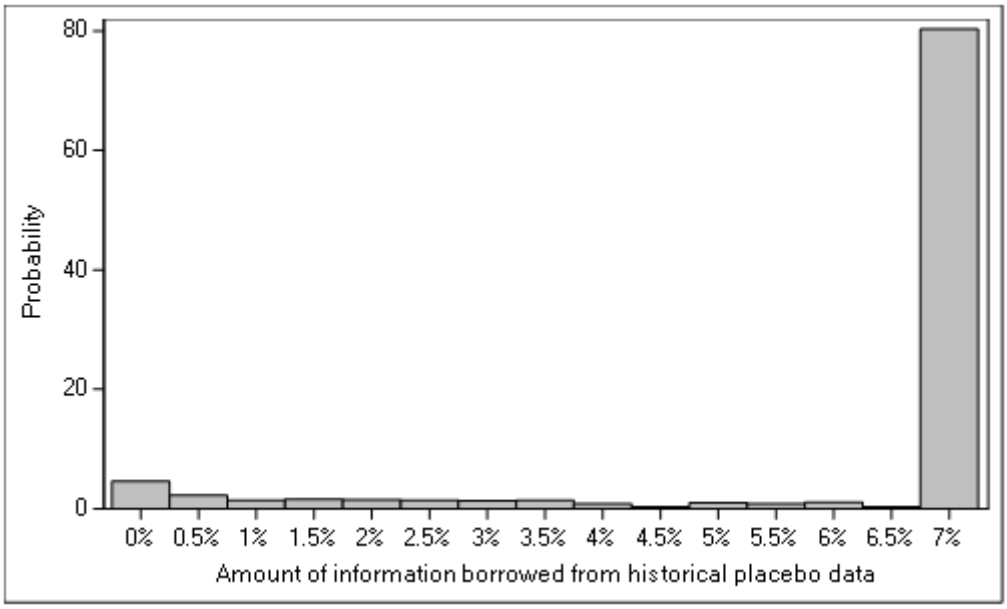
Based on 1,000 simulations.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

Figure 5 - Amount of information borrowed from the historical study, assuming percent change of -35% in the placebo arm of the current study



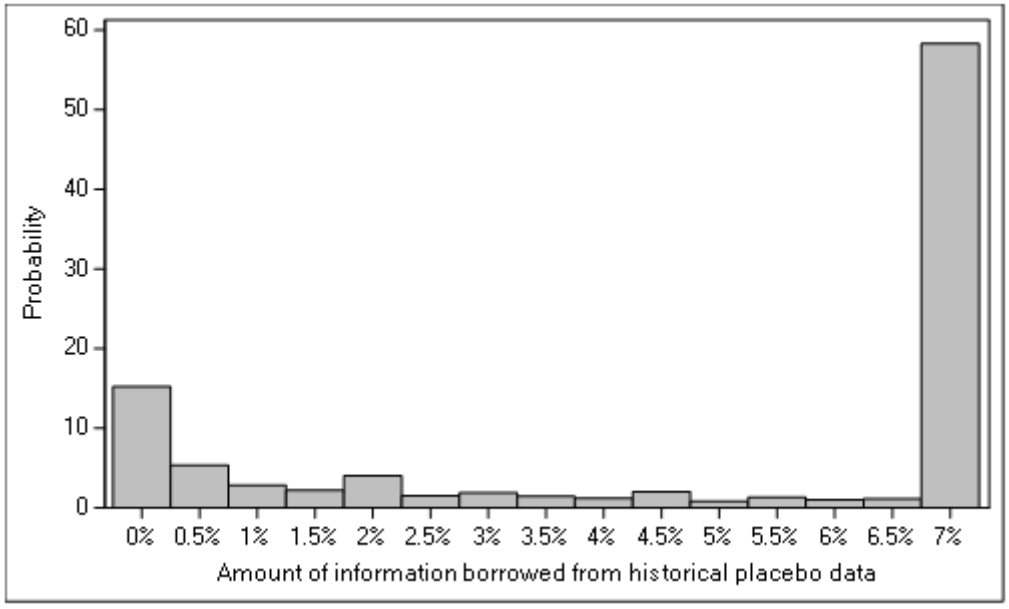
Based on 1,000 simulations.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

Figure 6 - Amount of information borrowed from the historical study, assuming percent change of -40% in the placebo arm of the current study



Based on 1,000 simulations.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

Figure 7 - Amount of information borrowed from the historical study, assuming percent change of -45% in the placebo arm of the current study

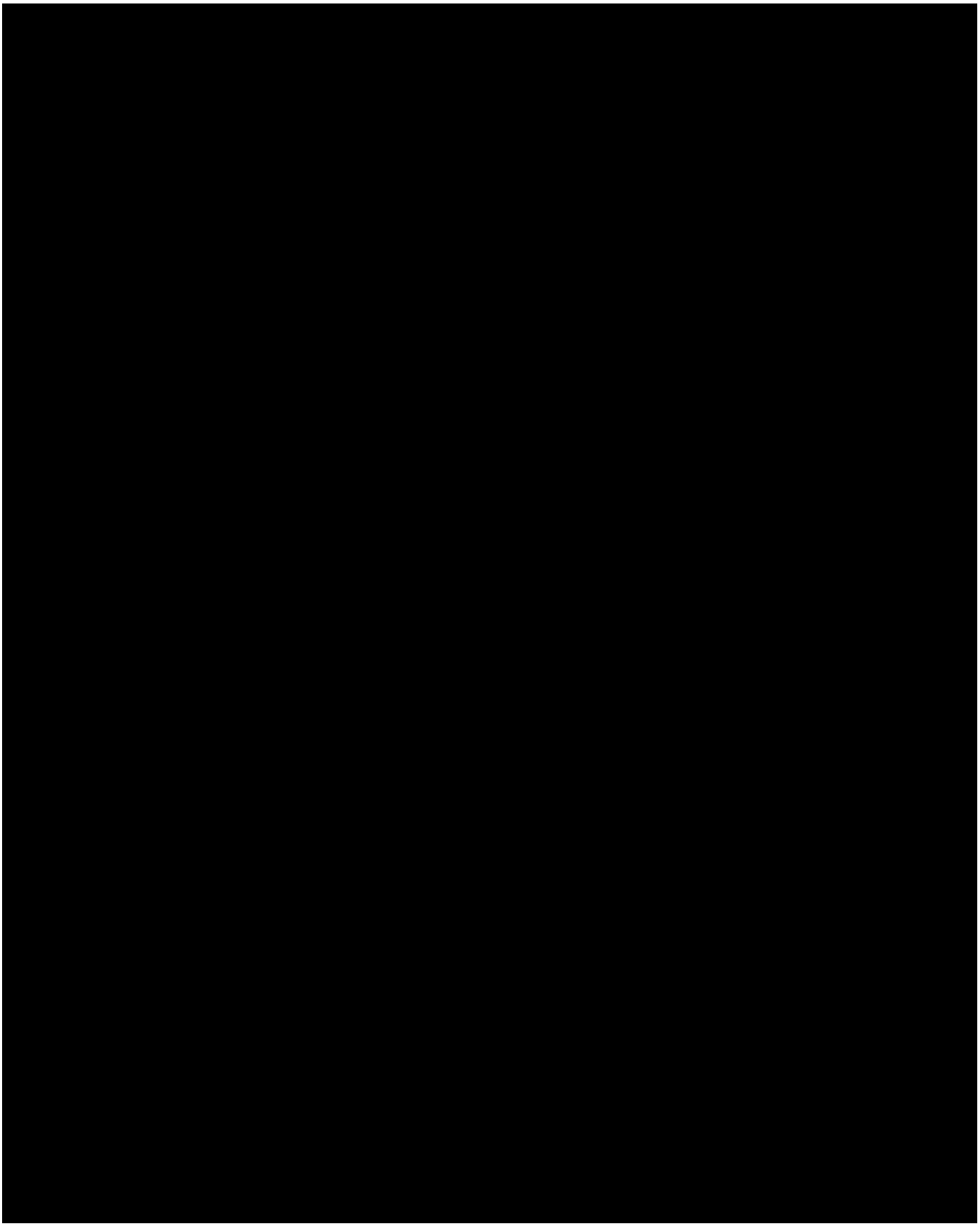


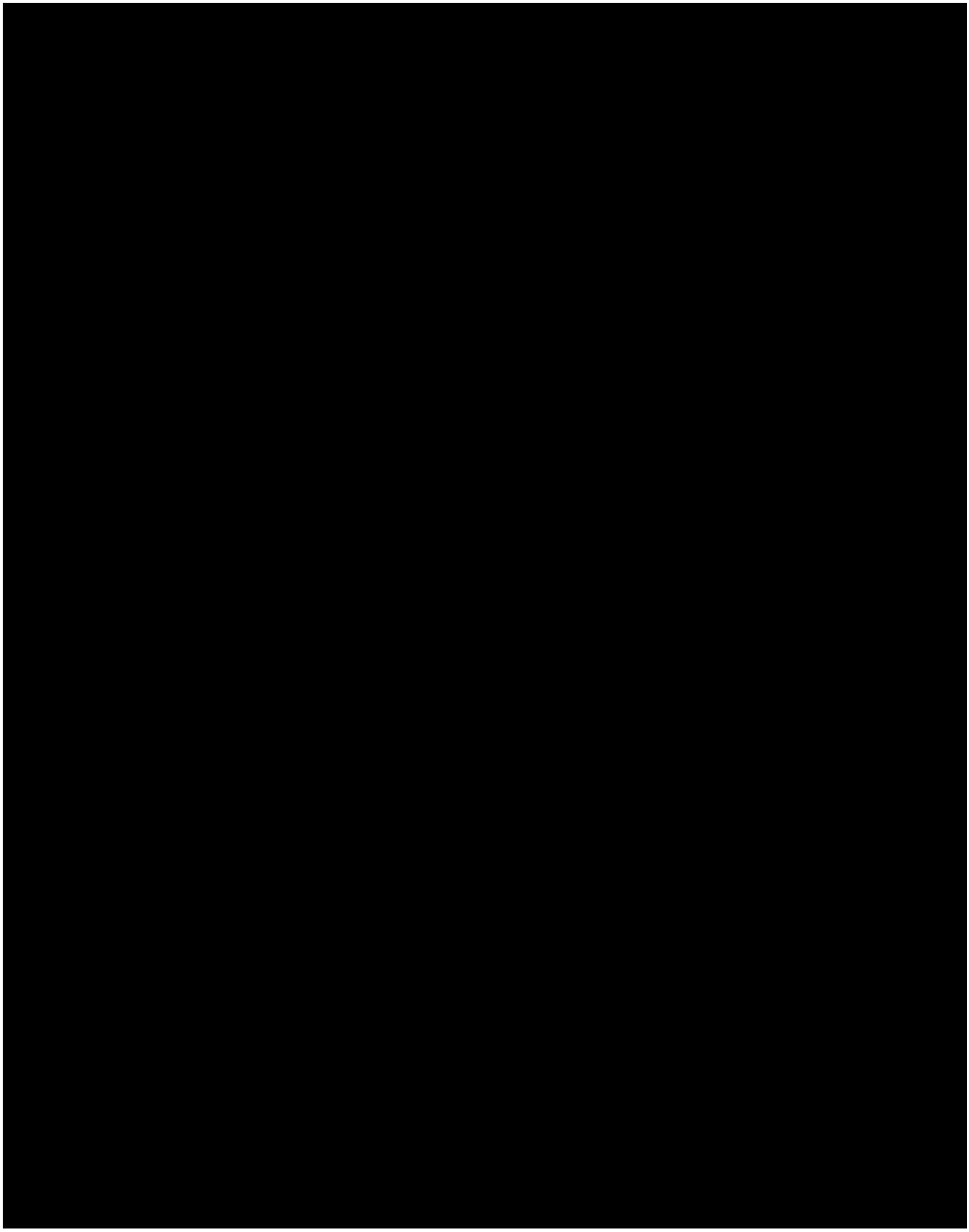
Based on 1,000 simulations.
Each simulated trial included 50 participants with simulated data on percent change in EASI score (rilzabrutinib: placebo = 30: 20). Data were simulated assuming a treatment difference of 25% and a common SD of percent change in EASI score of 35%.

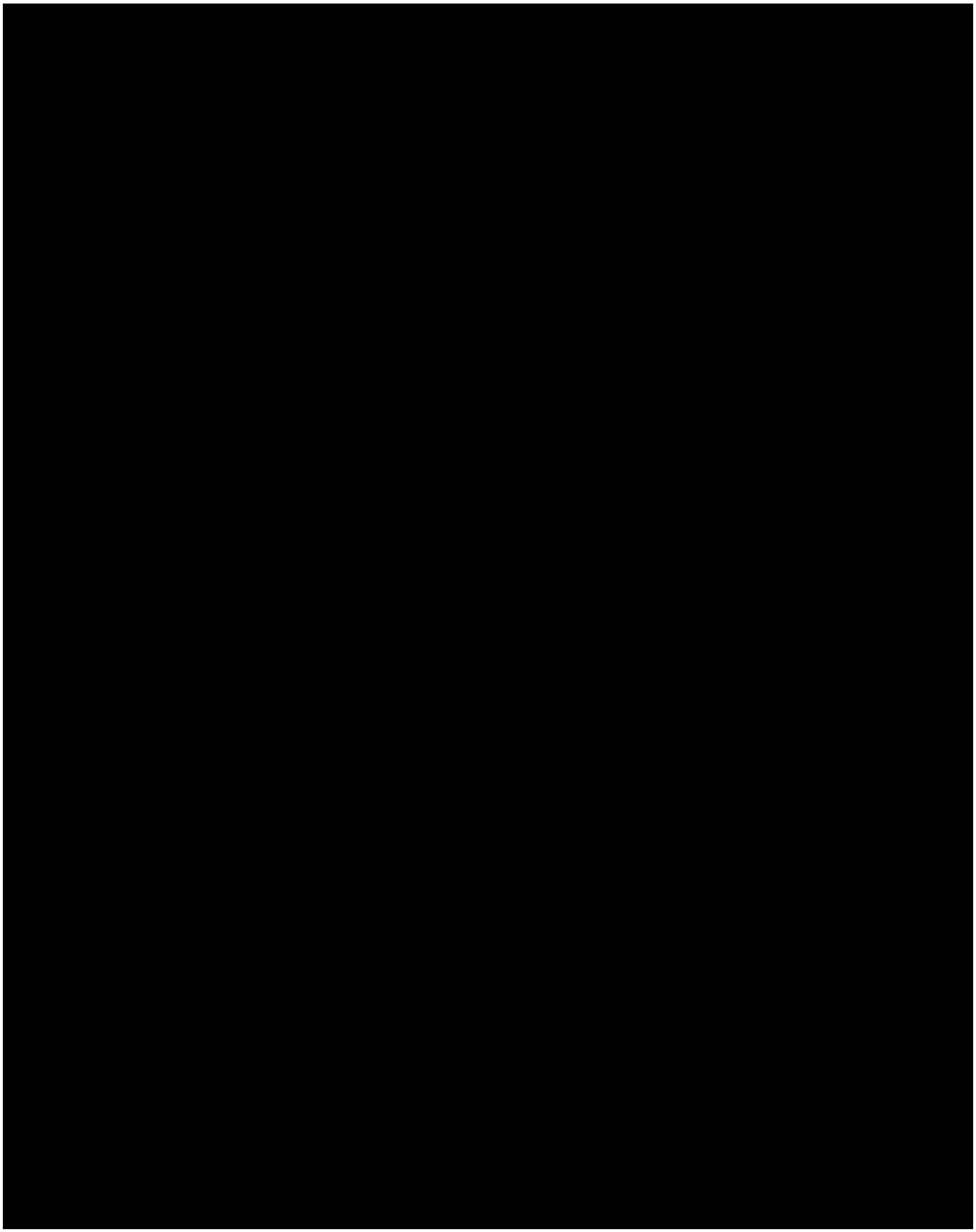
5.6.3 SAS code of simulations

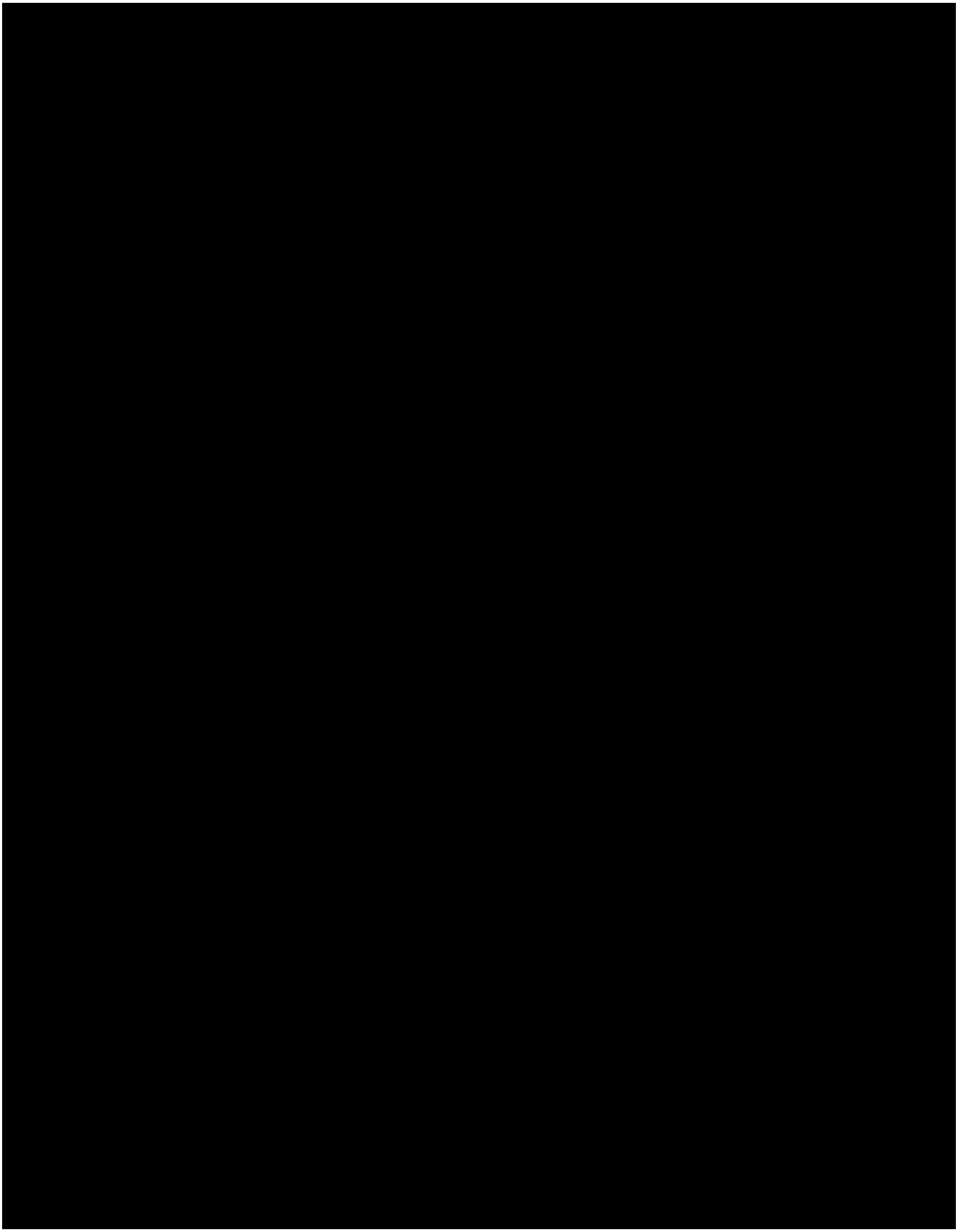
[redacted]

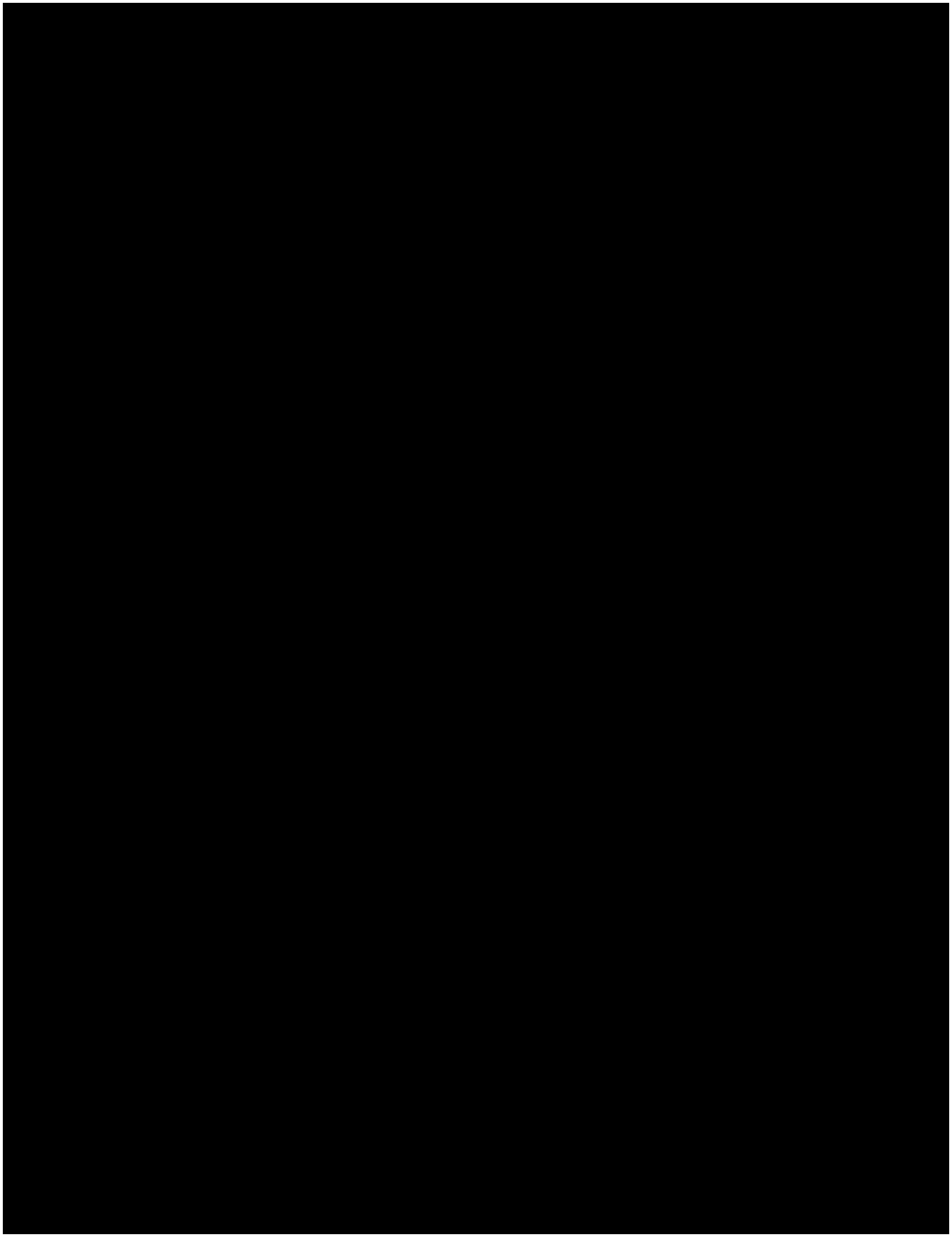
[redacted]

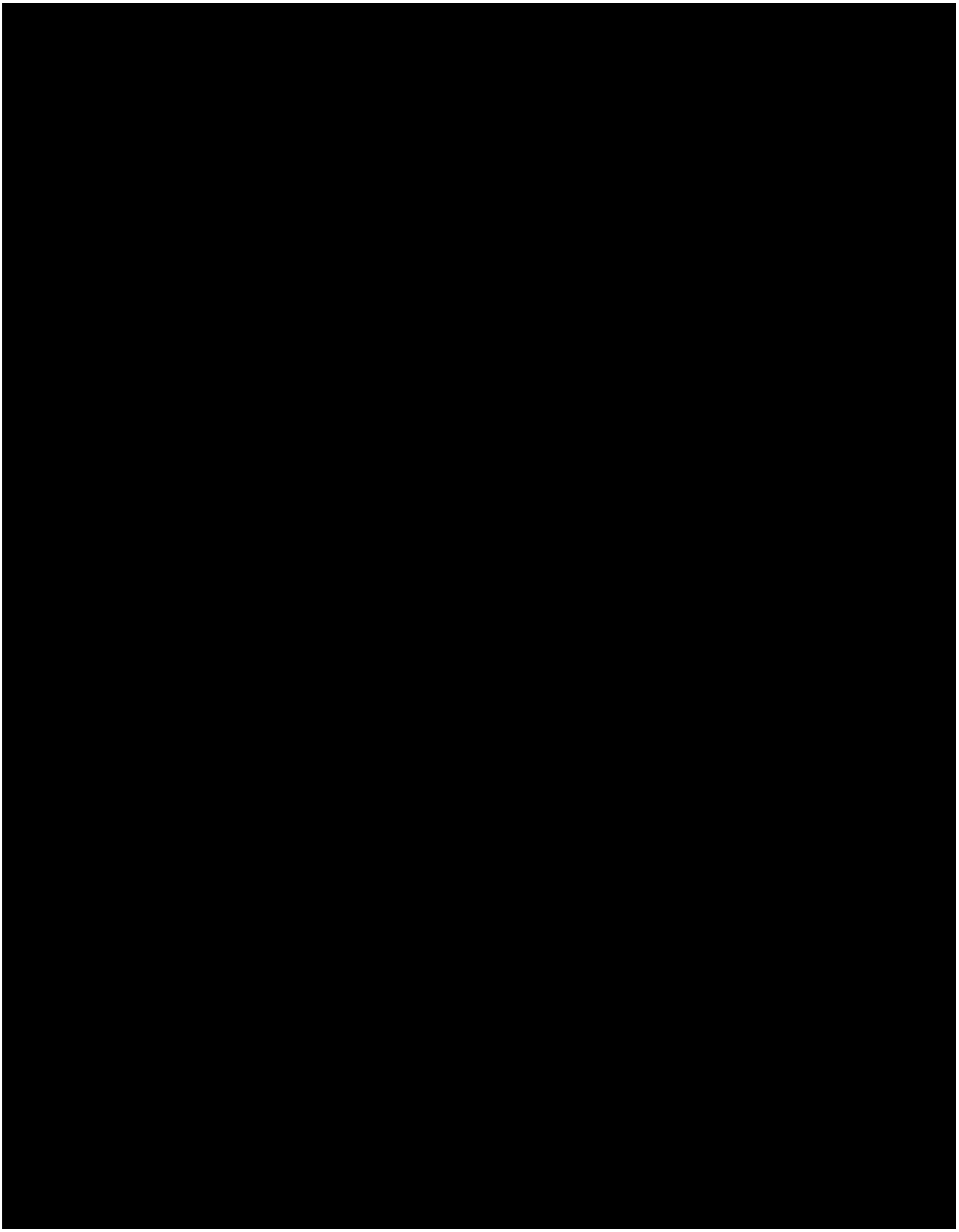












6 REFERENCES

1. https://www.ema.europa.eu/en/medicines/human/EPAR/rinvoq

2. Pocock S. The combination of randomized and historical controls in clinical trials, Journal of Chronic Diseases. 1976, 175-88.

3. Psioda M, Ibrahim J. Bayesian clinical trial design using historical data that inform the treatment effect. Biostatistics 2019:400-15.

Signature Page for VV-CLIN-0636783 v1.0
act17207-16-1-9-sap

Approve & eSign	[redacted]
Approve & eSign	[redacted]