Document Type:	Statistical Analysis Plan
Official Title:	A Phase III, randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of copanlisib in combination with rituximab in patients with relapsed indolent B-cell non-Hodgkin's lymphoma (iNHL) – CHRONOS-3
NCT Number:	NCT02367040
Document Date:	26-AUG-2020

Title page

A Phase III, randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of copanlisib in combination with rituximab in patients with relapsed indolent B-cell non-Hodgkin's lymphoma (iNHL) – CHRONOS -3

Copanlisib and rituximab in relapsed iNHL

Bayer study drug	BAY 80-6946 / Copanlisib		
Study purpose:	Efficacy and safety (Quality of Life and biomarker)		
Clinical study phase:	III	**Date:**	26 AUG 2020
Study No.:	17067	**Version:**	6.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Table of Tables

Table of Figures

List of Abbreviations

AE	Adverse event
AKT	Protein kinase B
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BMI	Body mass index
BSA	Body Surface Area
CHOP	Cyclophosphamide, hydroxydaunorubicin, vincristine, prednisolone
CI	Confidence interval
CSP	Clinical study protocol
CR	Complete response
CRF	Case Report Form
CRR	Complete tumor response rate
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria Adverse Event
CVP	Cyclophosphamide, vincristine, prednisolone
DMC	Data Monitoring Committee
DOR	Duration of response
DRS-E	Disease-related symptoms - emotional
DRS-P	Disease-related symptoms – physical
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
e.g.	For example (*exempli gratia*)
eGFR	Estimated Glomerular Filtration Rate
EOT	End-of-treatment
FAS	Full Analysis Set
FL	Follicular Lymphoma
FLymSI-18	NCCN-FACT Lymphoma Symptom Index-18 Questionnaire
FWB	Function and well-being
GCP	Good Clinical Practice
h	Hour(s)
H_0	Null hypothesis
H_1	Alternative hypothesis
HbA1c	Glycated hemoglobin
ID	Identifier
i.e.	That is (*id est*)
IgM	Immunoglobulin M
ImpDRSP	Improvement in disease-related symptoms - physical
iNHL	indolent non-Hodgkin's lymphoma
IRC	Independent review committee
ITT	Intention-to-treat
IV	Intravenous
IxRS	Interactive Voice/Web Response System
kg	Kilogram
LKAD	Last known alive date
LPL	Lymphoplasmacytic lymphoma
m	Meter
M-1	Metabolite 1
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
min	Minutes
MR	Minor response
MRI	Magnetic resonance imaging
mTOR	Mammalian target of rapamycin
MUGA	Multiple gated acquisition

MZL	Marginal-zone lymphoma
NCCN-FACT	National Comprehensive Cancer Network – Functional Assessment of Cancer
NCI	National Cancer Institute
NE	Non-evaluable
NHL	Non-Hodgkin's lymphoma
NYHA	New York Heart Association
OEE	Overall extent of exposure
ORR	Objective tumor response rate
OS	Overall survival
PBMQ	Product-specific Bayer Medical Queries
PD	Progressive disease
PFS	Progression-free survival
PI3K	Phosphatidylinositol 3-kinase
PK	Pharmacokinetics
PT	Preferred Term
PR	Partial response
SAC	Statistical Analysis Center
SAE	Serious adverse event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SLL	Small lymphocytic lymphoma
SOC	Primary System Organ Class
TEAE	Treatment-emergent adverse event
TSE	Treatment side-effect
TTP	Time to progression
US, USA	United States (of America)
VGPR	Very good partial response
VRM	Validity Review Meeting
VRR	Validity Review Report
vs.	As opposed to (*versus*)
WHO-DD	World Health Organization - Drug Dictionary
WM	Waldenström macroglobulinemia

1. Introduction

Non-Hodgkin's lymphomas (NHLs) are a heterogeneous group of lymphoproliferative malignancies typically originating in the lymph nodes, but can involve almost any organ tissue [1]. NHLs can be divided according to their clinical behavior in two main prognostic groups: indolent NHL (iNHL) and aggressive NHL. Aggressive lymphomas are characterized by an aggressive clinical course and may evolve into a lethal presentation if not immediately treated. Indolent NHLs have a relatively good prognosis with a median survival longer than 10 years, but they are incurable with current available therapeutic options, especially in advanced stages. While they are highly responsive to standard chemotherapy regimens and to radiotherapy, their natural history is characterized by a continuous pattern of relapses, which can be generally treated with success. But the time to next relapse progressively decreases each time, finally evolving into a refractory disease or in a transformation into an aggressive histologic type.

Optimal treatment of advanced stages of indolent NHL is controversial because of low cure rates with the current therapeutic options. The first-line standard therapy includes rituximab, usually administered together with cytotoxic combinations (CHOP, CVP) or single agents (alkylators, e.g. bendamustine, or purine nucleoside analogs such as fludarabine or 2-chlorodeoxyadenosine).

There is no acknowledged standard treatment for patients with recurrent disease. As long as disease appears to be responsive to rituximab (treatment-free intervals of > 6 months after the previous rituximab-containing treatment), a rituximab-based chemoimmunotherapy using non-cross resistant cytotoxic agents would be administered at the next relapse.

Rituximab is a CD20-directed cytolytic antibody indicated for the treatment of patients with NHL and widely used as a single drug or in combination regimens [2-6]. Retreatment with rituximab alone or in combination has been shown to be feasible and active [7,8].

Copanlisib is a small molecule phosphatidylinositol 3-kinase (PI3K) inhibitor and showed excellent anti-tumor activity in pre-clinical models with up-regulated PI3Kα pathway. The PI3K/AKT/mTOR pathway is one of the prominent pathways that promote cellular survival and constitutively is activated in many types of cancers [9,10].

Considering the pre-clinical profile of copanlisib and the promising preliminary efficacy data from the Phase I study 12871 and the ongoing Phase II study 16349, it is expected that in comparison to rituximab with placebo, copanlisib in combination with rituximab will lengthen progression-free survival (PFS) in patients with relapsed iNHL who have received one or more lines of treatment, including rituximab and who either had a treatment-free interval of ≥ 12 months after completion of the last rituximab-containing treatment, or are considered unwilling to receive chemotherapy/for whom chemotherapy is contraindicated.

This Statistical Analysis Plan (SAP) is based on the integrated clinical study protocol (CSP Amendment 11, version 7.0, dated 22 MAY 2020) and describes the primary analysis of study 17067 to be included in the clinical study report (CSR).

2. Study Objectives

The primary objective of this study is:

- To evaluate whether copanlisib in combination with rituximab is superior to placebo in combination with rituximab in prolonging PFS in patients with relapsed iNHL who have received one or more lines of treatment, including rituximab, and who either had a treatment-free interval of $\geq$ 12 months after completion of the last rituximab-containing treatment, or who are unwilling to receive chemotherapy/ for whom chemotherapy is contraindicated on reason of age, comorbidities, and/or residual toxicity

The secondary objectives of this study are to evaluate:

- The following characteristics of disease-related physical symptoms (DRS-P): 'time to deterioration' and 'time to improvement'
- Other radiological and clinical indicators of treatment efficacy (ORR, DOR, CRR, TTP, and OS)
- Safety and tolerability of copanlisib

The other objectives of this study are to evaluate:

- Pharmacokinetics
- Biomarkers
- Quality of life

Table 2–1 gives an overview of the primary and secondary efficacy objectives and the relevant variables to be analyzed.

Table 2–1: Overview of Efficacy Objectives and Variables

Objective	Variable
Primary objective	Progression-free survival (PFS)*, described in Section 6.2.1
Secondary objective: Response rate	Objective tumor response rate (ORR)*, described in Section 6.2.2
Secondary objective: Characteristics of disease-related physical symptoms	Time to deterioration in DRS-P of at least 3 points*, Time to improvement in DRS-P of at least 3 points*, described in Section 6.2.2

Table 2–1: Overview of Efficacy Objectives and Variables

Objective	Variable
Secondary objective: Other radiological and clinical indicators	Duration of response (DOR), Disease control rate (DCR), Complete response rate (CRR), Time to progression (TTP), Overall survival (OS), described in Section 6.2.2

*Variable is part of the confirmatory testing strategy (see Section 6.2.3)

For the analyses of primary efficacy endpoint and key secondary efficacy endpoints, separate test strategies will be used for the United States and Europe (see Section 6.2.3).

3. Study Design

This is a randomized, double-blind, placebo-controlled, two-arm, Phase III study to evaluate efficacy and safety of copanlisib in combination with rituximab, in comparison to placebo in combination with rituximab, in patients with relapsed iNHL.

Approximately 450 (including follicular lymphoma (FL) and other iNHL) patients who meet the eligibility criteria will be randomly assigned in a 2:1 ratio to one of the double blinded treatment arms: copanlisib plus rituximab or placebo plus rituximab. Patients will be stratified according to four factors based on baseline characteristics:

iNHL histology:

- FL histology
- Other iNHL histology (Small lymphocytic lymphoma (SLL), Marginal-zone lymphoma (MZL), Lymphoplasmacytic lymphoma (LPL)/Waldenström macroglobulinemia (WM))

Entry criterion:

- progression-free and treatment-free interval of ≥ 12 months following completion of the last rituximab- containing treatment
- considered unwilling/unfit to receive chemotherapy

Presence of bulky disease (as defined by the presence of a nodal or extranodal mass ≥ 7 cm in the longest diameter, with the exception of spleen):

- yes
- no

Previous treatment with PI3K inhibitors:

- yes
- no

Resulting from the combination of these four stratification factors, patients will be randomized into 16 different strata (see Table 4–2). The stratification factor 'entry criterion' is an abbreviation of the inclusion criterion 13 from the integrated CSP Amendment 10. The exact definition from the inclusion criterion should be used to assess which entry criterion is fulfilled. If the patient fulfills both entry criteria, patients will be stratified into the criterion "considered unwilling/unfit to receive chemotherapy for age or comorbidities".

The study is composed of the following periods:

- Screening
- Treatment
- Safety follow-up / Active follow-up
- Survival follow-up

The start of the screening period is defined by signing of the informed consent form. The maximum interval allowed between signature of informed consent and start of treatment is 28 days. The start of the treatment period is defined by the first administration of study treatment.

Copanlisib and placebo for copanlisib formulations are administered before rituximab, in a normal saline solution, intravenous (IV), over 1 hour. Dosing is weekly for the first 3 weeks of a 28-day cycle (on Days 1, 8, and 15), followed by a 1-week break (i.e., no infusion on Day 22).

Rituximab will be given IV at 375 mg/m^2 body surface weekly during Cycle 1 (on Days 1, 8, 15 and 22). For those patients who are still on treatment after evaluation at the end of Cycle 2, rituximab will be administered once in each of Cycles 3, 5, 7 and 9 (on Day 1).

Treatment with copanlisib or placebo will be continued until the occurrence of progressive disease (PD; per central independent blinded radiology review) as defined in the Lugano Classification [11] (for patients with WM according to the Owen Criteria [12]), unacceptable toxicity, or until another criterion is met for withdrawal from the study treatment (as described in protocol Section 5.2.1.1). Rituximab treatment will be continued until the same criteria as defined for copanlisib are met, for a maximum of 8 infusions (until Cycle 9).

An end-of-treatment (EOT) visit will be performed no later than 7 days after the decision is made to discontinue study treatment. Patients who discontinue study treatment because of PD will enter the Safety-follow up period and patients who discontinue study drug for reasons other than PD will enter the Active follow-up period (which also serves as a Safety follow-up), except for patients who object to follow-up data collection. The patients in the Active follow-up will have tumor assessments by central independent blinded review until the end of the Active follow-up period, defined as when either PD is documented or a new anti-tumor treatment is administered, whichever occurs first.

All patients will be followed off study for overall survival at 3-monthly intervals during the Survival follow-up period (up to 3 years after the last patient started study treatment), except for patients who object to follow-up data collection. During this period, patients are not considered to be 'on-study'.

The first radiological tumor assessments with IV contrast-enhanced computed tomography/magnetic resonance imaging (CT/MRI) scans of neck, chest, abdomen and pelvis

will be performed at Screening (including WM patients). During the treatment period as well as during the Active follow up period tumor assessments with the same modality will be performed every 8 weeks (± 7 days) during Year 1, every 12 weeks (± 7 days) during Year 2, and every 24 weeks (± 7 days) during Year 3 and onwards, starting from Cycle 1 Day 1. CT/MRI scans are not required at the EOT visit if the patient discontinues due to PD, which has been radiologically confirmed within the 4 weeks preceding EOT.

The response assessment will be done according to the Lugano Classification, and, for patients with WM, according to the Owen Criteria. In the event of progression, radiological real-time confirmation by central independent blinded evaluation is required before a final decision to stop the treatment is made. Radiological real-time confirmation will only be conducted until the database cut-off for the primary analysis. The final evaluation of treatment response (best response: objective tumor response rate and complete response rate) will be done by central blinded review retrospectively.

All efficacy analyses will be performed when at least 190 centrally evaluated PFS events are observed in the study. Evaluations from central blinded review will be used for the primary efficacy analyses of primary and secondary endpoints containing radiological tumor assessments.

4. General Statistical Considerations

4.1 General Principles

Statistical analyses will be conducted by or under the supervision of the sponsor's Study statistician, except for the analysis of biomarker data and pharmacokinetics/pharmacodynamics data, which will be performed by or under the direction of the sponsor's Genomics and Biomarker Statistical Expert and PK experts.

The statistical evaluation will be performed by using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA). All variables will be analyzed by descriptive statistical methods. The number of data available and missing data, mean, standard deviation, minimum, quartiles, median, and maximum will be calculated for metric data. Frequency tables will be generated for categorical data.

Definition of efficacy and safety endpoints, analysis strategies, structure of analysis datasets and layout of analysis data displays are following Bayer standards as documented in the Bayer standard system: Clinical Copanlisib Project Standards, Oncology Therapeutic Area Standards, and Global Medical Standards, respectively. The order reflects the priority of the different standards, where specifications of the latter ones have to be followed only if not specified in standards mentioned before. Study-specific specifications may be included in addition to the project standards, if needed.

4.2 Handling of Dropouts

A patient who discontinues study participation prematurely (i.e. prior to disease progression confirmed by central review or death) for any reason is defined as a 'dropout' if the patient has already been randomized, even if no study drug has been taken. Patients who drop out will not be replaced.

A subject who for any reason terminates the study before randomization is regarded a 'screening failure'.

Patients who discontinue study drug due to reasons other than death or PD confirmed by central review shall enter the Active follow-up period, except for those who object to follow-up data collection. The patients in the Active follow-up shall have follow-up tumor assessments until disease progression is documented or new anti-tumor treatment is administered, whichever occurs first. Tumor assessments from the Active follow-up will also be used for primary and secondary analyses.

4.3 Handling of Missing Data

In order to achieve a well conducted clinical trial in accordance with Good Clinical Practice (GCP), every effort will be made to collect all data. However, despite best efforts, it may be inevitable that missing or incomplete data are reported. All missing or partial data will be presented in the subject data listing as they are recorded on the Case Report Form (CRF). Unless specifically specified, missing data will not be carried forward or otherwise imputed in any statistical analysis.

The following rules will be implemented where appropriate so as not to exclude patients from statistical analyses due to missing or incomplete data.

4.3.1 Time-to-event variables

For time-to-event analyses the censoring mechanism is assumed to be non-informative. Patients will be handled as right-censored in time-to-event analyses, if applicable.

Missing or non-evaluable tumor observations (including scheduled assessment that was not done and incomplete assessment that does not result in an unambiguous tumor response) will not be used in the calculation of derived efficacy variables related to tumor assessments, e.g. response. No imputation will be performed for missing tumor assessment and response. For example, if a subject misses a scan visit and PD is documented at the next available scan visit, the actual visit date of the first documented PD will be used to calculate PD-related endpoints.

For complete definitions of efficacy parameters such as PFS, OS and TTP, and specific handling of missing data for time to event variables refer to Section 6.2.

4.3.2 Response rates

If a patient has no post-baseline tumor assessment available (or no post-baseline laboratory/clinical tests available for WM patients without radiologically measurable lesion(s)), i.e. the overall best response assessment is missing, the patient will be non-evaluable (NE), but will be included into denominator for calculation of objective tumor response rate (ORR) and complete response rate (CRR).

4.3.3 Patient-Reported Outcomes Questionnaire

Physical symptoms of lymphoma are assessed by the NCCN-FACT Lymphoma Symptom Index-18 (FLymSI-18) questionnaire. Missing individual items will be handled in accordance with the scoring instructions for the FLymSI-18 questionnaire (Appendix 9.1). Specifically, if there are missing items, subscale scores are prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered.

When there are missing data, prorating by subscale in this way is acceptable as long as at least 50% of the items were answered (e.g., a minimum of 5 of 9 items, 2 of 4 items, etc.).

The total score is considered to be an acceptable indicator of patient quality of life as long as overall item response rate is greater than 80%. This is not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if at least 50% of items are answered. In addition, a total score should only be calculated if all of the component subscales have valid scores.

Time to deterioration in disease-related symptoms – physical (DRS-P) of at least 3 points (See Appendix 9.2): Patients without deterioration will be censored at the date of their last tumor evaluation, if the reason for stopping treatment is not related to PD or deaths.

Time to improvement in DRS-P of at least 3 points (See Appendix 9.2): Patients will be censored at the date of their last tumor evaluation, if the reason for stopping the study (i.e. not being in treatment or active follow-up) is not related to PD. Patients dropping out due to progression-related reasons (e.g. an AE related to PD) or experiencing PD event or death due to any reason will be censored at the largest observation time (of events and censoring in all patients evaluated for improvement), plus 1 day.

4.4 Interim Analyses and Data Monitoring

No interim analysis is planned for this study before the primary evaluation. However, a Data Monitoring Committee (DMC) will be monitoring the safety of the drug combination while keeping the sponsor blinded. For more details see DMC discussion below.

The first analysis of overall survival (OS) will also be performed at the time of primary PFS analysis. The survival status will further be collected quarterly during the survival follow-up period up to 3 years after the last patient started study treatment. Twelve months after analysis of the primary and secondary efficacy variables, a further analysis of OS ('follow-up OS analysis') will be performed. This analysis will be an update of the analysis of OS performed on data available at the time of analysis of the primary endpoint, PFS.

After the follow-up analysis of OS, it will be determined whether a subsequent final OS analysis will be required, and the study duration will be adjusted accordingly. A subsequent final OS analysis will be required, if OS data are not yet mature at the time of the follow-up analysis. Mature OS data is available if the estimated median of the time-to OS will not change in further follow-up which is the case if no administrative censoring (i.e. no patients under risk are censored by the cut-off date) occurs before the estimated median.

A DMC is established for this study that is reviewing study data and providing an independent recommendation on the advisability of continuing the study as planned.

The report for the DMC, including tables, listings, and figures, is generated by an independent statistician from a Statistical Analysis Center (SAC). The format and content of these data summaries have been specified separately from this study SAP.

When the first 30 patients were recruited in the blinded study, the probability to have at least 15 patients in the copanlisib + rituximab arm was greater than 95%. Based on these 30 patients, a protocol pre-specified initial assessment of safety of the drug combination was performed by the DMC in these patients. The DMC agreed to proceed with the continued enrollment. The investigators, patients, and the sponsor remain blinded during the study. Further DMC reviews will take place as outlined in the DMC charter.

Decisions on trial termination, amendment, or cessation of patient recruitment based on the risk/benefit assessment will be made after recommendations from the DMC have been assessed by the sponsor.

4.5 Sample size justification for primary efficacy test on PFS in the total study population

Study AUGMENT [13] reported a hazard ratio of 0.46 (95% CI 0.34-0.62) for PFS and median PFS of 14.1 months in the rituximab + placebo arm in the indication of second line iNHL. Based on the results from this study, Bayer has decided to update the sample size justification.

With at least 190 PFS events (progression based on central review or death if death occurs before progression), a randomization ratio of 2:1 between the experimental and control arms, a 1-sided alpha of 0.025, a hazard ratio of 0.61 (i.e., median PFS of 23 months for copanlisib + rituximab arm and 14 months for placebo + rituximab arm) can be detected with a power of at least 89%. Based on the planned 450 patients for this study and the observed and projected recruitment as well as blinded monitoring of event times, the expected study duration to reach 190 events is 59 months, (with maximum accrual rate being approximately 10 patients per month). The rate of dropouts (i.e. lost to survival follow-up or withdrawal of consent before their PFS event) is assumed to be 135 (30%) patients over the duration of the study through the primary PFS analysis.

The primary analysis will be performed when at least 190 centrally evaluated PFS events are observed in the study.

As an additional criterion, a sufficient number of PFS events has to be observed in each of the two strata defined by the inclusion criterion (progression-free and treatment-free interval of ≥ 12 months after completion of the last rituximab-containing treatment vs. unwilling/unfit to receive chemotherapy) before the study is unblinded for primary evaluation. This especially ensures the contribution of events also from patients who have previously shown a good response to rituximab.

The required minimum number of PFS events in these strata will depend on the relative proportion of patients recruited into the respective cohorts and can be seen from Table 4–1. The observed proportions and event numbers will be determined during blind data review after the end of recruitment.

Table 4–1: Required number of PFS events per entry criteria stratum

Proportion of patients randomized in one strata	Required number of PFS events	Proportion of patients randomized in the other strata	Required number of PFS events
0% - 5%	no required numbers (in both strata)	95% - 100%	no required numbers (in both strata)
>5% - 10%	10	90% - <95%	122
>10% - 20%	20	80% - <90%	112
>20% - 30%	33	70% - <80%	99
>30% - 40%	46	60% - <70%	86
>40% - 50%	59	>50% - <60%	73

As an example: In case 32% of patients would be recruited in the "progression-free and treatment-free interval of ≥ 12 months after completion of the last rituximab-containing treatment" stratum, then at least 46 PFS events are required in this stratum. In addition, at least 86 PFS events from the complement stratum "considered unwilling/unfit to receive chemotherapy" are required. To fulfill the overall criterion of at least 190 PFS events, approximately 58 additional events from any of the two strata are required.

In case of an extreme proportion of censored/discontinued patients in a stratum that severely impacts the relation of "Proportion of recruited patients" ≈ "Proportion of PFS events", the study team might decide to deviate from the above rule during blind data review.

The number of PFS events will be monitored in the blinded database throughout the study. When the required numbers of events are reached in the FAS population and each of the two strata defined by the inclusion criterion (treatment-free interval after rituximab treatment vs. contraindication for chemotherapy) the database will be cut and cleaned for analysis. Data reported during cleaning for the time after the data cut will not be considered in the final analysis with the exception of any death reported during that time.

4.6 Data Rules

4.6.1 Time intervals

If time intervals are to be displayed other than days in statistical evaluations, then one year is considered to have 365.25 days (average length of a year, including leap years), one month is considered to have 30.44 days (average length of a month, including leap years), one week is considered to have 7 days, and one cycle is considered to have 28 days (i.e. 4 weeks).

4.6.2 Baseline

Baseline is defined as the last measurements performed prior to the first study drug administration in Cycle 1. If the actual time is not available but date is available for certain measurements, the baseline value is defined as the last non-missing value collected on or before the date of the first dose of study medication. For patients who have been randomized but not treated with any dose, randomization date will be used as the reference date for baseline value calculation.

Also consider:

- Consider the time part for baseline flagging. EX and LB both capture times. If the time in either one is missing, use the dates in that case.
- Baseline can be either a scheduled or unscheduled visit.
- If the patient has a measurement on Cycle 1 Day 1 (scheduled), then this measurement will be considered as the baseline;
- If the patient has no measurements on Cycle 1 Day 1, but has a measurement at screening visit (scheduled or unscheduled), then the screening visit measurement will be considered as the baseline;
- If multiple measurements were taken at the same cycle 1 day 1 visit or same screening visit and all with non-missing assessment dates, then the measurement with the latest assessment date/time will be considered as baseline, whether scheduled or unscheduled.
- If both scheduled and unscheduled measurements exist for the same visit and can't be decided by the above rules, then the measurement taken at scheduled visit will be considered as the baseline.

4.6.3 Repeated measures

If there are repeated measurements per time point (e.g. laboratory values, vital signs, etc.), the following rules will be used (unless otherwise specified):

- Before the start of the study drug administration (i.e., for screening and baseline value), the latest measurement at scheduled visits will be used. Unscheduled visits will be used, if there are no measurements at scheduled visits. If the latter is the case, the last unscheduled visit will be used.
- In case of repeated measurements at any post baseline time point, the first measurement at scheduled visits will be used. Unscheduled visits will be used, if there are no measurements at scheduled visits. If the latter is the case, the first unscheduled visit will be used.

4.6.4 Overall extent of exposure

As a general rule, and in accordance with the Oncology Therapeutic Area Standard, leading '0 mg' (prior to the first positive amount of drug) and trailing '0 mg' records (not followed by any positive amount of drug), will not be included in the calculation of any drug duration or amount. Similarly, the according trailing 'drug interruptions' will not be used in statistical tables. A footnote will be included, stating that 'Interruption becoming permanent study treatment discontinuation before resumption of study treatment is not accounted as an interruption'.

Overall extent of exposure (OEE) for copanlisib/placebo or rituximab, respectively, is defined as the time from first respective study drug intake (day_{first}) until last study drug intake (day_{last}), including 7 additional days in order to consider the weekly dosing regimen, and is calculated as:

$OEE = day_{last} - day_{first} + 7$

If the respective treatment ends with dose interruptions, the day of the last actual dose will be considered to be day_{last}.

4.6.5 Stratification

In case of discrepancies between stratification factors entered in the interactive voice/web response system (IxRS) and information entered in the CRF, the information from IxRS will be used for analysis. This especially includes analyses that are performed separately for the subgroup of FL subjects, which is based on one of the stratification factors.

The stratification from IxRS will be derived from the subjects' randomization number. The mapping of randomization numbers to strata is summarized in Table 4–2.

Table 4–2: Stratum per randomization number

Randomization Numbers	Stratum			
	iNHL Histology	**Entry Criterion**	**Presence of bulky disease**	**Previous treatment with PI3Ki**
PPD	FL	treatment-free interval	yes	yes
	FL	treatment-free interval	yes	no
	FL	treatment-free interval	no	yes
	FL	treatment-free interval	no	no
	FL	unfit for chemotherapy	yes	yes
	FL	unfit for chemotherapy	yes	no
	FL	unfit for chemotherapy	no	yes
	FL	unfit for chemotherapy	no	no
	other iNHL	treatment-free interval	yes	yes
	other iNHL	treatment-free interval	yes	no
	other iNHL	treatment-free interval	no	yes
	other iNHL	treatment-free interval	no	no
	other iNHL	unfit for chemotherapy	yes	yes
	other iNHL	unfit for chemotherapy	yes	no
	other iNHL	unfit for chemotherapy	no	yes
	other iNHL	unfit for chemotherapy	no	no

iNHL: indolent non-Hodgkin's Lymphoma; FL: Follicular Lymphoma; PI3Ki: Phosphatidylinositol 3-Kinase Inhibitor;
treatment-free interval: progression and treatment-free interval following the last rituximab-containing treatment ≥ 12 months unfit for chemotherapy: considered unwilling/unfit to receive chemotherapy

This study is using 16 strata (defined by 4 binary stratification variables) with 450 patients to be recruited and including a 2:1 randomization. In order to avoid a too low number of events, only stratification factors “iNHL histology” and “entry criterion” will be adjusted simultaneously in the statistical analyses, e.g., stratified log-rank test, and Cox proportional hazard model.

4.6.6 Region

For demographics overview and subgroup analyses, subjects are grouped together into regions. The following regions are defined:

Geographic Regions 1 (as of protocol):

- US
- Europe (Austria, Belgium, Bulgaria, Germany, Spain, France, Greece, Hungary, Ireland, Italy, Lithuania, Poland, Portugal, Romania, Slovakia, and Ukraine)
- Rest of the world (All other countries)

Geographic Regions 2:

- North America (Canada and United States),
- Asia Pacific (China, Japan, Korea, Taiwan, Hongkong, Malaysia, Singapore, Thailand, Vietnam, and Philippines), excluding Australia,

- Rest of the world (All other countries).

4.6.7 As treated

An IxRS will be used for drug distribution. Despite the obligatory use of the IxRS, patients may erroneously receive the wrong study medication (i.e. copanlisib instead of placebo or vice versa). For the 'as treated' analyses, using a conservative approach, subjects that have received at least one dose of copanlisib will be considered for the copanlisib treatment group. Subjects that have received exclusively placebo (and rituximab) and no copanlisib will be considered for the placebo treatment group.

4.7 Validity Review

The results of the validity review meeting (VRM) will be documented in the Validity Review Report (VRR) and may comprise decisions and details relevant for statistical evaluation. Any changes to the statistical analysis prompted by the results of the validity review meeting will be documented in an amendment and, if applicable, in a supplement to this SAP.

5. Analysis Sets

5.1 Assignment of analysis sets

Final decisions regarding the assignment of subjects to analysis sets will be made during the VRM and documented in the VRR (see Section 4.7).

Full analysis set (FAS): All subjects randomized.

Safety analysis set (SAF): All FAS subjects with at least one intake of copanlisib/placebo or rituximab.

Pharmacokinetics analysis set (PKS): All subjects with at least one valid pharmacokinetics (PK) measurement (i.e. subjects took at least one dose of study drug and had at least one PK sample collected and measured) will be included in the evaluation of PK concentration and parameters.

The FAS will be used for the display of efficacy variables. Following the intention-to-treat (ITT) principle, the analyses will be performed as randomized. The FAS will also be used to display demographics and baseline characteristics.

The SAF will be used for the analyses of safety variables. The analyses will be performed as treated.

In case the SAF and FAS are differing, displays for baseline characteristics and demographics will be repeated in the SAF.

Patients who signed the informed consent but were not assigned to treatment will be considered screening failures. They will be listed separately.

6. Statistical Methodology

6.1 Population characteristics

Population characteristics will be summarized by randomized treatment group and overall. Unless otherwise specified, analyses will be performed in the FAS and, in case it is differing, in the SAF. Patients will be analyzed as randomized for the FAS and as treated for the SAF.

6.1.1 Disposition

The number of subjects enrolled, randomized, and valid for the different analysis sets FAS, SAF will be summarized overall and by treatment group, country.

The number of subjects discontinuing the screening phase together with the primary reason for discontinuation will be presented overall. The number of subjects discontinuing the treatment, safety, active and survival follow-up phases together with the primary reason for discontinuation will be presented by treatment group and overall.

Kaplan-Meier plot for 'Time to end of study treatment' will be provided by treatment group.

In addition, the number of subjects with major, minor, and important protocol deviations will be presented overall and by country for treatment group and overall.

A separate subject listing of protocol deviations related to COVID-19 (e.g., delay of tumor assessments, interruption or discontinuation of the treatment due to the impact of COVID-19) will be provided. These protocol deviations will be identified from the investigators' specification with "COVID".

6.1.2 Demographics and other baseline characteristics

Demographics variables and baseline characteristics will be summarized by treatment group and overall. In addition, summary statistics will be presented for metric variables. Frequency tables will be presented for categorical variables.

Demographic variables include age, gender, race, ethnicity, body weight, body height, body mass index (BMI), systolic and diastolic blood pressure, heart rate and temperature. Age and BMI will each be summarized as continuous variable and categorized with the following categories:

- Age group (years): <80, ≥80
- Age group (years): < 65, >=65
- Age group (years): < 65, 65-74, 75-84, >=85
- BMI group (kg/m²): <18.5, ≥18.5- <30, ≥30

In addition, the four stratification factors as reported on the randomization page in the CRF will be summarized.

The following additional baseline characteristics will be analyzed:

- Most recent histology of tumor
- Most recent staging of tumor at study entry
- FL grade (for patients with FL) at study entry
- Eastern Cooperative Oncology Group (ECOG) performance status at baseline
- Time from initial diagnosis to the date of randomization (months)
- Time from most recent progression to the date of randomization (weeks)
- Time since first progression to the date of randomization (months)
- New York Heart Association (NYHA) classification

- Baseline tumor size (evaluated as sum of the product of the perpendicular diameters of up to 6 target measurable nodes and extranodal sites)
- Number of lesions (number of target and non-target lesions) at baseline
- Serum IgM level (for WM patients only) at baseline
- Estimated Glomerular Filtration Rate (eGFR) at baseline (continuous and categorization: eGFR Normal: ≥ 90 vs. Mild: 60 - <90 vs. Moderate: 30 - <60)
- HbA1c at baseline (< 5.7% vs. ≥ 5.7% – < 6.5% vs. ≥ 6.5%)
- Diabetic History: [PBMQ] Medical history of diabetes (no/yes)
- Hypertension History: [PBMQ] Medical history of hypertension (no/yes)
- Hepatic function at baseline
    - Normal: Total bilirubin and AST ≤ ULN
    - Mild impairment: Total bilirubin > ULN to 1.5 x ULN and AST any value; or Total bilirubin ≤ ULN and AST> ULN
    - Moderate impairment: Total bilirubin > 1.5 to 3 x ULN and AST any value
- Bilirubin at baseline
- Histology group by investigator at baseline (FL, MZL, SLL and LPL/WM)
- Reason for treatment allocation
    - Progression-free interval of at least 12 months after completion of the last rituximab-containing treatment
    - patients unfit or unwilling to receive chemotherapy
        - >= 80 years old
        - < 80 years with at least 3 G3 CIRS-G comorbidities
        - < 80 years with at least 1 G4 CIRS-G comorbidity
        - Unwilling to receive chemotherapy
- Smoking history

In addition, demographic and other baseline characteristics will be summarized for each Histology (FL, MZL, SLL and LPL/WM).

The retrospective evaluation of histopathological diagnosis at baseline will be performed centrally. These data are exploratory and will be listed for the CSR addendum.

6.1.3 Medical history

Medical history findings (as defined in protocol Section 7.2.2) will be coded by Medical Dictionary for Regulatory Activities (MedDRA) codes. Medical history will be presented for each MedDRA Primary System Organ Class (SOC) and Preferred Term (PT) by treatment group and overall.

6.1.4 Prior and concomitant medication

Prior and concomitant medications will be coded by Anatomical Therapeutic Chemical (ATC) classification system according to the World Health Organization Drug Dictionary (WHO-DD). Summaries will be provided by ATC class and subclass. Note that the same medication can appear multiple times in the table as it can have several ATC codes.

A medication that has been stopped after first administration of study treatment is considered as concomitant and otherwise as prior. Medications with missing start and stop date but flagged as being ongoing at end of study will be considered to have started prior to start of study medication and end after stop of study medication.

Prior and concurrent anti-cancer therapy

The minimum, median, and maximum number of prior systemic anti-cancer therapy lines as well as number of patients with 1, 2, 3, ≥4 lines of therapy will be summarized by treatment group and overall for FAS. Time since last systemic anti-cancer therapy will be summarized using descriptive statistics by treatment group and overall for FAS. The time between the start day of last course of systemic anti-cancer therapy and the day of confirmation of the most recent progression will also be displayed as a classification table, showing proportions of patients with ≤ 6 months vs. > 6 to <12 months vs. ≥ 12 months, by treatment group.

In addition, the following frequency tables will be provided by treatment group and overall for FAS:

- Prior radiotherapies
- Prior systemic anti-cancer therapies
- Type of prior systemic anti-cancer therapy
- Prior diagnostic and/or therapeutic procedures for iNHL
- Prior treatment with PI3K inhibitors
- Concurrent diagnostic and/or therapeutic procedures for iNHL
- Radiotherapies during follow-up
- Systemic anti-cancer therapies during follow-up
- Type of systemic anti-cancer therapy during follow-up

6.1.5 Treatment duration and exposure

Descriptive statistical summaries will be provided separately for copanlisib/placebo and rituximab by treatment group for the following variables:

- Overall extent of exposure (OEE, as defined in Section 4.6.4) for FL, MZL and all patients
- Overall extent of exposure by categories:
    - 0-180 days
    - 181-360 days
    - 361-540 days

- 541-720 days
- >720 days

- Number of cycles
- Number of infusions during treatment phase

In addition, the following analyses will be provided

- Total amount actually administered: sum of actual dose (mg) for each patient
  - For rituximab, the actual dose is defined as Prescribed dose [mg/m^2] × (Total amount administered [ml] / Total amount prior to administration [ml]) / Body surface Area (BSA) [m^2]
- Percent of planned dose received = Actual dose [mg] / Planned dose [mg] × 100%.
  - For copanlisib/placebo, a standard of 60 mg is set for all time points, i.e. infusion days (D1, D8 and D15 of each cycle), for planned dose (might however be modified by individual dose-reductions or interruptions). Completed cycles therefore have a planned dose of 180 mg. For incomplete cycles, the planned dose depends on the number of days (d) in that cycle:
    - 60 mg if $0 < d < 8$
    - 120 mg if $8 \leq d < 15$
    - 180 mg if $15 \leq d < 28$
  - For rituximab, a standard of 375 mg/m² BSA is set for all time points for planned dose.

Evaluations of this section will be done in SAF by treatment group.

For patients with dose reduction and re-escalation (only for copanlisib/placebo), interruption, or delay, the number of dose reductions, interruptions, or delays per patient and their reasons will be summarized separately for copanlisib/placebo and rituximab.

The duration of dose interruptions or delays in days will be summarized for copanlisib/placebo. The duration is defined as the difference between the date of interruption or delay and the date of next infusion. Due to the treatment regimen with one week break at the end of each cycle, interruptions or delays on day 15 are reduced by 7 days, if the next infusion is given in a subsequent cycle.

Subject listings will be provided for dose modification.

6.2 Efficacy

Descriptive evaluations of efficacy variables will be performed. Especially, summaries of tumor response by central review and investigator's assessment will be provided.

Disease progression in the context of statistical efficacy evaluation is considered to be radiological progression, as assessed by central review. For WM patients without radiologically measurable disease, clinical progression will be used.

All efficacy analyses will be performed when at least 190 PFS events are observed in the study. See Section 4.4.

Evaluations from central blinded review and Owen criteria for WM patients will be used for the primary efficacy analyses of primary and secondary endpoints containing radiological tumor assessments.

Efficacy analyses will be performed as randomized in the FAS. Censoring of efficacy variables will be handled according to Sections 4.3.1, 4.3.2, 6.2.1 and 6.2.2.

6.2.1 Primary Efficacy Variable Progression-Free Survival

Definition of progression-free survival (PFS)

The primary efficacy variable is PFS, defined as the time (in days) from randomization to PD or death from any cause (if no progression is documented) whichever occurs earlier. The primary PFS analysis will be based on central review. Radiological progression will be confirmed in real-time by central review. Biochemical progression in patients with WM without lesions evaluable by imaging will be assessed locally, and the confirmation of PD by central review is not needed. Progression-free survival (PFS) for patients without PD or death at the time of analysis will be censored at the last actual date of tumor assessment or last biochemical assessment for patients with WM without lesions evaluable by imaging.

For ease of reading, tumor assessment in the following context refers to either 1) tumor imaging assessment or 2) biochemical assessment for patients with WM without lesions evaluable by imaging.

The actual date of tumor assessments or date of death will be used for this calculation. If no assessment was done at all or the assessment was incomplete at a particular time point, in general, the patient is non-evaluable (NE) for this time point. If the examination was incomplete and only a subset of measurements could be made but fulfilling the criteria for progression and/or new lesions were detected, this patient would be considered evaluable with assessment of PD.

PFS for patients who have neither tumor assessments nor death after baseline will be censored at Day 1. PFS for patients without baseline tumor assessments (e.g. non-existence of a measurable lesion at baseline confirmed by central review) will be censored at Day 1.

If a tumor assessment date is partially missing, it will be imputed following the imputation rules below:

1. If only the day is missing, it will be imputed by the first day of the month.
2. If day and month are missing, it will be imputed by the date of the last previous tumor assessment plus 28 days unless the Last known active date (LKAD) is earlier than this date; in which case the LKAD will be used for imputation.

Table 6–1 lists the censoring rules for PFS events.

Table 6–1: Progression-free survival (PFS) censoring rules

Situation	End Date/Day	Censored	Reason for Censoring	Detailed Reference
No baseline tumor assessment	Reference Date/Day 1	Yes	No baseline tumor assessment	
No post-baseline tumor assessment and no death	Reference Date/Day 1	Yes	No post-baseline tumor assessment and no death	Appendix 9.3.1
Death during treatment and active follow-up period or early enough in survival follow-up to be within the time window	Death date	No	N/A	Appendix 9.3.2
Death after the time window	Last Tumor assessment	Yes	Death happened after time window specified in Appendix 9.3.2	Appendix 9.3.2
No post-baseline tumor assessment and died after the first 17 (16+1) weeks after randomization during survival follow-up period	Reference Date/Day 1	Yes	N/A	
Death occurs after two or more consecutive missed tumor assessments immediately during survival follow-up period	Last tumor assessment	Yes	Two or more consecutive missed tumor assessments immediately prior to death	
New anti-cancer therapy other than the study medication prior to observing progression	Last evaluable tumor assessment prior to the initiation of anti-cancer therapy	Yes	New anti-cancer therapy	Appendix 9.3.3
The progression occurs at the next tumor assessment after two or more consecutive missing or non-evaluable assessments	Last tumor assessment before the missing tumor assessments	Yes	Two or more consecutive missed tumor assessments immediately prior to progression	Appendix 9.3.1
Discontinued or withdraw early from the study without documented disease progression	Last tumor assessment date	Yes	Discontinued	Appendix 9.3.4
Progression without two or more consecutive missed tumor assessments immediately prior to death	Date of progression	No	N/A	
Neither PD nor death and no other criteria fulfilled	Last tumor assessment	Yes	Patient regularly ongoing in trial without PFS event occurred	

Primary analysis of PFS

The primary efficacy analyses will evaluate whether PFS in the copanlisib group is longer compared to PFS in the placebo group for FAS for the total study population, by means of a stratified log-rank test.

Stratification factors for this test will be "iNHL histology" and "entry criterion" used for randomization (see Section 4.6.5).

The following test procedure will be applied:

According to the size of this study it is justified to assume that under $H_{0, PFS}$, the one-sided log-rank test [14] is a sufficiently close approximation to the normal distribution. The actual normal distributed one-sided log-rank test will be developed by dividing the test statistic with the standard error and comparing the resulting p-value against the study-wise alpha of 2.5%.

The SAS code to generate the required test statistic will be similar to the following pseudo code, with variable names taking their obvious meaning:

```
PROC LIFETEST DATA=dataset;
  TIME event_time*censorny(cens_value);
  STRATA {strata variables}/group=treatmgr test=(logrank);
  ODS OUTPUT homtests=loguni;
RUN;
```

Technically, the binary treatment variable will be designed with "1" representing copanlisib treatment, to allow the correct direction of the one-sided log-rank test (and "0" representing the placebo group).

For the purpose of the analysis, patients who before protocol amendment 7 were stratified as unfit will be combined with those who after protocol amendment 7 were stratified as unfit and unwilling; and patients who before protocol amendment 9 were stratified as "treatment-free ≥ 12 months after completion of the last rituximab containing treatment" will be combined with those who after protocol amendment 9 were stratified as "progression-free and treatment-free ≥ 12 months after completion of the last rituximab containing treatment".

The PFS will be tested based on hierarchy test, refer to section 6.2.3.

Additional analyses of PFS

The hazard ratio (including 95% two-sided confidence interval) will be derived for the total study population from the Cox proportional hazards models that are stratified by the same factors as used for the primary efficacy analysis. SAS code similar to the following pseudo code will be used:

```
PROC PHREG DATA=dataset;
  MODEL event_time*censorny(cens_value) = treatmgr;
  STRATA {strata variables};
RUN;
```

Kaplan-Meier estimates of median times to PFS (including 95% two-sided confidence intervals) and Kaplan-Meier curves will be presented for each treatment group using the following pseudo code:

```
PROC LIFETEST DATA=dataset;
  TIME event_time*censorny(cens_value);
  STRATA treatmgr;
```

```
RUN;
```

Subgroup analysis for PFS will be done by histology (FL, MZL SLL, and LPL/WM), see section 6.2.5.

Exploratory analyses of PFS

Analysis of concordance and discordance between radiological progression evaluation by central blinded review and by investigator assessment during the blinded study phase will be performed via cross tabulations in FAS for overall and for each treatment group. To assess the impact of the stratification variables, Kaplan-Meier curves and median PFS times by treatment groups will be estimated, for each of the four binary stratification variables, similar to the following SAS pseudo-code:

```
PROC LIFETEST DATA=dataset;

BY { strata variable };

  TIME event_time*censorny(cens_value);

  STRATA treatmgr;

RUN;
```

To assess the potential impact of clinical progression, the investigator-based evaluation of PFS will also be performed including clinical progression (either direct "clinical progression" or "AE related to clinical progression") as a progression event[1]. The first PD assigned by the investigator will be used as event. It is acknowledged that due to a substantial amount of "PD non-confirmation" that might occur in the study, the investigator PFS times might be generally shorter, in both study arms. This will be attributable to the above effect and needs to be considered in the statistical interpretation of results.

Sensitivity analyses of PFS

To align with US health authority, a sensitivity analysis for PFS will be conducted using stratified log-rank test with all four stratification factors at randomization. The four stratification factors are iNHL histology, entry criterion, presence of bulky disease, and previous treatment with PI3Ki.

The censoring mechanism of subjects without PFS event at the time of analysis is assumed to be non-informative for the primary efficacy analysis. Therefore, additional sensitivity analyses will be performed, assessing the impact of a potential informative censoring of such patients. The following sensitivity analyses for the primary PFS analysis will be performed using similar methodology as the primary analysis by central review unless otherwise specified:

- Consider all patients that are censored to be events in both arms (Kaplan-Meier curves and Cox model for hazard ratio).

[1] If the clinical progression is at end of treatment, the last date of treatment or, if there is an AE related to clinical progression, the earlier date of either is used. If the clinical progression is at end of active follow-up, the last visit date (as provided on the CRF) or, if there is an AE related to clinical progression, the earlier date of either is used.

Further sensitivity analyses will be

- PFS will be evaluated with the unstratified log-rank test and Cox model.
- In case of many discrepancies between stratification factors entered in the interactive voice/web response system (IxRS) and information entered in CRF, PFS will be evaluated with the stratified log-rank test and Cox model based on stratification information entered in CRF. Note that some subjects will be excluded from this analysis (see Table 6–2) since these patients should not have been enrolled.

Table 6–2: Entry Criteria Stratum and Exclusions

	Treatment- and progression-free interval ≥ 12 months	Treatment- and progression-free interval between 6 and 12 months	Treatment- and progression-free interval < 6 months
Willing/Fit to receive chemotherapy (none of the unfit criteria is fulfilled)	Treatment and progression-free interval ≥ 12	**Excluded**	**Excluded**
Unwilling/Unfit to receive chemotherapy (at least one of the unfit criteria is fulfilled)	Unwilling/Unfit for chemotherapy	Unwilling/Unfit for chemotherapy	**Excluded**

- PFS will be evaluated considering any change as a progression event, e.g., treating initiation of a new anticancer agent as an event at the date of start of new anticancer agent, treating disease progression as an event at the date of progression ignoring scheduled missing assessments, treating treatment discontinuation as an event at the date of discontinuation [15]. Table 6–3 lists the rules for sensitivity analysis.

Table 6–3: Progression-free survival (PFS) censoring rules for sensitivity analysis (any change considered as a progression event)

Situation	End Date/Day	Censored	Reason for Censoring	Detailed Reference
No baseline tumor assessment	Reference Date/Day 1	Yes	No baseline tumor assessment	
No post-baseline tumor assessment and no death	Reference Date/Day 1	Yes	No post-baseline tumor assessment and no death	Appendix 9.3.1
Death or progression occurred during treatment period	Earliest of: • Death date or • Date of progression	No	N/A	
New anti-cancer therapy other than the study medication prior to observing progression	Date of start of new anticancer treatment	No	N/A	
Discontinued or withdraw early from the treatment without documented disease progression	Date of treatment discontinuation	No	N/A	
Neither PD nor death and no other criteria fulfilled	Last tumor assessment	Yes	Patient regularly ongoing in trial without PFS event occurred	

- PFS will be evaluated by log-rank test and Cox model and stratified by the following regions:
  - North America (Canada and United States),
  - Asia Pacific, excluding Australia (China, Japan, Korea, Taiwan, Hongkong, Malaysia, Singapore, Thailand, Vietnam, and Philippines)
  - Rest of the world (All other countries).
- Additional sensitivity analysis of PFS related to COVID-19 will be performed if treatment interruptions due to COVID-19 are observed and PFS evaluations are considered being affected.

6.2.2 Secondary Efficacy Variables

Definition of secondary efficacy variables

Objective tumor response rate (ORR), assessed in all patients up to the time of analysis of PFS. ORR is defined as the proportion of patients who have a best response rating over the whole duration of the study (i.e. until time of analysis of PFS) of complete response (CR) or partial response (PR) according to the Lugano Classification and for patients with WM a response rating of CR, very good partial response (VGPR), PR, or minor response (MR) according to the Owen Criteria.

Duration of response (DOR), defined as the time (in days) from first observed tumor response (CR, VGPR, PR, or MR) until PD or death from any cause, whichever is earlier. DOR will only be defined for patients with at least one CR, VGPR, PR, or MR.

Disease control rate (DCR), DCR is defined as the proportion of patients who have a best response rating of CR, VGPR, PR, MR or SD (excluding unconfirmed early SD) that is achieved during treatment or within 35 days after termination of study treatment. The unconfirmed early SD is defined as SD on or before Study Day 48.

Complete response rate (CRR), assessed in all patients up to the time of analysis of PFS. CRR is defined as the proportion of patients who have a best response rating over the whole duration of the study (i.e. until the time of analysis of PFS) of CR according to the Lugano Classification and for patients with WM a response rating of CR according to the Owen criteria.

Time to progression (TTP), defined as the time (in days) from randomization to PD or death related to PD, whichever is earlier. Death related to PD is considered to be any death except for:

a. Death due to an AE unrelated to progression
b. Death with a specification of "other" as reason (which excludes PD)

Time to progression (TTP) for patients without PD or with death not related to PD at the time of analysis will be censored at the date of their last tumor evaluation. TTP for patients who have neither tumor assessments nor death related to PD after baseline will be censored at Day 1. TTP for patients without baseline tumor assessments will be censored at Day 1. The conventions for calculation of TTP are the same as for PFS (See Section 6.2.1) except without considering death not related to PD as an event.

The actual dates of tumor assessments will be used for this calculation.

Overall survival (OS), defined as the time (in days) from randomization until death from any cause. Overall survival (OS) of patients alive at the time of analysis will be censored at the last date they were known to be alive (last known alive date; LKAD). Death after the data cut reported during cleaning will be considered for the OS analysis and will be censored at the date of cutoff.

The LKAD is derived from the main data sources. Except for death, data reported during cleaning for the time after the data cut will not be considered in the derivation of LKAD. The last available date across all key data panels will be picked as the LKAD by patient. Information from key data, e.g., visit dates, exposure information, laboratory measurements, tumor assessment dates and disposition events or follow up assessments will be used to determine survival status.

If a death date is partially or completely missing, it will be imputed following the imputation rules below:

1. If there is an adverse event (AE) with outcome 'Death', the date of death will be imputed by the end date of the AE.
2. If there is no AE with outcome 'Death' and only the day of death is missing, it will be imputed by the first day of the month unless the LKAD is later than this date; in which case the LKAD will be used for imputation. If both day and month are missing, then the LKAD will be used for imputation.

If a tumor assessment date is partially missing, it will be imputed following the imputation rules below:

3. If only the day is missing, it will be imputed by the first day of the month.
4. If day and month are missing, it will be imputed by the date of the last previous tumor assessment plus 28 days unless the LKAD is earlier than this date; in which case the LKAD will be used for imputation.

Time to deterioration in DRS-P of at least 3 points, as measured by the FLymSI-18 questionnaire, will be evaluated for all patients. It is defined as the time (in days) from randomization to DRS-P decline, PD or death due to any reason, whichever is earlier. DRS-P deterioration censoring rules are listed in the Table 6–4.

Table 6–4: DRS-P deterioration censoring rules

Situation	End Date/Day	Censored	Reason for Censoring
No baseline DRS-P assessment	Reference Date/Day 1	Yes	No baseline or post-baseline DRS-P assessment
No post-baseline DRS-P assessment and no PD or Death within time window	Reference Date/Day 1	Yes	No post-baseline DRS-P assessment and no PD/Death within time window
Patient has DRS-P deterioration and also has PD	Min(DRS-P deterioration date, PD)	No	N/A
Patient has DRS-P deterioration and no PD	DRS-P deterioration date	No	N/A
Patient has no DRS-P deterioration but has PD before the last DRS-P assessment	PD date	No	Has event, as PD occurred (even if no deterioration)
Patient has no DRS-P deterioration but has PD or death within the time window	PD date or death date	No	
Patient has no DRS-P deterioration, no PD or no death within time window	Last DRS-P assessment	Yes	
New anti-cancer therapy other than the study medication prior to observing DRS-P deterioration (and no PD before switch to new anti-cancer therapy)	Last evaluable DRS-P assessment prior to the initiation of anti-cancer therapy	Yes	New anti-cancer therapy
Discontinued or withdraw early from the study without documented DRS-P or PD or, Death within time window	Last DRS-P assessment date	Yes	Discontinued
Neither DRS-P, nor PD/death and no other criteria fulfilled	Last DRS-P assessment	Yes	Patient regularly ongoing in trial without DRS-P event occurred

Similar time window for PFS will be used for the derivation.

Time to improvement in DRS-P of at least 3 points, as measured by the FLymSI-18 questionnaire, will be evaluated for all patients. It is defined as the time (in days) from randomization to DRS-P improvement of at least 3 points. The patient will generally be censored at the last non-missing post-baseline DRS-P score, if no DRS-P improvement, PD or death occurs. If a PD or death occurs before a potential DRS-P improvement (or no DRS-P improvement is documented at all), the patient will be censored at the largest observation

time (of events and censoring in all patients evaluated for improvement), plus 1 day. Regular censoring rules, apart from the special rule for PD/Death will be similar to the censoring for deterioration of DRS-P. The DRS-P improvement censoring rules censoring rules are listed in Table 6–5:

Table 6–5: DRS-P improvement censoring rules

Situation	End Date/Day	Censored	Reason for Censoring
Patient has DRS-P improvement before PD or death	DRS-P improvement date	No	N/A
Patient has PD or death (including drop out due to PD) before DRS-P improvement	largest observation time (of events and censoring in all patients evaluated for improvement), plus 1 day	Yes	PD / Death occurs before improvement and thus improvement becomes impossible for these patients
Patient has PD or death (including drop out due to PD) after DRS-P improvement	DRS-P improvement date	No	Same rule as Line 1, as PD / Death after DRS-P improvement has no impact
No baseline or post-baseline DRS-P assessment	Reference Date/Day 1	Yes	No baseline or post-baseline DRS-P assessment
Patient has no DRS-P improvement, no PD and no death	Last DRS-P assessment	Yes	Event for DRS-P improvement not reached
New anti-cancer therapy other than the study medication prior to observing DRS-P improvement (and no PD before last DRS-P assessment before switch to other than study medication)	largest observation time (of events and censoring in all patients evaluated for improvement), plus 1 day	Yes	New anti-cancer therapy
Discontinued or withdraw early from the study without documented DRS-P or PD/Death	Last DRS-P assessment date	Yes	Discontinued
Neither DRS-P, nor PD/death and no other criteria fulfilled	Last DRS-P assessment	Yes	Patient regularly ongoing in trial without DRS-P event occurred

The FLymSI-18 questionnaire contains 18 items, each of which utilizes a Likert scale with 5 possible responses ranging from 0 'Not at all' to 4 'Very much'. Nine items reflect the DRS-P, and the responses to the items are summed to calculate a DRS-P subscale score. Four items represent disease-related emotional symptoms (DRS-E), and the responses to those items may be used to calculate a DRS-E subscale score. Three items represent treatment side effects, and the responses to these items may be summed to calculate a treatment side-effect (TSE) subscale score. Finally, two items represent function and well-being (FWB), and responses to those items may be summed to calculate an FWB subscale score. The questionnaire allows for calculation of total score and four subscales: DRS-P, DRS-E, TSE and FWB. A summary of the scoring guideline is included in Appendix 9.1.

Analyses of secondary efficacy variables

Analysis of ORR, Time to deterioration and Time to improvement in DRS-P

All secondary efficacy variables will be analyzed in the FAS at the time of the analysis of the primary variable.

The ORR will be analyzed using the Cochran-Mantel-Haenszel test, stratified for the same stratification factors as used for PFS. The null hypothesis is defined as:

$H_{0,\,ORR}$: $ORR_{Copanlisib+Rituximab} \leq ORR_{Placebo+Rituximab}$

The alternative hypothesis is:

$H_{1,\,ORR}$: $ORR_{Copanlisib+Rituximab} > ORR_{Placebo+Rituximab}$

In addition, the point estimate as well as 95% two-sided confidence intervals for the Mantel-Haenszel weighted treatment difference [16] will be calculated.

The time to deterioration and time to improvement in DRS-P subscale of FLymSI-18 will be analyzed analogously to PFS, using similar hypotheses, decision rules, and significance level.

Analysis of other secondary efficacy variables

OS, TTP, DOR as well as CRR will not be included into the confirmatory testing strategy but analyzed supportively only.

OS, TTP, and DOR will be analyzed using stratified log-rank tests analogue to the analysis for the primary analysis of the primary endpoint, PFS. CRR will be analyzed using the Cochran-Mantel-Haenszel test and estimated difference analogous to the analysis of ORR. The test will be adjusted for the same stratification factors as used for the primary analysis of PFS.

The first analysis of OS will be performed at the time of analysis of the primary endpoint, PFS. The survival status will further be collected quarterly during the survival follow-up period up to 3 years after the last patient started study treatment.

Twelve months after analysis of the primary and secondary efficacy variables, a further analysis of OS ('follow-up OS analysis') will be performed. This analysis will be an update of the analysis of OS performed on data available at the time of analysis of the primary endpoint, PFS.

After the follow-up analysis of OS, it will be determined whether a subsequent final OS analysis will be required, and the study duration will be adjusted accordingly.

Additional analyses of secondary efficacy variables

ORR

Subgroup analysis for ORR will be done by histology (FL, MZL SLL, and LPL/WM), see section 6.2.5.

DRS-P deterioration

As sensitivity analysis, an analysis of “worsening of DRS-P alone”, considering patients as 'censored' at the date of their last DRS-P measurement, ignoring PD or death (that is, not considering PD or death as events), will be performed.

6.2.3 Confirmatory statistical test strategy

Separate statistical test strategies will be conducted for the United States and Europe as outlined below.

United States

For the United States, to control the study-wise alpha, a fixed-sequence multiple testing strategy will be implemented. The sequence of the testing of primary and secondary endpoints is as follows and shown in Figure 6–1:

The primary endpoint PFS will be tested in the FAS population (tested at one-sided alpha=0.025).

If the PFS test in the FAS population is successful,

a) a second test on the PFS will be performed (tested at one-sided alpha=0.025), evaluating the subpopulation of combined FL and MZL population, if this is successful

b) a test on the ORR in the combined FL and MZL population will be performed (tested at one-sided alpha=0.025), if this is successful,

c) a test on the DRS-P deterioration in the combined FL and MLZ population will be performed (tested at one-sided alpha=0.025), if this is successful.

d) a test on the DSR-P improvement in the combined FL and MZL population will be performed (tested at one-sided alpha=0.025), it would conclude the confirmatory test procedure.

For the testing Families 1 to 2 of this confirmatory test strategy, an assessment of power based on Schoenfeld's formula [17] was performed for two scenarios as listed in Table 6–6:

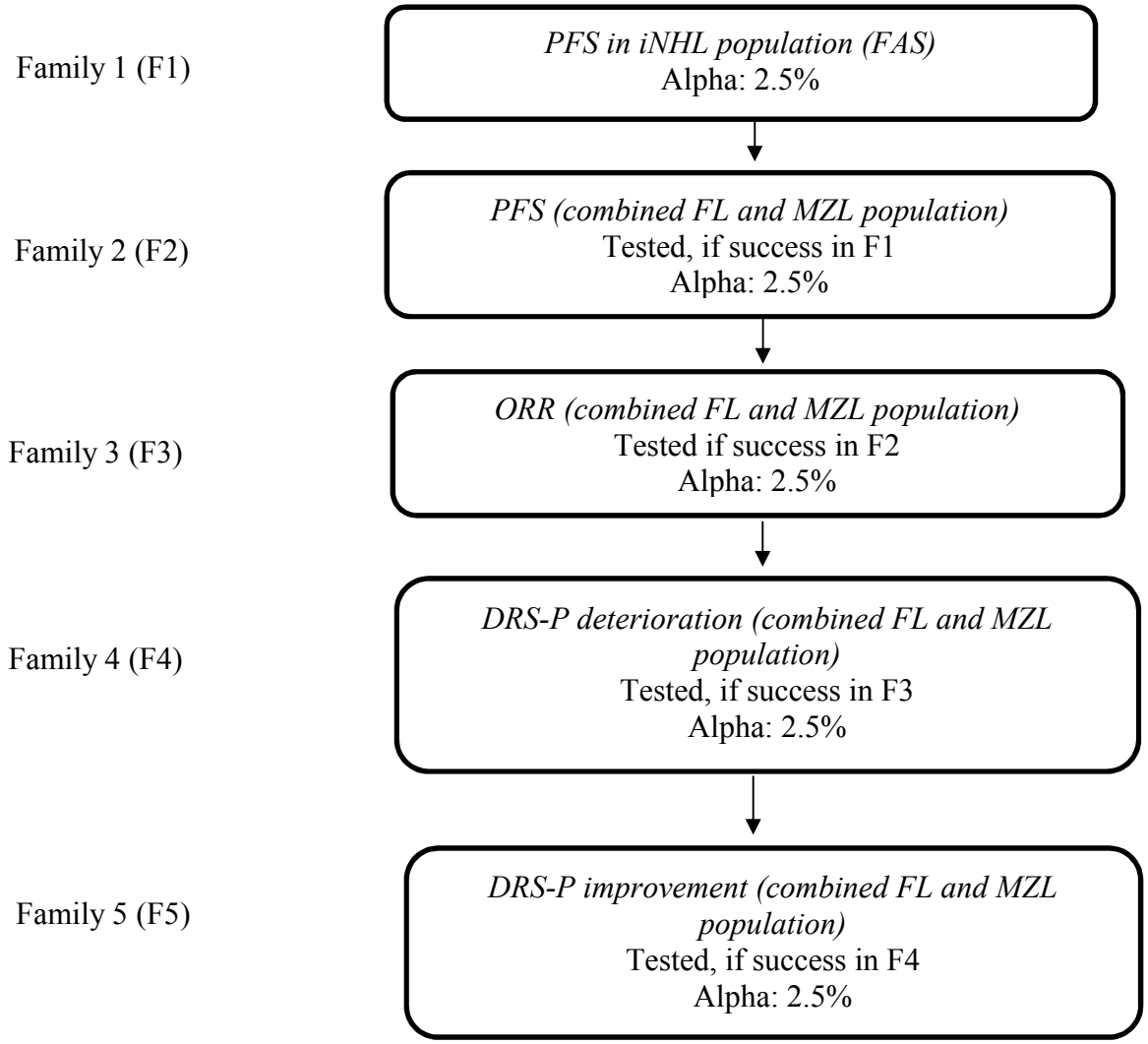
- Scenario 1: HR=0.61 in the study (according to the re-planned sample size planning)
- Scenario 2: HR=0.53 in the study (less conservative assumption)

Table 6–6: Assessment of Statistical Power for PFS Test in the Overall FAS Population and Combined FL and MZL Population

	Total Patients	Dropout rate	Events to be expected (E)	Hazard ratio	Power
Scenario 1:					
Family 1: overall FAS population	458	30%	190	0.61	89.5%
Family 2: combined FL and MZL population	370	30%	140	0.61	78.7%
Scenario 2:					
Family 1: overall FAS population	458	30%	190	0.53	98.5%
Family 2: combined FL and MZL population	370	30%	140	0.53	94.3%

The number of total patients in Table 6–6 is the number of patients having been randomized in the study.

Figure 6–1: Confirmatory Test Strategy Based on Five Test Families for United States



Europe

For Europe, to control the study-wise alpha, a fixed-sequence multiple testing strategy will be implemented. In this strategy, the sequence of the testing of primary and secondary endpoints is as follows and shown in Figure 6–2:

The primary endpoint PFS will be tested in the FAS population (tested at one-sided alpha=0.025).

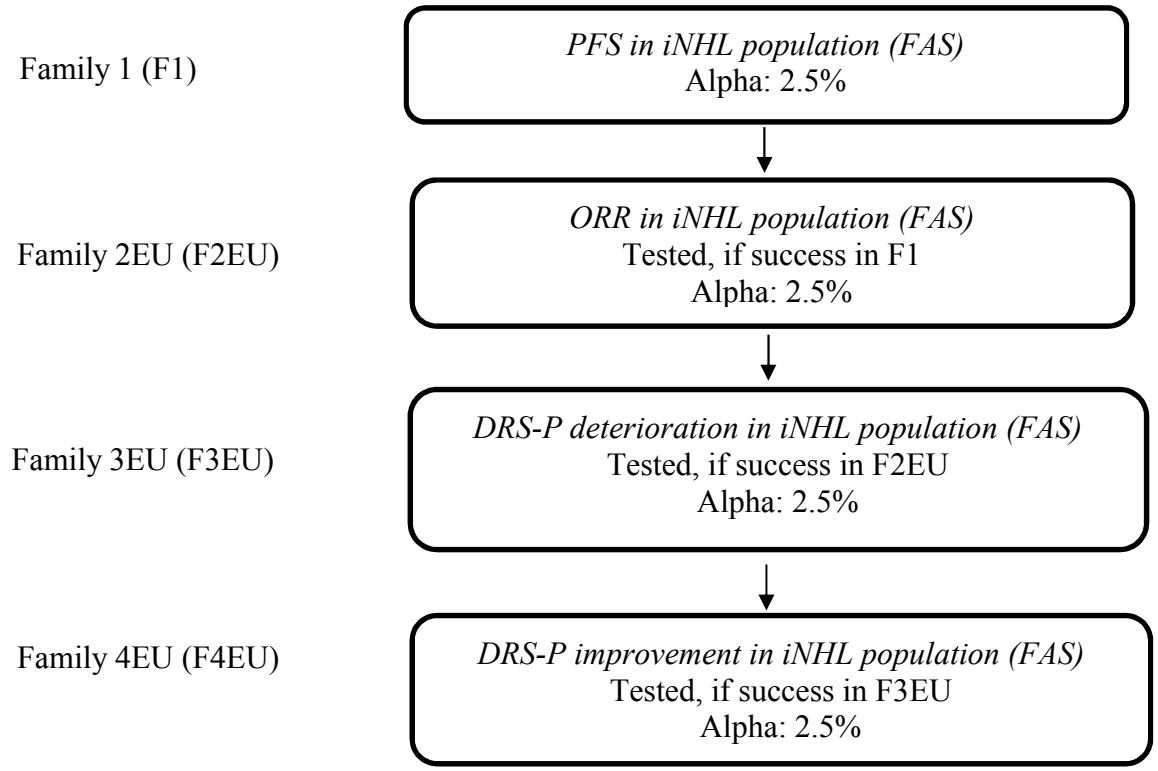
If the PFS test in the FAS population is successful,

a) a second test on the ORR in the FAS population will be performed (tested at one-sided alpha=0.025), if this is successful,

b) a test on the DRS-P deterioration in the FAS population will be performed (tested at one-sided alpha=0.025), if this is successful,

c) a test on the DSR-P improvement in the FAS population will be performed (tested at one-sided alpha=0.025), it would conclude the confirmatory test procedure.

Figure 6–2: Confirmatory Test Strategy Based on Four Test Families for Europe



Note: EU refers to Europe

6.2.4 Other Efficacy Variables

1. Area under the curve (AUC) across all data of FLymSI-18 DRS-P subscale score

An additional analysis of the physical symptoms of lymphoma (as measured using the FLymSI-18 DRS-P subscale) will be performed to assess differences between treatment arms in the FAS based on AUC. AUC will be derived using the linear trapezoidal rule:

$$AUC_{0-t_n} \approx \frac{1}{2}\sum_{i=1}^{n-1}(t_{i+1} - t_i)(C_i + C_{i+1})$$

where C_i denotes the non-missing DRS-P score at timepoint t_i. Missing DRS-P scores will not be replaced. If a patient has only the baseline value, the AUC will not be calculated.

Summary statistics will be provided for AUC and for the AUC normalized over the observation time.

2. FLymSI-18 total and subscale scores (DRS-P, DRS-E, TSE, FWB), and time to onset of physical symptoms of lymphoma based on the DRS-P subscale

Total FLymSI-18 and subscales will be evaluated descriptively. The number and percentage of patients with onset of physical symptoms defined as a decrease in DRS-P score ≥ 3 points will be presented by visit and overall. The time to onset of physical symptoms will be described.

3. Performance status (ECOG)

ECOG performance status will be summarized by visit using descriptive statistics for the original score, as well as for the change from baseline score by treatment group in the FAS.

6.2.5 Subgroup Analyses

6.2.5.1 Histology specific subgroup analysis

For selected subgroups of patients in the FAS, as defined by cancer histology, separate analyses of treatment effect for the primary and various secondary endpoints will be performed. Diagnosis of cancer histology from investigator's assessment will be used in statistical analyses.

The respective endpoints and histology subgroups can be seen from Table 6–7, which describes the planned output format.

In addition, for the time to event endpoints, histology-specific Kaplan Meier evaluations comparing treatment groups will be provided (i.e. KM plot, descriptive statistics including log-rank test).

Table 6–7: Overview of Histology Specific Subgroup Analyses

	Variable						
Histology Subgroup	**PFS**	**ORR**	**CRR**	**DCR**	**Time to deterioration* DRS-P**	**Time to improvement* DRS-P**	**OS**
FL	a)	b)	b)	b)	a)	a)	a)
MZL	a)	b)	b)	b)	a)	a)	a)
Extranodal	a)	b)	b)	b)	a)	a)	a)
Nodal	a)	b)	b)	b)	a)	a)	a)
Splenic	a)	b)	b)	b)	a)	a)	a)
SLL	a)	b)	b)	b)	a)	a)	a)
LPL/WM	a)	b)	b)	b)	a)	a)	a)

* of at least 3 points
FL: follicular lymphoma; MZL: marginal-zone lymphoma; SLL: small lymphocytic lymphoma; LPL:

lymphoplasmacytic lymphoma; WM: Waldenström macroglobulinemia
a) Evaluated by use of a non-stratified Cox-PH model.
Results displayed in table: Number of events per treatment group, HR, 95% CI.
b) Evaluated by use of a chi-square test.
Results displayed in table: Number patients per treatment group, estimated treatment difference and 95% CI.

Same subgroup analyses will be conducted for the secondary endpoints DOR, and TTP by histology subgroups FL and MZL separately.

6.2.5.2 Other subgroup analyses

Subgroup analyses will include forest plots of response/hazard ratios as well as treatment-interaction analyses. The forest plots of response/hazard ratios will be provided for the primary efficacy endpoint (PFS) as well as the secondary efficacy endpoints ORR, time to deterioration in DRS-P, time to improvement in DRS-P. The treatment-interaction analysis will be conducted for the primary efficacy endpoint (PFS), and the secondary efficacy endpoints time to deterioration in DRS-P and time to improvement in DRS-P. Kaplan-Meier estimates of median times to PFS (including 95% CI) and Kaplan-Meier curves will be provided for each subgroup level for each treatment group. Subgroup levels without sufficient number of events will be excluded from corresponding subgroup analysis. All subgroup analyses will be done on non-stratified Cox model and log-rank test. For efficacy endpoint ORR, chi-square test will be used.

For the primary efficacy endpoint (PFS), subgroup levels to be analyzed are:

- iNHL histology (FL vs. other iNHL histology)
- iNHL histology (FL vs. MZL vs. SLL vs. LPL/WM) treatment interaction test only will be performed; for subpopulation analysis see Section 6.2.5.1
- Entry criterion (progression-free and treatment-free interval following the last rituximab-containing treatment ≥ 12 months vs. considered unwilling/unfit to receive chemotherapy for age or comorbidities)
- Presence of bulky disease (yes vs. no)
- Previous treatment with PI3K inhibitors (yes vs. no)
- Prior lines of systemic anti-cancer therapy (1 vs. 2 vs. 3 vs. ≥4)
- Geographic regions 1 (US, Europe, Rest of world)
- Geographic regions 2 (North America, Asia Pacific, Rest of the world)
- Age (<65 vs. ≥ 65)
- ECOG performance status (0 vs. 1 vs. 2)
- Gender (male vs. female)
- Race
- Ethnicity (Hispanic or Latino vs. non-Hispanic and non-Latino)
- BMI group (<18.5 vs. 18.5 - <30 vs. ≥30 kg/m²)
- Renal function at baseline (eGFR Normal: ≥ 90 vs. Mild: 60 - < 90 vs. Moderate: 30 - <60 vs Severe:15- <30 mL/min/1.73m²)

For secondary efficacy endpoints ORR, time to deterioration of DRS-P, and time to improvement in DRS-P, the same subgroup levels will be analyzed excluding gender, BMI group, and renal function at baseline.

The descriptive statistics on ORR and DOR by treatment will also be provided for subjects with at least 2 prior therapies in the MZL subgroup.

6.3 Safety

6.3.1 Adverse Events (AEs)

Adverse events will be coded by Medical Dictionary for Regulatory Activities (MedDRA) codes and National Cancer Institute (NCI) Common Terminology Criteria Adverse Event (CTCAE) version 4.03 (or higher) dictionary. The final versions used will be stored in the clinical database and provided in a footnote to the respective tables. Severity of AEs and hematological/biochemical toxicities based on laboratory measurements will be graded using NCI-CTCAE dictionary. AEs will be classified by the investigator as related or not related to study drug.

A treatment-emergent AE (TEAE) is defined as any event arising or worsening after the start of study drug administration until 30 days after the last study drug administration (end of safety follow-up).

Descriptive summary tables (frequency and percentage of subjects, not of events) will be presented by treatment group, MedDRA SOC/PT and by NCI CTCAE worst grade for the following:

- An overall summary of TEAEs
- TEAEs
- Drug-related TEAEs
- Treatment-emergent serious AEs (TESAEs)
- Drug-related TESAEs
- TEAEs of special safety interest.

Exposure adjusted TEAEs will be summarized by treatment group and MedDRA SOC/PT. The exposure adjusted incidence rate per 100 subject years is defined as 100 times the number of subjects with at least one TEAE divided by the sum of exposure times in years, where exposure time is the time to first occurrence if an TEAE occurs, or the treatment duration plus time at risk after treatment ends if no TEAE occurs. The time at risk after treatment is the minimum of 30 days or the days of death relative to the last treatment.

Incidence rates of TEAEs, drug-related TEAEs (copanlisib/placebo or rituximab-related) and/or TESAEs will be repeated by the following criteria:

- entry criterion strata (progression-free and treatment-free interval following the last rituximab-containing treatment ≥ 12 months vs. considered unwilling/unfit to receive chemotherapy for age or comorbidities)
- renal function at baseline (GFR Normal: ≥ 90 vs. Mild: 60 - < 90 vs. Moderate: 30 - <60 vs. Severe: 15- < 30 mL/min/1.73m²)

- hepatic function at baseline (Normal: Total bilirubin and AST ≤ ULN vs. Mild hepatic impairment: Total bilirubin > ULN to 1.5 x ULN and AST any value; or Total bilirubin ≤ ULN and AST > ULN vs. Moderate impairment: Total bilirubin > 1.5 to 3 x ULN and AST Any value vs. Severe impairment: Total bilirubin > 3-10 x ULN, AST Any value)
- age (<65 vs. ≥ 65)
- gender
- history of hypertension: [PBMQ] Medical history of hypertension (n/y)
- history of diabetes: [PBMQ] Medical history of diabetes mellitus (n/y), and
- BMI group (<18.5 vs. 18.5 - <30 vs. ≥30 kg/m²).

Furthermore, all TEAEs and drug-related TEAEs with incidence of at least 10% will be summarized by MedDRA and by worst NCI-CTCAE grade. In addition, all TEAEs and drug-related TEAEs with incidence of at least 5% will be summarized by MedDRA for worst NCI-CTCAE grade, for NCI-CTCAE grade 3-4. All TEAEs will be summarized by MedDRA for NCI-CTCAE grade 5.

The number and percentage of patients who discontinued study treatment due to TEAE or required a dose reduction (copanlisib/placebo only) or interruption caused by a TEAE will be summarized. The incidences of these TEAEs will be presented also separately by drug relatedness (copanlisib/placebo and/or rituximab).

In addition, the following AE summary according to investigator pathology for FL and MZL patients will be provided:

- Overview of TEAEs (i.e. includes deaths, discontinuations and dose modifications)
- TEAEs
- drug-related TEAEs
- TESAEs
- drug-related TESAEs
- The Standardized MedDRA Query (SMQ) with narrow search (i.e. category 2A): Interstitial lung disease

Deaths will be summarized by treatment group and overall.

In addition, subject listings will be provided for the following AEs by treatment group:

- Subjects who died during study treatment or up to 30 days after the last dose of study medication: subject ID, histology, sponsor AE identifier, start and stop date of study medication, date of death, cause of death, MedDRA SOC and PT term and CTCAE toxicity grade.
- Subjects who died later than 30 days after the last dose of study medication: subject ID, histology, sponsor AE identifier, start and stop date of study medication, date of death, cause of death, MedDRA SOC and PT term and CTCAE toxicity grade.
- Subjects with treatment-emergent SAEs: subject ID, histology, sponsor AE identifier, investigator AE term, MedDRA SOC and PT term and toxicity grade, start and stop

dates of study drug administration, start and stop date of AE, drug related (yes/no), protocol required procedure related (yes/no), outcome, action taken. Listing of study drug-related SAEs will be provided separately for copanlisib/placebo and rituximab, respectively, with similar information.

- Subjects with TEAEs of special interest: subject ID, histology, sponsor AE identifier, investigator AE term, MedDRA SOC and PT term and toxicity grade, start and stop dates of study drug administration, start and stop date of AE, drug related (yes/no), protocol required procedure related (yes/no), outcome, action taken.
- Subjects who discontinued study treatment, required a dose reduction (copanlisib/placebo only), or interruption caused by a TEAE will be listed separately for copanlisib/placebo and rituximab, respectively: subject ID, histology, sponsor AE identifier, investigator AE term, MedDRA SOC and PT term and toxicity grade, start and stop dates of study drug administration, start and stop date of AE, drug related (yes/no), protocol required procedure related (yes/no), outcome.

6.3.2 Adverse events by time-interval

The categories of TEAEs, TE copanlisib-related AEs, and TE serious AEs will be presented according to time-of-onset (new onset or worsening).

The following categories will be presented:

- Day 1 – Day 90: to include all copanlisib-treated patients
- Day 91 – Day 180: to include all copanlisib-treated patients treated for at least 91 days
- Day 181 – Day 270: to include all copanlisib-treated patients treated for at least 181 days
- Day 271 – Day 360: to include all copanlisib-treated patients treated for at least 271 days
- Day 361 – Day 450: to include all copanlisib-treated patients treated for at least 361 days
- Day 451 – Day 540: to include all copanlisib-treated patients treated for at least 451 days
- Day 541 – Day 630: to include all copanlisib-treated patients treated for at least 541 days
- Day 631 and greater: to include all copanlisib-treated patients treated for at least 631 days

6.3.3 Non-infectious Pneumonitis/Interstitial lung disease

Non-infectious Pneumonitis is defined as AE of special safety interest (protocol Section 6.4.1). The Standardized MedDRA Query (SMQ) 'Interstitial lung disease' with narrow search (i.e. category 2A) will be used to select these AEs of special interest.

Summary of patients with Non- infectious pneumonitis/Interstitial lung disease requiring corticosteroids, antibiotics, or both will be displayed descriptively.

6.3.4 COVID-19 relevant adverse events

The COVID-19 relevant adverse events will be identified using a high-level term (HLT) of coronavirus infection via latest MedDRA version. A subject listing of these AEs will be provided. The serious AEs will be flagged in the listing.

6.3.5 Pregnancies

The results of pregnancy tests will be listed. Any pregnancy will be documented.

6.3.6 Clinical Laboratory Parameters

Laboratory toxicities by worst CTCAE grade post-baseline will be summarized. Summary statistics on the values and changes from baseline will also be presented for each quantitative clinical laboratory variable, at each post-baseline visit. For these variables, the number and percentage of patients with transitions from worst grade at baseline to worst grade at post-baseline visits relative to the respective laboratory's reference ranges will be presented by visit. Changes in worst grade of laboratory toxicities at last pre-treatment value compared to worst grade post-baseline value and Changes in worst grade of laboratory toxicities under treatment is also provided. For the change in worst grade of laboratory toxicities under treatment the latest observation will be taken. In case there is more than one observation for the same latest time, the worst grade will be taken.

In addition, for renal function (eGFR and CLcr) and hepatic function, worst grade during study will be summarized.

The eGFR, CLcr and hepatic function grade are defined as the following:

- eGFR, as defined per draft FDA guidance [18]

  Normal: eGFR $\geq$90 ml/min/1.73m²

  Mildly impaired: 60 $\leq$ eGFR $<$ 90 ml/min

  Moderately impaired: 30 $\leq$ eGFR $<$ 60 ml/min/1.73m²

  Severly impaired: 15 $\leq$ eGFR $<$ 30 ml/min/1.73m²

  End stage renal disease: eGFR $<$ 15 ml/min/1.73m²

- Creatinine Clearance (CLcr), as defined per draft FDA guidance [18]:

  Normal: CLcr $\geq$90 ml/min

  Mildly impaired: 60 $\leq$ CLcr $<$ 90 ml/min

  Moderately impaired: 30 $\leq$ CLcr $<$ 60 ml/min

  Severly impaired: 15 $\leq$ CLcr $<$30 ml/min

  End stage renal disease: CLcr $<$ 15

  CLcr will be calculated according to the Cockcroft-Gault formula:

$$CLcr\ (mL/min) = \frac{[140 - age\ (years)] \times weight\ (kg)}{72 \times serum\ creatinine\ (mg/dL)} \{\times 0.85\ for\ female\ patients\}$$

- Hepatic impairment [19] [20]:

  Normal: Total bilirubin and AST ≤ upper limit of normal (ULN)

  Mild impairment: Total bilirubin > ULN to 1.5 x ULN or (Total bilirubin ≤ ULN and AST > ULN)

  Moderate impairment: Total bilirubin > 1.5 to 3 x ULN, any AST

  Severe impairment: Total bilirubin > 3 x ULN, any AST

In addition to the above summary, specific glucose evaluation will be described in the next section.

6.3.7 Further Safety Parameters

6.3.7.1 Hyperglycemia

Hyperglycemia adverse events using specific MedDRA PT grouping 'MLG Hyperglycemia' will be summarized by treatment group and by NCI CTCAE worst grade for the following:

- TEAEs
- Drug-related TEAEs
- TESAE
- Drug-related TESAE

Glucose measurements will be displayed by CTCAE grade of pre copanilisib/placebo infusion and post-dose 1 h and 2 h after the end of copanlisib/placebo infusion, and at the end of rituximab infusion on Cycle 1 Day 1, and by CTCAE grade of pre copanlisib/placebo infusion and post-dose 1 h after the end of copanlisib/placebo infusion, and at the end of rituximab infusion on subsequent visits. Changes from respective pre-dose (defined as the last pre-dose glucose measurement in each visit) in glucose will be summarized using descriptive statistics at each visit and analysis time point by pre-dose fasting status.

Patients will be categorized according to their HbA1c baseline values as

- < 5.7%
- ≥ 5.7% – < 6.5%
- ≥ 6.5%.

Within each group, the maximum HbA1c values by cycle and end of treatment visit, as well as 3 months after last dose (if available) will be summarized by descriptive statistics. Shift table from baseline to end of treatment visit, and 3 months after last dose for each category of HbA1c will be presented.

For patients who required home glucose monitoring, home blood glucose information will be listed in Appendix 16 of the CSR.

For patients with hyperglycemia adverse events, the number of patients using antihyperglycemic treatment with drug start date on or after their first treatment-emergent

adverse event will be summarized in overall, and separately for the WHO-DD ATC classes for a) insulin and analogues and b) Blood glucose lowering drugs, excluding insulins. In addition, these summaries will be presented separately for the subgroups of patients with vs without history of diabetes (as defined in Section 6.3.1).

6.3.7.2 Electrocardiogram (ECG)

Results of electrocardiogram (ECG) will be summarized with descriptive statistics and/or frequency tables. ECG data will be summarized by visit and timepoint, including change from baseline where appropriate.

The overall interpretation of the 12-lead ECG and the ECG diagnosis, as well as the overall interpretation of the echocardiogram and the corresponding diagnosis will be summarized by visit.

Frequency and shift tables for number of patients by interpretation of electrocardiogram (ECG) as abnormal will be performed at all-time points.

6.3.7.3 Vital Signs

Results of vital signs will be summarized with descriptive statistics and/or frequency tables. Vital signs will be summarized by visit and timepoint, including change from baseline where appropriate.

6.3.7.4 Echocardiogram

The number of patients with abnormal cardiac function will be displayed for echocardiogram and multiple gated acquisition (MUGA) by visit.

6.3.7.5 Hypertension

Hypertension adverse events will be presented using MLG grouping of MLG Hypertension for the following:

- TEAEs
- Drug-related TEAEs
- TESAE
- Drug-related TESAE

For patients with hypertension adverse events, the number of patients using antihypertensive treatment with drug start date on or after their first treatment-emergent adverse event will be summarized descriptively. In addition, the number of patients requiring new antihypertensive treatment will be presented separately for the subgroups of patients with vs without history of hypertension (as defined as Section 6.3.1). The number of patients with abnormal blood pressure (according to CTCAE grading) will be displayed by visit and worst post-infusion hypertension grade.

6.4 Pharmacokinetics/pharmacodynamics

Individual concentration-time data of copanlisib and its metabolite M-1 as well as rituximab will be listed. Further analyses will be performed and reported under separate cover by PK experts (pharmacometrics) and include population pharmacokinetics, and exposure response analysis.

6.5 Biomarker evaluation

Biomarker data, if collected in the clinical database will be listed. Retrospective exploratory biomarker analyses including evaluation of relationship between biomarker data and clinical outcomes will be performed by or under the supervision of the sponsor's genomics and biomarker statistical expert and/or pharmacometrics expert. Results of these analyses will be reported separately, under a separate cover.

7. Document history and changes in the planned statistical analysis

7.1 Document history

- SAP final version 1.0, dated 13 FEB 2017
- SAP final version 2.0, dated 12 DEC 2018
- SAP final version 3.0, dated 15 JAN 2020
- SAP final version 4.0, dated 03 MAR 2020
- SAP final version 5.0, dated 28 JUL 2020
- SAP final version 6.0, dated 26 AUG 2020

7.2 Changes in the planned statistical analysis

7.2.1 Changes from the SAP V1.0

SAP version 2.0 was done to reflect Amendment 9 of the protocol.

Study objective

Wording of the study objective was adapted in order to be fully consistent with updates in the protocol Amendment 9 (in which changes in the inclusion criterion 13 and exclusion criterion 4 were aligned with the study objectives).

Sample size

The sample size estimation was updated to be based on the FAS instead of the follicular lymphoma subgroup (also changed in the protocol Amendment 9).

Handling of Missing Data

Reorganized the rules for handling of missing data for time to event variables under the definitions of individual time to event variables in Section 6.2.

Stratification

The assessment and handling of stratification factor 'entry criterion' is further clarified.

The definition entry criteria stratification factor was adjusted according to the clarification of inclusion criterion 13 in the protocol Amendment 9. Since the inclusion criterion 13 was modified to be consistent with exclusion criterion 4, the following wording was also harmonized in the definition of the entry criteria stratification factor: progression-free and treatment-free interval of ≥ 12 months after completion of the last rituximab-containing treatment.

For the purpose of the analysis, patients who were stratified before protocol Amendment 7 as unfit will be combined with those who were stratified after protocol Amendment 7 as unfit and unwilling; and patients who were stratified before protocol Amendment 9 as "treatment-free ≥ 12 months after completion of the last rituximab-containing treatment" will be combined with those who were stratified after protocol Amendment 9 as "progression-free and treatment-free ≥ 12 months after completion of the last rituximab-containing treatment".

Which stratum patients should be randomized to if a they fulfill both entry criteria was changed to match IxRS instructions and protocol Amendment 9.

Demography and baseline characteristics

Demographic and other baseline characteristics will be summarized for each Histology (FL, MZL, SLL and LPL/WM).

HbA1c, Diabetic and Hypertension history and Hepatic function at baseline were added to the baseline characteristics to be summarized.

Efficacy

There is no regulatory obligation of powering on a specific subtype of iNHL. As a result, the total sample size was reduced from 567 patients to 450 patients and the primary efficacy analysis will be performed in the FAS instead of both FAS and FL subgroup. The confirmatory statistical testing section was modified to reflect this change in strategy.

Based on more updated site information, the recruitment ramps up period was increased from 6 months to 29 months and the dropout rate was increased to 30%.

All analysis by FL and total study population will be done for the FAS population.

Additional conditions related to the minimum number of PFS events was included.

A potential pooling of strata strategy has been added.

Additional PFS sensitivity analysis were added.

Additional rules were added for censoring or calculating PFS in special case such as, if progression occurs at the next tumor assessment after 2 or more consecutive missing or non-evaluable assessments.

For the purpose of the analysis, patients who were stratified before protocol Amendment 7 as unfit will be combined with those who were stratified after protocol Amendment 7 as unfit and unwilling; and patients who were stratified before protocol Amendment 9 as treatment-free ≥ 12 months after completion of the last rituximab-containing treatment will be combined with those who were stratified after protocol amendment 8 as progression- and treatment-free ≥ 12 months after completion of the last rituximab-containing treatment.

The Time to deterioration in DRS-P of at least 3 points was changed to Time to deterioration in DRS-P of at least 3.5 points.

The Time to improvement in DRS-P of at least 3 points was changed to Time to deterioration in DRS-P of at least 3.5 points.

Efficacy Subgroup Analysis

Histology Specific Subgroup Analysis were added.

Safety

Adverse Events

Clarification that incidences of TEAE, drug-related and/or Serious TEAEs and TEAEs of special interest will be summarize by subgroup as specified in section 6.3.1.

The categories of TEAEs, drug-related TEAEs, and TESAEs will be presented according to time-of-onset (new onset or worsening)

Clinical Laboratory

Clinical laboratory use updated to reflect the current best practices.

7.2.2 Changes from the SAP V2.0

SAP version 3.0 was done to reflect Amendment 10 of the protocol.

- The statistical assumptions for the primary efficacy analysis of PFS were modified. The required number of PFS events was changed from 288 to 190
- The confirmatory testing hierarchy was modified to include a test for the PFS in the combined FL and MZL population after testing the overall iNHL population and a test for the ORR in the combined FL and MZL population after the testing on PFS.
- Added TEAE by histology (FL, MZL) analysis.
- Added ORR by histology analysis.

7.2.3 Changes from the SAP V3.0

The following updates are made:

- Subgroup analysis by history of safety and exposure analysis were updated to be done for FL/MZL and all patients.
- The Pooling strategy in Appendix 9.3 is updated.

- Removed concurrent procedures from the list of 'Prior and concurrent anti-cancer therapy' (Section 6.1.4).
- Added eGFR, CLcr, and hepatic function calculation and grade definition (Section 6.3.5).
- Updated censoring rules for DRS-P endpoints (Section 6.2.2).
- Remove PKS (pharmacokinetics analysis set) definition (Section 5.1).
- Updated disposition section by removing the analysis 'by investigator' (Section 6.1.1).
- Added age group/renal function/bilirubin/histology by investigator/histology by independent assessor in demographic and baseline characteristics section (Section 6.1.2).
- Removed concurrent procedure in 'prior and concurrent anti-cancer therapy' (Section 6.1.4).
- Removed 'by hepatic function at baseline' from other subgroup analysis (Section 6.2.5.2).

7.2.4 Changes from the SAP V4.0

SAP version 5.0 was done to reflect Amendment 11 of the protocol.

- Power estimation is updated based on blinded study information using Schoenfeld's formula. (Section 4.5 & Section 6.2.3)
- Removed potential pooling of strata. In order to avoid a too low number of events, only stratification factors "iNHL histology" and "entry criterion" will be adjusted simultaneously in the statistical analyses. (Section 4.6.5)
- Added a separate subject listing of protocol deviations related to COVID-19. (Section 6.1.1)
- Added PKS (pharmacokinetics analysis set) definition because listings will be provided for PK data. (Section 5.1)
- Removed the Kaplan-Meier plot for 'time to the end of study'. (Section 6.1.1)
- Removed histology group from independent assessor at baseline from the baseline summary tables because all the subgroup analyses are based on the histology group by investigator at baseline. (Section 6.1.2)
- Removed 'Concurrent radiotherapies' from the list of 'Prior and concurrent anti-cancer therapy'. (Section 6.1.4).
- Removed evaluations for treatment duration and exposure in FAS. (Section 5.1 & Section 6.1.5)
- Modified one of the sensitivity analyses for PFS based on the stratification information entered in IxRS to be based on the stratification information entered in CRF. (Section 6.2.1)

- Added additional sensitivity analysis for the primary efficacy endpoint PFS considering any change as a progression event. (Section 6.2.1)
- Added possible sensitivity analysis for the primary efficacy endpoint PFS if the impact of COVID-19 on the study treatment was observed. (Section 6.2.1)
- Added DCR and its definition as one of the secondary efficacy variables. (Section 6.2.2)
- Removed a test of the PFS in SLL population from the confirmatory statistical test strategy for the United States. (Section 6.2.3)
- Added a separate confirmatory statistical test strategy for Europe. (Section 6.2.3).
- Added subgroup analyses for the secondary endpoints DOR, and TTP by histology subgroups FL and MZL. (Section 6.2.5.1)
- Added subgroup analyses on ORR and DOR for subjects with at least 2 prior therapies in the MZL subgroup. (Section 6.2.5.2)
- Added Section 6.3.4 for COVID-19 relevant adverse events.
- Modifications were made on the summary of hyperglycemia adverse events and laboratory hyperglycemia glucose measurements. (Section 6.3.7.1)
- Modifications were made on the summary of hypertension adverse events. (Section 6.3.7.5)

7.2.5 Changes from the SAP V5.0

- Added a statement to indicate the evaluation of histopathological diagnosis from central reviewer at baseline is retrospective. Data will be presented in the CSR addendum and are considered as exploratory. (Section 6.1.2)
- Added a sensitivity analysis of PFS to align with US health authority feedback received on August 3, 2020. (Section 6.2.1)
- Removed stratification factors from the subgroup analyses. Subgroup analyses will be conducted using unstratified log-rank test or unstratified Cox regression model for time to event outcome. For binary efficacy endpoints, e.g. ORR, chi-square test will be used. (Section 6.2.5)

8. References

1. Shankland KR, Armitage JO, Hancock BW. Non-Hodgkin lymphoma. Lancet. 2012; 380(9844):848-57.
2. Hiddemann W, Kneba M, Dreyling M, Schmitz N, Lengfelder E, Schmits R, et al. Frontline therapy with rituximab added to the combination of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) significantly improves the outcome for patients with advanced-stage follicular lymphoma compared with therapy with CHOP alone: results of a prospective randomized study of the German Low-Grade Lymphoma Study Group. Blood. 2005;106(12):3725-32.
3. Marcus R, Imrie K, Solal-Celigny P, Catalano JV, Dmoszynska A, Raposo JC, et al. Phase III study of R-CVP compared with cyclophosphamide, vincristine, and prednisone

alone in patients with previously untreated advanced follicular lymphoma. J Clin Oncol. 2008;26(28):4579-86.

4. Rummel M, Kaiser U, Balser C. Bendamustine plus rituximab versus fludarabine plus rituximab in patients with relapsed follicular, indolent and mantle cell lymphomas - final results of the randomized phase III study NHL 2-2003 on behalf of the StiL (Study Group Indolent Lymphomas, Germany) Blood. 2010(116).
5. Hainsworth JD, Litchy S, Burris HA, 3rd, Scullin DC, Jr., Corso SW, Yardley DA, et al. Rituximab as first-line and maintenance therapy for patients with indolent non-hodgkin's lymphoma. J Clin Oncol. 2002;20(20):4261-7.
6. McLaughlin P, Grillo-Lopez AJ, Link BK, Levy R, Czuczman MS, Williams ME, et al. Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program. J Clin Oncol. 1998;16(8):2825-33.
7. Johnston A, Bouafia-Sauvy F, Broussais-Guillaumot F, Michallet AS, Traulle C, Salles G, et al. Retreatment with rituximab in 178 patients with relapsed and refractory B-cell lymphomas: a single institution case control study. Leuk Lymphoma. 2010;51(3):399-405.
8. Hainsworth JD, Litchy S, Shaffer DW, Lackey VL, Grimaldi M, Greco FA. Maximizing therapeutic benefit of rituximab: maintenance therapy versus re-treatment at progression in patients with indolent non-Hodgkin's lymphoma--a randomized phase II trial of the Minnie Pearl Cancer Research Network. J Clin Oncol. 2005;23(6):1088-95.
9. Ihle NT, Powis G. Take your PIK: phosphatidylinositol 3-kinase inhibitors race through the clinic and toward cancer therapy. Molecular cancer therapeutics. 2009;8(1):1-9.
10. Yuan TL, Cantley LC. PI3K pathway alterations in cancer: variations on a theme. Oncogene. 2008;27(41):5497-510.
11. Cheson, BD, Fisher, RI, Barrington, SF, Cavalli, F, Schwartz, LH, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. J Clin Oncol. 2014; 32(27):3059-67.
12. Owen, RG, Kyle, RA, Stone, MJ, Rawstron, AC, Leblond, V, et al. Response assessment in Waldenstrom macroglobulinaemia: update from the VIth International Workshop. Britisch journal of haematology. 2013; 160(2):171-6.
13. Leonard JP, Trneny M, Izutsu K, Fowler NH, Hong X, Zhu J, Zhang H, Offner F, Scheliga A, Nowakowski GS, Pinto A. AUGMENT: a phase III study of lenalidomide plus rituximab versus placebo plus rituximab in relapsed or refractory indolent lymphoma. Journal of Clinical Oncology. 2019;37(14):1188-+.
14. Collett D. Modelling Survival Data in Medical Research: Chapman and Hall; 2003.
15. FDA. Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry, April 2015. Available from: https://www.fda.gov/downloads/drugs/guidances/ucm259421.pdf (accessed date: 18 MAY 2020)

16. Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. 'Categorical Data Analysis' in Statistical methodology in the pharmaceutical sciences /edited by Berry DA, Marcel Dekker; 1990.
17. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika. 1981;68(1): 316–319.
18. FDA. Guidance for Industry Pharmacokinetics in Patients with Impaired Renal Function – Study Design, Data Analysis, and Impact on Dosing and Labeling, Draft guidance, March 2010
19. FDA. Guidance for Industry Pharmacokinetics in Patients with Impaired Hepatic Function: Study Design, Data Analysis, and Impact on Dosing and Labeling, May 2003
20. CTEP Simplified Disease Classification v1.0 (MedDRA v10.0), Available from: https://ctep.cancer.gov/protocolDevelopment/codes_values.htm (accessed date: 21 MAY 2020)

9. Appendix

9.1 Scoring Instruction for FLymSI-18

Record answers of each item response (0 = 'Not at all' to 4 = 'Very much'). In order to have a common order for the scores with 0 reflecting worst outcome and 4 reflecting best outcome, reverse the coding for the following items by subtracting the response from '4':

GP1: I have lack of energy

GP4: I have pain

C2: I am losing weight

Leu1: I am bothered by lumps or swelling in certain parts of my body

BMT6: I get tired easily

BP1: I have bone pain

HI8: I have trouble concentrating

GE6: I worry that my condition will get worse

BRM9: I have emotional ups and downs

Leu4: Because of my illness, I have difficulty planning for the future

Leu5: I feel uncertain about my future health

GP2: I have nausea

N3: I worry about getting infections

GP5: I am bothered by side effects of treatment

FLymSI-18 Total score (range 0-72)

Sum the (reversed, if applicable) responses of all items. Multiply by 18 (number of items in the FLymSI-18) and divide by the number of items answered.

FLymSI-18 DRS-P score (range 0-36)

Sum the (reversed, if applicable) responses of the items GP1, GP4, C2, Leu1, BMT6, BP1, HI8, GF5, C6. Multiply by 9 (number of items in the FLymSI-18 DRS-P score) and divide by the number of items answered.

FLymSI-18 DRS-E score (range 0-16)

Sum the reversed responses of the items GE6, BRM9, Leu4, Leu5. Multiply by 4 (number of items in the FLymSI-18 DRS-E score) and divide by the number of items answered.

FLymSI-18 TSE score (range 0-12)

Sum the reversed responses of the items GP2, N3, GP5. Multiply by 3 (number of items in the FLymSI-18 TSE score) and divide by the number of items answered.

FLymSI-18 FWB score (range 0-8)

Sum the responses of the items GF3, GF7. Multiply by 2 (number of items in the FLymSI-18 FWB score) and divide by the number of items answered.

NCCN/FACT Lymphoma Symptom Index-18 (FLymSI-18)
Scoring Guidelines (Version 2)

Instructions:*
1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the symptom index score.
4. As with all FACIT questionnaires, a high score is good. Therefore, a score of "0" is a severely symptomatic patient and the highest possible score is an asymptomatic patient.

Scale	Item Code	Reverse item?		Item response	Item Score
FLymSI-18 Total	GP1	4	-	_____	=_____
	GP4	4	-	_____	=_____
	C2	4	-	_____	=_____
Score range: 0-72	LEU1	4	-	_____	=_____
	BMT6	4	-	_____	=_____
	BP1	4	-	_____	=_____
	HI8	4	-	_____	=_____
	GF5	0	+	_____	=_____
	Ca6	0	+	_____	=_____
	GE6	4	-	_____	=_____
	BRM9	4	-	_____	=_____
	LEU4	4	-	_____	=_____
	LEU5	4	-	_____	=_____
	GP2	4	-	_____	=_____
	N3	4	-	_____	=_____

GP5	4	-	_____	=_____
GF3	0	+	_____	=_____
GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 18: _____
Divide by number of items answered: _____
=FLymSI-18 score

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
FLymSI-DRS-P (Disease Related Symptoms-Physical) *Score range:* 0-36	GP1	4	-	_____	=_____
	GP4	4	-	_____	=_____
	C2	4	-	_____	=_____
	LEU1	4	-	_____	=_____
	BMT6	4	-	_____	=_____
	BP1	4	-	_____	=_____
	HI8	4	-	_____	=_____
	GF5	0	+	_____	=_____
	Ca6	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 9: _____
Divide by number of items answered: _____
=FLymSI-DRS-P score

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
FLymSI-DRS-E (Disease Related Symptoms-Emotional) *Score range:* 0-16	GE6	4	-	_____	=_____
	BRM9	4	-	_____	=_____
	LEU4	4	-	_____	=_____
	LEU5	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 4: _____
Divide by number of items answered: _____
=FLymSI-DRS-E score

FLymSI-TSE (Treatment Side Effects) *Score range:* 0-12	GP2	4	-	_____	=_____
	N3	4	-	_____	=_____
	GP5	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 3: _____
Divide by number of items answered: _____
=FLymSI-TSE score

FLymSI-F/WB (Function/ Well-Being)	GF3	0	+	_____	=_____
	GF7	0	+	_____	=_____

Score range: 0-8

Sum individual item scores: _____
Multiply by 2: _____
Divide by number of items answered: _____
=FLymSI-F/WB score

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.

9.2 Patient Reported Outcomes

Please refer to Patient Reported Outcomes Dossier in Support of the National Comprehensive Cancer Network—Functional Assessment of Cancer Therapy Lymphoma Symptom Index-18 (NFLymSI-18) in Indolent B-Cell Non-Hodgkin's Lymphoma (iNHL) Patients by (Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University)

9.3 PFS Rules

9.3.1 Progression and PD after Two or More Consecutive Missed tumor assessments

If the progression occurs at the next tumor assessment after 2 or more consecutive missing or non-evaluable assessments, PFS will be censored at the date of the last evaluable scan before the consecutive missing assessments. Considering tumor assessment are scheduled at every 8 weeks (± 7 days) during Year 1, every 12 weeks (± 7 days) during Year 2, and every 24 weeks (± 7 days) during Year 3 and onwards), the following conditions will apply to account for the tumor assessment schedule change starting at year 2 and year 3:

For subjects with progression observed before death with an evaluable post-baseline evaluation before two missing assessments.

- if date of progression is ≤ 64 weeks (56 weeks+8 weeks) after randomization and > last evaluable assessment + 2*8 + 1 weeks, PFS will be censored at the date of the last evaluable scan.
- if date of progression is > 64 (56+8 weeks) and ≤ 76 (64+12 weeks) weeks after randomization and > last evaluable tumor assessment + 8+12 + 1 weeks, PFS will be censored at the date of the last evaluable scan BEFORE the two missing assessments.
- if date of progression is > 76 and ≤ 120 weeks (108+12 weeks) after randomization and > last evaluable tumor assessment + 2*12 + 1 weeks, PFS will be censored at the date of the last evaluable scan BEFORE the two missing assessments.
- if date of progression is > 120 and ≤ 144 weeks (120+24 weeks) after randomization and > last evaluable tumor assessment + 12+24 + 1 weeks, PFS will be censored at the date of the last evaluable scan BEFORE the two missing assessments.

- if date of progression is > 144 weeks after randomization and > last evaluable tumor assessment + 2*24 + 1 weeks, PFS will be censored at the date of the last evaluable scan BEFORE the two missing assessments.
- otherwise the PFS time will not be censored and be calculated as the date of progression minus the date of randomization.

For subjects with progression observed before death without any evaluable post-baseline tumor evaluation before disease progression,

- if progression occurs within the 17 (16+1) weeks after randomization, PFS time will not be censored and will be calculated as the date of progression minus the date of randomization
- if progression occurs later than 17 (16+1) weeks after randomization, PFS will be censored at Day 1.

9.3.2 Death

Considering tumor assessments are scheduled at every 8 weeks (± 7 days) during Year 1, every 12 weeks (± 7 days) during Year 2, and every 24 weeks (± 7 days) during Year 3 and onwards), the following conditions will apply:

For subjects who die after the last evaluable tumor assessment, in the absence of progression, the death event is a PFS event in the following conditions:

- if date of death is ≤ 64 weeks after randomization, and within the 16+1 weeks of the last evaluable tumor assessment,
- if date of death is > 64 and ≤ 76 weeks after randomization, and within the 20+1 weeks of the last evaluable tumor assessment,
- if date of death is > 76 and ≤ 120 weeks after randomization, and within the 24 +1 weeks of the last evaluable tumor assessment,
- if date of death is > 120 and ≤ 144 weeks after randomization, and within the 36 +1 weeks of the last evaluable tumor assessment,
- if date of death is > 144 weeks after randomization, and within the 48 +1 weeks of the last evaluable tumor assessment.

9.3.3 Subjects with New Anti-Cancer Therapy

For subjects who change anti-cancer therapy to something other than the study medication prior to observing progression, PFS will be censored at the date of the last evaluable tumor assessment prior to the initiation of anti-cancer therapy. Independent review committee (IRC) assessments will be used in the event of IRC and investigator disagreement on the last response assessment prior to a switch to a new anticancer therapy. For Waldenström's macroglobulinemia (WM) patients, if PD was assessed by an investigator based on laboratory parameters alone, no independent confirmation of PD by independent blinded review is necessary.

9.3.4 Subjects Discontinued or Withdrew

For subjects who discontinue or withdraw early from the study without documented disease progression, PFS will be censored at the date of the last evaluable tumor assessment unless

the subject dies and death is considered as a PFS event mentioned above. Note that this applies to subjects without progression based on blinded central radiological review but who discontinued because according to the investigator assessment he/she had progresses (if such a subject exists).

9.4 Imputation rule for Initial Diagnosis Date and Progression Date

The following rules are used to impute partial initial diagnosis date and progression date,

A. If partial date has day and month missing, then January 01 will be assigned to the missing fields.

B. If partial date has missing day only, then the first day of the month will be assigned to the missing day.

C. If imputed progression date from A or B above is before the Randomization date, then Randomization date will be assigned to overwrite the imputed date from A or B.

D. If the date is completely missing, no imputation will be applied for initial diagnosis date. Use day 1 for progression date.