Official Title: A Study to Evaluate the Benefit of Venetoclax Plus Rituximab Compared With Bendamustine Plus Rituximab in Participants With Relapsed or Refractory Chronic Lymphocytic Leukemia (CLL) (MURANO)

NCT Number: **NCT02005471**

Document Dates: SAP V1.0: 27-Jul-15
SAP V4.0: 18-May-17

STATISTICAL ANALYSIS PLAN

TITLE: **A MULTICENTER, PHASE III, OPEN-LABEL, RANDOMIZED STUDY IN RELAPSED/REFRACTORY PATIENTS WITH CHRONIC LYMPHOCYTIC LEUKEMIA TO EVALUATE THE BENEFIT OF VENETOCLAX (GDC-0199/ABT-199) PLUS RITUXIMAB COMPARED WITH BENDAMUSTINE PLUS RITUXIMAB**

PROTOCOL NUMBER: GO28667

STUDY DRUG: Venetoclax (GDC-0199/ABT-199) (RO5537382)

VERSION NUMBER: 1

IND NUMBER: 110159

EUDRACT NUMBER: 2013-002110-12

SPONSOR: F. Hoffmann-La Roche Ltd and AbbVie Inc. will act as co-sponsors of this trial globally.

PLAN PREPARED BY: [redacted] Ph.D.

DATE FINAL: See electronic date stamp below.

STATISTICAL ANALYSIS PLAN APPROVAL

Name	Reason for Signing	Date and Time (UTC)
[redacted]	Company Signatory	27-Jul-2015 19:42:58

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. BACKGROUND

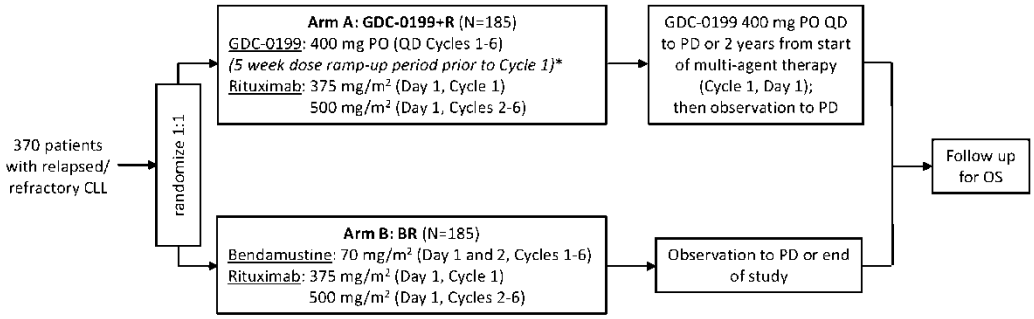
This document is based on the Statistical Considerations and Analysis Plan section of the study protocol and will provide more details on the planned statistical analyses. For purposes of registration, the analyses outlined in this Statistical Analysis Plan will supersede those specified in the protocol.

2. STUDY DESIGN

This is an open-label, international, multicenter, randomized, Phase III study to investigate the efficacy and safety of venetoclax (GDC-0199/ABT-199) in combination with rituximab (venetoclax + R) compared with bendamustine in combination with rituximab (BR) in patients with relapsed or refractory chronic lymphocytic leukemia (CLL; see Figure 1).

The primary objective of the study is to evaluate the efficacy of venetoclax + R compared with BR in patients with relapsed or refractory CLL, as measured by investigator-assessed progression-free survival (PFS).

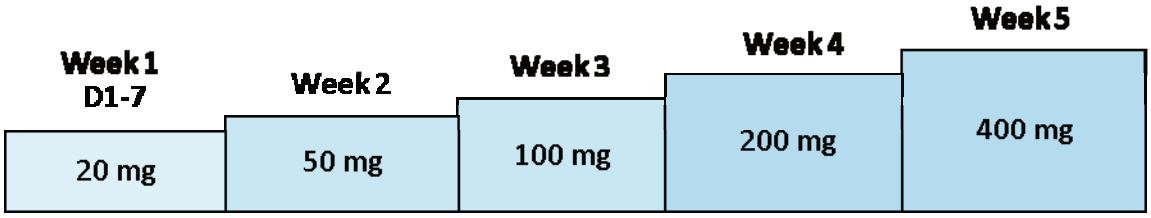
Figure 1 Design of the Study



Arm A = venetoclax (GDC-0199) and rituximab (venetoclax + R); Arm B = bendamustine and rituximab (BR); 1 cycle = 28 days; CLL = chronic lymphocytic leukemia; OS = overall survival; PD = progressive disease; PO = per os; QD = once daily.

* Patients will receive venetoclax starting on Day 1 (venetoclax dose ramp-up period) as delineated in Figure 2. Venetoclax will then be self-administered at 400 mg per day for a maximum of 2 years from Cycle 1 Day 1 or until disease progression (whichever is earlier). Combination therapy consisting of 6 cycles of rituximab and daily venetoclax dosing will start after completion of the venetoclax ramp-up period.

Figure 2 Venetoclax Dosing Scheme during the Ramp-Up Period



D = day.

Note: For all patients enrolled after Protocol Version 4. Prior to Version 4, ramp up was 4 or 5 weeks.

Approximately 370 patients will be recruited and randomly assigned in 1:1 ratio to receive either venetoclax+R (Arm A) or BR (Arm B). Randomization will be stratified according to the following 3 factors:

- 17p deletion: yes or no
- Risk status: high risk or low risk

  High risk: defined as harboring 17p deletion, no response to front-line chemotherapy-containing regimen, relapsed within 12 months after chemotherapy, or relapsed within 24 months after chemoimmunotherapy

  Low risk: defined as relapse more than 12 months after chemotherapy or 24 months after chemoimmunotherapy.
- Geographic region: United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia

Patients randomized to Arm A (venetoclax + R) will have a 5-week venetoclax dose ramp-up period to reach the target dose of 400 mg daily, followed by 6 cycles of rituximab consisting of a single infusion on the first day of each 28-day cycle. Patients will continue to take their daily dose of venetoclax during the rituximab cycles. Patients whose disease has not progressed following the completion of the 6 cycles will continue to receive venetoclax until disease progression or for a maximum of 2 years from Cycle 1 Day 1. Patients randomized to Arm B (BR) will receive 6 cycles of BR consisting of a single infusion of rituximab on Day 1 and bendamustine infusions on Days 1 and 2 of each 28-day cycle.

After 6 cycles of combination therapy, patients in both arms will be followed clinically every 3 months through Year 3 from Cycle 1, Day 1, after which they will be followed every 6 months for an additional 2 years, until withdrawal of consent, or until end of study, whichever comes first.

Patients who discontinue all components of study treatment for other reasons, or receive a new anti-CLL therapy at any time during follow-up in the absence of disease

progression will also be followed for progression and survival according to the schedule defined in the protocol.

2.1 PROTOCOL SYNOPSIS

The protocol synopsis of Amendment 5 is provided in Appendix 1. For additional details, see the Schedule of Assessments in Appendix 2.

2.2 OUTCOME MEASURES

All patients will have baseline tumor assessment at the screening visit and will be assessed for response to treatment at the scheduled response assessment follow-up visits using standard clinical and laboratory examinations and/or computed tomography (CT) scans according to the International Workshop on CLL (iwCLL) guidelines (Hallek et al. 2008).

The assessment of response and progression by the investigator will be considered the primary analysis for all of the endpoints described in the study. Response and progression will also be assessed by an Independent Review Committee (IRC).

2.2.1 Primary Efficacy Outcome Measures

The primary efficacy outcome measure for this study is investigator-assessed PFS. [redacted]

Duration of PFS is defined as the time from randomization to the first occurrence of progression or relapse, determined using standard iwCLL guidelines, or death from any cause, whichever comes first. Data from patients without disease progression, relapse, or death will be censored at the time of the last response assessment. If no response assessments were performed after the baseline visit, the PFS date will be censored at the randomization date plus one day. Patients who have initiated new anti-CLL therapy without documented disease progression will not be censored.

2.2.2 Secondary Efficacy Outcome Measures

The secondary efficacy outcome measures for this study are as follows:

- IRC-assessed PFS in the subset of CLL patients with 17p deletion identified by fluorescence in situ hybridization (FISH) testing at a central laboratory
- Investigator-assessed PFS in the subset of CLL patients with 17p deletion identified by FISH testing at a central laboratory
- Overall response rate (ORR, which is defined as the percentage of patients with complete response [CR], complete response with incomplete marrow recovery [CRi], nodular partial response [nPR], or partial response [PR]) as assessed by the investigator. In cases where no post-baseline response assessment is available, patients will be considered non-responders.

- ORR, CR, CRi, nPR, and PR rates at the end of combination treatment response visit, as assessed by the investigator. In cases where no post-baseline response assessment is available, patients will be considered non-responders.
- ORR, CR, CRi, nPR, and PR rates at end of combination treatment response visit, as assessed by the IRC. In cases where no post-baseline tumor assessment is available, patients will be considered non-responders.
- Overall survival (OS), which is defined as the time from randomization to death from any cause. Patients who were not reported as having died at the time of the analysis will be censored at the date when they were last known to be alive.
- Event-free survival (EFS), which is defined as the time between date of randomization and the date of disease progression/relapse, death, or start of a new non-protocol-specified anti-CLL therapy. If the specified event (disease progression/relapse, death, start of a new anti-CLL treatment) does not occur, patients will be censored at the date of last response assessment. In cases where no post-baseline response assessment is available, patients without an event will be censored at the randomization date plus one day.
- Duration of response (DOR), which is defined for patients with a best overall response of CR, CRi, nPR, or PR as the time from first occurrence of a documented CR/CRi/nPR/PR to disease progression/relapse, as assessed by the investigator, or death from any cause. Patients with no documented progression or death after CR, CRi, nPR, or PR will be censored at the last date at which they are known to have had the CR, CRi, nPR, or PR. Patients who have never responded will not be included in this analysis.
- Time to next anti-CLL treatment (TTNT), which is defined as the time from randomization to start of new non-protocol–specified anti-CLL therapy or death from any cause. Patients who were reported as not having started new non-protocol anti-CLL therapy or death will be censored at the last visit date for this outcome analysis.
- Minimal residual disease (MRD) response rate at the end of combination treatment response visit as measured at a central laboratory on peripheral blood samples and/or bone marrow aspirate samples. In cases where no post-baseline MRD assessment is available, patients will be considered as MRD-positive in this analysis.

2.2.3 Exploratory Efficacy Outcome Measures

The exploratory outcome measures for this study are as follows:

- Assessment of potential biomarkers that are prognostic and/or predictive of response and resistance to treatment with venetoclax + R or BR
- Evaluation of the relationship between clinical response and PFS and various potential biomarkers for patients treated with venetoclax + R or BR
- MRD response rate as measured at a central laboratory on peripheral blood samples and/or bone marrow aspirate samples over time.

2.2.4 Pharmacokinetic Efficacy Outcome Measures

The pharmacokinetic (PK) outcome measures for this study are as follows:

- Apparent clearance, apparent volume of distribution, and other appropriate PK parameters of venetoclax characterized using population PK techniques

2.2.5 Pharmacodynamic Efficacy Outcome Measure

The pharmacodynamic outcome measure for this study is as follows:

- Serial assessment of B-cell, T-cell, and NK-cell lymphocyte subsets by flow cytometry

2.2.6 Safety Outcome Measures

The safety outcome measures for this study are as follows:

- Incidence, nature, and severity of adverse events and serious adverse events, including deaths
- Changes in vital signs, physical findings, and clinical laboratory results (including hematology and chemistry) during and following administration of study treatment
- Incidence of adverse events of special interest:

  Grade $\geq$3 tumor lysis syndrome (TLS) and infusion-related reactions (IRRs)
- Measures of immune function, including serial immunoglobulin levels (IgG, IgM, IgA) following treatment with venetoclax+R or BR

2.2.7 Health-Related Quality of Life (HRQOL) and Symptom Measures

The HRQOL and disease- and treatment-related symptom measures for this study are as follows:

- MD Anderson Symptom Inventory (MDASI) (see Appendix 5 of the protocol)
- European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30) and Quality of Life Questionnaire CLL module (QLQ-CLL16) (see Appendix 3 and Appendix 4, respectively, of the protocol)

2.2.8 Health Economic Outcome Measure

The health economic outcome measure for this study is as follows:

- The EQ-5D-5L questionnaire (see Appendix 6 of the protocol)

2.3 DETERMINATION OF SAMPLE SIZE

The primary endpoint of PFS was used to determine the sample size for the study based on the following assumptions:

- Two-sided log-rank test at the 0.05 level of significance
- 80% power to detect a hazard ratio (HR) for venetoclax+R versus BR of 0.66, corresponding to an approximate median improvement of 15.2 months (Fischer et al. 2011) to 23 months (34% reduction in risk of a PFS event)

- Exponential distribution of PFS
- An annual dropout rate of 5%
- One interim analysis for efficacy at approximately 75% of total investigator-assessed PFS events (140 investigator-assessed PFS events).

With these assumptions, 186 investigator-assessed PFS events are required to achieve 80% power for the primary analysis of PFS in all patients. It is planned to enroll 370 patients across the two arms, with 1:1 randomization ratio.

Sample size calculations, including the calculations, were performed with Insightful S+Seq Trial S 2.0.6.

2.4 ANALYSIS TIMING

An efficacy interim analysis is planned at approximately 140 investigator-assessed PFS events (75% of the 186 events required for the final primary efficacy analysis) have occurred for both treatment arms combined (see Section 4.11). The stopping boundary follows a unified family with parameter $P=2$ (Kittelson et al. 1999). Based on 140 events, this corresponds to approximately a 2-sided p-value of $2\times0.0013=0.0026$.

If the p-value of the two-sided log-rank test is less than or equal to 0.0026, the trial will have met its primary efficacy endpoint (corresponding to a HR of approximately 0.60 or better). Based on the assumption of HR=0.66 (venetoclax+R arm vs. BR arm), and the duration of enrollment of 20 months, the interim analysis will happen at around 26 months after first patient in. If the study results are not released in the IA, the final analysis for the study will be conducted when approximately 186 PFS events based on the investigator assessment have occurred. Based on the aforementioned assumptions of HR and the duration of enrollment, the final analysis will happen at around 32 months after first patient in.

3. STUDY CONDUCT

3.1 RANDOMIZATION SPECIFICATIONS

Randomization will be performed by an interactive voice/Web-based system (IxRS). Patients will be assigned in 1:1 ratio to one of the two treatment arms through a block stratified randomization procedure. The randomization scheme will ensure approximately equal sample sizes in the two treatment groups in regard to a total of 15 strata by the following stratification factors:

- 17p deletion by local testing (yes or no)
- Risk status: high risk or low risk
- Geographic region (United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia)

A unique patient number will be assigned at randomization. This patient number will be used to identify the patient in the electronic data capture system and all other data sources.

3.2 INDEPENDENT REVIEW COMMITTEE

An IRC composed of board-certified radiologists and board-certified oncologists with experience in CLL will assess all patients for response and progression on the basis of imaging results, bone marrow biopsy results, and relevant clinical data. The IRC assessment will be blinded with respect to treatment arm and investigator assessment of response.

A charter for the IRC will further describe the responsibilities, methods of response evaluation, and determination of progression.

3.3 INDEPENDENT DATA MONITORING COMMITTEE

This trial includes an independent Data Monitoring Committee (iDMC) for periodic review of safety and efficacy data collected during the study. Reviews by the iDMC will be conducted according to a charter written and approved prior to study initiation. Members of the iDMC will be external to the Sponsors and the study team and will follow a charter that outlines their roles and responsibilities.

At the beginning of the study, intensive monitoring and analysis of all clinically significant safety events will be performed. The iDMC will assemble to review a safety analysis of significant safety events approximately 1 month after the first patient is enrolled depending on the rate of initial patient enrollment, then approximately every 2 months until 40 patients have completed 2 cycles of treatment (with approximately 20 patients in each arm). Thereafter, the iDMC will meet approximately every 6 months and subsequently at a frequency determined by the iDMC and the Sponsors according to the emerging safety profile. In addition, either the Sponsors or the iDMC can request ad hoc iDMC meetings at any time that potential safety concerns arise.

An interim analysis of efficacy data will be conducted and further reviewed by the iDMC when approximately 140 (75%) of the 186 investigator-assessed PFS events required for the primary efficacy analysis have occurred. Recommendations to release the study results early because of significant evidence of efficacy will be based on the specified interim analysis methodology.

An independent Data Coordinating Center (iDCC), which is independent of the Sponsors, will prepare all summaries and analyses for the iDMC review.

Further details about the definition, the role, as well as the responsibility of the iDMC are provided in a separate document, the iDMC Charter.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATION

4.1.1 Randomized Patients

By following the intent-to-treat principle, efficacy analyses will be performed using all randomized patients. Patients will be analyzed in the treatment arm that they were randomized to by the IxRS, regardless the actual treatment received.

Safety analyses will include all randomized patients who received at least one dose of study treatment (venetoclax, rituximab, or bendamustine), with patients grouped according to the treatment arm as treated. If patients in the control (BR) arm received venetoclax, they will be assigned to venetoclax-containing arm in the safety analysis population. If patients in the venetoclax-containing arm accidentally missed some doses of venetoclax, they will remain in the venetoclax-containing arm in the safety analysis population.

4.1.2 Pharmacokinetic Evaluable Population

PK analyses and the evaluable population will be defined in a separate analysis plan.

4.2 ANALYSIS OF STUDY CONDUCT

The number of patients who are randomized will be tabulated by treatment group, center, and country. Major eligibility violations, major protocol deviations, patient disposition, and reasons for study discontinuation will be summarized by treatment group for all randomized patients. Duration of follow-up will also be assessed.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic characteristics, including but not limited to age, sex, race, ethnicity, baseline weight, height, baseline Eastern Cooperative Oncology Group (ECOG) Performance Status, baseline disease characteristics (e.g., disease staging at diagnosis, time from first diagnosis, fludarabine-refractory status, prior number of oncology therapies), cytogenetic abnormalities (including 17p and P53 mutation), IgVH status, bulky disease (nodes $\geq$5 cm or $\geq$10 cm), absolute lymphocyte count at baseline ($\geq 25 \times 10^9$/L or $\geq 50 \times 10^9$/L), stratification factors (17p deletion, risk status, and geographic region), and TLS risk category (low, medium, or high) identified upon study entry based on both case report form (CRF) and programming will be summarized by treatment group for all randomized patients.

The number of non-missing observations, mean, median, standard deviations, minimum, and maximum values of the continuous variables will be summarized by treatment group. Percentages of patients within each subgroup of categorical variables will be provided. Number of patients with missing information will also be summarized.

4.4 MEDICAL HISTORY

Medical history data will be summarized and presented using body systems and conditions/diagnoses as captured on the CRF. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of patients with a particular condition/diagnosis will be summarized. Patients reporting more than one condition/diagnosis within a body system will be counted only once for that body system.

4.5 PREVIOUS TREATMENT AND CONCOMITANT MEDICATIONS

A prior medication is defined as any medication started and ended prior to the first dose of study drug. Concomitant therapy includes any medication (e.g., prescription drugs, over-the-counter drugs, herbal/homeopathic remedies, nutritional supplements) used by a patient from 28 days prior to the initiation of study treatment through the end of treatment. Concomitant medication will be summarized by treatment groups.

The number and percentage of patients who have taken medications will be summarized for prior medications, concomitant medications, and prior cancer therapies.

For summaries of concomitant medications, if an incomplete start date was collected for a medication, the medication will be assumed to be a concomitant medication only if the partial date information indicates that the drug was used by a patient from 28 days prior to the initiation of study treatment.

4.6 EFFICACY ANALYSIS

4.6.1 Primary Efficacy Endpoint

The primary efficacy endpoint of this study is PFS. [redacted] the primary efficacy endpoint is the IRC-assessed PFS, which is defined as the time from randomization to the first occurrence of progression or relapse as assessed by the IRC (determined using standard iwCLL guidelines [Hallek et al. 2008]) or death from any cause, whichever comes first.

The primary analysis of the study will test the equality of PFS distributions in the venetoclax and rituximab combination (venetoclax + R, Arm A) and in bendamustine and rituximab combination (BR, Arm B) arms, as follows:

H_0: $PFS_{venetoclax+R} = PFS_{BR}$ vs. H_1: $PFS_{venetoclax+R} \neq PFS_{BR}$

The treatment comparison will be performed using a two-sided stratified log-rank test (at the 0.05 significance level, appropriately adjusted for an interim analysis if it is conducted) stratified by 17p deletion status (yes or no), Risk status (high risk or low risk), and geographic region (United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia). If the null hypothesis is rejected and the observed

HR is favorable for the venetoclax+R combination, then it is shown that the venetoclax+R treatment arm has statistically significantly longer PFS than the BR arm. The investigator-assessed PFS will also be analyzed using the same models.

For cases in which a patient is misrandomized with respect to a stratification factor (i.e., there is a discrepancy between the IxRS-recorded stratification factor level and the eCRF-recorded stratification factor level), the IxRS data will be used in the primary analysis. Discordances between the IxRS and the eCRF data will be summarized.

Further estimates of the treatment effect will be expressed as HRs from a stratified Cox proportional-hazards analysis, including 95% confidence intervals. The stratification factors in the Cox model will include aforementioned stratification factors.

The Kaplan–Meier plot will be used to provide a visual description of the differences in duration of PFS across treatment arms. Median PFS for both treatment groups (if available) and the 95% confidence intervals will be estimated based on the method of Brookmeyer and Crowley (1982).

Sensitivity analyses for PFS will also be performed as described in Section 4.6.4.

4.6.2 Secondary Efficacy Endpoints

Secondary endpoints are as defined in Section 2.2.2. Formal statistical tests of duration of PFS between the two arms for the population of patients determined to have 17p deletion by the central laboratory, the ORR based on investigator assessment as well as the OS will be conducted.

To adjust for multiple testing of key secondary efficacy endpoints, thereby controlling the overall type I error rate at a two-sided significance level of 0.05, a fixed sequence testing procedure will be used (Westfall and Krishen 2001). For registration purposes in the United States, the following endpoints will be tested in the order given (see also Section 2.2.2 for secondary endpoints not included in the fixed sequence testing procedure):

- PFS in randomized patients based on IRC assessment
- PFS based on IRC assessment in the population of patients with 17p deletion mutation detected based on the central laboratory FISH test results
- ORR based on investigator assessment
- OS

Specifically, if the study meets its primary efficacy endpoint of prolonged PFS in the treatment arm (venetoclax+R) in all randomized patients, then a formal statistical test of IRC-assessed PFS between the two arms will be performed at the two-sided significance level of 0.05 for the patients with 17p deletion mutation detected based on

the central laboratory FISH test results. An unstratified log-rank test will be used as the main test for treatment effect comparison.

If the study does not meet its primary endpoint, then this test will not be performed.

A stratified log-rank test with early or late relapse or progression after prior chemotherapy-containing and geographic region will also be performed as a supportive analysis.

The ORR based on investigator assessment will only be tested once the test for PFS in the population of patients with 17p deletion mutation detected based on the central laboratory FISH test results has been rejected at a two-sided level significance of 0.05.

If the test for ORR has been rejected at a two-sided significance level of 0.05, the duration of OS will be further tested in the final analysis (when approximately 186 investigator-assessed PFS events have occurred) at a two-sided significance level of 0.05.

This fixed-sequence testing of the duration of PFS in 17p-deleted patients, ORR by investigator assessment or OS maintains the type I error level at 0.05 (Alosh et al. 2010). No further adjustment for multiplicity will be made for the other secondary endpoints analyses.

Investigator-assessed PFS will be analyzed in the subset of CLL patients with 17p deletion identified by central laboratory FISH testing by methods similar to those used for the IRC based PFS analyses for the subset of CLL patients with 17p deletion.

ORR, CR, CRi, nPR, and PR rate at the end of combination treatment response visit as assessed by the investigator or by the IRC will be compared between the two arms using the stratified Cochran–Mantel–Haenszel (CMH) tests. Stratification factors to be used are identical to those used for the primary PFS endpoint analyses. Various response rates and 95% confidence intervals will be reported for each treatment group.

Both bone marrow aspirate- and peripheral blood-based MRD response rates from the two treatment arms at the end of combination treatment response visit will be compared by using the Chi-square test. In cases where no post-baseline MRD assessment is available at this time point, patients will be considered as MRD-positive in this analysis.

Time-to-event endpoints, including OS, EFS, and TTNT, will be analyzed using the same statistical log-rank tests as described for the primary analysis of PFS.

Time-to-event analysis of DOR will incorporate data only from the subset of patients in both treatment arms that achieved a CR/CRi/nPR/PR. As there is no expectation for treatment arm balance, analyses of DOR will not incorporate stratification factors and will

not produce a p-value but only summarize the treatment arm estimates and confidence intervals.

4.6.3 Exploratory Efficacy Endpoints

Relationship between various baseline biomarkers and clinical outcome parameters in patients from both arms of the study may be assessed using appropriate laboratory measures.

The MRD response rates based on the peripheral blood/bone marrow aspirate samples at each disease response assessment visit will be summarized over time for both treatment arms. In cases where no post-baseline MRD assessment is available at a specific time point, patients will be considered as MRD-positive, except for those who have not been followed up long enough to have MRD data collected for that timepoint.

4.6.4 Sensitivity Analyses

To check robustness of the statistical models and underlying assumptions to be implemented in the primary endpoint analyses, the following two sensitivity analyses for PFS will be performed:

- An unstratified log-rank test for the primary PFS comparison between treatment arms will be conducted.
- The impact of patients' initiation of non-protocol–specified anti-CLL therapy without meeting the criteria of disease progression on PFS will be assessed by censoring these patients at the start date of the non-protocol–specified anti-CLL treatment. Stopping only one component of the randomized study treatment will not be considered a reason for censoring patients.

4.6.5 Subgroup Analyses

Subgroup analyses of investigator-assessed PFS, IRC-assessed PFS, and ORR will be performed according to prognostic factors to assess internal consistency. The odds ratio of response and their 95% confidence intervals, hazard ratio of time-to-event endpoint and their 95% confidence intervals, as well as the sample sizes will be reported separately for each level of the following subgroups in forest plot:

- Baseline characteristics (including, but not limited to age at randomization, sex, race, 17p deletion status or P53 mutation status)
- Stratification factors (17p deletion by local testing, risk status, geographic region)

4.7 SAFETY ANALYSES

Safety of the study will be assessed through adverse events (AE) reporting, including serious adverse events (SAE) and non-serious adverse events of special interest, safety laboratory assessments, measurement of vital signs, and other protocol-specified tests that are deemed critical to the safety evaluation.

The safety analyses will be performed based on the Treated Patients.

4.7.1 Exposure to Study Medication

Information on study drug administration such as the duration of treatment, dose intensity (percentage of planned dosage received) and number of dose/cycles will be summarized by treatment arm and by study medication by using the following statistics: sample size, mean, standard deviation, median, minimum, and maximum.

Withdrawals of patients from study treatment will be reported as listings and summary tables by treatment arms as well.

4.7.2 Adverse Events

After informed consent has been obtained but prior to initiation of study drug, only SAEs caused by a protocol-mandated intervention should be reported. After initiation of study drug, all AEs will be reported until 28 days after the last dose of study drug, or 90 days after last dose of rituximab, whichever is longer. Verbatim descriptions of AEs will be mapped to the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms.

AEs and SAEs will be summarized by treatment arm, by body system and by the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI–CTCAE) version 4.0 grading system. For all AEs, the most extreme intensity will be used for reporting. Individual listings and summary tables will be presented by body system and intensity. In tables showing the overall incidence of AEs, patients who experienced the same event on more than one occasion will be counted only once in the calculation of the event frequency.

AEs of special interest (Grade $\geq$3 TLS and IRR) will be summarized by treatment arm and listings will be provided.

AEs leading to dose interruptions, early treatment discontinuation and early study withdrawal will be summarized by treatment arm.

Summary tables of AEs by age, gender, and race will also be provided.

Deaths reported during the study treatment period and those reported after treatment completion/discontinuation will be summarized by treatment arm and by AE versus non-AE.

4.7.3 Laboratory Data

Selected laboratory measurements and change from baseline will be summarized over time by treatment arm, where baseline measurement is defined as the last valid measurement before first dose of any study medication. For example, TLS-related lab measurements (including uric acid, potassium, inorganic phosphorus, and calcium) and changes from baseline will be summarized by treatment arm over time. In addition, laboratory measurements from individual patients who developed an AE of TLS during

the study will be presented by listings. Important hematology parameters, such as absolute lymphocyte count, absolute neutrophil count, Hg, Hct, and WBC will be categorized according to NCI CTCAE grading version 4.0. Shift tables will be generated to cross-tabulate the number of patients grouped into each grade at baseline versus post-baseline observations.

4.7.4 Vital Signs

Vital signs will be summarized over time for absolute values and changes from baseline without imputation for missing data. Descriptive statistics will be provided by treatment arm.

ECG data at the screening visit will be summarized by treatment arm for number of patients that fall into each category of status as collected in the CRF.

4.8 MISSING DATA

For PFS, patients who do not have documented disease progression or death will have observations censored on the date of the last tumor assessment or, if no response assessments were performed after the baseline visit, at the time of randomization plus one day.

For OS, patients for whom death has not been documented will have observations censored on the last date at which they were known to be alive.

For response endpoints, patients with no response assessments (for any reason) will be considered non-responders.

For DOR, patients with no documented progression after CR, CRi, nPR, or PR will be censored at the last date at which they are known to have had the CR, CRi, nPR, or PR. Patients who have never responded will not be included in this analysis.

4.9 PHARMACODYNAMIC ANALYSES

The pharmacodynamic endpoint in this study is serial assessment of B-cell, T-cell, and NK-cell lymphocyte subsets by flow cytometry. Proportions of each lymphocyte subset and change from baseline in the proportions will be summarized by treatment arm over time as warranted by the data.

4.10 PHARMACOKINETIC ANALYSES

Venetoclax concentrations will be summarized by nominal visit and timepoint of collection in the Clinical Study Report (CSR). Population PK methods will be used to characterize the PK of venetoclax in this study in conjunction with appropriate historical data to calculate apparent clearance, apparent volume of distribution, and other appropriate PK parameters of venetoclax. Potential correlations of exposure with dose, demographics, pharmacodynamic variables, safety, and efficacy outcomes may be

explored as warranted by the data. PK analyses will be defined in a separate PK report. The results from the population PK analysis may be reported separately from the CSR.

4.11 INTERIM ANALYSES

One interim analysis is planned during the conduct of the study to assess the efficacy of venetoclax + R combination treatment compared with BR treatment, and to allow for release of the results earlier than the planned final analysis in case of significant differences.

The interim analysis will be performed when approximately 140 investigator-assessed PFS events have occurred in both treatment arms combined (75% of the 186 events required for the final primary efficacy analysis is available). The stopping boundary follows a unified family with parameter $P = 2$ (Kittelson et al. 1999). Based on 140 events, the duration of PFS will be tested at the interim analysis, which corresponds to approximately a 2-sided p-value of $2 \times 0.0013 = 0.0026$. If the number of events is not exactly 140, then the boundary will be updated to reflect the number of events. Table 1 summarizes the probability of passing boundary at the planned interim analysis given a hazard ratio.

Table 1 Probability of Passing Boundary at the Interim Analysis for a Given Hazard Ratio

True Hazard Ratio	0.66	0.42	0.35
Approximate median PFS in RV arm [a] (months)	23.0	36.2	43.4
Probability of passing boundary at interim analysis	29%	98%	99%

PFS = progression-free survival; RV venetoclax + rituximab.

[a] Assumes a median duration of PFS of 15.2 months for the BR arm.

Both investigator-assessed PFS analysis and a corresponding analysis on the basis of IRC-assessed PFS events will be conducted in this interim analysis. The same stratification factors as specified in the primary efficacy endpoint analysis will be used. If the p-value of the stratified two-sided log-rank test is less than or equal to 0.0026 (approximately corresponding to a HR of 0.60) for both investigator-assessed PFS analysis and IRC-assessed PFS analysis and the observed HR is favorable for the venetoclax + R combination treatment, the trial will have met its primary efficacy endpoint. An iDMC will evaluate interim analysis results and provide a recommendation as to whether to release the trial results early. All summaries and analyses according to treatment arm will be prepared by an iDCC for the iDMC review. [redacted]

The final primary efficacy analysis will be performed when approximately 186 investigator-assessed PFS events have been observed. The statistical test level will be adjusted to incorporate the alpha spent at the interim analysis (if it was conducted) so that the overall type I error rate will be maintained at the two-sided 0.05 level.

4.12 REFERENCES

Alosh M, Huque M. A consistency-adjusted alpha-adaptive strategy for sequential testing. Stat Med 2010;29:1559–71.

Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29-41.

Fischer K, Cramer P, Busch R, et al. Bendamustine combined with rituximab in patients with relapsed and/or refractory chronic lymphocytic leukemia: a multicenter Phase II trial of the German Chronic Lymphocytic Leukemia Study Group. J Clin Oncol 2011;29:3559–66.

Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood 2008;111:5446–56.

Kittelson J, and Emerson S. A Unifying Family of Group Sequential Test Designs. Biometrics 1999;55:874-82

Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. J Stat Plan Inference 2001;99:25–40.

STATISTICAL ANALYSIS PLAN

TITLE: **A MULTICENTER, PHASE III, OPEN-LABEL, RANDOMIZED STUDY IN RELAPSED/REFRACTORY PATIENTS WITH CHRONIC LYMPHOCYTIC LEUKEMIA TO EVALUATE THE BENEFIT OF VENETOCLAX (GDC-0199/ABT-199) PLUS RITUXIMAB COMPARED WITH BENDAMUSTINE PLUS RITUXIMAB**

PROTOCOL NUMBER:	GO28667
STUDY DRUG:	Venetoclax (GDC-0199/ABT-199, RO5537382)
VERSION NUMBER:	4
IND NUMBER:	110159
EUDRACT NUMBER:	2013-002110-12
SPONSOR:	F. Hoffmann-La Roche Ltd and AbbVie Inc. will act as co-sponsors of this trial globally.
PLAN PREPARED BY:	[redacted] Ph.D.
DATE FINAL:	Version 1: 27 July 2015
DATES AMENDED	Version 2: 20 June 2016 Version 3: 21 December 2016 Version 4: See electronic date stamp below.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Name	Reason for Signing	Date and Time (UTC)
[redacted]	Company Signatory (Clinical)	18-May-2017 17:23:32

STATISTICAL ANALYSIS PLAN, VERSION 4:

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
BR	bendamustine in combination with rituximab
CLL	chronic lymphocytic leukemia
CMH	Cochran–Mantel–Haenszel
CR	complete response
CRi	complete response with incomplete marrow recovery
DOR	duration of response
eCRF	electronic Case Report Form
ECOG	Eastern Cooperative Oncology Group
EFS	event-free survival
FISH	fluorescence in situ hybridization
HR	hazard ratio
HRQOL	health-related quality of life
IA	interim analysis
iDCC	independent Data Coordinating Center
iDMC	independent Data Monitoring Committee
IRC	Independent Review Committee
IRR	infusion-related reaction
iwCLL	International Workshop on CLL
IxRS	interactive voice/Web-based system
MDASI	MD Anderson Symptom Inventory
MRD	Minimal residual disease
NCI CTCAE	National Cancer Institute Common Terminology Criteria for adverse Events
NK	natural killer
nPR	nodular partial response
ORR	objective response rate
OS	overall survival
PFS	progression-free survival
PK	pharmacokinetic
PR	partial response
QLQ-C30	Quality of Life Questionnaire Core 30
QLQ-CLL16	Quality of Life Questionnaire CLL module
TLS	tumor lysis syndrome
TTNT	time to next anti-CLL treatment
venetoclax + R	venetoclax (GDC-0199/ABT-199) in combination with rituximab

1. BACKGROUND

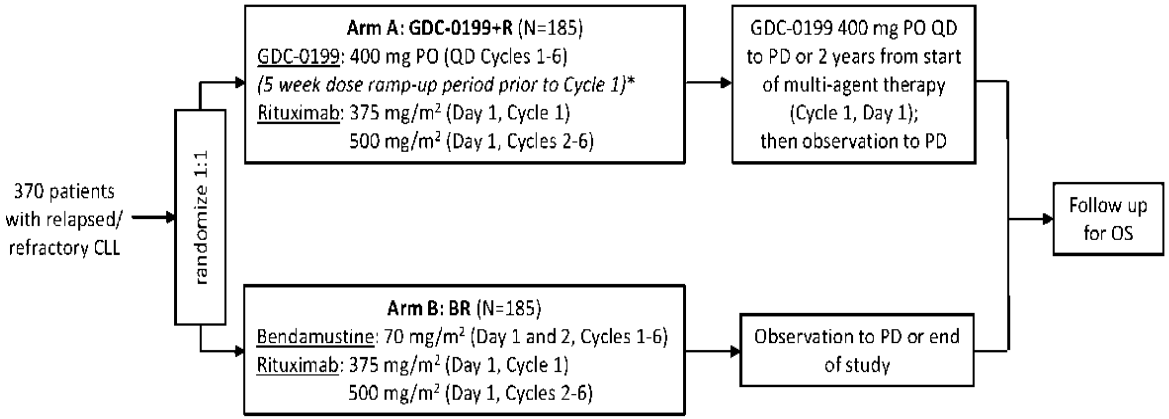
This document is based on the Statistical Considerations and Analysis Plan section of the study GO28667 (MURANO) protocol and will provide more details on the planned statistical analyses. For purposes of registration, the analyses outlined in this Statistical Analysis Plan will supersede those specified in the protocol.

2. STUDY DESIGN

This is an open-label, international, multicenter, randomized, Phase III study to investigate the efficacy and safety of venetoclax (GDC-0199/ABT-199) in combination with rituximab (venetoclax + R) compared with bendamustine in combination with rituximab (BR) in patients with relapsed or refractory chronic lymphocytic leukemia (CLL) (see Figure 1).

The primary objective of the study is to evaluate the efficacy of venetoclax + R compared with BR in patients with relapsed or refractory CLL, as measured by investigator-assessed progression-free survival (PFS). In the United States, Independent Review Committee (IRC)-assessed PFS will be analyzed to support the primary efficacy analysis and will be the basis for regulatory decisions.

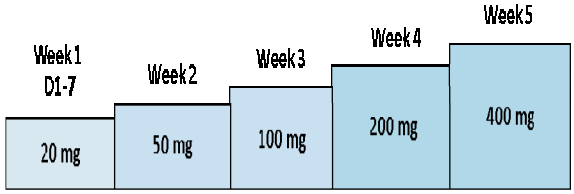
Figure 1 Design of the Study



1 cycle = 28 days; Arm A = venetoclax (GDC-0199) and rituximab (venetoclax = R); Arm B = bendamustine and rituximab (BR); CLL = chronic lymphocytic leukemia; OS = overall survival; PD = progressive disease; PO = per os; QD = once daily.

*Patients will receive venetoclax starting on Day 1 (venetoclax dose ramp-up period) as delineated in Figure 2. Venetoclax will then be self-administered at 400 mg per day for a maximum of 2 years from Cycle 1 Day 1 or until disease progression (whichever is earlier). Combination therapy consisting of 6 cycles of rituximab and daily venetoclax dosing will start after completion of the venetoclax ramp-up period.

Figure 2 Venetoclax Dosing Scheme during the Ramp-Up Period



D = day.

Note: For all patients enrolled after Protocol Version 4. Prior to Version 4, ramp up was 4 or 5 weeks

Approximately 370 patients will be recruited and randomly assigned in 1:1 ratio to receive either venetoclax+R (Arm A) or BR (Arm B). Randomization will be stratified according to the following 3 factors:

- 17p deletion: yes or no
- Risk status: high risk or low risk

  High risk: defined as harboring 17p deletion, no response to front-line chemotherapy-containing regimen, relapsed within 12 months after chemotherapy, or relapsed within 24 months after chemoimmunotherapy

  Low risk: defined as relapse more than 12 months after chemotherapy or 24 months after chemoimmunotherapy.
- Geographic region: United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia

Patients randomized to Arm A (venetoclax + R) will have a 5-week venetoclax dose ramp-up period to reach the target dose of 400 mg daily, followed by 6 cycles of rituximab consisting of a single infusion on the first day of each 28-day cycle. Patients will continue to take their daily dose of venetoclax during the rituximab cycles. Patients whose disease has not progressed following the completion of the 6 cycles will continue to receive venetoclax until disease progression or for a maximum of 2 years from Cycle 1 Day 1. Patients randomized to Arm B (BR) will receive 6 cycles of BR consisting of a single infusion of rituximab on Day 1 and bendamustine infusions on Days 1 and 2 of each 28-day cycle.

After 6 cycles of combination therapy, patients in both arms will be followed clinically every 3 months through Year 3 from Cycle 1, Day 1, after which they will be followed every 6 months for an additional 2 years, until withdrawal of consent, or until end of study, whichever comes first.

Patients who discontinue all components of study treatment for other reasons, or receive a new anti-CLL therapy at any time during follow-up in the absence of disease progression will also be followed for progression and survival according to the schedule defined in the protocol.

2.1 PROTOCOL SYNOPSIS

The protocol synopsis of Version 7 is provided in Appendix 1. For additional details, see the Schedule of Assessments in Appendix 2.

2.2 OUTCOME MEASURES

All patients will have baseline tumor assessment and will be assessed for response to treatment at the scheduled response assessment follow-up visits using standard clinical and laboratory examinations and/or computed tomography scans according to the International Workshop on CLL (iwCLL) guidelines (Hallek et al. 2008).

The assessment of response and progression by the investigator will be considered the primary analysis for all of the endpoints described in the study. Response and progression will also be assessed by an IRC.

2.2.1 Primary Efficacy Outcome Measures

The primary efficacy outcome measure for this study is investigator-assessed PFS. [redacted]

Duration of PFS is defined as the time from randomization to the first occurrence of progression or relapse, determined using standard iwCLL guidelines, or death from any cause, whichever comes first. Data from patients without disease progression, relapse, or death will be censored at the time of the last adequate response assessment. If no response assessments were performed after the baseline visit, PFS will be censored at the date of randomization. Patients who have initiated new anti-CLL therapy prior to the documented disease progression will not be censored at the date of initiation of the new anti-CLL therapy.

An adequate response assessment visit must include the following assessments:

1. Physical examination (nodal and extranodal disease as applicable, spleen, and liver)
2. Hematology labs (lymphocytes, platelets, and hemoglobin).

[redacted]

2.2.2 Secondary Efficacy Outcome Measures

The secondary efficacy outcome measures for this study are as follows:

- IRC-assessed PFS in the subset of CLL patients with 17p deletion identified by fluorescence in situ hybridization (FISH) testing at a central laboratory

- Investigator-assessed PFS in the subset of CLL patients with 17p deletion identified by FISH testing at a central laboratory
- Investigator-assessed best overall response rate (ORR), which is defined as the percentage of patients with the best response of complete response (CR), complete response with incomplete marrow recovery (CRi), nodular partial response (nPR), or partial response (PR), at any time point during the study as assessed by the investigator. In cases where no post-baseline response assessment is available, patients will be considered non-responders.
- IRC-assessed ORR, which is defined as the percentage of patients with the best response of CR, CRi, nPR, or PR at any time point during the study as assessed by the IRC. In cases where no post-baseline response assessment is available, patients will be considered non-responders.
- IRC-assessed CR rate, which is defined as the percentage of patients with the best response of CR or CRi at any time point during the study as assessed by the IRC.
- ORR, CR, CRi, nPR, and PR rates at the end of combination treatment response visit, as assessed by the investigator. In cases where no post-baseline response assessment is available, patients will be considered non-responders.
- ORR, CR, CRi, nPR, and PR rates at end of combination treatment response visit, as assessed by the IRC. In cases where no post-baseline tumor assessment is available, patients will be considered non-responders.
- Overall survival (OS), which is defined as the time from randomization to death from any cause. Patients who were not reported as having died at the time of the analysis will be censored at the date when they were last known to be alive.
- Investigator-assessed event-free survival (EFS), which is defined as the time between date of randomization and the date of disease progression/relapse, death, or start of a new non-protocol-specified anti-CLL therapy as assessed by the investigator. If the specified event (disease progression/relapse, death, start of a new anti-CLL treatment) does not occur, patients will be censored at the date of last adequate response assessment. In cases where no post-baseline response assessment is available, patients without an event will be censored at the randomization date.
- Duration of response (DOR), which is defined for patients with a best overall response of CR, CRi, nPR, or PR as the time from first occurrence of a documented CR/CRi/nPR/PR to disease progression/relapse, as assessed by the investigator, or death from any cause. Patients with no documented progression or death after CR, CRi, nPR, or PR will be censored at the last date of adequate response assessment. Patients who have never responded will not be included in this analysis.
- Time to next anti-CLL treatment (TTNT), which is defined as the time from randomization to start of new non-protocol–specified anti-CLL therapy or death from any cause. Patients who were reported as not having started new non-protocol anti-CLL therapy or death will be censored at the last visit date for this outcome analysis.

- Minimal residual disease (MRD) response rate at the end of combination treatment response visit as measured at a central laboratory on peripheral blood samples.

2.2.3 Exploratory Efficacy Outcome Measures

The exploratory outcome measures for this study are as follows:

- Assessment of potential biomarkers that are prognostic and/or predictive of response and resistance to treatment with venetoclax + R or BR.
- Evaluation of the relationship between clinical response and PFS and various potential biomarkers for patients treated with venetoclax + R or BR.
- MRD response rate as measured at a central laboratory based on peripheral blood samples over time.
- The best MRD response rate as measured by a central laboratory based on peripheral blood samples collected during the study.
- MRD response rate as measured by a central laboratory based on bone marrow aspirate samples in responders (CR/CRi, nPR/PR).

2.2.4 Pharmacokinetic Efficacy Outcome Measures

The pharmacokinetic (PK) outcome measures for this study are as follows:

- Plasma venetoclax concentrations at the specified timepoints
- Apparent clearance, apparent volume of distribution, and other appropriate PK parameters of venetoclax characterized using population PK techniques, as data allow

2.2.5 Pharmacodynamic Efficacy Outcome Measure

The pharmacodynamic outcome measure for this study is as follows:

- Serial assessment of B-cell, T-cell, and natural killer (NK)-cell lymphocyte subsets by flow cytometry

2.2.6 Safety Outcome Measures

The safety outcome measures for this study are as follows:

- Incidence, nature, and severity of adverse events and serious adverse events, including deaths
- Mean of changes in vital signs, physical findings, and clinical laboratory results (including hematology and chemistry) during and following administration of study treatment
- Incidence of adverse events of special interest:

  Grade > 3 tumor lysis syndrome (TLS) and infusion-related reactions (IRRs)
- Measures of immune function, including serial immunoglobulin levels (IgG, IgM, and IgA) following treatment with venetoclax+R or BR

2.2.7 Patient-Reported Outcome Measures

The health-related quality of life (HRQOL) and disease- and treatment-related symptom measures for this study are as follows:

- MD Anderson Symptom Inventory (MDASI) (see Appendix 5 of the protocol)
- European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (QLQ-C30) and Quality of Life Questionnaire CLL module (QLQ-CLL16) (see Appendix 3 and Appendix 4, respectively, of the protocol)

2.2.8 Health Economic Outcome Measure

The health economic outcome measure for this study is as follows:

- The EuroQol five-dimension five-level questionnaire (see Appendix 6 of the protocol)

2.3 DETERMINATION OF SAMPLE SIZE

The primary endpoint of PFS was used to determine the sample size for the study based on the following assumptions:

- Two-sided log-rank test at the 0.05 level of significance
- 80% power to detect a hazard ratio (HR) for venetoclax + R versus BR of 0.66, corresponding to an approximate median improvement of 15.2 months (Fischer et al. 2011) to 23 months (34% reduction in risk of a PFS event)
- Exponential distribution of PFS
- An annual dropout rate of 5%
- One interim analysis (IA) for efficacy at approximately 75% of total investigator-assessed PFS events (approximately 140 investigator-assessed PFS events).

With these assumptions, 186 investigator-assessed PFS events are required to achieve 80% power for the primary analysis of PFS in all patients. It is planned to enroll 370 patients across the two arms with 1:1 randomization ratio.

Sample size calculations were performed with Insightful S + Seq Trial S 2.0.6.

2.4 ANALYSIS TIMING

An efficacy IA is planned when approximately 140 investigator-assessed PFS events (75% of the 186 events required for the final primary efficacy analysis) have occurred for both treatment arms combined (see Section 4.12). The stopping boundary follows a unified family with parameter $P = 2$ (Kittelson and Emerson 1999). Based on 140 events, this corresponds to approximately a 2-sided p-value of $2 \times 0.0013 = 0.0026$. The exact crossing boundary for the IA will be determined by the actual number of events observed.

If the p-value of the two-sided log-rank test is less than or equal to the crossing boundary, for example, 0.0026, the trial will have met its primary efficacy endpoint (corresponding to a HR of approximately 0.60 or better). Based on the assumption of $HR = 0.66$ (venetoclax + R arm vs. BR arm), and the duration of enrollment of 20 months, the IA will take place approximately 26 months after first patient is enrolled. If the study

results are not released in the IA, the final analysis for the study will be conducted when approximately 186 PFS events based on the investigator assessment have occurred. Based on the aforementioned assumptions of HR and the duration of enrollment, the final analysis will take place approximately 32 months after the first patient is enrolled.

3. STUDY CONDUCT

3.1 RANDOMIZATION SPECIFICATIONS

Randomization will be performed by an interactive voice/Web-based system (IxRS). Patients will be assigned in 1:1 ratio to one of the two treatment arms through a block stratified randomization procedure. The randomization scheme will ensure approximately equal sample sizes in the two treatment groups in regard to a total of 15 strata by the following stratification factors:

- 17p deletion by local testing (yes or no)
- Risk status: high risk or low risk
- Geographic region (United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia)

A unique patient number will be assigned at randomization. This patient number will be used to identify the patient in the electronic data capture system and all other data sources.

3.2 INDEPENDENT REVIEW COMMITTEE

An IRC composed of board-certified radiologists and board-certified oncologists with experience in CLL will assess all patients for response and progression on the basis of imaging results, bone marrow biopsy results, and relevant clinical data. The IRC assessment will be blinded with respect to treatment arm and investigator assessment of response.

The responsibilities, methods of response evaluation, and determination of disease response/progression can be found in the IRC Charter.

3.3 INDEPENDENT DATA MONITORING COMMITTEE

This trial includes an independent Data Monitoring Committee (iDMC) for periodic review of safety and efficacy data collected during the study. Reviews by the iDMC will be conducted according to a charter written and approved prior to study initiation. Members of the iDMC will be external to the Sponsors and the study team and will follow a charter that outlines their roles and responsibilities.

At the beginning of the study, intensive monitoring and analysis of all clinically significant safety events will be performed. The iDMC will assemble to review a safety analysis of significant safety events approximately 1 month after the first patient is enrolled depending on the rate of initial patient enrollment, then approximately every 2 months until 40 patients have completed 2 cycles of treatment (with approximately 20 patients in

each arm). Thereafter, the iDMC will meet approximately every 6 months and subsequently at a frequency determined by the iDMC and the Sponsors according to the emerging safety profile. In addition, either the Sponsors or the iDMC can request ad hoc iDMC meetings at any time that potential safety concerns arise.

An IA of efficacy data will be conducted and further reviewed by the iDMC when approximately 140 (75%) of the 186 investigator-assessed PFS events required for the primary efficacy analysis have occurred. Recommendations to release the study results early because of significant evidence of efficacy will be based on the specified IA methodology.

An independent Data Coordinating Center (iDCC), which is independent of the Sponsors, will prepare all summaries and analyses for the iDMC review.

Further details about the definition, the role, as well as the responsibility of the iDMC are provided in a separate document, the iDMC Charter.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATION

4.1.1 Randomized Patients

By following the intent-to-treat principle, efficacy analyses will be performed using all randomized patients. Patients will be analyzed in the treatment arm that they were randomized to by the IxRS, regardless of the actual treatment received.

4.1.2 Treated Patients

Safety analyses will include all randomized patients who received at least one dose of study treatment (venetoclax, rituximab, or bendamustine), with patients grouped according to the treatment arm as treated. If patients in the control (BR) arm received venetoclax, they will be assigned to venetoclax-containing arm in the safety analysis population. If patients in the venetoclax-containing arm accidentally missed some doses of venetoclax, they will remain in the venetoclax-containing arm in the safety analysis population.

4.1.3 Pharmacokinetic Evaluable Population

Pharmacokinetic analyses and the evaluable population will be defined separately.

4.2 ANALYSIS OF STUDY CONDUCT

The number of patients who are randomized will be tabulated by treatment group, center, and country. Major eligibility violations, major protocol deviations, patient disposition, and reasons for study discontinuation will be summarized by treatment group for all randomized patients. Duration of follow-up will also be assessed.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic characteristics, including but not limited to age, sex, race, ethnicity, baseline weight, height, baseline Eastern Cooperative Oncology Group (ECOG) Performance Status, baseline disease characteristics (e.g., disease stage at diagnosis, time from first diagnosis, fludarabine-refractory status, prior number of oncology therapies), cytogenetic abnormalities (including 17p and P53 mutation), IgVH status, bulky disease (nodes ≥ 5 cm or ≥ 10 cm), absolute lymphocyte count at baseline ($\geq 25 \times 10^9$/L or $\geq 100 \times 10^9$/L), stratification factors (17p deletion, risk status, and geographic region), and TLS risk category (low, medium, or high) identified upon study entry based on both the electronic Case Report Form (eCRF) and programming based on raw data will be summarized by treatment group for all randomized patients.

The number of non-missing observations, mean, median, standard deviations, minimum, and maximum values of the continuous variables will be summarized by treatment group. Percentages of patients within each subgroup of categorical variables will be provided. Number of patients with missing information will also be summarized.

4.4 MEDICAL HISTORY

Medical history data will be summarized and presented using body systems and conditions/diagnoses as captured on the eCRF. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of patients with a particular condition/diagnosis will be summarized. Patients reporting more than one condition/diagnosis within a body system will be counted only once for that body system.

4.5 PREVIOUS TREATMENT AND CONCOMITANT MEDICATIONS

A prior medication is defined as any medication started and ended prior to the first dose of study drug. Concomitant therapy includes any medication (e.g., prescription drugs, over-the-counter drugs, herbal/homeopathic remedies, or nutritional supplements) used by a patient from 28 days prior to the initiation of study treatment through the end of treatment. Concomitant medication will be summarized by treatment groups.

The number and percentage of patients who have taken medications will be summarized for prior medications, concomitant medications, and prior cancer therapies.

For summaries of concomitant medications, if an incomplete start date was collected for a medication, the medication will be assumed to be a concomitant medication only if the partial date information indicates that the drug was used by a patient from 28 days prior to the initiation of study treatment.

4.6 EFFICACY ANALYSIS

4.6.1 Primary Efficacy Endpoint

The primary efficacy endpoint of this study is PFS as assessed by investigator. [redacted] the primary efficacy endpoint is the IRC-assessed PFS, which is defined as the time from randomization to the first occurrence of progression or relapse as assessed by the IRC (determined using standard iwCLL guidelines [Hallek et al. 2008]) or death from any cause, whichever comes first.

The primary analysis of the study will test the equality of PFS distributions in the venetoclax and rituximab combination (venetoclax + R, Arm A) and in bendamustine and rituximab combination (BR, Arm B) arms, as follows:

H_0: $PFS_{venetoclax+R} = PFS_{BR}$ vs. H_1: $PFS_{venetoclax+R} \neq PFS_{BR}$

The treatment comparison will be performed using a two-sided stratified log-rank test (at the 0.05 significance level, appropriately adjusted for an IA (if it is conducted) stratified by 17p deletion status (yes or no), risk status (high risk or low risk), and geographic region (United States/Canada, Australia/New Zealand, Western Europe, Central and Eastern Europe, or Asia). If the null hypothesis is rejected and the observed HR is favorable for the venetoclax + R combination, then it has shown that the venetoclax + R treatment arm has statistically significantly longer PFS than the BR arm. The investigator-assessed PFS will also be analyzed using the same models. Agreement/disagreement between investigator assessment and assessment by the IRC will be summarized.

For cases in which a patient is misrandomized with respect to a stratification factor (i.e., there is a discrepancy between the IxRS-recorded stratification factor level and the eCRF-recorded stratification factor level), the IxRS data will be used in the primary analysis. Discordances between the IxRS and the eCRF data will be summarized.

Further estimates of the treatment effect will be expressed as HRs from a stratified Cox proportional-hazards analysis, including 95% confidence intervals. The stratification factors in the Cox model will include aforementioned stratification factors.

The Kaplan–Meier plot will be used to provide a visual description of the differences in duration of PFS across treatment arms. Median PFS for both treatment groups (if available) and the 95% confidence intervals will be estimated based on the method of Brookmeyer and Crowley (1982).

Sensitivity analyses for PFS will also be performed as described in Section 4.6.4.

4.6.2 Secondary Efficacy Endpoints

Secondary endpoints are as defined in Section 2.2.2.

To adjust for multiple testing of key secondary efficacy endpoints, a fixed sequence testing procedure will be used (Westfall and Krishen 2001). The following endpoints will be tested in the order listed below (see also Section 2.2.2 for secondary endpoints not included in the fixed sequence testing procedure):

- IRC-assessed CR rate
- IRC-assessed best ORR
- OS

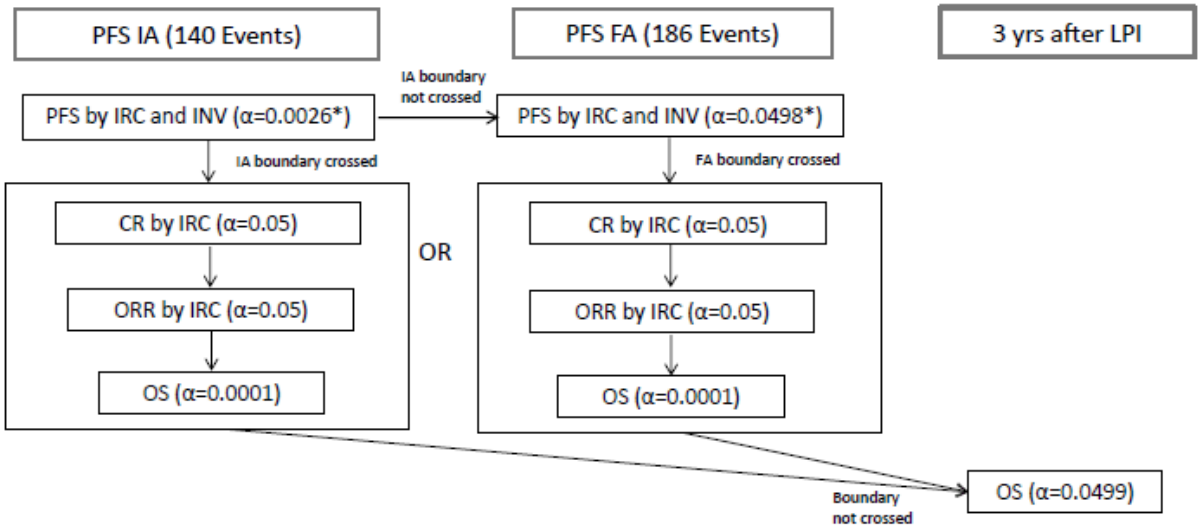
If the study meets its primary efficacy endpoint of prolonged PFS in the treatment arm (venetoclax+R) in all randomized patients, then a formal statistical test of IRC-assessed CR rate between the two arms will be performed at the two-sided significance level of 0.05 by using a stratified Cochran–Mantel–Haenszel (CMH) test with stratification factors in it. If the study does not meet its primary endpoint, then this test will not be performed.

The IRC-assessed best ORR will only be tested by the same stratified CMH test once the null hypothesis for IRC-assessed CR rate in all randomized patients has been rejected at a two-sided significance level of 0.05.

If the null hypothesis for IRC-assessed best ORR has been rejected at a two-sided significance level of 0.05, the duration of OS will be analyzed at a nominal alpha spend of 0.0001. Approximately 3 years after the last patient is enrolled, OS will be tested at a two-sided significance level of 0.0499.

Refer to Figure 3 for a schematic of the hierarchical testing of primary and key secondary endpoints. Of note, CR and ORR endpoints will only be tested once either at interim or final analysis depending on the outcome of the PFS analysis.

Figure 3 Diagram of hierarchical testing of primary and key secondary endpoints



IA = interim analysis, FA = final analysis, INV = investigator assessment, IRC = independent review committee assessment, LPI = last patient in, PFS = progression-free survival, CR = complete response, ORR = overall response rate, CR = complete response, OS = overall survival

*Assumes 140 events at interim and 186 events at final. Exact boundaries for IRC and INV may differ based on actual number of events observed respectively at each analysis.

Investigator-assessed best ORR, as well as CR, CRi, nPR, and PR rate at the end of combination treatment response visit as assessed by the investigator or by the IRC will be compared between the two arms using the stratified CMH tests. Stratification factors to be used are identical to those used for the primary PFS endpoint analyses. Various response rates and 95% confidence intervals will be reported for each treatment group.

Investigator-assessed and IRC-assessed PFS will be analyzed in the subset of CLL patients with 17p deletion identified by central laboratory FISH testing. An unstratified log-rank test will be used as the main test for treatment effect comparison. A stratified log-rank test with geographic region will also be performed as a supportive analysis.

Peripheral blood-based MRD response rates from the two treatment arms at the end of combination treatment response visit will be compared by using the Chi-square test. In cases where no post-baseline MRD assessment is available at this time point, patients will be considered as MRD-positive in this analysis, except for those who have not been followed up long enough to have MRD data collected for this timepoint.

Time-to-event endpoints, including OS, EFS, and TTNT, will be analyzed using the same log-rank tests as described in the primary analysis of PFS.

Time-to-event analysis of DOR will incorporate data only from the subset of patients in both treatment arms that achieved a CR/CRi/nPR/PR. As there is no expectation for treatment arm balance, analyses of DOR will not incorporate stratification factors and will not produce a p-value but only summarize the treatment arm estimates and confidence intervals.

4.6.3 Exploratory Efficacy Endpoints

Relationship between various baseline biomarkers and clinical outcome parameters in patients from both arms of the study may be assessed.

The MRD response rates in peripheral blood samples and bone marrow aspirate samples will each be assessed separately. The MRD response rates at each response assessment visit will be summarized over time for both treatment arms. The MRD response rates based on the bone marrow aspirate and based on peripheral blood samples in responders (CR, CRi, nPR, and PR) will be summarized. The best MRD response rate in two treatment arms based on the peripheral blood samples collected in the study will be analyzed. In cases where no post-baseline MRD assessment is available at a specific time point, patients will be considered as MRD-positive, except for those who have not been followed up long enough to have MRD data collected for that timepoint.

4.6.4 Sensitivity Analyses

To check robustness of the statistical models and underlying assumptions to be implemented in the primary endpoint analyses, the following sensitivity analyses for PFS (both investigator-assessed and IRC-assessed) will be performed:

- An unstratified log-rank test for the primary PFS comparison between treatment arms will be conducted.
- The impact of patients' initiation of non-protocol–specified anti-CLL therapy without meeting the criteria of disease progression on PFS will be assessed by censoring these patients at the start date of the non-protocol–specified anti-CLL treatment. Stopping only one component of the randomized study treatment will not be considered a reason for censoring patients.
- To assess the impact of missing assessments on PFS, an analysis on PFS will be performed by censoring those patients who progressed or died after missing more than one visit consecutively at their last adequate response assessment date before the missed visits.

4.6.5 Subgroup Analyses

Subgroup analyses of investigator-assessed PFS, IRC-assessed PFS, and ORR as assessed by the investigator or by IRC will be performed to assess internal consistency. The odds ratio of response and their 95% confidence intervals, HR of time-to-event endpoint and their 95% confidence intervals, as well as the sample sizes will be reported separately for each level of the following subgroups in forest plot:

- Baseline characteristics (including, but not limited to age at randomization, sex, race, IgVH mutational status, relapsed versus refractory disease to most recent prior therapy; response duration [$<$12 months vs greater] to most recent prior therapy, 17p deletion detected by central laboratory test, 11q deletion detected by central laboratory test, P53 mutation status, *TP53* mutation and/or 17p deletion detected by central laboratory test, number of prior CLL therapy [1, 2, 3 or greater], Fludarabine refractory, baseline nodal size, baseline ALC counts, renal impairment, hepatic impairment, etc.)
- Stratification factors (17p deletion by local test, risk status, and geographic region)

4.7 SAFETY ANALYSES

Safety data of the study will be assessed through adverse events reporting, including serious adverse events and non-serious adverse events of special interest, safety laboratory assessments, measurement of vital signs, and other protocol-specified tests that are deemed critical to the safety evaluation.

The safety analyses will be performed on the treated patients.

4.7.1 Exposure to Study Medication

Information on study drug administration such as the duration of treatment, dose intensity (percentage of planned dosage received) and number of dose/cycles will be

summarized by treatment arm and by study medication by using the following statistics: sample size, mean, standard deviation, median, minimum, and maximum.

Withdrawals of patients from study treatment will be reported as listings and summary tables by treatment arms as well.

4.7.2 Adverse Events

After informed consent has been obtained but prior to initiation of study drug, only serious adverse events caused by a protocol-mandated intervention should be reported. After initiation of study drug, all adverse events will be reported in accordance with the protocol. Verbatim descriptions of adverse events will be mapped to the most recent version of the Medical Dictionary for Regulatory Activities thesaurus terms.

Treatment-emergent adverse events and serious adverse events will be summarized by treatment arm, by body system and by the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 4.0 grading system. For all adverse events, the most extreme intensity will be used for reporting. Individual listings and summary tables will be presented by body system and severity. In tables showing the overall incidence of adverse events, patients who experienced the same event on more than one occasion will be counted only once in the calculation of the event frequency.

Adverse events of special interest (including elevated ALT or AST in combination with either elevated bilirubin or clinical jaundice, suspected transmission of an infectious agent by the study drug, Grade >3 TLS, and Grade >3 IRR) will be summarized by treatment arm and listings will be provided.

Adverse events leading to dose interruptions, treatment discontinuation, and study withdrawal will be summarized by treatment arm.

Summary tables of adverse events by age (<65 or ≥ 65 years), gender, and race will also be provided.

Deaths reported during the study treatment period and those reported after treatment completion/discontinuation will be summarized by treatment arm and by adverse event versus non-adverse event.

4.7.3 Laboratory Data

All applicable laboratory measurements and mean of changes from baseline will be summarized over time by treatment arm, where baseline measurement is defined as the last valid measurement before first dose of any study medication. For example, TLS-related lab measurements (including uric acid, potassium, inorganic phosphorus, and calcium) and changes from baseline will be summarized by treatment arm over time. In addition, laboratory measurements from individual patients who developed an adverse

event of TLS during the study will be presented by listings. Important hematology parameters, such as absolute lymphocyte count, absolute neutrophil count, hemoglobin, hematocrit, and WBC will be categorized according to NCI CTCAE grading version 4.0. Shift tables will be generated to cross-tabulate the number of patients grouped into each grade at baseline versus post-baseline observations.

4.7.4 Vital Signs

Vital signs will be summarized over time for absolute values and changes from baseline without imputation for missing data. Descriptive statistics will be provided by treatment arm.

ECG data at the screening visit will be summarized by treatment arm.

4.8 MISSING DATA

For PFS, patients who do not have documented disease progression or death will have observations censored on the date of the last adequate response assessment or, if no response assessments were performed after the baseline visit, at the time of randomization.

For OS, patients for whom death has not been documented will have observations censored on the last date at which they were known to be alive.

For response endpoints, patients with no response assessments (for any reason) will be considered non-responders.

For DOR, patients with no documented progression after CR, CRi, nPR, or PR will be censored at the last date of adequate response assessment. Patients who have never responded will not be included in this analysis.

4.9 PHARMACODYNAMIC ANALYSES

The pharmacodynamic endpoint in this study is serial assessment of B-cell, T-cell, and NK-cell lymphocyte subsets by flow cytometry. Proportions of each lymphocyte subset and change from baseline in the proportions will be summarized by treatment arm over time as warranted by the data.

4.10 PHARMACOKINETIC ANALYSES

Individual plasma concentrations of venetoclax will be tabulated after appropriate grouping, summarized (e.g., mean, standard deviation, coefficient of variation, median, minimum, and maximum), and plotted.

Population PK methods will be used to characterize the pharmacokinetics of venetoclax in this study in conjunction with appropriate historical data. Apparent clearance, apparent volume of distribution, and other appropriate PK parameters of venetoclax may be calculated and summarized as data allow. Potential correlations of exposure with

dose, demographics, pharmacodynamic variables, safety, and efficacy outcomes may be explored as warranted by the data. The results from the population PK analysis may be reported separately from the Clinical Study Report.

4.11 PATIENT-REPORTED OUTCOMES ANALYSES

Unless otherwise specified, patient reported outcome endpoint analyses will include all randomized patients who have a non-missing baseline and at least one post-baseline patient report outcome assessment. Patients in this subset will be analyzed according to their randomized treatment assignment, irrespective of the treatment received.

Compliance, summary statistics of HRQOL and CLL symptom scores and change from baseline as measured by MDASI, QLQ-C30, and QLQ-CLL16 questionnaires will be provided. Mean of changes in scores from baseline, as well as time to disease-related symptom progression in both treatment arms will be assessed as data allow. As deemed necessary, additional analyses may be performed.

4.12 INTERIM ANALYSES

One IA is planned during the conduct of the study to assess the efficacy of venetoclax + R combination treatment compared with BR treatment, and to allow for release of the results earlier than the planned final analysis in case of significant differences.

The IA will be performed when approximately 140 investigator-assessed PFS events have occurred in both treatment arms combined (75% of the 186 events required for the final primary efficacy analysis is available). The stopping boundary follows a unified family with parameter P = 2 (Kittelson and Emerson 1999). Based on 140 events, the duration of PFS will be tested at the IA, which corresponds to approximately a 2-sided p-value of $2 \times 0.0013 = 0.0026$. Table 1 summarizes the probability of crossing boundary at the planned IA given assumed true HRs.

Table 1 Probability of Passing Boundary at the Interim Analysis for a Given True Hazard Ratio

True Hazard Ratio	0.66	0.42	0.35
Approximate median PFS in venetoclax + R arm (months) [a]	23.0	36.2	43.4
Probability of passing boundary at interim analysis	29%	98%	99%

PFS = progression-free survival; venetoclax + R = venetoclax + rituximab.

[a] Assumes a median duration of PFS of 15.2 months for the BR arm and exponential distributions for PFS in both treatment arms.

Both investigator-assessed PFS analysis and a corresponding analysis on the basis of IRC-assessed PFS events will be conducted in this IA. In case the observed number of events is not exactly 140, the boundaries will be updated to reflect the number of events based on investigator assessment and IRC assessment, respectively. The same stratification factors as specified in the primary efficacy endpoint analysis will be used. If the p-value of the stratified two-sided log-rank test is less than or equal to the respective boundary for both investigator-assessed PFS analysis and IRC-assessed PFS analysis and the observed HR is favorable for the venetoclax + R combination treatment, the trial will have met its primary efficacy endpoint.

An iDMC will evaluate IA results and provide a recommendation as to whether to release the trial results early. All summaries and analyses according to treatment arm will be prepared by an iDCC for the iDMC review. [redacted]

The final primary efficacy analysis will be performed when approximately 186 investigator-assessed PFS events have been observed. The statistical test level will be adjusted to incorporate the alpha spent at the IA (if it was conducted) so that the overall type I error rate will be maintained at the two-sided 0.05 level.

5. REFERENCES

Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29-41.

Fischer K, Cramer P, Busch R, et al. Bendamustine combined with rituximab in patients with relapsed and/or refractory chronic lymphocytic leukemia: a multicenter Phase II trial of the German Chronic Lymphocytic Leukemia Study Group. J Clin Oncol 2011;29:3559–66.

Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood 2008;111:5446–56.

Kittelson J, Emerson S. A Unifying Family of Group Sequential Test Designs. Biometrics 1999;55:874-82

Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. J Stat Plan Inference 2001;99:25–40.