Official Title: A Randomized, Open-Label, Multicenter Phase III Study Evaluating Efficacy and Safety of Mosunetuzumab in Combination with Polatuzumab Vedotin in Comparison with Rituximab in Combination with Gemcitabine Plus Oxaliplatin in Participants with Relapsed or Refractory Aggressive B-Cell Non-Hodgkin's Lymphoma

NCT Number: NCT05171647

Document Date: Statistical Analysis Plan Version 5: 06-Mar-2024

STATISTICAL ANALYSIS PLAN

STUDY TITLE: **A RANDOMIZED, OPEN-LABEL, MULTICENTER PHASE III STUDY EVALUATING EFFICACY AND SAFETY OF MOSUNETUZUMAB IN COMBINATION WITH POLATUZUMAB VEDOTIN IN COMPARISON WITH RITUXIMAB IN COMBINATION WITH GEMCITABINE PLUS OXALIPLATIN IN PARTICIPANTS WITH RELAPSED OR REFRACTORY AGGRESSIVE B-CELL NON-HODGKIN'S LYMPHOMA**

STUDY NUMBER: GO43643

STUDY NAME: SUNMO

VERSION NUMBER: 1

TEST COMPOUND(S): Mosunetuzumab (RO7030816; BTCT4465A)
Polatuzumab vedotin (RO5541077; DCDS4501S)
Tocilizumab (RO4877533)

EUDRACT NUMBER: To be determined

IND NUMBER: 120651

NCT NUMBER: NCT05171647

PLAN PREPARED BY: [redacted], M.Sc.

STATISTICAL ANALYSIS PLAN APPROVAL

SPONSOR: F. Hoffmann-La Roche Ltd
LEGAL REGISTERED ADDRESS: Grenzacherstrasse 124
4070 Basel, Switzerland

DATE FINAL: See electronic date stamp on the last page of this document

STATISTICAL ANALYSIS PLAN VERSION HISTORY

This Statistical Analysis Plan (SAP) was developed based on Roche SAP model document Version 2, 28 February 2022.

SAP Version	Approval Date	Based on Protocol (Version, Approval Date)
1	See electronic date stamp on the last page of this document	Version 4 (Global), 20 February 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Description
AE	adverse event
AESI	adverse events of special interest
BMI	body mass index
CCOD	clinical cut-off date
CMR	complete metabolic response
CR	complete response
CRS	cytokine release syndrome
CSR	Clinical Study Report
CT	computed tomography
[redacted]	
DLBCL	diffuse large B-cell lymphoma
DOR	duration of response
DOCR	duration of complete response
ECG	electrocardiogram
eCRF	electronic Case Report Form
EFS	event-free survival
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire −Core 30
FACT	Functional Assessment of Cancer Therapy
FL3b	follicular lymphoma grade 3b
HR	hazard ratio
IA	interim analysis
ICH	International Council on Harmonization
iDCC	independent Data Coordinating Center
iDMC	independent Data Monitoring Committee
IRF	Independent Review Facility
ITT	intent-to-treat
IV	intravenous
IxRS	interactive voice/web-based response system
LBCL	large B-cell lymphoma
LDH	lactate dehydrogenase
Lym LymS	lymphoma questionnaire, subscale for lymphoma symptoms
MDD	minimally detectable difference
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation or Term	Description
NALT	new anti-lymphoma therapy
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NHL	non-Hodgkin Lymphoma
NMR	no metabolic response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PET	positron emission tomography
PFS	progression-free survival
PMD	progressive metabolic disease
PMR	partial metabolic response
PR	partial response
PRO	patient-reported outcomes
PD	pharmacodynamics
PK	pharmacokinetic
R/R	relapsed or refractory
RMST	restricted mean survival time
SAE	serious adverse events
SAP	Statistical Analysis Plan
SC	subcutaneous
SD	stable disease
SMQs	standardized MedDRA queries
trFL	transformed follicular lymphoma

1. INTRODUCTION

The purpose of Study GO43643 (SUNMO) is to assess the efficacy and safety of mosunetuzumab in combination with polatuzumab vedotin (M+P) in participants with relapsed or refractory (R/R) diffuse-large B-cell lymphoma (DLBCL), high-grade B-cell lymphoma, transformed follicular lymphoma (FL) and FL Grade 3B (FL3b) in comparison with a commonly used regimen in this participant population, rituximab, gemcitabine and oxaliplatin (R-GemOx). While there have been recent advances in treatments for participants with R/R DLBCL, including those with novel mechanisms of action, a high unmet medical need continues to exist particularly for those who are not eligible for autologous stem cell transplant (ASCT).

This Statistical Analysis Plan (SAP) provides details of the planned analyses and statistical methods for Study GO43643 (SUNMO), a randomized, open-label, multicenter Phase III study evaluating efficacy and safety of mosunetuzumab in combination with polatuzumab vedotin in comparison with rituximab in combination with gemcitabine plus oxaliplatin in participants with relapsed or refractory aggressive B-cell non-Hodgkin lymphoma.

The analyses described in this SAP will supersede those specified in Protocol GO43643 for the purposes of a regulatory filing.

The detailed background information for the study can be found in the study protocol. Specific references to the study protocol are based on study Protocol Version 4 (Global), dated 20 February 2024.

There are no changes to the planned analyses described in the protocol.

1.1 OBJECTIVES AND ENDPOINTS OR ESTIMANDS

This study will evaluate the efficacy and safety of M+P compared with R-GemOx in participants with R/R large B-cell lymphoma (LBCL), including DLBCL, high-grade B-cell lymphoma, trFL, and FL3B, who received at least one prior systemic therapy and are not candidates for ASCT. Specific objectives and corresponding endpoints for the study are outlined in Table 1 and Table 2.

Table 1 Primary and Key Secondary Objectives and Corresponding Estimands

Primary Objective	Estimand Definition
• To evaluate the efficacy of M+P (Arm A) compared with R-GemOx (Arm B)	• Population: The first [redacted] randomized participants. • Endpoint: ORR, defined as the proportion of participants in whom an objective response (CR or PR) was observed at any time during the study as determined by the IRF, according to Lugano 2014 response criteria. • Treatment: o Experimental arm: M+P o Control arm: R-GemOx • Intercurrent events and handling strategies: o [redacted] o [redacted] o [redacted] o [redacted] • Population-level summary: difference in proportion for ORR
• To evaluate the efficacy of M+P (Arm A) compared with R-GemOx (Arm B)	• Population: All randomized participants. • Endpoint: PFS, defined as the time from randomization to the first occurrence of disease progression or death from any cause, whichever occurs first, as determined by the IRF, according to Lugano 2014 Response criteria. • Treatment: as defined above • Intercurrent events and handling strategies: o [redacted] o [redacted] o [redacted] o [redacted] o [redacted] • Population-level summary: stratified hazard ratio

Table 1 Primary and Key Secondary Objectives and Corresponding Estimands (Cont.....)

Secondary Objective	Estimand Definition
• To evaluate the efficacy of M+P (Arm A) compared with R-GemOx (Arm B)	• Population: All randomized participants. • Endpoint: OS, defined as the time from randomization to death from any cause. • Treatment: as defined above • Intercurrent events and handling strategies: ○ [redacted] ○ [redacted] ○ [redacted] • Population-level summary: stratified hazard ratio

CR= complete response; IRF=Independent Review Facility; M=mosunetuzumab; ORR=objective response rate; OS=overall survival; P=polatuzumab vedotin; PFS=progression-free survival; PR=partial response; R-GemOx=rituximab, gemcitabine, oxaliplatin.

Table 2 Other Secondary and Exploratory Objectives and Endpoints

Other Secondary Objectives	Corresponding Endpoints
• To evaluate the efficacy of M+P (Arm A) compared with R-GemOx (Arm B)	• ORR, defined as the proportion of participants in whom an objective response (CR or PR) was observed at any time during the study as determined by the investigator • DOR, defined as the time from the first occurrence of a documented objective response to disease progression, or death from any cause, whichever occurs first as determined by the investigator and IRF • PFS, defined as the time from randomization to the first occurrence of disease progression as determined by the investigator, or death due to any cause, whichever occurs first • CRR, defined as the proportion of participants in whom CR was observed at any time during the study as determined by IRF and by the investigator • DOCR, defined as the time from the first occurrence of a documented CR to disease progression, or death from any cause, whichever occurs first. This will be determined by IRF and by the investigator. • Time to deterioration in physical functioning and fatigue scales, as measured by the EORTC QLQ-C30 • Time to deterioration in lymphoma symptoms, as measured by FACT-Lym LymS

Table 2 Other Secondary and Exploratory Objectives and Endpoints (Cont.….)

Other Secondary Objectives	Corresponding Endpoints
• To evaluate the safety and tolerability of M+P (Arm A) compared with R-GemOx (Arm B)	• Incidence and severity of adverse events, with severity determined according to the NCI CTCAE v5.0, including CRS, with severity determined according to the ASTCT CRS Consensus grading criteria • Change from baseline in targeted vital signs • Change from baseline in targeted clinical laboratory test results • Tolerability, as assessed by dose interruptions, dose reductions, and dose intensity, and study treatment discontinuation because of adverse events • Change from baseline in peripheral neuropathy, as measured by the FACT/GOG-Ntx
Exploratory Objectives	**Corresponding Endpoints**
• [redacted]	• [redacted]
• [redacted]	• [redacted]
• [redacted]	• [redacted]
• [redacted]	• [redacted]

Table 2 Other Secondary and Exploratory Objectives and Endpoints (Cont.….)

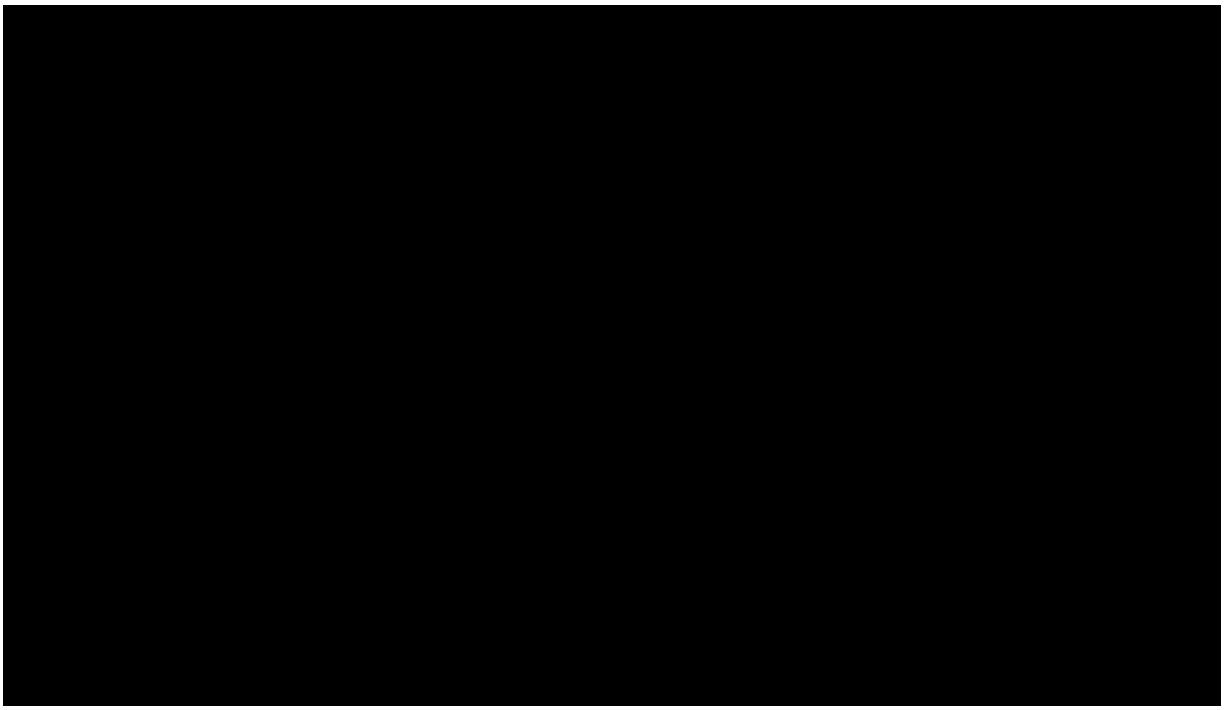
Exploratory Objectives	Corresponding Endpoints
• [redacted]	• [redacted]
• [redacted]	
• [redacted]	
• [redacted]	

ADA=anti-drug antibody [redacted] ASTCT=American Society for Transplantation and Cellular Therapy; CAR=chimeric antigen receptor; [redacted]; COO=cell-of-origin; CR=complete response; CRR=complete response rate; CRS=cytokine release syndrome; [redacted] DH=double hit lymphoma; DLBCL=diffuse-large B-cell lymphoma; DOCR=duration of complete response; DOR=duration of response; EORTC QLQ-C30=European Organization for Research and Treatment of Cancer Quality of Life Questionnaire −Core 30; FACT=Functional Assessment of Cancer Therapy; GOG-Ntx=Gynecologic Oncology Group−Neurotoxicity; [redacted]; IRF=Independent Review Facility; Lym LymS=lymphoma questionnaire, subscale for lymphoma symptoms; M=mosunetuzumab; [redacted]; NCI CTCAE=National Cancer Institute Common Terminology Criteria for Adverse Events; ORR=objective response rate; OS=overall survival; P=polatuzumab vedotin; PFS=progression-free survival; PR=partial response; R-GemOx=rituximab, gemcitabine, oxaliplatin; R/R=relapsed or refractory; TBNK=T cell, B cell, natural killer cell; [redacted]

1.2 STUDY DESIGN

Figure 1 illustrates the study design of this Phase III, open-label, multicenter, randomized, controlled trial in participants with R/R DLBCL, high-grade B-cell lymphoma, trFL, or FL3B, who are not candidates for ASCT.

Figure 1 Study Schema



Approximately 222 eligible participants will be randomized in a 2:1 ratio to receive either M+P (Arm A) or R-GemOx (Arm B); this is the population from which the primary analysis will be performed. China is included as a participating country, additional participants may be enrolled in an extended China enrollment cohort at China's sites, including [redacted], to ensure a total of approximately [redacted] participants with R/R aNHL in a China subpopulation. The global population will include all participants enrolled during the global enrollment phase (including approximately 10 participants enrolled at China's sites during that phase), and the China subpopulation will include all participants enrolled at China's sites (i.e., approximately [redacted] participants in the global cohort and [redacted] participants in the China extension cohort).

The Chinese participants recruited at China's sites into the global study will be analyzed together with all other participants enrolled in the global study and will be reported as part of the global study Clinical Study Report (CSR).

The China subpopulation includes all Chinese participants enrolled in the global study and the China extension cohort. The China subpopulation analysis will be summarized in a CSR, separate from the global CSR after the completion of the China extension cohort.

Participants will be randomized in a 2:1 ratio to either Arm A or Arm B. Arms A and B are defined as:

- Arm A, mosunetuzumab+polatuzumab vedotin: One cycle of treatment is 21 days. Mosunetuzumab will be administered [redacted] mg SC on Cycle 1, Day 1; [redacted] mg on

Cycle 1, Day 8; Cycle 1, Day 15; and Day 1 of Cycles 2−8. Polatuzumab vedotin will be administered IV at 1.8 mg/kg on Day 1 of Cycles 1−6.

- Arm B, rituximab, gemcitabine, and oxaliplatin: One cycle of treatment is 14 days. Rituximab 375 mg/m^2, gemcitabine 1000 mg/m^2, and oxaliplatin 100 mg/m^2 will be administered IV on Day 1. Gemcitabine should be administered before administration of oxaliplatin.

During randomization, permuted blocks will be employed using the following stratification factors:

- Number of previous lines of systemic therapy for aggressive lymphoma (1 vs. ≥2)
- Outcome after last systemic therapy (relapsed vs. refractory)
    - Relapsed disease in this study is defined as disease that has recurred after having a documented history of response (CR or PR) ≥6 months in duration from completion of the last treatment.
    - Refractory disease is defined as disease that either did not respond to, or progressed within 6 months (<6 months) of last treatment.

No crossover to the experimental arm is allowed. Participants will undergo tumor assessments at screening, every 8 weeks (±1 week) for the first 6 months following treatment initiation, and every 3 months (±1 month) for the first two years on study, then at month 30 (± 2 months) after Cycle 1 Day 1, regardless of dose delay, until disease progression, start of new anti-lymphoma therapy, or study discontinuation, whichever is earlier, per Lugano Criteria 2014 (Cheson et al. 2014; Appendix 1). At the investigator's discretion, tumor assessments may be repeated at any time if PD is suspected. Response assessments will be performed according to the 2014 Lugano Response Criteria (Cheson et al. 2014; Appendix 1), as assessed on positron emission tomography (PET)/computed tomography (CT) scans.

All participants will be monitored for adverse events, clinical laboratory test results and vital signs throughout the study and for at least 90 days after the final dose of study treatment or the initiation of next anti-lymphoma therapy, whichever is earlier. Adverse events will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 5.0, except for CRS severity, which will be determined per the American Society for Transplantation and Cellular Therapy (ASTCT) CRS grading criteria (Lee et al. 2019). To characterize the pharmacokinetic (PK) profile and immune response in response to study treatment, blood samples will be taken at various time points before and after dosing. See protocol for detailed methodology.

An independent Data Monitoring Committee (iDMC) monitors safety data.

1.2.1 Treatment Assignment

This is a randomized, open-label study. After initial written informed consent has been obtained, all screening procedures and assessments have been completed, and eligibility has been established for a participant, the study site will obtain the participant's identification number and treatment assignment from an interactive voice/web-based response system (IxRS).

Participants will be randomly assigned to one of two treatment arms: Arm A or Arm B. Randomization will occur in a 2:1 ratio through use of a permuted-block randomization method to ensure a balanced assignment to each treatment arm. Randomization will be stratified by the number of prior treatment regimens for aggressive lymphoma (1 vs. ≥2) and the outcome after the last therapy (relapsed vs. refractory). See Section 1.2 for details.

The randomization method implemented in the China extension cohort was the same as that implemented in the global population.

1.2.2 Independent Review Facility

An Independent Review Facility (IRF) will perform a centralized, independent review of images, and other clinical data as needed, prior to the efficacy analyses. The IRF membership and procedures will be detailed in an IRF charter. The IRF will be used to evaluate the study endpoints of progression-free survival (PFS), objective response rate (ORR), complete response rate (CRR), duration of response (DOR), and duration of complete response (DOCR) in a blinded manner.

1.2.3 Data Monitoring

An independent Data Monitoring Committee (iDMC) will evaluate safety and efficacy data periodically during the study. The analysis supporting iDMC review will be conducted by an independent Data Coordinating Center (iDCC) and provided to the iDMC. Sponsor affiliates will be excluded from iDMC membership. The iDMC will follow a charter that outlines the iDMC roles and responsibilities. Refer to the iDMC charter and protocol for further details.

2. STATISTICAL HYPOTHESES AND SAMPLE SIZE DETERMINATION

2.1 STATISTICAL HYPOTHESES

Analysis for ORR by IRF, one of the dual primary endpoints, will be conducted for the first [redacted] randomized participants (interim analysis population [IAP]) to test the following hypothesis (Section 4.2.1):

[redacted]

Analysis for PFS by IRF, one of the dual primary endpoints, will test the equality of PFS distribution in M+P versus R-GemOx in the intent-to-treat (ITT) population:

[redacted]

2.2 SAMPLE SIZE DETERMINATION

Assuming a median PFS of [redacted] months in the R-GemOx arm (Mounier et al. 2013; Cazelles et al. 2019; Schade et al. 2019), and a randomization ratio of 2:1, [redacted] events are required to detect a between-group difference of [redacted] months in median PFS (hazard ratio=[redacted]) assuming an exponential distribution of PFS, with use of a log-rank test with [redacted]% power and a [redacted]-sided α of [redacted]. At the interim efficacy analysis for Study GO40516 (clinical cut-off date (CCOD) of 15 March 2021), a median PFS of [redacted] months ([redacted]% CI: [redacted], NE) was observed in [redacted] participants with R/R DLBCL receiving M+P. Although immature, these results are consistent with the assumed median PFS of [redacted] months for M+P.

Assuming a median overall survival (OS) of [redacted] months in the R-GemOx arm (Mounier et al. 2013; Cazelles et al. 2019; Schade et al. 2019), and considering an interim OS efficacy analysis at the time of the primary PFS analysis, a between-group difference of [redacted] months in median OS (hazard ratio=[redacted]), assuming an exponential distribution of OS, would be detected, with use of a log-rank test with [redacted]% power and a [redacted]-sided α of [redacted] (Section 4.3.1).

A total of approximately 222 participants will be randomized in this study. The actual numbers of events for PFS and OS, and the analysis timings as described above have changed due to an extended enrollment timeline (See Section 4.2.3 and Section 4.3.1 updated projections). SUNMO completed enrollment except for the US and China, leading to extended follow up of participants enrolled outside the US and China. Initially, the primary PFS analysis was powered at [redacted]%, and the final OS analysis was powered at [redacted]%. The extended enrollment timeline increased the power for the primary analysis as PFS events continue to accumulate beyond the target number of events throughout trial enrollment. Therefore, the study has adequate power to reclassify objective response rate as a dual primary endpoint along with PFS, and add an earlier analysis based on objective response rate to detect clinically meaningful improvement based on this intermediate efficacy endpoint (Section 2.2.1 and Section 4.2.1).

2.2.1 Type I Error Control

The overall type I error rate for this study is strictly controlled at [redacted]% ([redacted]-sided) using the Fallback Method (see Section IV C in FDA Guidance on Multiple Endpoints in Clinical Trials 2017). A hierarchical testing procedure including possible α recycling (Bretz et al. 2009) will be used to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints. The test hierarchy and α spending plan for key secondary efficacy endpoints are described below and shown in Figure 2.

The ORR analysis will occur when the first [redacted] randomized participants have a minimum of [redacted] months of follow-up from the first response assessment. ORR will be tested at the [redacted]% [redacted]-sided α level.

The first PFS by IRF analysis will be performed by the iDCC for iDMC review to assess futility, at the time of the interim ORR analysis, on the first [redacted] randomized participants. The PFS futility boundary is set at PFS HR greater than [redacted] for mosunetuzumab and polatuzumab in comparison with R-GemOx. The second PFS analysis will be conducted at the planned primary PFS analysis [redacted] months after the last patient in (LPI), which is expected approximately [redacted] months after first patient in (FPI). If the ORR analysis is positive, the primary PFS analysis will use a [redacted]% [redacted]-sided α level. If the ORR analysis is negative, the primary PFS analysis will use a [redacted]% [redacted]-sided α level.

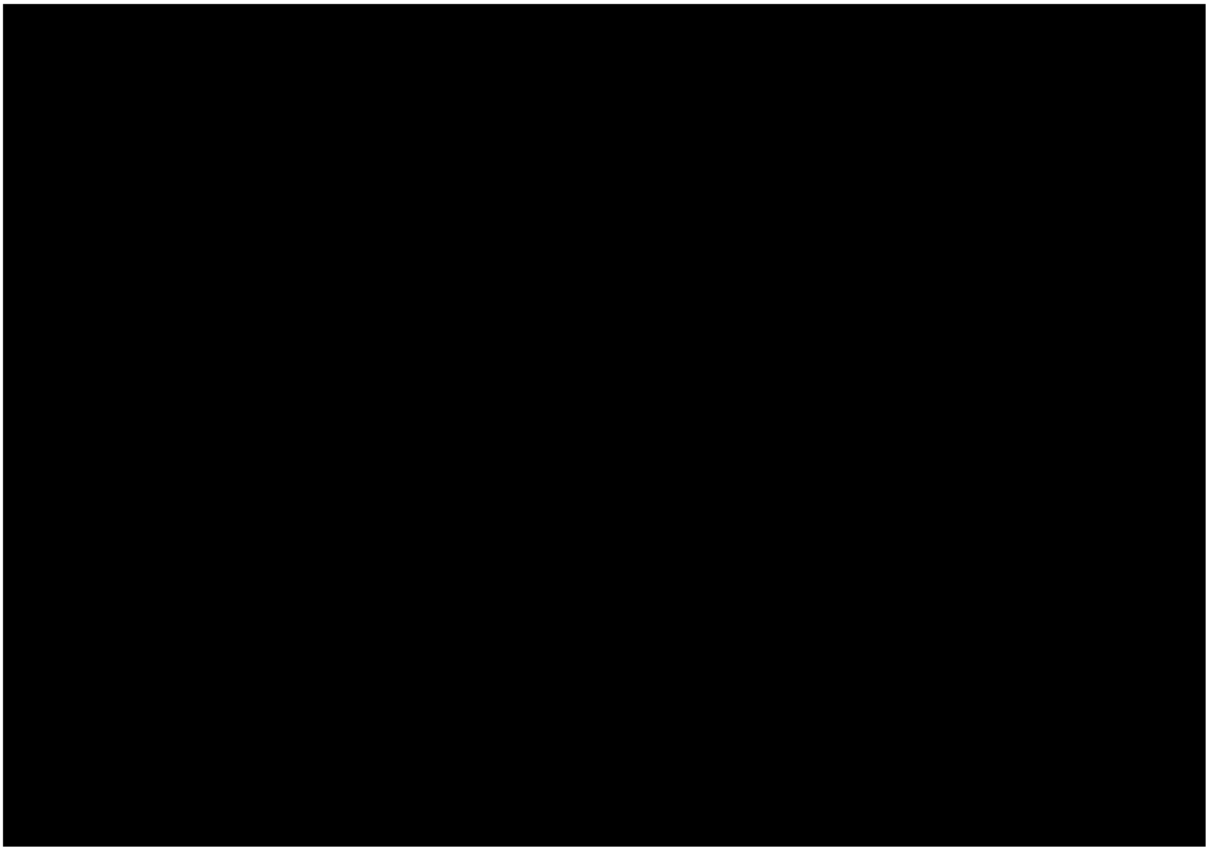
Depending on the number of events, there will be three or four analyses for the key secondary endpoint of OS based on hierarchical testing order to ensure the final OS analysis was powered at [redacted]% (Figure 2). The first and second OS analyses will be performed by the iDCC for iDMC review to assess futility. The first OS analysis will be performed when approximately [redacted] OS events are observed. The second OS analysis will be performed at the time of the dual primary endpoint ORR analysis. For the first and second OS analyses, OS effect is considered detrimental if OS HR is greater than [redacted] for mosunetuzumab and polatuzumab vedotin compared to R-GemOx.

The third OS analysis will be conducted at the time of the primary PFS analysis to assess safety and efficacy. OS effect is considered detrimental if the observed OS HR is greater than [redacted]. OS will be formally tested for efficacy only if PFS is positive. Depending on the number of events, if the power for the third OS analysis is at least [redacted]% at the same α level as PFS (either at α=[redacted] or [redacted] the third OS analysis is considered final. If the power for the third OS analysis is less than [redacted]%, a group sequential test will be used, and a fourth OS analysis may be needed. The fourth OS analysis will be conducted only if the previous OS are not considered final and when there are sufficient OS events to achieve a power of [redacted]%. The α levels for OS analyses depend on the spending function boundaries and actual number of OS events.

An overview of the type I error rate control strategy is shown in Figure 2. This figure is based on the projected number of events and analysis timings.

[redacted]

[redacted]



2.2.2 Sample Size for China Population

This study plans to enroll approximately 222 participants across all sites during the global enrollment phase, among which approximately [redacted] participants are from China. Additional [redacted] participants are enrolled in an extended China enrollment cohort to ensure a total of approximately [redacted] participants in the China subpopulation. The global population includes all participants enrolled during the global enrollment phase, and the China subpopulation includes all China participants enrolled during both the global enrollment phase and the extended China enrollment cohort.

3. ANALYSIS SETS

The participant analysis sets for the purposes of analyses are defined in Table 3.

Table 3 Participant Analysis Sets

Participant Analysis Set	Description
Enrolled	All participants who sign the Informed Consent Form.
IAP	The first [redacted] randomized participants; participants will be included in the analyses according to the treatment they were assigned.
ITT	All randomized participants; participants will be included in the analyses according to the treatment they were assigned.
Safety-evaluable	All participants exposed to study treatment; participants will be analyzed according to the treatment that they actually received.
PK-evaluable	All participants who have at least received 1 dose mosunetuzumab and have at least 1 evaluable PK sample post-dose for at least 1 analyte.
Immunogenicity-evaluable	All participants who have at least 1 pre-dose or 1 post-dose ADA assessment.

ADA = anti-drug antibody; IAP = interim analysis population; ITT = intent-to-treat; PK = pharmacokinetic.

4. STATISTICAL ANALYSES

Unless otherwise specified, the analyses described in this section are based on participants enrolled during the global enrollment phase. Details of the planned analyses, including any additional analyses needed to support country-specific or regional marketing applications, will be provided in Section 4.7.6.

The timing for interim and final analyses are provided in Section 2.2.1.

Unless specified otherwise, all continuous variables will be summarized with descriptive statistics (e.g., number of non-missing values, mean, median, standard deviation, minimum, and maximum). All categorical variables will be summarized with frequency counts and percentages. Data will be presented by treatment arm.

4.1 GENERAL CONSIDERATIONS

For the efficacy analyses, participants will be analyzed according to the treatment to which they were assigned; whereas for safety analyses, participants will be included in the analyses according to the treatment they actually received.

Hypothesis tests will be [redacted]-sided, unless otherwise indicated. The type I error (α) for this study is [redacted]-sided).

Baseline for non-safety analyses is defined as the last available measurement obtained on or prior to randomization. Baseline for safety analyses is defined as the last available

measurement prior to first exposure to any of the study drugs. Participants with missing baseline assessments will not be imputed.

For OS, data from participants who did not have death documented will be censored on the last date they were known to be alive. Participants who do not have information after baseline will be censored at the date of randomization. For PFS, data for participants who do not have documented disease progression or who have not died will be treated as censored observations on the date of the last response assessment by CT, or positron emission tomography (PET)-CT. If no response assessments were performed after the baseline visit, PFS data will be censored at the date of randomization. Detailed censoring rules for PFS are provided in Table 5. For the response endpoints, participants with no response assessments (for any reason) will be considered non-responders.

4.1.1 Missing Response Data Handling Strategy

In real world clinical practice, once a patient achieves complete metabolic response (CMR) by PET-CT/CT scans, the complete response continues until progressive disease is confirmed by PET-CT, CT scan, or clinical deterioration/assessment. At each scheduled or unscheduled response assessment visit, if the PET-CT scan result is missing or not evaluable, the final overall response for that visit will be a combination of the metabolic response at prior visit and the change of the current CT assessment to the CT assessment at the last time point of FDG-PET with a metabolic response evaluation, as described in Table 4.

Table 4 Rules for Response Data Handling

Response reported in the last evaluation visit with PET-CT scan	PET-CT scan reported in current visit	CT data reported in current visit compared to the last CT evaluation with PET-CT scan	PET-CT/CT endpoint based overall response assessment
Any (NE, PMD, NMR, PMR, CMR)	Missing/ND/NE	No change of response assessment	PET response from last visit
Any (NE, PMD, NMR, PMR, CMR)	Missing/ND/NE	Unequivocal PD (New lesion / significant progression of preexisting lesions)	PD
NE, NMR, PMR	Missing/ND/NE	Improvement to PR	PR
NE, NMR, PMR	Missing/ND/NE	Improvement to CR	CR
CMR	Missing/ND/NE	Improvement to PR or CR	CR

CMR = complete metabolic response; CR = complete response; CT = computed tomography; ND = not done; NE = not evaluable; NMR = no metabolic response; PD = progressive disease; PET = positron emission tomography; PMD = progressive metabolic disease; PMR = partial metabolic response; PR = partial response

4.2 DUAL PRIMARY ENDPOINTS ANALYSES

The primary objective of this study is to evaluate the efficacy of M + P in comparison with R-GemOx in participants with R/R LBCL, as measured by the dual primary endpoints of ORR by the IRF in the IAP and PFS by the IRF in the ITT, according to Lugano 2014 Criteria (Cheson et al. 2014; Appendix 1).

4.2.1 Definition of Dual Primary Endpoint of ORR by IRF

The dual primary endpoint of ORR is defined as the proportion of participants in whom an objective response (CR or PR) was observed at any time during the study, based on PET-CT and/or CT scan, as determined by the IRF, according to Lugano 2014 Response Criteria (Cheson et al. 2014; Appendix 1).

The ORR analysis will be conducted after the first [redacted] randomized participants (i.e., IAP) have a minimum of 9 months of follow-up from the first response assessment by the IRF.

ORR by the IRF will be evaluated once more at the time of primary PFS analysis. However, at the time of primary PFS analysis, ORR will be a supportive secondary endpoint that will not be included in the overall Type I error control strategy.

Assuming the ORR is [redacted]% in R-GemOx versus [redacted]% in M + P, this analysis has approximately [redacted]% power to detect a difference in ORR at [redacted]% [redacted]-sided α level. The

following hypothesis will be tested at the [redacted] % [redacted]-sided α level for the first [redacted] R/R LBCL participants:

[redacted]

It is projected that an observed ΔORR of [redacted] % or better will result in a statistically significant difference between treatment arms. That is, a ΔORR of [redacted] % will be the minimal detectable difference (MDD) for the analysis.

4.2.2 Main Analytical Approach for Dual Primary Endpoint of ORR by IRF

To provide clarity on the target of estimation, the 5 attributes in the estimand framework are defined below:

Population:

- At the time of interim ORR analysis, the first [redacted] randomized participants (IAP population), analyzed by randomized treatment group.
- At the time of primary PFS analysis, all randomized participants (ITT population), analyzed by randomized treatment group.

Variable: ORR by IRF

Treatment: Participants will be randomized into either Arm A (M+P) or Arm B (R-GemOx). During the conduct of the study participants may also receive concomitant medications as detailed in Section 6.8 of the Protocol.

Intercurrent events and handling strategies:

- [redacted]
- [redacted]
- [redacted]
- [redacted]

Population-level summary for the variable: Difference in proportion for ORR.

ORR will be compared between treatment arms using the Cochran-Mantel-Haenszel test stratified by the IxRS-recorded stratification factors. Responses after initiation of NALT will not be included in the analysis of ORR. For cases in which a participant is mis-randomized with respect to a stratification factor (i.e., there is a discrepancy between the IxRS-recorded stratification factor level and the electronic Case Report Form [eCRF]-recorded stratification factor level for previous lines of systemic therapy or outcome after the last systemic therapy), the IxRS data will be used in the primary analysis. Discordances between the IxRS and the eCRF data will be summarized.

4.2.3 Definition of Dual Primary Endpoint of PFS by IRF

The dual primary endpoint of PFS is defined as the time from randomization to the first occurrence of disease progression, as determined by the IRF with use of the Lugano 2014 Response criteria (Cheson et al. 2014; Appendix 1), or death from any cause, whichever occurs first. For participants who have neither progressed nor died as of the CCOD for analysis, PFS will be censored on the date of last disease assessment when the participant is known to be progression-free. For participants who do not have any evaluable post-baseline tumor assessments, PFS will be censored on the date of randomization. Additional censoring rules for participants who initiated new anti-lymphoma therapy or participants with two or more consecutive missed assessments are detailed in Table 5.

The primary PFS analysis will be conducted approximately [redacted] months after LPI on the ITT population when approximately [redacted] events are observed. This analysis will have [redacted]% power at [redacted]% [redacted] sided α level if ORR analysis is positive or [redacted]% power at [redacted]% [redacted]-sided α level if ORR analysis is negative to detect a hazard ratio of [redacted] in M+P in comparison with R-GemOx (Figure 2). The timing of the primary PFS analysis is no longer event-driven because of the delay in US enrollment. An event driven primary PFS analysis would likely result in sufficient events prior to LPI on study. Instead, the primary PFS analysis is conducted [redacted] months after LPI (i.e., [redacted] months from FPI) to allow all participants to complete treatment.

The following assumptions are made for PFS:

- PFS curve follows an exponential distribution.
- Median PFS of [redacted] months in R-GemOX arm and [redacted] months in M+P arm (corresponding to a target HR of [redacted]
- The dropout rate is [redacted]% over a [redacted]-month period.

The primary PFS analysis of the study will test the equality of PFS distribution in M+P compared with R-GemOx:

[redacted]

Treatment comparisons will be made with use of a stratified log-rank test with a [redacted]-sided [redacted] level if ORR analysis is positive or [redacted] level if ORR analysis is negative. The

randomization stratification factors to be used in the efficacy analyses are the number of previous lines of systemic therapy for aggressive lymphoma [redacted] and the outcome after the last systemic therapy (relapsed vs. refractory).

An observed HR of [redacted] or better for PFS will result in a statistically significant difference between treatment arms if ORR is positive (i.e., an HR of [redacted] will be the minimal detectable difference for the analysis). Alternatively, an observed HR of [redacted] or better for PFS will result in a statistically significant difference between treatment arms if ORR is negative.

For cases in which a participant is mis-randomized 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 for previous lines of systemic therapy or outcome after the last systemic therapy), the IxRS data will be used in the primary analysis due to the ITT population. Discordances between the IxRS and the eCRF data will be summarized.

4.2.4 Main Analytical Approach for Dual Primary Endpoint of PFS by IRF

To provide clarity on the target of estimation, the 5 attributes in the estimand framework are defined below:

Population: All randomized participants (ITT population), analyzed by randomized treatment group.

Variable: Progression-free survival, defined as time from randomization to the first occurrence of disease progression, as determined by the IRF with use of the Lugano 2014 Response Criteria (Cheson et al. 2014; Appendix 1), or death from any cause, whichever occurs first.

Treatment: Participants will be randomized into either Arm A (M+P) or Arm B (R-GemOx). During the conduct of the study, participants may also receive concomitant medications as detailed in Section 6.8 of the Protocol.

Intercurrent events and handling strategies:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Population-level summary for the variable: Stratified hazard ratio.

Detailed event and censoring rules for PFS are summarized in Table 5.

Table 5 Event and Censoring Rules for PFS

Scenario[1]	Date of PFS Event or Censoring	Outcome
Disease progression or death without missing response assessments	Date of earliest disease progression or death before the CCOD	Event
No death and no disease progression before the CCOD	Date of last adequate[2] assessment before the CCOD. If no adequate post baseline assessment is available, or no baseline assessment is available, then date of randomization.	Censored
Two or more consecutive missed assessments, followed by disease progression or death	Date of last adequate assessment before the missed assessments. If no adequate post baseline assessment is available, or no baseline assessment is available, then date of randomization.	Censored
Two or more missed consecutive assessments, followed by no disease progression or no death	Date of last adequate assessment before the missed assessments. If no adequate post baseline assessment is available, or no baseline assessment is available, then date of randomization.	Censored
Early treatment discontinuation, followed by disease progression or death	Date of earliest disease progression or death before the CCOD	Event
Early treatment discontinuation, followed by no disease progression or no death	Date of last adequate assessment before the CCOD. If no adequate post baseline assessment is available prior to treatment discontinuation date, then date of randomization.	Censored
Study discontinuation prior to disease progression or death	Date of last adequate assessment before the CCOD. If no adequate post baseline assessment is available prior to treatment discontinuation date, then date of randomization.	Censored
Start of new anti-lymphoma therapy (NALT) before documentation of disease progression or death	Date of last adequate assessment prior to start of NALT. If no adequate post baseline assessment is available prior to start of NALT, then date of randomization.	Censored

CCOD = clinical cut-off date; PFS = progression-free survival.

[1] Sensitivity analyses may be performed for other situations if significant imbalances between arms are observed.

[2] To be considered adequate, a tumor assessment not including PET should have CR, PR, SD, and PD as outcome; and/or a tumor assessment including PET-CT should have CMR, PMR, NMR, or PMD using Lugano 2014 criteria (Cheson et al. 2014; Appendix 1). Assessments that are “not evaluable” (NE) or “not done” (ND) are considered not adequate.

The Kaplan-Meier method will be used to estimate the median PFS, if reached, and PFS distribution for each treatment arm. Kaplan-Meier estimates will be provided at [redacted] months. The Brookmeyer-Crowley methodology (Brookmeyer and Crowley 1982) will be used to construct the [redacted]% CI for the median PFS for each treatment arm.

Stratified Cox proportional-hazards model will be used to estimate the hazard ratio and its [redacted]% CI.

4.2.5 Sensitivity Analyses

The following analysis approaches will be used as sensitivity analyses for the dual primary endpoints of ORR and PFS by the IRF, using Lugano 2014 criteria (Cheson et al. 2014; Appendix 1).

4.2.5.1 Sensitivity Analysis Related to Stratification Factors Discordance

If 5% or more participants have stratification discrepancies between the eCRF and IxRS, eCRF recorded stratification factors will be performed for the dual primary endpoints of ORR and PFS by IRF.

4.2.5.2 Sensitivity Analysis Related to NALT and Missed Assessments

Additional sensitivity analysis will also be performed on PFS without censoring for NALT or two or more missed assessments. The event and censoring rules for NALT and missed assessments sensitivity analysis for PFS are described in Table 6.

Table 6 Event and Censoring Rules for PFS Sensitivity Analysis Related to NALT and Missed Assessments

NALT Scenario	Date of PFS Event or Censoring	Outcome
Start of NALT, followed by death or disease progression	Date of earliest disease progression or death before the CCOD	Event
Start of NALT, followed by no death or no disease progression	Date of last adequate[1] assessment before the CCOD. If no adequate post baseline assessment is available prior to treatment discontinuation date, then date of randomization.	Censored
Two or more consecutively missed scheduled visits, followed by disease progression or death	Date of earliest disease progression or death before the CCOD	Event
Two or more consecutively missed scheduled visits, followed by no disease progression or no death	Date of last adequate[1] assessment before the CCOD. If no adequate post baseline assessment is available prior to treatment discontinuation date, then date of randomization.	Censored

CCOD = clinical cut-off date; NALT = new anti-lymphoma therapy; PFS = progression-free survival.

[1] To be considered adequate, a tumor assessment not including PET should have CR, PR, SD, and PD as outcome; and/or a tumor assessment including PET-CT should have CMR, PMR, NMR, or PMD using Lugano 2014 criteria (Cheson et al. 2014; Appendix 1). Assessments that are "not evaluable" (NE) or "not done" (ND) are considered not adequate.

4.2.5.3 Sensitivity Analysis Related to COVID-19

As a COVID-19 sensitivity analysis for ORR, participants who died due to COVID-19 or discontinued from treatment due to COVID-19 prior to having any tumor assessments will be removed in the ORR analysis.

Additional sensitivity analysis will also be performed on PFS where death due to COVID-19 will be censored. The event and censoring rules for COVID-19 sensitivity analysis for PFS are described in Table 7. Other event and censoring rules are the same as in Table 5.

Table 7 Event and Censoring Rules for PFS Sensitivity Analysis Related to COVID-19

COVID-19 Scenario	Date of PFS Event or Censoring	Outcome
Death due to COVID-19	Date of death before the CCOD	Censored
Treatment discontinuation due to COVID-19	Date of treatment discontinuation	Censored

CCOD = clinical cut-off date; COVID-19 = coronavirus disease; PFS = progression free survival

4.2.5.4 Evaluation of Proportional Hazards Assumption

The proportional hazards assumption on PFS may be examined using both graphical and analytical methods if hazards are not proportional. The log [-log] of the survival function versus time for PFS may be plotted for the comparison between M + P and R-GemOx. If the curves are not parallel, indicating that hazards are not proportional, supportive analyses may be conducted to account for the possible non-proportional hazards effect using the restricted mean survival time (RMST) method (Royston and Parmar 2011).

The RMST will be computed for PFS using the area under the curve from baseline to several time points ([redacted] and [redacted] months). The RMST will be computed for each treatment arm and the difference with its [redacted]% CI (by Greenwood method) and p-values (by Z test) will be provided for descriptive purpose.

4.2.5.5 Concordance Analysis of ORR and PFS

Agreement/disagreement between the investigator assessment and assessment by the IRF of ORR and PFS will be summarized.

For PFS, a specific analysis which outlines whether the IRF event/censoring was earlier or later than investigator assessment will also be provided. The two assessments will be

considered in agreement if the two time-to-event determinations do not differ by more than 30 days and both agree on whether there is disease progression or not. Note that PFS events due to death are handled separately from disease progression events within the summary table.

4.2.6 Supplementary Analyses

4.2.6.1 Subgroup Analyses for Dual Primary Endpoints

The generalizability of ORR and PFS results when comparing M+P to R-GemOx is investigated by estimating the treatment effect in subgroups based on stratification factors, key baseline demographics, disease characteristics, etc. Summaries of primary endpoints by these subgroups will be provided in forest plots. Table 8 specifies the subgroups that will be explored; other subgroups may be included in the analysis. Subgroup analyses will not be adjusted for multiplicity; all subgroup analyses will be exploratory only.

Table 8 Subgroups for Subgroup Analyses

Subgroup	Grouping
Age	(<65, ≥ 65); (<75, ≥ 75)
Sex	Male, Female
Ethnicity	Hispanic or Latino, Not Hispanic or Latino, Not stated or Unknown
Race	White, Black/ African American, Asian, American Indian or Alaska Native, Middle Eastern, Native Hawaiian or Pacific Islander, North African, Not stated or Unknown
Region	US and Canada, Latin America, East Asia, ROW
Baseline Body Mass Index	$<$ median, $\geq$ median
Baseline Eastern Cooperative Oncology Group (ECOG) status	$0, \geq 1$
Baseline CD20 status	Positive, negative
Prior lines of therapy (recorded by IxRS and eCRF)	(1, 2, 3+); (1, 2+)
Relapsed/refractory to last line of therapy (recorded by IxRS and eCRF)	Refractory, non-refractory
Relapsed/refractory to first line of therapy	Refractory, non-refractory
Received prior CAR-T therapy	Yes, No
Relapsed/refractory to prior CAR-T therapy	Refractory, non-refractory
Relapsed/refractory to prior anti-CD20 therapy	Refractory, non-refractory
Time since last anti-CD20	≤ 3 months, >3 months

Subgroup	Grouping
Cell of origin	GCB, non-GCB, unknown
Double-hit (MYC and BCL2 rearrangements)	Yes, No
IPI score at study entry	Low (0-1), Low-intermediate (2), High-intermediate (3), High (4-5)
Prior autologous stem-cell transplant (ASCT)	Yes, No
Early relapse from ASCT (PD < 12 months from completion)	Yes, No
Bulky disease (> 7.5 cm)	Yes, No
Type of non-Hodgkin lymphoma (NHL)	FL3b, DLBCL, HGBCL
trFL	Yes, No
Primary refractory disease or relapse within one year of first line therapy	Yes, No
Lactate dehydrogenase (LDH) level	Not high (≤ ULN), high (> ULN)
Ann Arbor stage at study entry	I, II, III, IV

BCL2 = B-cell lymphoma 2; CAR = chimeric antigen receptor; CD = cluster of differentiation; eCRF = electronic case report form; DLBCL = diffuse large B-cell lymphoma; FL3b = follicular lymphoma 3b; GCB = germinal center B cell type; HGBCL = high grade B cell lymphoma; IPI = International Prognostic Index; IxRS = interactive voice/web-based response system; NHL = Non-Hodgkin lymphoma; trFL = transformed follicular lymphoma; ULN = upper limit normal.

4.3 SECONDARY ENDPOINT ANALYSES

4.3.1 Key Secondary Endpoint: Overall Survival in Intent-to-Treat (ITT) Population

Overall survival, defined as the time from randomization to date of death from any cause. For participants who have not died at the clinical cutoff date (CCOD) for analysis, OS will be censored on the last date when the participants are known to be alive. Participants who do not have information after baseline will be censored at the date of randomization.

To provide clarity on the target of estimation, the 5 attributes in the estimand framework are defined below:

Population: All randomized participants (ITT population), analyzed by randomized treatment group.

Variable: Overall survival, defined as the time from randomization to date of death from any cause.

Treatment: Participants will be randomized into either Arm A (M+P) or Arm B (R-GemOx). During the conduct of the study participants may also receive concomitant medications as detailed in Section 6.8 of the Protocol.

Intercurrent event and handling strategies:

- [redacted]
- [redacted]
- [redacted]

Population-level summary for the variable: Stratified hazard ratio.

The methodologies used for the PFS analysis will be used for the OS analysis. Event and censoring rules for OS are summary in Table 9.

Table 9 Event and Censoring Rules for OS

Scenario	Date of PFS Event or Censoring	Outcome
Death	Death date	Event
No death	Last known alive date[1] before CCOD	Censored
No death and no post-baseline survival information available	Randomization date	Censored

CCOD=clinical cut-off date; PFS=progression-free survival; OS=overall survival.

[1] Last known alive date is defined as the last date patient has documented data to show that they are alive. Scenarios considered in this definition may include last survival follow-up date with patient status of "alive", date of last of tumor assessment with a valid response, date of last treatment administration with a valid dose, date of last lab assessment with valid results, and date of last update of adverse event information.

The analysis timing and stopping boundaries for OS are in Table 10. Depending on the number of events, there will be three or four analyses for the key secondary endpoint of OS based on hierarchical testing order to ensure the final OS analysis is powered at [redacted]% (Figure 2). The first and second OS analyses will be performed by the iDCC for iDMC review to assess futility. The first OS analysis will be performed when approximately [redacted] OS events are observed. The second OS analysis will be performed at the time of the dual primary endpoint ORR analysis for all randomized patients by the CCOD. For the first and second OS analyses, OS effect is considered detrimental if OS HR is greater than [redacted] for mosunetuzumab and polatuzumab vedotin compared to R-GemOx.

The third OS analysis will be conducted at the time of the primary PFS analysis to assess safety and efficacy. OS effect is considered detrimental if the observed OS HR is greater than [redacted]. OS will be formally tested for efficacy only if PFS is positive.

If ORR analysis in IAP (Section 4.2.1) is positive and PFS is positive, the third OS analysis will be conducted at the time of the primary PFS analysis when approximately

[redacted] OS events are observed. This may be the only one OS analysis needed for efficacy, as power is approximately [redacted]%, at [redacted]-sided α of [redacted]%, to detect a hazard ratio of [redacted] in M+P in comparison with R-GemOx.

If ORR is negative and PFS is positive, a third OS analysis will be conducted at the time of the PFS analysis when approximately [redacted] OS events are observed, with [redacted]% power and a [redacted]-sided local α of [redacted]%. If the third OS analysis is negative, a fourth OS analysis will be conducted approximately [redacted] months after LPI when approximately [redacted] events are observed, with [redacted]% power and a [redacted]-sided local α of [redacted]%, to detect a hazard ratio of [redacted] in M+P in comparison with R-GemOx.

In general, depending on the number of events, if the power for the third OS analysis is at least [redacted]% at the same α level as PFS (either at α=[redacted] or [redacted]); the third OS analysis is considered final. If the power for the third OS analysis is less than [redacted]%, a group sequential test will be used, and a fourth OS analysis may be needed. The fourth OS analysis will be conducted when there are sufficient OS events to achieve a power of [redacted]%. The actual α levels for OS analyses depend on the spending function boundaries and actual number of OS events.

The following assumptions are made for OS:

- OS curve follows an exponential distribution
- Median OS of [redacted] months in the R-GemOx arm and [redacted] months in the M+P arm (corresponding to a target HR of [redacted]
- The drop-out rate is [redacted]% over a [redacted]-month period

An observed HR of [redacted] or better for OS will result in a statistically significant difference between the treatment arms if ORR is positive and PFS is positive. Alternatively, an observed HR of [redacted] at interim OS analysis and [redacted] at final OS analysis will result in a statistically significant difference between treatment arms if ORR is negative and PFS is positive.

Table 10 Analysis Timing and Stopping Boundaries for Overall Survival

[redacted]

4.3.2 Supportive Secondary Endpoints

Apart from the key secondary endpoint of OS in Section 4.3.1, all other secondary endpoints, sensitivity analyses and exploratory analyses are not considered key secondary endpoints. All supportive secondary endpoints are tested in an exploratory manner at the [redacted] level, at the time of dual primary endpoints analyses of ORR and PFS. PFS by INV will only be analyzed by the iDCC for iDMC review, at the time of dual primary endpoint ORR analysis, as part of the PFS by IRF futility analysis.

4.3.2.1 Complete Response Rate (CRR) and ORR

CRR is defined as the proportion of participants whose objective response is a CR during the study, based on PET-CT and/or CT scan, according to the Lugano 2014 Response Criteria (Cheson et al. 2014; Appendix 1), as determined by the IRF and the investigator (INV).

ORR is defined as the proportion of participants in whom an objective response (CR or PR) was observed at any time during the study, based on PET-CT and/or CT scan, as determined by the INV, according to Lugano 2014 Response Criteria (Cheson et al. 2014; Appendix 1).

The same estimand framework and statistical methods used for ORR by IRF (Section 4.2.2) will be used for CRR by IRF and INV and ORR by INV.

As a COVID-19 sensitivity analysis for CRR and ORR, participants who died due to COVID-19 or discontinued from treatment due to COVID-19 prior to having any tumor assessments will be removed in the CRR and ORR analyses.

4.3.2.2 Duration of Objective Response (DoR) and Duration of Complete Response (DoCR)

Duration of objective response is defined as the time interval from the date of the first occurrence of an objective response (PR or CR) until the first date that progressive disease or death is documented, whichever occurs first. Duration of CR is defined as the time interval from the date of the first occurrence of CR until the first date that progressive disease or death is documented, whichever occurs first. The same PFS event and censoring rules as described in Table 5 will be applied to the DoR and DoCR, except all responders will have at least one adequate post-baseline assessment.

To provide clarity on the target of estimation, the 5 attributes in the estimand framework are defined below:

Population:

- At the time of interim ORR analysis: For DoR, all responders in the first [redacted] randomized participants (IAP population), analyzed by randomized treatment group. For DoCR, all complete responders in the first [redacted] randomized participants (IAP population), analyzed by randomized treatment group.

- At the time of primary PFS analysis: For DoR, all responders in all randomized participants (ITT population), analyzed by randomized treatment group. For DoCR, all complete responders in all randomized participants (ITT population), analyzed by randomized treatment group.

Variable:

- Duration of objective response, defined as the time interval from the date of the first occurrence of an objective response (PR or CR) until the first date that progressive disease or death is documented, whichever occurs first, as determined by the IRF and the investigator.
- Duration of CR, defined as the time interval from the date of the first occurrence of CR until the first date that progressive disease or death is documented, whichever occurs first, as determined by the IRF and the investigator.

Treatment: Participants will be randomized into either Arm A (M+P) or Arm B (R-GemOx). During the conduct of the study participants may also receive concomitant medications as detailed in Section 6.8 of the Protocol.

Intercurrent events and handling strategies:

- [redacted]

Population-level summary for the variable: Hazard ratio.

The methodologies detailed for the PFS analysis will be used for the DoR and DoCR analyses, except that these analyses will not be stratified. The sensitivity analyses related to NALT/missed assessments and COVID-19 (Sections 4.2.5.2 and 4.2.5.3) will also be conducted for DoR and DoCR.

Concordance analysis will be done for DoR and DoCR, following the same method as concordance analysis for PFS (Section 4.2.5.5).

4.3.2.3 Clinical Outcome Assessments

Compliance

Completion rates will be summarized by the number and proportion of participants among those expected to complete the EORTC QLQ-C30, FACT-Lym LymS, and FACT/GOG-NTX at each time point for each treatment arm. The questionnaire is considered completed if at least 50% questions have been answered. This analysis will be based on the ITT. Reasons for non-completion will be summarized at each time point by treatment arm.

Time to Deterioration

Time to deterioration analyses will be performed on the EORTC QLQ-C30 physical functioning and fatigue scales and FACT-Lym lymphoma subscale. For the EORTC QLQ-C30 physical functioning and fatigue scales, time to deterioration in physical functioning and/or fatigue is defined as the time from randomization to the first documentation of a 10-point or more decrease or increase, respectively, from baseline (Osoba et al. 1998; Cocks et al. 2012). For the FACT-Lym LymS, time to deterioration in lymphoma-specific symptoms is defined as the time from randomization to the first documentation of a 3-point or more decrease, from baseline (Carter et al. 2008; Hlubocky et al. 2013). The hazard ratio for deterioration will be estimated using a stratified Cox proportional hazards model. The 95% CI for the hazard ratio will be provided. Kaplan-Meier methodology will be used to estimate 1-year and 2-year rates, as well as the median time to deterioration (if reached) for each treatment arm, and Kaplan-Meier curves will be presented. Participants who do not have an observed deterioration at the time of the CCOD will be censored at the last non-missing PRO assessment date. Participants without a post-baseline PRO assessment will be censored at randomization.

4.3.2.4 Secondary Safety Endpoints

The following secondary safety endpoints will be analyzed on the safety evaluable population, defined as all randomized participants who receive any amount of any study treatment, group according to treatment received:

- Incidence and severity of adverse events, with severity determined according to the NCI CTCAE v5.0, including CRS, with severity determined according to the ASTCT CRS grading criteria (Lee et al. 2019; Appendix 2)
- Change from baseline in targeted vital signs
- Change from baseline in targeted clinical laboratory test results
- Tolerability, as assessed by dose interruptions, dose reductions, and dose intensity, and study treatment discontinuation because of adverse events
- Change from baseline in peripheral neuropathy, as measured by the FACT/GOG-Ntx

All verbatim adverse event terms occurring on or after first study treatment will be mapped to the Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms, and adverse event severity will be graded according to NCI CTCAE v5.0, except for CRS severity, which will be graded according to the ASTCT CRS grading criteria (Lee et al. 2019; Appendix 2).

4.4 EXPLORATORY ENDPOINTS ANALYSIS

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

4.5 ADVERSE EVENTS

All verbatim adverse event terms will be mapped to the most recent version of the MedDRA thesaurus terms. Other than adverse events of CRS, which will be graded according to ASTCT CRS Consensus Grading (see Appendix 2; Lee et al. 2019), the adverse event severity grading scale for the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 5.0 will be used for assessing adverse event severity. All adverse events, serious adverse events, adverse events leading to death, adverse events of special interest, and adverse events leading to study treatment discontinuation that occur on or after the first dose of study treatment (i.e., treatment-emergent adverse events), adverse events leading to dose modifications (including dose interruptions or delays, and dose reductions) will be summarized by mapped term, appropriate thesaurus level, and severity grade. For events of varying

severity, the highest grade will be used in the summaries. Deaths and cause of death will be summarized. All recorded adverse events will be listed by treatment, patient number, and schedule. The most extreme grade will be used for reporting. All serious adverse events will be listed separately and summarized. Deaths reported during the study treatment period and during follow up after treatment discontinuation will be listed and summarized. Additional listings may be provided if deemed relevant by Safety Science.

4.5.1 Extent of Exposure

Information on each study drug administration will be summarized by treatment duration, number of cycles, number of doses, and cumulative dose. The descriptive analysis will include number of participants, mean, standard deviation, median, minimum, and maximum. Listings and summary tables for participants who discontinued study treatment will be reported.

4.5.2 Cytokine Release Syndrome Analysis

Cytokine release syndrome will be graded according to ASTCT CRS Consensus Grading (see Appendix 2; Lee et al. 2019). Summaries for CRS events will be produced, including signs and symptoms graded by NCI CTCAE, study day of onset, treatment cycle, outcome, management, and treatment for adverse event. Along with this summary, a line list of concomitant medications taken on the day of the dose before any CRS events will be presented, and these events will be summarized by dose and grade (ASTCT CRS grading criteria per Lee et al. 2019). Additional tabular and graphic representations of CRS events may also be produced.

In addition, CRS algorithm previously requested by the FDA (BLA 761263 FDA Label Round 2 – Serial No. 0078, 22 November 2022) will be performed and reported separately from the CSR.

4.5.3 Laboratory Data

All clinical laboratory data will be stored in the database in the units in which they were reported. Patient summary statistics at each assessment time will be presented by using the International System of Units (Système International d'Unités [SI]). Laboratory data not reported in SI units will be converted to SI units before processing.

Laboratory test values outside the normal ranges will be presented by individual summaries.

For all laboratory parameters included, a Roche predefined standard reference range exists. Laboratory values that fall outside this standard reference range will be labeled "H" for high or "L" for low in patient summaries of laboratory data. In addition to the standard reference range, a marked reference range has been predefined by the Sponsor for each laboratory parameter. The marked reference range is broader than the standard reference range. Values outside the marked reference range that also

represent a defined change from baseline will be considered marked laboratory abnormalities (i.e., potentially clinically relevant). If a baseline value is not available for a patient, the midpoint of the standard reference range will be used as the patient's baseline value to determine marked laboratory abnormalities. Marked laboratory abnormalities will be labeled in the patient summaries as "HH" for very high or "LL" for very low.

4.5.4 Vital Signs

Vital signs data (pulse rate, respiratory rate, blood pressure, pulse oximetry, and temperature) will be presented by individual summaries, and values outside the normal ranges and marked abnormalities will be flagged. In addition, tabular summaries will be used as appropriate. Changes in vital signs will be summarized.

4.6 EXPLORATORY PRO ENDPOINTS

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

1. [redacted]

2. [redacted]

3. [redacted]
4. [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

4.7 OTHER ANALYSES

4.7.1 Summaries of Conduct of Study

Enrollment, study treatment administration, and discontinuation from the study will be summarized by treatment arm. The reasons for study treatment discontinuation will also be tabulated. Major protocol deviations, including major deviations with regard to the inclusion and exclusion criteria, will be summarized by treatment arm.

Pre-treatment and concomitant medications will be summarized separately by treatment group.

4.7.2 Summaries of Demographics and Baseline Characteristics

Demographics and baseline characteristics (including age, sex, and race/ethnicity) will be summarized by treatment arm. Baseline data are the last data obtained prior to initiation of study treatment. Descriptive statistics (mean, standard deviation, median, and range) will be presented for continuous variables and counts and percentages will be presented for categorical variables.

4.7.3 Pharmacokinetic Analysis

The PK population for analysis will include all participants who have received at least 1 dose of mosunetuzumab and polatuzumab vedotin and have at least 1 evaluable PK sample post-dose for at least 1 analyte.

Individual and mean concentrations of mosunetuzumab (serum) and polatuzumab total antibody (serum), acMMAE (plasma) and unconjugated MMAE (plasma) versus time data will be tabulated and plotted. The pharmacokinetics of the above analytes will be summarized as the data will allow for by estimating selected PK parameters. The data from this study may be pooled with data from other studies or analyzed via population PK approach, as data will allow for and at the Sponsor's discretion.

Exposure-response analysis may be conducted using plasma/serum concentrations or relevant PK parameters and available clinical outcome data, per the Sponsor's discretion and as the data will allow for.

To assess for potential PK drug-drug interactions, PK parameters for mosunetuzumab and each analyte of polatuzumab vedotin may be compared with historical data, as the data will allow for.

4.7.4 Immunogenicity Analyses

The numbers and proportions of ADA-positive participants and ADA-negative participants at baseline (baseline prevalence) and after drug administration (post-baseline incidence) will be summarized by treatment group. When determining post-baseline incidence, participants are considered to be ADA-positive if they are ADA-negative or have missing data at baseline but develop an ADA response following study drug exposure (treatment-induced ADA response), or if they are ADA-positive at baseline and the titer of one or more post-baseline samples is at least 0.60 titer unit greater than the titer of the baseline sample (treatment-enhanced ADA response). Participants are considered to be ADA-negative if they are ADA-negative or have missing data at baseline and all post-baseline samples are negative, or if they are ADA-positive at baseline but do not have any post-baseline samples with a titer that is at least 0.60 titer unit greater than the titer of the baseline sample (treatment unaffected).

The relationship between ADA status and safety, efficacy, PK, and biomarker endpoints may be analyzed and reported with use of standard language/terminology.

4.7.5 Biomarker Analyses

Participants who harbor translocations in BCL2 and MYC are considered high-risk and are defined as double-hit, high grade B cell lymphoma (HGBCL). Central confirmation by FISH (fluorescence in situ hybridization) will be performed to assess status of BCL2 and MYC. Association of this exploratory subset relative to efficacy measurements may be analyzed.

Although no formal statistical analysis of other exploratory biomarkers will be performed, data may be analyzed in the context of this study and in aggregate with data from other studies.

4.7.6 Analyses of China Subpopulation

The Asian subpopulation analysis will be conducted only for China to meet local regulatory requirements. The objective of this subgroup analysis is to assess the treatment effect of M+P compared with R-GemOx in the Asia subpopulation, and to investigate the consistency in treatment effect between this Asian subpopulation and the global population for the purpose of registration in China.

The Asian subpopulation will include all participants enrolled at China's sites (i.e., during both the global enrollment phase and the extended China enrollment phase) and the participants enrolled from other Asian countries/regions (e.g., [redacted], etc. during the global enrollment phase). Results from these analyses will be summarized in a separate CSR.

The analysis of PFS in the Asian subpopulation will be performed when at least half of the Asian participants have had PFS events and no earlier than the primary analysis for the global population.

4.8 INTERIM ANALYSES

[redacted]

[redacted]

5. SUPPORTING DOCUMENTATION

The following documents support this SAP:

- iDMC Charter
- Lugano Independent Read Charter
- Data Access Memo

Appendix 1 Lugano Response Criteria for Malignant Lymphoma (Cheson et al. 2014)

TARGET AND NON-TARGET LESIONS

Up to 6 of the largest target nodes, nodal masses, or other lymphomatous lesions that are measurable in 2 diameters should be identified from different body regions representative of the participant's overall disease burden and include mediastinal and retroperitoneal disease, if involved. At baseline, a measurable node must be greater than 15 mm in longest diameter (LDi). Measurable extranodal disease may be included in the 6 representative, measured lesions. At baseline, measurable extranodal lesions should be greater than 10 mm LDi.

All other lesions (including nodal, extranodal, and assessable disease) should be followed as non-measured disease as non-target lesions (e.g., cutaneous, gastrointestinal, bone, spleen, liver, kidneys, pleural or pericardial effusions, ascites, bone, bone marrow).

SPLIT LESIONS AND CONFLUENT LESIONS

Lesions may split or may become confluent over time. In the case of split lesions, the individual product of the perpendicular diameters (PPDs) of the nodes should be summed together to represent the PPD of the split lesion; this PPD is added to the sum of the PPDs of the remaining lesions to measure response. If subsequent growth of any or all of these discrete nodes occurs, the nadir of each individual node is used to determine progression. In the case of confluent lesions, the PPD of the confluent mass should be compared with the sum of the PPDs of the individual nodes, with more than 50% increase in PPD of the confluent mass compared with the sum of individual nodes necessary to indicate progressive disease. The LDi and smallest diameter (SDi) are no longer needed to determine progression.

Revised Criteria for Response Assessment		
Response and Site	**PET–CT-Based Response**	**CT-Based Response**
Complete	Complete metabolic response	Complete radiologic response (all of the following)
Lymph nodes and extralymphatic sites	Score 1, 2, or 3 [a] with or without a residual mass on 5PS [b] It is recognized that in Waldeyer's ring or extranodal sites with high physiologic uptake or with activation within spleen or marrow (e.g., with chemotherapy or myeloid colony-stimulating factors), uptake may be greater than normal mediastinum and/or liver. In this circumstance, complete metabolic response may be inferred if uptake at sites of initial involvement is no greater than surrounding normal tissue even if the tissue has high physiologic uptake.	Target nodes/nodal masses must regress to ≤ 1.5 cm in LDi No extralymphatic sites of disease
Non-measured lesion	Not applicable	Absent
Organ enlargement	Not applicable	Regress to normal
New lesions	None	None
Bone marrow	No evidence of FDG-avid disease in marrow	Normal by morphology; if indeterminate, IHC negative
Partial	Partial metabolic response	Partial remission (all of the following)
Lymph nodes and extralymphatic sites	Score 4 or 5 [b] with reduced uptake compared with baseline and residual mass(es) of any size	≥ 50% decrease in SPD of up to 6 target measurable nodes and extranodal sites
	At interim, these findings suggest responding disease	When a lesion is too small to measure on CT, assign 5 mm × 5 mm as the default value
	At end-of-treatment, these findings indicate residual disease	When no longer visible, 0 × 0 mm For a node > 5 mm × 5 mm, but smaller than normal, use actual measurement for calculation

Revised Criteria for Response Assessment		
Response and Site	**PET–CT-Based Response**	**CT-Based Response**
Non-measured lesion	Not applicable	Absent/normal, regressed, but no increase
Organ enlargement	Not applicable	Spleen must have regressed by > 50% in length beyond normal
New lesions	None	None
Bone marrow	Residual uptake higher than uptake in normal marrow but reduced compared with baseline (diffuse uptake compatible with reactive changes from chemotherapy allowed). If there are persistent focal changes in the marrow in the context of a nodal response, consideration should be given to further evaluation with MRI or biopsy or an interval scan.	Not applicable
No response or stable disease	No metabolic response	Stable disease
Target nodes/nodal masses, extranodal lesions	Score 4 or 5 [b] with no significant change in FDG uptake from baseline at interim or end-of-treatment	< 50% decrease from baseline in SPD of up to six dominant, measurable nodes and extranodal sites; no criteria for progressive disease are met
Non-measured lesion	Not applicable	No increase consistent with progression
Organ enlargement	Not applicable	No increase consistent with progression
New lesions	None	None
Bone marrow	No change from baseline	Not applicable

Revised Criteria for Response Assessment		
Response and Site	**PET–CT-Based Response**	**CT-Based Response**
Progressive disease	Progressive metabolic disease	Progressive disease requires at least 1 of the following
Individual target nodes/nodal masses	Score 4 or 5 [b] with an increase in intensity of uptake from nadir[c] (baseline can be the nadir when there has not been prior regression of the involved lesions) and/or	PPD progression
Extranodal lesions	New FDG-avid foci consistent with lymphoma at interim or end-of-treatment assessment	An individual node/lesion must be abnormal with: LDi $>$ 1.5 cm and Increase by $\geq$ 50% from PPD nadir and An increase in LDi or SDi from nadir 0.5 cm for lesions $\leq$ 2 cm 1.0 cm for lesions $>$ 2 cm In the setting of splenomegaly ($>$ 13 cm), the splenic length must increase by $>$ 50% of the extent of its prior increase beyond baseline (e.g., a 15-cm spleen must increase to $>$ 16 cm). If no prior splenomegaly, must increase by at least 2 cm from baseline. New or recurrent splenomegaly New or clear progression of preexisting non-measured lesions
New lesions	New FDG-avid foci consistent with lymphoma rather than another etiology (e.g., infection, inflammation); if uncertain regarding etiology of new lesions, biopsy or interval scan may be considered	Regrowth of previously resolved lesions A new node $>$ 1.5 cm in any axis A new extranodal site $>$ 1.0 cm in any axis; if $<$ 1.0 cm in any axis, its presence must be unequivocal and must be attributable to lymphoma Assessable disease of any size unequivocally attributable to lymphoma
Bone marrow	New or recurrent FDG-avid foci	New or recurrent involvement

5PS = 5-point scale; CT = computed tomography; FDG = fluorodeoxyglucose; GI = gastrointestinal; IHC = immunohistochemistry; LDi = longest transverse diameter of a lesion; MRI = magnetic resonance imaging; PET = positron emission tomography; PPD = cross product of the LDi and perpendicular diameter; SDi = shortest axis perpendicular to the LDi; SPD = sum of the product of the perpendicular diameters for multiple lesions.

[a] A score of 3 in many participants indicates a good prognosis with standard treatment, especially if at the time of an interim scan. However, in trials involving PET where de-escalation is investigated, it may be preferable to consider a score of 3 as inadequate response (to avoid undertreatment).
Measured dominant lesions: Up to 6 of the largest dominant nodes, nodal masses, and extranodal lesions selected to be clearly measurable in 2 diameters. Nodes should preferably be from disparate regions of the body and should include, where applicable, mediastinal and retroperitoneal areas. Non-nodal lesions include those in solid organs (e.g., liver, spleen, kidneys, lungs), gastrointestinal involvement, cutaneous lesions, or those noted on palpation.
Non-measured lesions: Any disease not selected as measured; dominant disease and truly assessable disease should be considered not measured. These sites include any nodes, nodal masses, and extranodal sites not selected as dominant or measurable or that do not meet the requirements for measurability but are still considered abnormal, as well as truly assessable disease, which is any site of suspected disease that would be difficult to follow quantitatively with measurement, including pleural effusions, ascites, bone lesions, leptomeningeal disease, abdominal masses, and other lesions that cannot be confirmed and followed by imaging. In Waldeyer's ring or in extranodal sites (e.g., GI tract, liver, bone marrow), FDG uptake may be greater than in the mediastinum with complete metabolic response, but should be no higher than surrounding normal physiologic uptake (e.g., with marrow activation as a result of chemotherapy or myeloid growth factors).

[b] PET 5PS: 1 = no uptake above background; 2 = uptake ≤ mediastinum; 3 = uptake > mediastinum but ≤ liver; 4 = uptake moderately > liver; 5 = uptake markedly higher than liver and/or new lesions; X = new areas of uptake unlikely to be related to lymphoma.

[c] Deviation from Lugano Criteria (Cheson et al. 2014): Progressive Metabolic Disease (PMD) assessment to be compared to nadir, not baseline (unless the nadir is the baseline visit)

Appendix 2 ASTCT Cytokine Release Syndrome Consensus Grading

CRS Parameter	Grade 1	Grade 2	Grade 3	Grade 4
Fever [a]	Temperature ≥38°C	Temperature ≥38°C	Temperature ≥38°C	Temperature ≥38°C
Hypotension	None	Not requiring vasopressors	with Requiring a vasopressor with or without vasopressin	Requiring multiple vasopressors (excluding vasopressin)
Hypoxia	None	Requiring low-flow nasal cannula [c] or blow-by	and/or [b] Requiring high-flow nasal cannula, facemask, nonrebreather mask or Venturi mask	Requiring positive pressure (e.g., CPAP, BiPAP, intubation and mechanical ventilation)

ASTCT = American Society for Transplantation and Cellular Therapy; BiPAP = bi-level positive airway pressure; CPAP = continuous positive airway pressure; CRS = cytokine release syndrome; CTCAE = Common Terminology Criteria for Adverse Events.

Grade 5 CRS is defined as death due to CRS in which another cause is not the principal factor leading to this outcome.

Organ toxicities associated with CRS may be graded according to CTCAE v5.0 but they do not influence CRS grading.

[a] Fever is defined as a temperature ≥ 38ºC not attributable to any other cause. In patients who have CRS then receive antipyretic or ant cytokine therapy such as tocilizumab or steroids, fever is no longer required to grade subsequent CRS severity. In this case, CRS grading is determined by hypotension and/or hypoxia.

[b] CRS grade is determined by the more severe event, hypotension or hypoxia not attributable to any other cause. For example, a patient with temperature of 39.5 ºC, hypotension requiring 1 vasopressor, and hypoxia requiring low-flow nasal cannula is classified as Grade 3 CRS.

[c] Low-flow nasal cannula is defined as oxygen delivered at ≤ 6 L/minute. Low flow also includes blow-by oxygen delivery, sometimes used in pediatrics. High-flow nasal cannula is defined as oxygen delivered at > 6 L/minute.

6. REFERENCES

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

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics1982;38:29−41.

Carter GC, Liepa AM, Zimmermann AH, et al. Validation of the Functional Assessment of Cancer Therapy-Lymphoma (FACT-LYM) in patients with relapsed/refractory mantel cell lymphoma. Blood 2008;112:2376.

Cazelles C, Belhadj K, Vellemans H, et al. Rituximab plus gemcitabine and oxaliplatin (R-GemOx) in refractory/relapsed (R/R) DLBCL. A real life study in patients ineligible for autologous transplantation. Blood. 2019;134(Supplement 1):4115.

Cheson BD, Fisher IR, Barrington SF, et al. Recommendations for initial evaluation, staging and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano Classification. J Clin Oncol 2014;32:1–9.

Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer 2012;48:1713–21.

Hlubocky FJ, Webster K, Cashy J, et al. The Development and Validation of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma: The Functional Assessment of Cancer Therapy—Lymphoma (FACT-Lym). Lymphoma 2013;2013:1-9.

Lee DW, Santomasso BD, Locke FL et al. ASTCT consensus grading for cytokine release syndrome and neurologic toxicity associated with immune effector cells. Biol Blood Marrow Transplant 2019;25:625–38.

Mounier N, El Gnaoui T, Tilly H, et al. Rituximab plus gemcitabine and oxaliplatin in patients with refractory/relapsed diffuse large B-cell lymphoma who are not candidates for high-dose therapy. A Phase II Lymphoma Study Association trial. Haematologica 2013;98:1726–31.

Osoba D, Rodrigues G, Myles J, et al. Intrepreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139–44.

Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statistics in medicine. 2011;30(19):2409-21

Schade JR, Kim C, Drill E, et al. Retrospective analysis of gemcitabine and oxaliplatin (GemOx)-based treatment in patients with relapsed/refractory aggressive B-cell non-Hodgkin lymphoma. Blood 2019;134(Supplement 1):2904.

Signature Page for Study GO43643 (SUNMO) --Statistical Analysis Plan (SAP) V1-
System identifier: RIM-CLIN-523095

Approval Task	[redacted] Company Signatory 06-Mar-2024 03:05:04 GMT+0000