Official Title: A Phase III, Open-Label, Multicenter, Randomized, Study Evaluating the Safety and Efficacy of Polatuzumab Vedotin in Combination with Rituximab Plus Gemcitabine Plus Oxaliplatin (R-GEMOX) Versus R-GEMOX Alone in Patients with Relapsed/Refractory Diffuse Large B-Cell Lymphoma

NCT Number: NCT04182204

Document Date: SAP Version 3: 14-May-2024

STATISTICAL ANALYSIS PLAN

STUDY TITLE: **A PHASE III, OPEN-LABEL, MULTICENTER, RANDOMIZED STUDY EVALUATING THE SAFETY AND EFFICACY OF POLATUZUMAB VEDOTIN IN COMBINATION WITH RITUXIMAB PLUS GEMCITABINE PLUS OXALIPLATIN (R-GEMOX) VERSUS R-GEMOX ALONE IN PATIENTS WITH RELAPSED/REFRACTORY DIFFUSE LARGE B-CELL LYMPHOMA**

STUDY NUMBER: MO40598

STUDY NAME: POLARGO

VERSION NUMBER: 3

ROCHE COMPOUND(S): Polatuzumab vedotin (RO5541077)

EUDRACT NUMBER: 2018-003727-10

IND NUMBER: 109409

NCT NUMBER: NCT04182204

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

STATISTICAL ANALYSIS PLAN AMENDMENT 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 2.0, 28 February 2022.

SAP Version	Approval Date	Based on Protocol (Version, Approval Date)
3	See electronic date stamp on the last page of this document	Version 8.0, 19 October 2022
2	7 August 2023	Version 8.0, 19 October 2022
1	6 May 2019	Version 2.0, 16 April 2019

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

Key changes to the SAP, along with the rationale(s) for each change, are summarized below.

Section	Description of Change	Rationale for Change
1.1	Safety objectives clarified for the randomized arms.	Clarity
4.2.4.2.1	The supplementary analyses related to high impact further therapies have been updated to add bispecific therapies to the therapies of interest (stem cell transplantation [SCT] and chimeric antigen receptor [CAR]T cell)	To adapt to the evolving therapeutic landscape.
4.2.4.2.1	Inverse Probability of Censoring Weighting (IPCW) added as a potential supplementary analysis of overall survival [OS] to handle high impact further therapies	To further complement the supplementary analysis of OS related to high impact further therapies
4.3.1.1	Progression free survival (PFS) main estimand updated to handle intercurrent events related to non-protocol specified anti-lymphoma therapy [NALT]/ two or more missing tumor assessments with hypothetical strategy rather than treatment policy. PFS estimand handling NALT/ two or more missing tumor assessments with treatment policy is kept as a supplementary analysis for PFS.	To address FDA request
4.3.1.1.1	IPCW added as a potential sensitivity analysis of the main estimand for PFS	In response to the main estimand for PFS (key secondary endpoint) being changed from treatment policy to hypothetical strategy.
4.3.2	Best overall response [BOR] main estimand updated to handle intercurrent events related to NALT with hypothetical strategy rather than treatment policy	To be consistent with PFS update following FDA request
4.3.1.2	The rationale for the supplementary analyses using the “last observation carry forward” method has been revised	The clinical cut off is shifting so all ‘end of treatment’ visits will have taken place (if any) prior to the clinical cutoff date. Nevertheless, some patients did not perform an ‘end of treatment’ visit as their last assessment.

Section	Description of Change	Rationale for Change
4.3.2.4 and 4.3.2.5	Complete response rate [CRR] and objective response rate [ORR] at end of treatment based on response including PET-CT or CT data (composite response) have been added.	To complement the CRR/ORR based on response including PET-CT data only.
4.3.2.6	Duration of response [DOR] main estimand updated to handle intercurrent events related to NALT/ two or more missing tumor assessments with hypothetical strategy rather than treatment policy.	To be consistent with PFS update following FDA request.
4.3.2.9 and 4.4	Patient-reported outcomes [PRO] analysis will only include data up to initiation of any NALT	To be consistent with the PFS update (change from treatment policy to hypothetical strategy to handle intercurrent event related to NALT) following FDA request
4.6.6	Analyses of China subpopulations added	To meet local regulatory requirements.

Additional minor changes have been made throughout to improve clarity and consistency.

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE 3

1. INTRODUCTION 10

1.1 Objectives and Endpoint and Estimands 10

1.1.1 Expression of Objectives and Endpoints Using the Estimand Framework 16

1.2 Study Design 19

1.2.1 Treatment Assignment and Blinding 20

1.2.2 Independent Review Facility 21

1.2.3 Data Monitoring 21

1.2.3.1 Internal Monitoring Committee 21

1.2.3.2 Independent Data Monitoring Committee 22

2. STATISTICAL HYPOTHESES AND SAMPLE SIZE DETERMINATION 22

2.1 Statistical Hypotheses 22

2.2 Sample Size Determination 22

3. ANALYSIS SETS 22

4. STATISTICAL ANALYSES 23

4.1 General Considerations 23

4.2 Primary Endpoint/Estimands Analysis 24

4.2.1 Definition of Primary Endpoint/Estimand 24

4.2.2 Main Analytical Approach for Primary Endpoint 25

4.2.3 Sensitivity Analyses 26

4.2.3.1 Unstratified Analysis 26

4.2.3.2 Stratifications Errors 26

4.2.4 Supplementary Analyses 26

4.2.4.1 Subgroup Analyses for Primary Endpoint 26

4.2.4.2 Other Supplementary Analyses for Primary Endpoint 26

4.2.4.2.1 Impact of Stem Cell Transplant, CAR-T Therapy and Bispecific Therapy 26

4.2.4.2.2 Impact of Switch to Polatuzumab Vedotin 27

4.2.4.2.3 Impact of COVID-19 28

4.2.4.2.4 Impact of Proportional Hazards Assumption 28
4.3 Secondary Endpoints/Estimands Analyses 28
4.3.1 Key Secondary Endpoints/Estimands 28
4.3.1.1 Progression-Free Survival by Investigator 28
4.3.1.1.1 Sensitivity Analysis for PFS 30
4.3.1.1.2 Supplementary Analyses for PFS 31
4.3.1.2 Complete Response Rate at End of Treatment by IRC 32
4.3.1.3 Objective Response Rate at End of Treatment by IRC 33
4.3.2 Supportive Secondary Endpoints 33
4.3.2.1 Best Overall Response by Investigator 33
4.3.2.2 Complete Response Rate at End of Treatment by Investigator 34
4.3.2.3 Objective Response Rate at End of Treatment by Investigator 34
4.3.2.4 Complete Response Rate (Composite) at End of Treatment 34
4.3.2.5 Objective Response Rate (Composite) at End of Treatment 35
4.3.2.6 Duration of Response by Investigator 35
4.3.2.7 Event-Free Survival by Investigator 36
4.3.2.8 Efficacy in Safety Run-In Stage 37
4.3.2.9 Patient-Reported Outcomes 37
4.3.2.9.1 EORTC QLQ-C30 37
4.3.2.9.2 FACT-Lym Subscale 38
4.4 Exploratory Endpoints Analysis 38
4.5 Safety Analyses 39
4.5.1 Extent of Exposure 39
4.5.2 Adverse Events 39
4.5.3 Additional Safety Assessments 40
4.5.3.1 Laboratory Data 40
4.5.3.2 Vital Signs 40
4.5.3.3 ECGs 40
4.6 Other Analyses 40
4.6.1 Summaries of Conduct of Study 40

4.6.2 Summaries of Demographics and Baseline Characteristics 40
4.6.3 Pharmacokinetic Analyses 41
4.6.4 Immunogenicity Analyses 41
4.6.5 Biomarker Analyses 42
4.6.6 Analyses of China Subpopulation 42
[redacted] 42
[redacted] 42
4.8 Changes to Protocol-Planned Analyses 46
5. SUPPORTING DOCUMENTATION 47
6. REFERENCES 53

LIST OF TABLES

Table 1 Objectives and Corresponding Endpoints: Safety Run-In (Stage 1) 11
Table 2 Objectives and Corresponding Endpoints: Randomized-Controlled Trial (Stage 2) 13
Table 3 Primary and Key Secondary Objectives and Corresponding Estimands 17
Table 4 Participant Analysis Sets 23
Table 5 Censoring Rules Analysis for OS (Primary Endpoint) 25
Table 6 Event and Censoring Rules for PFS (Censoring for NALT/ Two or More Missing Tumor Assessments) 29
Table 7 Event and Censoring Rules for PFS (Treatment Policy Strategy) 31
Table 8 Event and Censoring Rules for DOR (Censoring for NALT / Two or More Missing Tumor Assessments) 36
[redacted] 44
[redacted] 45

LIST OF FIGURES

Figure 1 Study Design 20
Figure 2 Safety Run-in Schema (Stage 1) 20

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Description
ADA	anti-drug antibody
AE	adverse event
BOR	best overall response
CAR -T	chimeric antigen receptor T cell
CI	confidence interval
COVID-19	coronavirus disease 2019
CR	complete response
CRR	complete response rate
CSR	Clinical Study Report
CT	computed tomography
DLBCL	diffuse large B-cell lymphoma
DOR	duration of response
DMC	Data Monitoring Committee
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EFS_{eff}	event-free survival
EORTC	European Organization for the Research and Treatment of Cancer
EORTC QLQ-30	European Organization for the Research and Treatment of Cancer Quality-of-Life Questionnaire, Core 30
EQ-5D-5L	EuroQoL 5 Dimension questionnaire
FACT/GOG-NTX	Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity Subscale
FACT-Lym	Functional Assessment of Cancer Therapy-Lymphoma
HR	hazard ratio
[redacted]	
ICH	International Council on Harmonization
iDMC	independent Data Monitoring Committee
IMC	Internal Monitoring Committee
IPCW	Inverse probability of censoring weighting
IPI	international prognostic index
IRC	independent review committee
ITT	intent-to-treat
IxRS	interactive voice/web-based response system
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation or Term	Description
MDD	Minimal Detectable Difference
NALT	non-protocol specified anti-lymphoma therapy
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events,
ORR	objective response rate
OS	overall survival
PET	positron emission tomography
PFS	progression-free survival
Pola-R-GemOx	polatuzumab vedotin plus rituximab plus gemcitabine plus oxaliplatin
PK	pharmacokinetic
PN	peripheral neuropathy
PR	partial response
PRO	patient-reported outcomes
RCT	randomized controlled trial
R-GemOx	rituximab plus gemcitabine plus oxaliplatin
RMST	restricted mean survival time
SAE	serious adverse events
SAP	Statistical Analysis Plan
SC	Steering Committee
SCT	stem cell transplantation

1. INTRODUCTION

This Statistical Analysis Plan (SAP) provides details of the planned analyses and statistical methods for Study MO40598 (POLARGO), a Phase III, multicenter, open-label, randomized controlled trial designed to evaluate safety and efficacy of polatuzumab vedotin in combination with rituximab, gemcitabine and oxaliplatin (Pola-R-GemOx) compared to rituximab, gemcitabine and oxaliplatin (R-GemOx) in patients with relapsed or refractory diffuse large B-Cell lymphoma (DLBCL) who have received one or more ($\geq$1) prior lines of therapy.

The analysis plan and the endpoints specified in this document supersede the analysis plan described in the study protocol for the purposes of a regulatory filing.

Changes to the protocol-planned analyses are described in Section 4.8.

1.1 OBJECTIVES AND ENDPOINT AND ESTIMANDS

This study will evaluate the safety and efficacy of Pola-R-GemOx compared R-GemOx in patients with relapsed or refractory DLBCL.

The study will be conducted in two stages:

1. An initial safety run-in stage assessing Pola-R-GemOx; and
2. A randomized-controlled trial (RCT) stage comparing Pola-R-GemOx versus R-GemOx.

Specific objectives and corresponding endpoints for the study are outlined in Table 1. A subset of objectives and endpoints is expressed using the estimand framework in Section 1.2.1.

Table 1 Objectives and Corresponding Endpoints: Safety Run-In (Stage 1)

Primary Safety Objective	Corresponding Endpoints
• To evaluate the safety and tolerability of Pola-R-GemOx as a combination therapy	• Incidence, nature and severity of physical findings and AEs, with a specific focus on PN, according to the NCI CTCAE v5.0
Secondary Safety Objectives	**Corresponding Endpoints**
• To evaluate the safety and tolerability of Pola-R-GemOx as a combination therapy and to assess the immunogenicity of polatuzumab vedotin	• Incidence and assessment of PN, as measured by FACT/GOG-NTX-12 • Tolerability, as measured by dose interruptions, dose reductions and dose intensity • Prevalence of ADAs at baseline and incidence of ADAs during the study
Secondary Efficacy Objective	**Corresponding Endpoints**
• To evaluate the efficacy of Pola-R-GemOx	• CRR, defined as the proportion of patients who achieve complete metabolic response based on PET-CT, according to Lugano 2014 response criteria (Appendix 1), at the end of treatment as determined by the investigator • ORR, defined as the proportion of patients who achieve complete or partial metabolic responses, according to Lugano 2014 response criteria (Appendix 1), at end of treatment as determined by the investigator • BOR, defined as the best response while on study, according to Lugano 2014 response criteria (Appendix 1), as determined by the investigator • PFS, defined as the time from enrollment to the first occurrence of disease progression as determined by the investigator according to Lugano 2014 response criteria (Appendix 1) or death from any cause • OS, defined as time from enrollment to death from any cause

Secondary Efficacy Objective	Corresponding Endpoints
	• EFS_{eff}, defined as time from enrollment to the earliest occurrence of the below cases: – Disease progression or relapse – Death due to any cause • Initiation of any NALT
Exploratory Pharmacokinetic Objective	**Corresponding. Endpoints**
• To further evaluate the PK of polatuzumab vedotin	• PK of polatuzumab vedotin in combination with R-GemOx in patients with relapsed or refractory DLBCL

ADA = anti-drug antibody; AE = adverse event; BOR = best overall response; CR = complete response; CRR = complete response rate; DLBCL = diffuse large B-cell lymphoma; EFS_{eff} = event-free survival; FACT/GOG-NTX-12 = Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity 12-Item Scale; NALT = non-protocol specified anti-lymphoma therapy; NCI CTCAE v5.0 = National Cancer Institute Common Terminology Criteria for Adverse Events, Version 5.0; ORR = objective response rate; OS = overall survival; PET-CT = positron emission tomography-computed tomography; PFS = progression-free survival; PK = pharmacokinetics; PN = peripheral neuropathy; Pola-R-GemOx = polatuzumab vedotin, rituximab, gemcitabine, and oxaliplatin; R-GemOx = rituximab plus gemcitabine plus oxaliplatin.

Table 2 Objectives and Corresponding Endpoints: Randomized-Controlled Trial (Stage 2)

Primary Efficacy Objective	Corresponding Endpoint
• To evaluate the efficacy of Pola-R-GemOx compared with R-GemOx alone	• OS, defined as time from randomization to death from any cause
Secondary Efficacy Objective	**Corresponding Endpoints**
• To evaluate the efficacy of Pola-R-GemOx compared with R-GemOx alone	Key secondary endpoints included in the hierarchical testing procedure: • PFS, defined as the time from randomization to the first occurrence of disease progression (based on either response: including PET-CT data or not including any PET data), as determined by the investigator according to Lugano 2014 response criteria (Appendix 1), or death from any cause • CRR, defined as the proportion of patients who achieve complete metabolic response (based on response including PET-CT data), according to Lugano 2014 response criteria (Appendix 1), at the end of treatment as determined by an IRC • ORR, defined as the proportion of patients who achieve complete or partial metabolic responses (based on response including PET-CT data), according to Lugano 2014 response criteria (Appendix 1), at the end of treatment as determined by an IRC Secondary endpoints that will not be adjusted for testing multiplicity: • BOR, defined as the best response while on study, according to Lugano 2014 response criteria (Appendix 1), as determined by the investigator • CRR, defined as the proportion of patients who achieve complete metabolic response (based on response including PET-CT data), according to Lugano 2014 response criteria (Appendix 1), at the end of treatment as determined by the investigator

Secondary Efficacy Objective	Corresponding Endpoints
	• ORR, defined as the proportion of patients who achieve complete or partial metabolic responses (based on response including PET-CT data), according to Lugano 2014 response criteria (Appendix 1), at the end of treatment as determined by the investigator • DOR, defined as the time from the first occurrence of a documented objective response (based on response including PET-CT or CT data) to disease progression (based on either response: including PET-CT data or not including any PET data), as determined by the investigator according to Lugano 2014 response criteria (Appendix 1), or death from any cause, whichever occurs first • Event-free survival (EFS_{eff}), defined as time from randomization to the earliest occurrence of the below cases: – Disease progression or relapse (based on either response as determined by the investigator: including PET-CT data or not including any PET data) – Death due to any cause – Initiation of any NALT
Secondary PRO Objective	**Corresponding Endpoint**
• To evaluate impact of treatment and disease on aspects of health-related quality of life	• Time to deterioration in physical functioning and fatigue as measured by the EORTC QLQ-C30 • Time to deterioration in lymphoma symptoms as measured by the FACT-Lym lymphoma subscale • Descriptive summary statistics and the change from baseline for: – EORTC QLQ-C30 physical functioning and fatigue subscales – FACT-Lym subscale

	• Clinically meaningful improvement in: – EORTC QLQ-C30 physical functioning and fatigue subscales – FACT-Lym subscale
Primary Safety Objective	**Corresponding Endpoints**
• To evaluate the safety and tolerability of Pola-R-GemOx compared with R-GemOx	• Incidence, nature, and severity of AEs (including PN) according to NCI CTCAE v5.0 and physical findings
Secondary Safety Objective	**Corresponding Endpoints**
• To evaluate the safety and tolerability of Pola-R-GemOx compared with R-GemOx and to assess the immunogenicity of polatuzumab vedotin	• Tolerability, as assessed by dose interruptions, dose reductions and dose intensity • Prevalence of ADAs at baseline and incidence of ADAs during the study
Exploratory Biomarker Objectives	**Corresponding Endpoints**
• To identify biomarkers that: – Are prognostic of response to polatuzumab vedotin (i.e., predictive biomarkers) – Are associated with progression to a more severe disease (i.e., prognostic biomarkers) – Can provide evidence of polatuzumab vedotin activity, or can increase the knowledge and understanding of disease biology • To explore MRD as a prognostic marker in R/R DLBCL	• Associations between efficacy endpoints, including OS, PFS and CR rate, and exploratory biomarkers, which may include but are not limited to histological and molecular prognostic markers and profiles of circulating immune cells
Exploratory Pharmacokinetic Objective	**Corresponding Endpoints**
• To further evaluate the PK of polatuzumab vedotin	• PK of polatuzumab vedotin in combination with R-GemOx in patients with relapsed or refractory DLBCL

Exploratory PRO Objectives	Corresponding Endpoints
• To evaluate impact of treatment and disease on aspects of health-related quality of life	• Descriptive summary statistics and the change from baseline for: – All other scales for the EORTC QLQ-C30 – All other scales for the FACT-Lym – FACT/GOG-NTX-12 – EQ-5D-5L

ADA = anti-drug antibody; AE = adverse event; BOR = best overall response; CR = complete response; CRR = complete response rate; DLBCL = diffuse large B-cell lymphoma; DOR = duration of response; EFS_{eff} = event-free survival; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire, Core 30; EQ-5D-5L = EuroQol 5-Dimension Questionnaire, 5-Level Version; FACT/GOG-NTX-12 = Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity 12-Item Scale; FACT-Lym = Functional Assessment of Cancer Therapy-Lymphoma; IRC = Independent Review Committee; MRD = minimal residual disease; NALT = non-protocol specified anti-lymphoma therapy; NCI CTCAE v5.0 = National Cancer Institute Common Terminology Criteria for Adverse Events, Version 5.0; ORR = objective response rate; OS = overall survival; PET-CT = positron emission tomography-computed tomography; PFS = progression-free survival; PK = pharmacokinetics; PN = peripheral neuropathy; Pola-R-GemOx = polatuzumab vedotin, rituximab, gemcitabine, and oxaliplatin; R-GemOx = rituximab plus gemcitabine plus oxaliplatin; R/R = relapsed/refractory.

1.1.1 Expression of Objectives and Endpoints Using the Estimand Framework

For the RCT stage of the trial, the primary study objective and corresponding endpoint, as well as the secondary efficacy objective and the subset of corresponding key secondary efficacy endpoints, are expressed using the estimand framework in Table 3 in accordance with the International Conference for Harmonization (ICH) E9 (R1) statistical principles for clinical trials (ICH 2020).

Table 3 Primary and Key Secondary Objectives and Corresponding Estimands

Primary Objective	Estimand Definition
• To evaluate the efficacy of Pola-R-GemOx compared with R-GemOx alone	• Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population) • Variable: OS (as defined in Table 2) • Treatment: – Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles – Control arm: R-GemOx every 21 days for up to 8 cycles • Intercurrent events and handling strategies: – Early discontinuation from study treatment: treatment policy strategy – Stem cell transplant, start of CAR-T, or start of bispecific therapies at any time: treatment policy strategy – Switch to polatuzumab vedotin-containing therapy for patients randomized in R-GemOx: treatment policy strategy – Start of any other non-protocol anti-lymphoma therapy at any time: treatment policy strategy – Death due to COVID-19: composite strategy • Population-level summary: hazard ratio for OS
Secondary Objective	**Estimand Definition**
• To evaluate the efficacy of Pola-R-GemOx compared with R-GemOx alone	Note: All response assessments will be based on the 2014 Lugano Response Criteria. Estimand for the key secondary endpoints included in the testing strategy will be as follows: Estimand for PFS: • Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population)

Secondary Objective	Estimand Definition
	• Variable: PFS, as per investigator (as defined in Table 2) • Treatment: – Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles – Control arm: R-GemOx every 21 days for up to 8 cycles • Intercurrent events and handling strategies: – Early discontinuation from study treatment: treatment policy strategy – Start of non-protocol anti-lymphoma therapy prior to disease progression: hypothetical strategy – Missing two or more consecutive tumor response assessments: hypothetical strategy • Population-level summary: hazard ratio for PFS Estimand for CRR at end of treatment: • Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population) • Variable: CRR (as defined in Table 2) • Treatment: – Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles – Control arm: R-GemOx every 21 days for up to 8 cycles • Intercurrent events and handling strategies: – Missing tumor assessment due to early study withdrawal: composite strategy – Death while still on treatment: composite strategy • Population-level summary: difference in proportion for CRR

Secondary Objective	Estimand Definition
	Estimand for ORR at end of treatment: • Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population) • Variable: ORR (as defined in Table 2) • Treatment: – Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles – Control arm: R-GemOx every 21 days for up to 8 cycles • Intercurrent events and handling strategies: – Missing tumor assessment due to early study withdrawal: composite strategy – Death while still on treatment: composite strategy • Population-level summary: difference in proportion for ORR

CAR-T = chimeric antigen receptor T cell; COVID-19 = coronavirus disease 2019; CRR = complete response rate; DLBCL = diffuse large B-cell lymphoma; ITT = intent-to-treat; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; Pola-R-GemOx = polatuzumab vedotin, rituximab, gemcitabine, and oxaliplatin; R-GemOx = rituximab plus gemcitabine plus oxaliplatin.

1.2 STUDY DESIGN

Study MO40598 is a Phase III, multicenter, open-label RCT in patients with relapsed or refractory DLBCL. The study will consist of a screening period, a treatment period, and a post-treatment period.

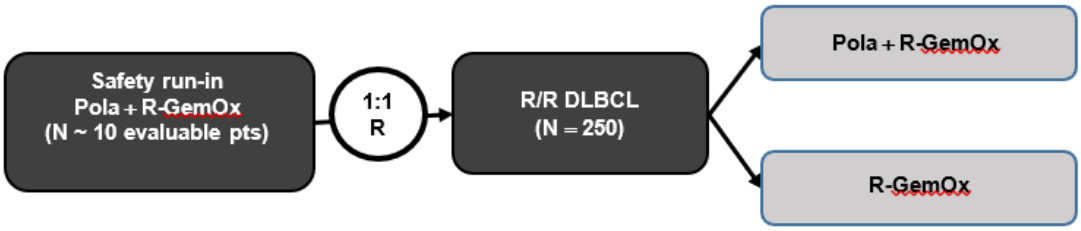
The Treatment Period will occur in two stages (Figure 1):

- A safety run-in, where approximately 13 patients will receive experimental study treatment with Pola-R-GemOx in order to reach at least 10 evaluable patients. The patients will be enrolled subsequently in 3 cohorts (Figure 2).
- A RCT, where approximately 250 patients will be randomly assigned in a 1:1 ratio to receive either Pola-R-GemOx (experimental arm) or R-GemOx (control arm). Randomization will be stratified by three factors: number of previous lines of systemic treatment (1 vs. ≥2), outcome of last systemic treatment (relapsed vs. refractory) and age (≤70 years vs. >70 years).

Further details on safety stopping rules for both stages are provided in Section 4.7.

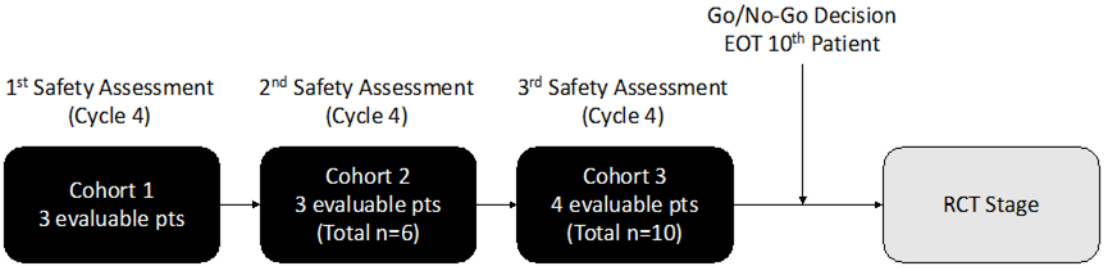
In both stages of the treatment period, patients will receive up to 8 cycles of Pola-R-GemOx or 8 cycles of R-GemOx, each administered on 21-day cycles.

Figure 1 Study Design



DLBCL = relapsed or refractory diffuse large B-cell lymphoma; Pola = polatuzumab vedotin; pts = patients; R = randomize; R-GemOx = rituximab + gemcitabine + oxaliplatin; R/R = relapsed/refractory

Figure 2 Safety Run-in Schema (Stage 1)



EOT = end of treatment; evaluable = treated for at least 4 cycles of therapy or discontinued due to ≥ Grade 3 peripheral neuropathy; pts = patients; RCT = randomized controlled trial.

1.2.1 Treatment Assignment and Blinding

Study MO40598 is an open-label trial, including two stages:

Stage 1: Safety Run-In:

During the safety run-in (Figure 2), approximately 13 patients will be treated with Pola-R-GemOx.

- Accrual of these patients will be staggered across three cohorts:
  - Cohort 1: 3 evaluable patients
  - Cohort 2: 3 evaluable patients
  - Cohort 3: 4 evaluable patients. Approximately 7 patients will be recruited in Cohort 3, with the aim of having at least 4 evaluable.
- Within each cohort, safety will be evaluated when the number of intended evaluable patients complete 4 cycles of treatment, with a focus on acute PN toxicities. Once all safety evaluations have been conducted within a cohort and further subject accrual is cleared, the next cohort will open to recruitment.

- Once at least 10 evaluable patients in the safety run-in have received the last dose of Pola-R-GemOx, the safety and tolerability of the Pola-R-GemOx regimen will be assessed by the Internal Monitoring Committee [IMC], and a decision will be made whether to continue into the RCT stage of the study. The Steering Committee (SC) will be available for consultation during this time.
- Only a limited number of sites (approximately 18 centers) will be open for accrual into the safety run-in.

Stage 2: Randomized-Control Trial

- If Pola-R-GemOx combination therapy is deemed tolerable in Stage 1, newly enrolled patients will be randomized in a 1:1 ratio to receive either Pola-R-GemOx or R-GemOx.
- Randomization will be performed by interactive voice/web-based response system (IxRS) using stratified permuted blocks. The randomization will be stratified on the following factors:
  - Number of previous lines of systemic therapy for DLBCL (1 vs. $\geq$2)
  - Outcome of last systemic therapy (relapsed vs. refractory)
  - Age ($\leq$70 years vs. $>$70 years)

1.2.2 Independent Review Facility

An Independent Review Committee (IRC) will be used to evaluate the study endpoints of complete response rate (CRR) and objective response rate (ORR) in a blinded manner. IRC membership and procedures will be detailed in an IRC Charter.

1.2.3 Data Monitoring

1.2.3.1 Internal Monitoring Committee

An IMC will be established to monitor patient safety during the safety run-in (Stage 1) and provide recommendations on whether the next cohorts or RCT should open. Within each cohort in safety run-in stage, safety will be evaluated at the end of the fourth cycle in all subjects, with a focus on acute peripheral neuropathy toxicities. Once all safety evaluations have been conducted within a cohort and further subject accrual is cleared, the next cohort will open for recruitment. This procedure will continue until all three cohorts have been fully recruited. Safety evaluations will be made by the IMC. The SC will be available for consultation during this time.

After the tenth subject has completed his or her treatment regimen, the IMC will assess the current safety and tolerability profiles of the Pola-R-GemOx regimen and provide a recommendation whether to continue into the RCT stage, which is scheduled to occur at the end of treatment for the tenth subject. The SC will be available for consultation during this time. The Sponsor will evaluate this recommendation, as well as any other subsequent safety signals, and make the final decision whether to continue into the RCT.

1.2.3.2 Independent Data Monitoring Committee

Although this is an open-label study, the Sponsor and the study team performing the primary analysis will not have access to aggregated statistical outputs by treatment arm as well as to anti-drug antibody (ADA) and pharmacokinetic (PK) data, during the RCT (Stage 2). An Independent Data Monitoring Committee (iDMC) will review the unblinded safety data periodically during the RCT stage. Frequency of interim safety analysis as well as specific study-wide stopping rules are described in Section 4.7. Any changes in study conduct will be communicated in a timely manner to the investigators for notification of the Institutional Review Boards and/or Ethics Committees (IRB/IEC). Further details will be given in the iDMC Charter.

2. STATISTICAL HYPOTHESES AND SAMPLE SIZE DETERMINATION

2.1 STATISTICAL HYPOTHESES

The primary objective of the randomized part of this study is to evaluate the efficacy of Pola-R-GemOx versus R-GemOx in patients with relapsed or refractory DLBCL as measured by overall survival (OS).

The null (H0) and alternative (H1) hypotheses regarding OS can be phrased in terms of the OS survival distribution function (SDF) in Pola-R-GemOx and SDF in R-GemOx, respectively:

H0: $SDF_{Pola\text{-}R\text{-}GemOx} = SDF_{R\text{-}GemOx}$ versus H1: $SDF_{Pola\text{-}R\text{-}GemOx} \neq SDF_{R\text{-}GemOx}$

Hypothesis tests will be two-sided, unless otherwise indicated. The type I error (α) for this study is 0.05 (two-sided).

2.2 SAMPLE SIZE DETERMINATION

Assuming a median OS of [redacted] months in the R-GemOx arm and a randomization ratio of 1:1, [redacted] events are required to detect a between-group difference of [redacted] months in the median OS (hazard ratio [HR]=[redacted], Minimal Detectable Difference [MDD]: [redacted]) with [redacted] power and a 2-sided α of 0.05. Based on the above statistical assumptions and anticipating a recruitment period of approximately 19 months and a follow-up of 12 months after the last patient was randomized, a total of approximately 250 patients will be randomized taking into account an estimated drop-out rate of [redacted].

In addition, approximately 10 patients will be enrolled in the safety run-in stage.

3. ANALYSIS SETS

The participant analysis sets for the purposes of analyses are defined in Table 4.

Table 4 Participant Analysis Sets

Participant Analysis Set	Description
Enrolled population	All enrolled participants (regardless of the stage). Participants will be included in the analyses according to the treatment assigned at randomization for the randomized part and to Pola-R-GemOx for patients in the safety run-in stage, whether or not the assigned study treatment was received.
Intent-To-Treat (ITT) population	All randomized participants. Participants will be included in the analyses according to the treatment assigned at randomization, whether or not the assigned study treatment was received.
Safety run-in population	All participants who received any amount of any study drug during safety run-in stage.
Safety-evaluable population	All participants who received any amount of any study drug (regardless of the stage). Participants will be included in the analyses according to the treatment actually received, and all participants who received any dose of polatuzumab vedotin will be included in the Pola-R-GemOx arm.
Pharmacokinetic-evaluable population	All participants who have received at least one dose of study drug and have at least one post-dose concentration result.
Immunogenicity-evaluable population	All participants who received at least one dose of polatuzumab vedotin with at least one evaluable post-baseline ADA sample.

ADA = anti-drug antibody; Pola-R-GemOx = polatuzumab vedotin plus rituximab plus gemcitabine plus oxaliplatin.

4. STATISTICAL ANALYSES

4.1 GENERAL CONSIDERATIONS

All analyses related to baseline characteristics and study conduct will be performed on the enrolled population. Patients will be analyzed according to the treatment they were assigned at randomization for the randomized part and to Pola-R-GemOx for patients in the safety run-in stage.

All efficacy analyses related to safety run-in stage will be performed on the safety run-in population.

All efficacy analyses (including patient-reported outcomes [PROs] related to European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire, Core 30 [EORTC QLQ-C30] and Functional Assessment of Cancer Therapy-Lymphoma [FACT-Lym]) for the RCT stage will be performed on the intent-to-treat (ITT) population, unless otherwise specified. Patients will be analyzed according to the treatment assigned at randomization.

All safety analyses (regardless of the stage), as well as PROs analyses related to Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity Subscale (FACT/GOG-NTX) will be performed on the safety-evaluable population, unless otherwise specified. Patients will be analyzed according to the treatment actually received. Specifically, for the RCT stage, a patient who received any dose of polatuzumab vedotin will be included in the Pola-R-GemOx arm, and all other treated patients will be included in the R-GemOx arm, regardless of the initial treatment assignment by the IxRS.

The baseline value of any non-efficacy variable or efficacy variable related to safety run-in stage will be defined as the last available value recorded on or prior to the first administration of any study medication. The baseline value of efficacy variable related to tumor assessment during RCT stage will be defined as the last available value recorded prior to randomization. Patients with missing baseline assessments will not be imputed.

Continuous variables will be summarized using means, standard deviations (SDs), medians, ranges and inter-quartile ranges. Categorical variables will be summarized with frequency counts and percentages. Data will be presented by treatment arm.

Throughout the statistical analysis, two-sided tests will be performed at a significance level of 5%, unless otherwise stated. To control the overall type I error rate at a two-sided 0.05 level of significance, a hierarchical testing procedure will be used to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints.

4.2 PRIMARY ENDPOINT/ESTIMANDS ANALYSIS

4.2.1 Definition of Primary Endpoint/Estimand

Overall survival (OS) is the primary endpoint for RCT stage and will be analyzed in the ITT population.

OS is defined as the time from randomization to death due to any cause. Patients who are not reported as having died at the time of analysis will be censored at the date when they were last known to be alive. Patients who do not have post-baseline information will be censored at the date of randomization.

The primary estimand is as defined in Table 3:

- Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population)
- Variable: OS
- Treatment:
  - Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles
  - Control arm: R-GemOx every 21 days for up to 8 cycles

- Intercurrent events and handling strategies:
  - Early discontinuation from study treatment: treatment policy strategy

    Stem cell transplant, start of chimeric antigen receptor T cell (CAR-T), or start of bispecific therapies at any time: treatment policy strategy
  - Switch to polatuzumab vedotin for patients randomized in R-GemOx: treatment policy strategy
  - Start of any other non-protocol anti-lymphoma therapy at any time: treatment policy strategy
  - Death due to coronavirus disease 2019 (COVID-19): composite strategy
- Population-level summary: hazard ratio for OS

Censoring rules for OS are summarized in Table 5.

Table 5 Censoring Rules Analysis for OS (Primary Endpoint)

Situation	Date of OS event or censoring	Outcome
Death	Death date	Event
No death and no post-baseline survival information available	Randomization date	Censored
No death	Last known alive date before data cutoff[1]	Censored

[1]Last known alive date is defined as the last date the patient has documented clinical data to show him/her alive. Scenarios considered in this definition may include last survival follow-up date with patient status of "alive", date of last tumor assessment with a valid response (i.e., not "unevaluable" or "not done"), 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.

4.2.2 Main Analytical Approach for Primary Endpoint

Treatment comparisons will be based on the stratified log-rank test. The stratification factors will be the randomization stratification factors: number of previous lines of systemic therapy for DLBCL (1 vs. ≥2), outcome of last systemic therapy (relapsed vs. refractory) and age (≤70 years vs. >70 years) and will be as entered in the IxRS.

The hazard ratio (HR) will be estimated using a stratified Cox regression model with the same stratification variables used for the stratified log-rank test, and the 95% CI for the HR will be provided.

Kaplan-Meier methodology will be used to estimate median OS for each treatment arm and to construct survival curves for each treatment arm. The Brookmeyer-Crowley methodology will be used to construct the 95% CI for the median OS for each treatment arm (Brookmeyer et al. 1982).

4.2.3 Sensitivity Analyses

4.2.3.1 Unstratified Analysis

To assess the impact of stratification, results from an unstratified log-rank test and the unstratified HR will also be provided.

4.2.3.2 Stratifications Errors

To assess the impact of stratification errors, the analysis may be repeated by using the stratification factors as entered in the electronic Case Report Form (eCRF).

4.2.4 Supplementary Analyses

4.2.4.1 Subgroup Analyses for Primary Endpoint

The generalizability of OS results when comparing Pola-R-GemOx to R-GemOx will be investigated by estimating the treatment effect in subgroups based on key baseline demographics (e.g., age, sex, race, ethnicity and geographic region) and disease characteristics (e.g., number of previous lines of systemic therapy for DLBCL, Ann Arbor stage at study entry, international prognostic index [IPI] score at study entry and Eastern Cooperative Oncology Group [ECOG] Performance Status, prior autologous stem cell transplant, relapsed/refractory to last line of therapy, primary refractory disease, bulky disease, extra nodal involvement at study entry, subtype of DLBCL at study entry, double expresser (MYC and B-cell leukemia/lymphoma 2 protein [BCL2] overexpression)). Summaries of OS by these subgroups will be provided in forest plots including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of the median provided separately for each level of the subgroups.

4.2.4.2 Other Supplementary Analyses for Primary Endpoint

4.2.4.2.1 Impact of Stem Cell Transplant, CAR-T Therapy and Bispecific Therapy

Patients may receive stem cell transplantation (SCT) or CAR-T therapy or bispecific therapy.

Supplementary analyses will be performed for OS using different methods to investigate how the OS results would have looked if those therapies were not available (hypothetical strategy):

- The censoring method. Patients who undergo a transplant or received CAR-T/bispecific therapy will be censored at the time of transplant/start of therapy.
- The discount method. More specifically, the time interval from when patients received transplant/CAR-T or bispecific therapy until the event or censoring time will be discounted at 10%, 30%, and 50% for both arms.
- Inverse probability of censoring weighting (IPCW) (Robins and Finkelstein 2000). The IPCW approach is appropriate for analyzing data after censoring for non-protocol specified anti-lymphoma therapy (NALT) because, by weighting remaining observations using baseline and time-dependent covariates, it can help correct to some extent the bias caused by non-independent censoring. This approach has also been recommended in Manitz et al. 2022 as a potential method

for a hypothetical strategy for any treatment switching. The following listed baseline and time-dependent covariates may be used in the IPCW approach for estimating the censoring weights:

- Baseline covariates:
  - Age at enrollment
  - Ann Arbor Stage
  - ECOG performance status
  - Bulky Disease (≥7.5cm)
  - Refractory status
  - Number of prior lines of therapy for DLBCL
- Time-dependent covariates:
  - Overall response assessment at each visit.

4.2.4.2.2 Impact of Switch to Polatuzumab Vedotin

Patients on R-GemOx may receive a polatuzumab vedotin-containing regimen as further therapy.

Supplementary analyses will be performed for OS using different methods to investigate how the OS results would have looked if this switch to a polatuzumab vedotin-containing regimen for patients randomized in R-GemOx was not available (hypothetical strategy):

- The censoring method. Patients randomized on R-GemOx who received a polatuzumab vedotin-containing regimen as further therapy will be censored at the start of therapy.
- The discount method. More specifically, the time interval from when patients randomized on R-GemOx received a polatuzumab vedotin-containing regimen (as further therapy) until the event or censoring time will be discounted at 10%, 30%, and 50%.
- The rank-preserving structural failure time (RPSFT) method. This method was introduced by Robins and Tsiatis, 1991. It estimates OS time measured from the time of a polatuzumab vedotin-containing regimen (as further therapy) by applying an estimate of the benefit of the polatuzumab vedotin. The adjusted OS time (sum of time to a polatuzumab vedotin-containing regimen (as further therapy) and the estimated survival time after a polatuzumab vedotin-containing regimen (as further therapy) will then be analyzed together with the OS times of the patients who did not receive a polatuzumab vedotin-containing regimen (as further therapy) by using the same methodology as for the primary analysis of OS, provided there is a sufficient number of patients randomized on R-GemOx with a polatuzumab vedotin-containing regimen as further therapy.

4.2.4.2.3 Impact of COVID-19

To assess impact of coronavirus COVID-19 on OS, OS analysis may be performed by considering death due to COVID-19 as hypothetical strategy. In this analysis, patients who die due to COVID-19 will be censored to date of death.

4.2.4.2.4 Impact of Proportional Hazards Assumption

The restricted mean survival time (RMST) (Royston and Parmar 2011) method will be used as an additional supplementary analysis to account for the possible non-proportional hazards effect. It will measure the difference in the average event-free survival time between treatment and control arm from the randomization through a prespecified timepoint. Specifically, unstratified non-parametric Kaplan–Meier estimate of restricted mean survival time (RMST) by arm as well as the difference of RMST between arms will be evaluated. The 95% confidence intervals (by Greenwood method) will be provided for descriptive purpose. The RMST of OS will be estimated at Month 12 and 18.

4.3 SECONDARY ENDPOINTS/ESTIMANDS ANALYSES

To control the overall type I error rate at a two-sided 0.05 level of significance for the RCT stage, a hierarchical testing procedure will be used to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints. Key secondary endpoints will be tested in the following order:

- PFS as determined by investigator (main estimand)
- CRR at end of treatment (based on response including PET-CT data) as determined by an IRC (main estimand)
- ORR at end of treatment (based on response including PET-CT data) as determined by an IRC (main estimand)

A given hypothesis in the bulleted list above will only be rejected if its p-value is less than 0.05, and all previous hypotheses have been rejected at a 2-sided 0.05 level of significance.

No multiplicity adjustment will be performed for the testing of other endpoints; all other p-values and 95% confidence intervals (CIs) will be given in an exploratory manner.

4.3.1 Key Secondary Endpoints/Estimands

4.3.1.1 Progression-Free Survival by Investigator

PFS is defined as the time from randomization to the first occurrence of disease progression (based on either response: including PET-CT data or not including any PET data), as determined by the investigator according to Lugano 2014 response criteria (Appendix 1) or death due to any cause, whichever occurs first. Event and censoring rules for PFS are summarized in Table 6. The main estimand for PFS is defined as indicated in Table 3.

- Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population)
- Variable: PFS, as per investigator
- Treatment:
  - Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles
  - Control arm: R-GemOx every 21 days for up to 8 cycles
- Intercurrent events and handling strategies:
  - Early discontinuation from study treatment: treatment policy strategy
  - Start of any non-protocol anti-lymphoma therapy prior to disease progression: hypothetical strategy
  - Missing two or more consecutive tumor response assessments: hypothetical strategy
- Population-level summary: hazard ratio for PFS

PFS will be analyzed in the ITT population. The methodology (as described in Section 4.2.2) used for OS will be applied for PFS.

Table 6 Event and Censoring Rules for PFS (Censoring for NALT/ Two or More Missing Tumor Assessments)

Situation	Date of PFS event or censoring	Outcome
No baseline disease assessments	Date of Randomization	Censored
Alive and without disease progression documentation	Date of last adequate[1] disease assessment	Censored
Death or disease progression after initiation of NALT	Date of last adequate[1] disease assessment before NALT. If no adequate post baseline assessment is available, then date of randomization.	Censored
Initiation of NALT in absence of death or disease progression	Date of last adequate[1] disease assessment before NALT. If no adequate post baseline assessment is available, then date of randomization.	Censored
Death or disease progression after two or more consecutive missed response assessments	Date of last adequate[1] disease assessment before the missed assessments. If no adequate post baseline assessment is available, then date of randomization.	Censored

Situation	Date of PFS event or censoring	Outcome
Two or more consecutive missed response assessments in absence of death of disease progression	Date of last adequate[1] disease assessment before the missed assessments. If no adequate post baseline assessment is available, then date of randomization.	Censored
Death or disease progression after treatment discontinuation due reasons other than progressive disease or death	Date of death or first disease assessment showing disease progression, whichever occurs first	Event
Death or disease progression before or without initiation of NALT	Date of death or first disease assessment showing disease progression, whichever occurs first	Event

[1]To be considered adequate, the results of a response assessment should not be "unevaluable" or "not done".

NALT = non-protocol specified anti-lymphoma therapy; PFS = progression-free survival.

4.3.1.1.1 Sensitivity Analysis for PFS

If there is a substantial amount in censoring due to NALT, a sensitivity analysis using a Cox model with inverse probability of censoring weighting (IPCW) (Robins and Finkelstein 2000) may be performed. The IPCW approach is appropriate for analyzing data after censoring for NALT because, by weighting remaining observations using baseline and time-dependent covariates, it can help correct to some extent the bias caused by non-independent censoring. This approach has also been recommended in Manitz et al. 2022 as a potential method for a hypothetical strategy for any treatment switching.

The following listed baseline and time-dependent covariates may be used in the IPCW approach for estimating the censoring weights:

- Baseline covariates:
  - Age at enrollment
  - Ann Arbor Stage
  - ECOG performance status
  - Bulky Disease (≥7.5cm)
  - Refractory status
  - Number of prior lines of therapy for DLBCL
- Time-dependent covariates:
  - Overall response assessment at each visit.

4.3.1.1.2 Supplementary Analyses for PFS

A supplementary analysis for PFS without censoring for initiation of NALT or two or more missing tumor assessments (treatment policy strategy) will be performed.

The corresponding estimand is defined as:

- Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population)
- Variable: PFS, as per investigator
- Treatment:
  - Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles
  - Control arm: R-GemOx every 21 days for up to 8 cycles
- Intercurrent events and handling strategies:
  - Early discontinuation from study treatment: treatment policy strategy
  - Start of any non-protocol anti-lymphoma therapy prior to disease progression: treatment policy strategy
  - Missing two or more consecutive tumor response assessments: treatment policy strategy
- Population-level summary: hazard ratio for PFS

Event and censoring rules for PFS (treatment policy strategy) are summarized in Table 7.

Table 7 Event and Censoring Rules for PFS (Treatment Policy Strategy)

Situation	Date of PFS event or censoring	Outcome
No baseline disease assessments	Date of Randomization	Censored
Alive and without disease progression documentation	Date of last adequate[1] disease assessment	Censored
Death or disease progression	Date of death or first disease assessment showing disease progression, whichever occurs first	Event

PFS = progression-free survival

[1]To be considered adequate, the results of a response assessment should not be "unevaluable" or "not done".

The impact of missing scheduled tumor assessments on PFS (as described in Table 7) will be assessed by performing a sensitivity analysis based on the interval censoring analysis methods. For each patient, the left and the right boundaries of the interval will be derived based on the following rules:

Situations	Left Boundary	Right Boundary
Patients who had disease progression prior to death	The date of the last assessment that showed a progression-free* status	The date of the first assessment that showed disease progression
Patients who died without disease progression	The date of the last assessment that showed a progression-free* status	Death date
Patients who did not die nor had disease progression	The date of the last assessment that showed a progression-free* status	Not applicable (Missing)

*For patients who did not have any post-baseline assessment with progression-free status, the left boundary is the date of randomization.

The PFS survival curves will be estimated using the nonparametric maximum likelihood estimate (NPMLE, Turnbull 1976) for each treatment arm. The median PFS of each treatment arm will be reported and its 95% confidence interval will be constructed based on the Brookmeyer-Crowley method (Brookmeyer et al. 1982).

Hypothesis testing will be performed based on the stratified log-rank test proposed by Sun (Sun 1996) to compare the PFS in the treatment arms. The treatment effect will be estimated using a stratified proportional hazard regression model (Finkelstein 1986) with a parametric assumption of piecewise Exponential distribution for the baseline hazard function (Friedman 1982, Royston and Parmar 2002). Results from an unstratified analysis will also be provided.

The generalizability of PFS results as per main estimand (Table 6) when comparing Pola-R-GemOx to R-GemOx will be investigated by estimating the treatment effect in subgroups (as defined in Section 4.2.4.1). Summaries of PFS censoring for NALT/ two or more missing tumor assessments by these subgroups will be provided in forest plots including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of the median provided separately for each level of the subgroups.

4.3.1.2 Complete Response Rate at End of Treatment by IRC

CRR is defined as the proportion of patients who had a complete metabolic response (based on response including PET-CT data) according to Lugano 2014 response criteria at the end of treatment, as determined by IRC. Patients not meeting these criteria, including patients without any tumor assessment at end of treatment, will be considered non-responders.

The estimand for CRR is defined as indicated in Table 3. CRR will be analyzed using the ITT population.

An estimate of CRR will be calculated for each treatment arm, and its 95% CI will be calculated using the Clopper-Pearson method. The difference in CRR between treatment arms will be calculated, and its 95% CI will be calculated using the normal approximation to the binomial distribution. CRR will be compared between treatment arms using the stratified Cochran-Mantel-Haenszel test. The stratification factors will be the same as those described for the analysis of the primary endpoint OS.

A supplementary analysis for CRR at end of treatment may be performed in which patients with missing response including PET-CT data at end of treatment will be imputed using the "last observation carried forward" method.

4.3.1.3 Objective Response Rate at End of Treatment by IRC

An objective response is defined as either a complete or partial metabolic response (based on response including PET-CT data) according to Lugano 2014 response criteria (Appendix 1) at end of treatment, as determined by IRC. Patients not meeting these criteria, including patients without any tumor assessment at end of treatment will be considered non-responders.

ORR is defined as the proportion of patients who had an objective response.

The estimand for ORR is defined as indicated in Table 3. ORR will be analyzed using the ITT population.

An estimate of ORR will be calculated for each treatment arm, and its 95% CI will be calculated using the Clopper-Pearson method. The difference in ORR between treatment arms will be calculated, and its 95% CI will be calculated using the normal approximation to the binomial distribution.

ORR will be compared between treatment arms using the stratified Cochran-Mantel-Haenszel test. The stratification factors will be the same as those described for the analysis of the primary endpoint OS.

As for CRR at end of treatment, a supplementary analysis for ORR at end of treatment may be performed in which patients with missing response including PET-CT data at end of treatment will be imputed using the "last observation carried forward" method.

4.3.2 Supportive Secondary Endpoints

4.3.2.1 Best Overall Response by Investigator

BOR is defined as the best response while on study (based on response including PET-CT or CT data) according to Lugano 2014 (Appendix 1) response criteria, as determined by investigator.

ORR is defined as the proportion of patients whose best overall response (BOR) is a partial response (PR) or a CR.

The main estimand for BOR is defined as:

- Population: Participants with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria (ITT population)
- Variable:
- BOR, as per investigator
- ORR, as per investigator
- Treatment:
  - Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles
  - Control arm: R-GemOx every 21 days for up to 8 cycles
- Intercurrent events and handling strategies:
  - Start of any non-protocol anti-lymphoma therapy prior to disease progression: hypothetical strategy
  - Missing tumor assessment due to early study withdrawal: composite strategy
  - Death while still on treatment: composite strategy
- Population-level summary: difference in response rate for ORR

BOR will be analyzed using the ITT population. An estimate of BOR rates will be calculated for each treatment arm, and its 95% CI will be calculated using the Clopper-Pearson method. The difference in ORR between treatment arms will be calculated, and its 95% CI will be calculated using the normal approximation to the binomial distribution. ORR will be compared between treatment arms using the stratified Cochran-Mantel-Haenszel test. The stratification factors will be the same as those described for the analysis of the primary endpoint OS. Responses after initiation of NALT will not be included in the analysis of BOR.

4.3.2.2 Complete Response Rate at End of Treatment by Investigator

The same analysis as done for CRR at end of treatment as determined by IRC will be repeated for the CRR at end of treatment, as determined by investigator.

4.3.2.3 Objective Response Rate at End of Treatment by Investigator

The same analysis as done for ORR at end of treatment as determined by IRC will be repeated for the ORR at end of treatment, as determined by investigator.

4.3.2.4 Complete Response Rate (Composite) at End of Treatment

The same analysis as done for CRR at end of treatment as determined by IRC and investigator will be repeated for the CRR at end of treatment based on response including PET-CT or CT data (composite response) for both IRC and Investigator.

4.3.2.5 Objective Response Rate (Composite) at End of Treatment

The same analysis as done for ORR at end of treatment as determined by IRC and investigator will be repeated for the ORR at end of treatment based on response including PET-CT or CT data (composite response) for both IRC and Investigator.

4.3.2.6 Duration of Response by Investigator

Duration of response (DOR) is defined as the time interval from the date of the first occurrence of a complete or partial response (based on response including PET-CT or CT data until the first date that progressive disease (based on either response: including PET-CT data or not including any PET data) or death is documented, whichever occurs first. Event and censoring rules for DOR are summarized in Table 8.

The main estimand for DOR is defined as:

- Population: Responder participant with relapsed/refractory DLBCL as defined by the study inclusion and exclusion criteria
- Variable: DOR, as per investigator
- Treatment:
  - Experimental arm: Pola-R-GemOx every 21 days for up to 8 cycles
  - Control arm: R-GemOx every 21 days for up to 8 cycles
- Intercurrent events and handling strategies:
  - Early discontinuation from study treatment: treatment policy strategy
  - Start of any non-protocol anti-lymphoma therapy prior to disease progression: hypothetical strategy
  - Missing two or more consecutive tumor response assessments: hypothetical strategy
- Population-level summary: hazard ratio for DOR

DOR will be assessed in patients who had a BOR as CR or PR, as determined by the investigator, using Lugano 2014 response criteria (Appendix 1).

The analysis of DOR is based on a non-randomized event-free survival subset of patients (specifically, patients who achieved an objective response), therefore, comparisons between treatment arms will be made for descriptive purposes. The methodologies detailed for the PFS analysis will be used for the DOR analysis, except that the analysis will not be stratified.

Table 8 Event and Censoring Rules for DOR (Censoring for NALT / Two or More Missing Tumor Assessments)

Situation	Date of DOR event or censoring	Outcome
No further disease assessments after the date of first occurrence of complete or partial response	Date of first occurrence of complete or partial response	Censored
Alive and without disease progression documentation	Date of last adequate[1] disease assessment	Censored
Death or disease progression after initiation of NALT	Date of last adequate[1] disease assessment before NALT. If no adequate post baseline assessment is available, then date of randomization.	Censored
Initiation of NALT in absence of death or disease progression	Date of last adequate[1] disease assessment before NALT. If no adequate post baseline assessment is available, then date of randomization.	Censored
Death or disease progression after two or more consecutive missed response assessments	Date of last adequate[1] disease assessment before the missed assessments. If no adequate post baseline assessment is available, then date of randomization.	Censored
Two or more consecutive missed response assessments in absence of death or disease progression	Date of last adequate[1] disease assessment before the missed assessments. If no adequate post baseline assessment is available, then date of randomization.	Censored
Death or disease progression before or without initiation of NALT	Date of death or first disease assessment showing disease progression, whichever occurs first	Event

DOR = Duration of response; NALT = non-protocol specified anti-lymphoma therapy

[1]To be considered adequate, the results of a response assessment should not be “unevaluable” or “not done”.

4.3.2.7 Event-Free Survival by Investigator

Event-free survival (EFS_{eff}) is defined as the time from randomization to first to the earliest occurrence of the below cases:

- Disease progression or relapse using Lugano 2014 response criteria (Appendix 1) (based on either response: including PET-CT data or not including any PET data)
- Death due to any cause
- Initiation of any NALT

Patients with no EFS_{eff} events reported at the time of analysis (clinical-cut off) will be censored i) on the date of the last evaluable tumor assessment if post-baseline tumor assessment ii) on the date of randomization if no post-baseline tumor assessment. EFS_{eff} will be analyzed using the ITT population and the methodologies detailed for the PFS analysis will be used for the EFS_{eff} analysis.

4.3.2.8 Efficacy in Safety Run-In Stage

Similar analyses related to CRR/ORR at end of treatment, BOR, OS, PFS (as per main estimand) and EFS_{eff} as defined in Table 1 for safety run-in stage will be performed for the safety run-in population but will be restricted to descriptive statistics only.

4.3.2.9 Patient-Reported Outcomes

PRO will be analyzed using the ITT population for EORTC QLQ-C30 and FACT-Lym, and will include data up to initiation of any NALT

4.3.2.9.1 EORTC QLQ-C30

Time to deterioration is defined as the time from randomization to the first documentation of a 10-point decrease in EORTC QLQ-C30 physical functioning scale from baseline. For fatigue, time to deterioration is defined as the time from randomization to the first documentation of a 10-point increase from baseline. Patients who do not have an observed deterioration at the time of clinical data cut-off will be censored i) at the last non-missing assessment date if post-baseline assessment ii) on the date of randomization if no post-baseline assessment. The hazard ratio for time to 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 the median time to deterioration for each treatment arm, and Kaplan-Meier curves will be produced. Median and 95% CI will be estimated.

The EORTC QLQ-C30 data will be scored according to the EORTC scoring manual (Fayers et al. 2001). Missing data will be assessed and reported by time point. In the event of incomplete data, for all questionnaire subscales, if more than 50% of the constituent items are completed, a pro-rated score will be computed consistent with the scoring manuals and published validation reports. For subscales with less than 50% of the items completed, the subscale will be considered as missing. Completion rates will be summarized by number and proportion of patients among those expected to complete the EORTC QLQ-C30 at each timepoint and by treatment arm.

For EORTC QLQ-C30 physical functioning and fatigue subscales, descriptive statistics at each visit and changes from baseline will be reported by treatment arm.

The number and proportion of patients with a clinically meaningful improvement will be summarized by treatment arm, for the EORTC QLQ-C30 physical functioning and fatigue scales. The 95% CI around the proportion will be calculated using the Clopper-Pearson method for each treatment arm. The difference in the proportions between the two

treatment arms will be presented with a two-sided 95% CI based on a normal approximation to the binomial distribution.

For the EORTC QLQ-C30 physical functioning scale, a clinically meaningful improvement is defined as at least a 7-point increase; for the fatigue scale, it is defined as at least a 9-point decrease (Cocks et al. 2012).

4.3.2.9.2 FACT-Lym Subscale

Time to deterioration is defined as the time from randomization to the first documentation of a > 3-point decrease from baseline in FACT-Lym LymS (Carter et al 2008; Hlubocky et al. 2013). Patients who do not have an observed deterioration at the time of clinical data cut-off will be censored i) at the last non-missing assessment date if post-baseline assessment ii) on the date of randomization if no post-baseline assessment. The hazard ratio for time to 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 the median time to deterioration for each treatment arm, and Kaplan-Meier curves will be produced. Median and 95% CI will be estimated. Supplemental item-level analyses will be conducted with the individual B-symptom items of the FACT-Lym LymS using a raw 1-point worsening.

For missing items within the questionnaire, prorated scores will be calculated according to developer guidance (Webster et al. 2003). PRO completion rates will be summarized at each timepoint by treatment arm.

For FACT-Lym LymS, descriptive statistics at each visit and changes from baseline will be reported by treatment arm.

The number and proportion of patients with a clinically meaningful improvement in lymphoma symptoms will be summarized by treatment arm. The 95% CI around the proportion will be calculated using the Clopper-Pearson method for each treatment arm. The difference in the proportions between the two treatment arms will be presented with a two-sided 95% CI based on a normal approximation to the binomial distribution. For the FACT-Lym LymS, clinically meaningful improvement is defined as a 3-point increase (Carter et al. 2008, Hlubocky et al. 2013). A raw 1-point change will be used for the individual B-symptom items (fever, weight loss, and night sweats) of the FACT-Lym LymS.

4.4 EXPLORATORY ENDPOINTS ANALYSIS

The exploratory efficacy objective is to evaluate the impact of treatment and disease on aspects of health-related quality of life related to:

- All other subscales for the EORTC QLQ-C30
- All other subscales for the FACT-Lym
- FACT/GOG-NTX-12

- EQ-5D-5L

Analyses will include data up to initiation of any NALT.

For EORTC QLQ-C30 other subscales, descriptive statistics at each visit and changes from baseline will be reported by treatment arm.

For the FACT/GOG-NTX-12 and FACT/GOG NTX-4, descriptive statistics at each visit and changes from baseline will be reported by treatment arm using the safety-evaluable population. For missing items within the questionnaire, prorated scores will be calculated according to developer guidance (Calhoun et al. 2003). PRO completion rates will be summarized at each timepoint by treatment arm.

For EQ-5D-5L, data analysis will be reported separately from the Clinical Study Report (CSR).

4.5 SAFETY ANALYSES

All safety analyses will be performed on the safety-evaluable population, (unless specified otherwise), by treatment arm for each stage separately and overall for the Pola-R-GemOx arm.

4.5.1 Extent of Exposure

Each study drug (polatuzumab vedotin, rituximab, gemcitabine, oxaliplatin) exposure, including treatment duration, dose intensity, number of cycles received and dose modifications (when applicable) will be summarized with descriptive statistics by treatment arm.

4.5.2 Adverse Events

All verbatim adverse event terms will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events, (NCI CTCAE) v5.0. Treatment-emergent adverse events, serious adverse events, adverse events leading to death, adverse events of particular interest, and adverse events leading to study treatment discontinuation will be summarized by mapped term, appropriate thesaurus level, NCI CTCAE grade (when specified) and treatment arm. For events of varying severity, the highest grade will be used in the summaries.

For reporting purposes, “treatment emergent” is defined as new or worsening adverse event through 90 days after the last dose of any study drug or prior to NALT, whichever is earlier.

Peripheral neuropathy will be further described overall and according to baseline peripheral neuropathy status and history. In particular, time to first onset and time to resolution will be summarized by treatment arm.

AEs associated with COVID-19, AEs associated with COVID-19 resulting in death, AEs associated with COVID-19 leading to study treatment discontinuation will be summarized by treatment arm.

All deaths and cause of death will be summarized by treatment arm.

4.5.3 Additional Safety Assessments

4.5.3.1 Laboratory Data

Laboratory data will be classified according to NCI CTCAE v5.0. Summary tables of shifts in NCI CTCAE v5.0 grades from baseline to the worst post baseline value will be presented by treatment arm for relevant laboratory data. In addition, post-baseline laboratory abnormality will also be summarized by treatment arm.

4.5.3.2 Vital Signs

For vital signs, change from baseline over time will be summarized by treatment arm.

For the ECOG performance status, shift table from baseline versus worst post-baseline will be presented by treatment arm.

4.5.3.3 ECGs

For ECG, results will be summarized by visit and treatment arm. In addition, a shift table of ECG interpretation from baseline versus worst post-baseline will be presented by treatment arm.

4.6 OTHER ANALYSES

4.6.1 Summaries of Conduct of Study

The following analyses will be performed on the enrolled population, by treatment arm for each stage separately and overall for the Pola-R-GemOx arm.

Study enrollment, patient disposition, reasons for discontinuation from the study treatment and reason for study termination will be summarized overall and by treatment arm.

Major protocol deviations, including violations of inclusion/exclusion criteria and deviations during study conduct, as well as major protocol deviations related to COVID-19 will be summarized overall and by treatment arm.

4.6.2 Summaries of Demographics and Baseline Characteristics

The following analyses will be performed on the enrolled population, by treatment arm for each stage separately and overall for the Pola-R-GemOx arm.

Demographic variables such as age, sex, race/ethnicity, stratification variables and other relevant baseline characteristics including disease history will be summarized overall

and by treatment arm. Reason for transplant ineligibility will also be summarized overall and by treatment arm.

Previous and concurrent medical history, as well as peripheral neuropathy history, will be summarized overall and by treatment arm.

Prior anti-lymphoma therapies/radiotherapies/surgeries, follow-up NALT will be summarized overall and by treatment arm. Time to first NALT will also be summarized.

4.6.3 Pharmacokinetic Analyses

Exploratory PK analyses will be performed on the PK-evaluable population at the Sponsor's discretion as appropriate.

Individual and mean serum and plasma concentrations of polatuzumab vedotin, gemcitabine, and oxaliplatin versus time data will be tabulated and plotted. Summary statistics of concentration data will be computed for each scheduled sampling time for each analyte. PK parameters, maximum concentrations (C_{max}), and trough concentration (C_{trough}), may be estimated (as appropriate for the data collected). Estimates for these parameters will be tabulated and summarized (mean and SD). PK parameters will be determined using the appropriate technique based on available data.

This study incorporates sparse PK sampling of polatuzumab vedotin to enable population PK analysis and to potentially enable additional exploratory correlative analyses of PK with pharmacodynamic, safety and/or efficacy endpoints. These analyses will be performed at the Sponsor's discretion as appropriate and may involve pooling of data from other clinical studies. If performed, the results of these analyses may be reported separately from the CSR.

4.6.4 Immunogenicity Analyses

Immunogenicity analyses will be performed on the immunogenicity–evaluable population.

The numbers and proportions of ADA-positive and ADA-negative patients at baseline (baseline prevalence) and after baseline (post-baseline incidence) will be summarized. Patients are considered to be ADA positive if they are ADA negative or have missing data at baseline but are tested positive for ADAs 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 [redacted] titer unit greater than the titer of the baseline sample (treatment-enhanced ADA response). Patients 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 [redacted] titer unit greater than the titer of the baseline sample (treatment unaffected).

The relationship between ADA status and safety, efficacy, and PK endpoints may be analyzed. If performed, the results of these analyses may be reported separately from the CSR.

4.6.5 Biomarker Analyses

Exploratory analyses of biomarkers related to tumor biology and the mechanisms of action of polatuzumab vedotin and rituximab may be conducted. Analyses will assess prognostic and/or predictive value of candidate biomarkers. The association between candidate biomarkers and OS, PFS, and PET-CT CR rate and potentially other measures of efficacy and safety, with treatment and independent of treatment, may be explored to assess potential predictive and prognostic value, respectively. The effects of baseline prognostic characteristics, including DLBCL subtypes (i.e., BCL2/MYC DH and DE, cell of origin) and mutation profiles on efficacy, may be evaluated using univariate and/or multivariate statistical methods. If performed, the results of these analyses may be reported separately from the CSR.

4.6.6 Analyses of China Subpopulation

The China subpopulation analysis will be conducted for China to meet local regulatory requirements (if primary analysis for the global population is positive).

The objective of those subgroups analysis is to assess the treatment effect of Pola-R-GemOx compared with R-GemOx in the China subpopulation, and to investigate the consistency in treatment effect between this China subpopulation and the global population for the purpose of registration in China.

The China subpopulation will include all participants enrolled at China sites.

Results from these analyses will be summarized in a separate CSR. The analysis of the China subpopulation will be performed at time of primary analysis for the global population.

Other subpopulations might be of interest to meet local regulatory requirements, as relevant.

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

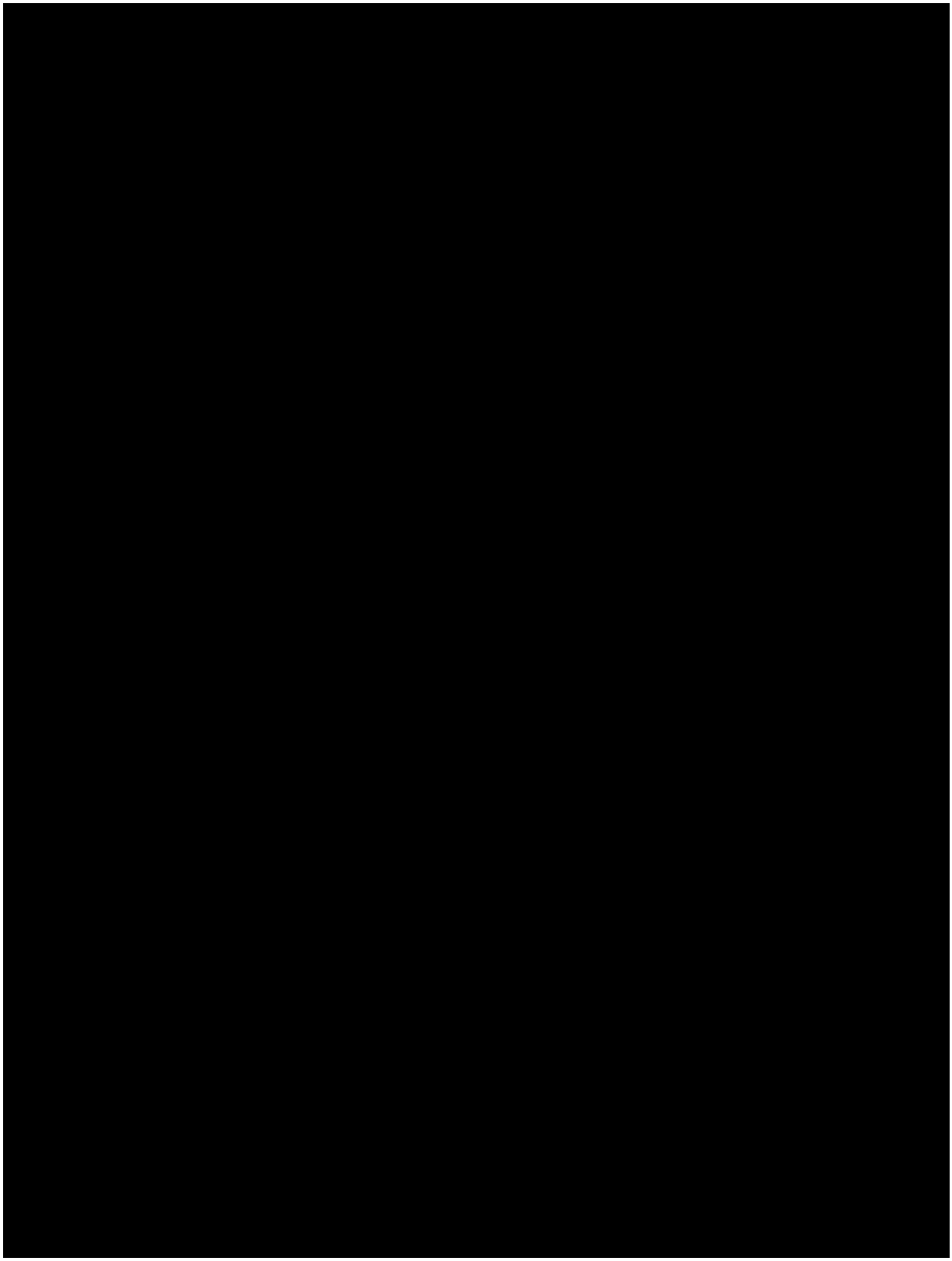
[redacted]

[redacted]

[redacted]

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

[redacted]



[redacted]

- [redacted]
- [redacted]

[redacted]

4.8 CHANGES TO PROTOCOL-PLANNED ANALYSES

Analysis populations have been revised with the addition of the enrolled population (all enrolled participants [regardless of the stage]), the modification of the safety evaluable population (all participants who received any amount of any study drug [regardless of the stage]) and the removal of PRO-evaluable population.

All analyses related to baseline characteristics and study conduct will be performed on the enrolled population, by treatment arm for each stage separately and overall for the Pola-R-GemOx arm.

All safety analyses (regardless of the stage), as well as PROs analyses related to FACT/GOG-NTX,will be performed on the safety evaluable population, by treatment arm for each stage separately and overall for the Pola-R-GemOx arm.

All efficacy analyses (including PROs related to EORTC QLQ-C30 and FACT-Lym) for the RCT stage will be performed on the ITT population, unless otherwise specified.

For PFS, the handling of intercurrent events related to NALT/ two or more missing tumor assessments have been changed in the main estimand from treatment policy strategy to hypothetical strategy (as per FDA request). Further supplementary analyses have been considered, in particular PFS without censoring for NALT/ two or more missing tumor assessments (treatment policy strategy for those intercurrent events).

PFS analyses related to further therapies are considering any NALT rather than specific one (like SCT or CAR-T therapy as mentioned in protocol).

CRR and ORR at end of treatment (not including PET data) will not be summarized.

BOR and DOR estimands have been updated to have handling of intercurrent events consistent with PFS main estimand.

DOR will be based on response as derived for BOR.

Further analyses have been added in SAP compared to the protocol.

5. SUPPORTING DOCUMENTATION

Appendix 1 2014 Lugano Response Criteria for Malignant Lymphoma

Response should be determined on the basis of radiographic and clinical evidence of disease according to Lugano response assessment criteria [1], as summarized below.

Target and Non-Target Lesions

Up to six of the largest target nodes, nodal masses, or other lymphomatous lesions that are measurable in two diameters should be identified from different body regions representative of the patient'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 six 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 nonmeasured disease as non-target lesions (e.g. cutaneous, GI, 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* with or without a residual mass on 5PS†	Target nodes/nodal masses must regress to ≤ 1.5 cm in LDi No extralymphatic sites of disease

[1] Cheson BD, Fisher RI, 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(27):3059-3067.

Revised Criteria for Response Assessment		
Response and Site	**PET-CT–Based Response**	**CT-Based Response**
	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	
Nonmeasured 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† with reduced uptake compared with baseline and residual mass(es) of any size At interim, these findings suggest responding disease At end of treatment, these findings indicate residual disease	≥50% decrease in SPD of up to 6 target measurable nodes and extranodal sites When a lesion is too small to measure on CT, assign 5 mm ×5 mm as the default value When no longer visible, 0 ×0 mm For a node >5 mm ×5 mm, but smaller than normal, use actual measurement for calculation
Nonmeasured lesions	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

Revised Criteria for Response Assessment		
Response and Site	**PET-CT–Based Response**	**CT-Based Response**
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 with no significant change in FDG uptake from baseline at interim or end of treatment	$<$50% decrease from baseline in SPD of up to 6 dominant, measurable nodes and extranodal sites; no criteria for progressive disease are met
Nonmeasured lesions	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
Progressive disease	Progressive metabolic response	Progressive disease requires at least 1 of the following
Individual target nodes/nodal masses	Score 4 or 5 with an increase in intensity of uptake from baseline 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 $>$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

Revised Criteria for Response Assessment		
Response and Site	**PET-CT–Based Response**	**CT-Based Response**
		In the setting of splenomegaly, 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
Non-measured lesions	None	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; 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 score of 3 in many patients 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 six of the largest dominant nodes, nodal masses, and extranodal lesions selected to be clearly measurable in two 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 (eg, liver, spleen, kidneys, lungs), GI involvement, cutaneous lesions, or those noted on palpation. Nonmeasured 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).

† 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.

6. REFERENCES

Berry SM, Carlin BP, Lee JJ, and Muller P. Bayesian Adaptive Methods for Clinical Trials. Chapman & Hall/CRC Biostatistics Series. Taylor & Francis, 2011.

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

Calhoun EA, Welshman EE, Chang CH et al. "Psychometric evaluation of the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (Fact/GOG-Ntx) questionnaire for patients receiving systemic chemotherapy." International Journal of Gynecological Cancer 13.6 (2003): 741-748.

Carter GC, Liepa AM, et al. Validation of the functional assessment of cancer therapy-lymphoma (FACT-LYM) in patients with relapsed/refractory mantle cell lymphoma. Blood. 2008;112(11):2376.

Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. Journal of clinical oncology. 2014;32(27):3059.

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.

Devlin NJ, Shah KK, Feng Y, et al. Valuing health‐related quality of life: An EQ-5 D-5 L value set for England. Health economics. 2018;27(1):7-22.

Fayers PM, Aaronson NK, Bjordal K, et al. on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 scoring manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics. 1986:845-854.

Friedman M. Piecewise exponential models for survival data with covariates. The Annals of Statistics. 1982;10(1):101-113.

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.

International Council on Harmonisation. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2020 [updated: 18 February 2020; cited: 12 January 2021].

Manitz, J., Kan-Dobrosky, N., Buchner, H., Casadebaig, M.L., Degtyarev, E., Dey, J., Haddad, V., Fei, J., Martin, E., Mo, M., Rufibach, K., Shentu, Y., Stalbovskaya, V., Tang, R., Yung, G., Zhu, J. Estimands for Overall Survival in Clinical Trials with Treatment Switching in Oncology (2022). Pharmaceutical Statistics, 21, 150-162.

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(11):1726.

Robins, Finkelstein. Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests. Biometrics. 2000, 56:779-788.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991;20(8):2609-31.

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):2409-2421.

Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in medicine. 2002;21(15):2175-2197.

Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. Statistics in medicine. 1996;15(13):1387-1395.

Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society: Series B (Methodological). 1976;38(3):290-295.

Webster K, Cella D, Yost K. The F unctional A ssessment of C hronic I llness T herapy (FACIT) Measurement System: properties, applications, and interpretation. Health and quality of life outcomes. 2003;1(1):1-7

Signature Page for Study MO40598-Statistical Analysis Plan (SAP) -Version 3 - Pu
System identifier: RIM-CLIN-534642

Approval Task	[redacted] Company Signatory 14-May-2024 16:30:46 GMT+0000