Statistical Analysis Plan

Study Code	D9670C00001
Edition Number	3.0
Date	20 Mar 2024

A Phase 3, Randomized, Multi-center, Open-label Study of Trastuzumab Deruxtecan (T-DXd) Versus Investigator's Choice Chemotherapy in HER2-Low, Hormone Receptor Positive Breast Cancer Patients whose Disease has Progressed on Endocrine Therapy in the Metastatic Setting (DESTINY-Breast 06)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or Special Term	Explanation
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
AJCC	American Joint Committee on Cancer
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BICR	blinded independent central review
BoR	best objective response
BP	blood pressure
CDK	cyclin-dependent kinase
CI	confidence interval
COVID-19	coronavirus disease 2019
CR	complete response
CRF	case report form
CSP	clinical study protocol
CSR	clinical study report
CT	computed tomography
ctDNA	circulating tumor DNA
CTCAE	Common Terminology Criteria for Adverse Event
DCO	data cut-off
DoR	duration of response
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
ECHO	echocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC QLQ-BR45	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire – Breast Module 45

Abbreviation or Special Term	Explanation
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30
EOT	end of treatment
EQ-5D-5L	European Quality of Life 5-Domain 5-Level Scale
EQ-VAS	European Quality of Life Visual Analog Scale
ER	estrogen receptor
ET	endocrine therapy
ERBB	erythroblastic oncogene B
FAS	full analysis set
HDU	high dependency unit
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
HR+	hormone receptor-positive
HRQoL	health-related quality of life
ICF	informed consent form
ICU	intensive care unit
IHC	immunohistochemistry
IDMC	independent data monitoring committee
ILD	interstitial lung disease
IP	investigational product
IPCW	inverse probability of censoring weighting
IPD	important protocol deviation
IRT	interactive response technology
ISH	in situ hybridization
ITT	intent-to-treat
IV	intravenous
KM	Kaplan-Meier
LD	longest diameter
LVEF	left ventricular ejection fraction
MAAA-1181a	deruxtecan

Abbreviation or Special Term	Explanation
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measures
MRI	magnetic resonance imaging
mRNA	messenger RNA
MSSO	Maintenance and Support Services Organization
MUGA	multiple gated acquisition scans
NA	not applicable
nAb	neutralizing antibody
NE	not evaluable
NED	no evidence of disease
NTL	non-target lesions
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PFS2	time from randomization to second progression or death
PGI-BR	Patient Global Impression – Benefit/Risk
PGIC	Patient Global Impression – Change
PGIS	Patient Global Impression – Severity
PGI-TT	Patient Global Impression –Treatment Tolerability
PgR	progesterone receptor
PK	pharmacokinetics
PR	partial response
PRO	patient-reported outcome
PRO-CTCAE	Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
q3w	every 3 weeks
q6w	every 6 weeks
q9w	every 9 weeks
QoL	quality of life

Abbreviation or Special Term	Explanation
QTcF	QT interval corrected for heart rate using Fridericia's formula
qw	every week
RDI	relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RMST	Restricted Mean Survival Time
RPSFT	rank preserving structural failure time
SAE	serious adverse event
SAP	statistical analysis plan
SAF	safety analysis set
SD	stable disease
SpO2	pulse oximetry
T-DXd	trastuzumab deruxtecan
TFST	time to first subsequent treatment or death
TL	target lesion
TOPO	topoisomerase
TSST	time to second subsequent treatment or death
ULN	upper limit of normal

AMENDMENT HISTORY

Version	Date	Description of change	Change from CSP?	Rationale
Original Version 1.0	21-JUL-2020	Initial approved SAP	N/A	N/A
Version 2.0	18-OCT-2022	- Protocol deviations section has been updated based on the final PD Plan V4.0 dated on 04Oct2022. - Clarification about the source data of HER2 IHC expression to be used for derivation of HER2-low population. - New ultra-low population on safety (HER2 IHC >0 <1+ SAF) added. - Handling partial dates for subsequent therapy - Updated AE section to clarify the AE summaries. - Figure 3 about Testing Strategy for Primary and Key Secondary Endpoints has been removed. - Objective response rate estimates are now computed along with the exact Clopper-Pearson 95% CIs. - Table 17 about significance level for interim and final OS Analyses has been updated to account the updated performed on step 3 of the MTP.	Yes CSP version 4.0	- Updates based on Protocol amendment 3 (Version 4.0). - Some clarification added based on comments on previous delivery.
Version 3.0	20-MAR-2024	- Section 1.2: Tumor assessments updated based on new CSP. - Table 5: Added HER2 IHC >0 <1+ population, HER2 IHC >0 <1+ PK set and HER2 IHC >0 <1+ ADA set. Table 6, Table 14 and Table 15; Section 4.2.4, 4.2.5, 4.2.8, 4.2.12 and 4.2.13: Added analysis of PFS, OS, ORR, DoR, PFS2, PK, ADA, laboratory measurements and vital signs for HER2 IHC >0 <1+ population/set. - Section 2.2: Section has been updated based on the final PD Plan V5.0 dated on 22Feb2023	Yes No No	To align with CSP V5.0. To further characterize the efficacy and safety profile of T-DXd compared with investigator's choice chemotherapy in the subgroup of patients with HER2 IHC >0 <1+. Updates based on Protocol Deviation Plan (Version 5.0).

		- Table 9 and Section 3.1: Added NED as possible overall visit response.	No	Some clarification added to handle special cases of visit responses.
		- Section 3.2.2: Clarified dates included in the derivation of last date.	No	To add clarity.
		- Section 3.3.3: Clarified concept of time to deterioration for EORTC QLQ-C30 and EORTC QLQ-BR45.	No	To add some more clarity on concept of time to deterioration and confirmation.
		- Section 3.5.1: Clarified concept of drug interruption for capecitabine.	No	To add some more clarity on concept of drug interruption for capecitabine.
		- Section 3.5.2: Clarified concept of intended cumulative dose.	No	To add some more clarity on concept of intended cumulative dose.
		- Section 3.5.3 and 4.2.8.1: LVEF re-labelled as 'Left ventricular dysfunction' as the undesirable outcome of LVEF reductions.	No	In accordance with the Revision 2 of the European Medicines Agency guidelines on Good Pharmacovigilance Practice
		- Section 4.1.4: Added rules for derivation of baseline in case time of first dose is not captured.	No	Some clarification added to handle cases where time of first drug is not expected to be collected.
		- Section 4.1.4: Added rules for derivation of baseline for PRO endpoints.	No	To add more clarity on baseline derivation for PRO endpoints.
		- Section 4.1.4: Specified rules for handling of end of treatment visit.	No	To add more clarity on handling of end of treatment visit.
		- Table 13: Added visit windows for LVEF and PROs.	No	More examples given to add clarity.
		- Section 4.2.6.4: Added summaries for compliance for PRO-CTCAE, PGIS, PGIC, PGI-TT and PGI-BR. Added summaries for response to PRO-CTCAE, PGI-TT. Added graphical representation of PRO-CTCAE responses.	No	To further characterize the impact of treatment on symptoms, patient-perceived treatment tolerability, and benefit/risk from the patients' perspective.
		- Section 4.1.6: Added more details in the imputation rules for end date of prior concomitant treatment and start date of subsequent anti-cancer	No	Some clarification added to handle special cases of missing dates.

		therapy, to cover more cases of missing dates.		
		- Section 4.2.3.1: Added analysis for piecewise HR and RMST.	No	To further characterize the variation of HR in case of lack of proportionality.
		- Sections 3.3.1, 3.3.2 and 4.2.6.1: Removed Best Overall Response for EORTC QLQ-C30 and EORTC QLQ-C45.	No	Deemed as no needed given the other summaries foreseen.
		- Section 4.2.3.1: Updated language to consider evaluable scans only.	No	To add consistency with censoring rule.
		- Section 4.2.3.1: Subgroups added.	No	To further characterize efficacy results on subgroups of patients.
		- Section 4.2.4.2: Added reporting of percentages of patients alive at 12 and 18 months.	No	To further characterize response in terms of overall survival.
		- Section 4.2.8.1 and 4.2.8.6: Added analyses for exposure-adjusted incidence rate analysis, ILD/pneumonitis by adjudication outcome and maximum reported CTCAE grade, infusion-related reactions, LVEF values and changes from baseline.	No	To further characterize the safety and tolerability profile of T-DXd compared with investigator's choice chemotherapy.
		- Section 4.2.8.2: Removed prothrombin complex INR CTCAE grade shift output, narratives for ALT/AST/BILI greater than a specified limit, plot of maximum post-baseline ALT/AST vs. maximum post-baseline total bilirubin.	No	Mentioned here in the previous version for error.
		- Section 4.2.11: Removed summaries for dose interruptions, reductions, cycle delays and RDI for HER2-low safety set.	No	Deemed as no needed given the other summaries foreseen.
		- Section 5.1: Added details on IDMC meetings after unblinding.	No	To add more clarity on the process.
		- Section 2.2: Added more details on IPD on missed safety assessments. Added HER2 IHC >0 <1+ SAF among the population affected by protocol deviations.	No	For completeness.
		- Section 3.1.1: Added more details on lymph node side after CR in target lesion.	No	To add more clarity.
		- Section 3.3.1: Added details on derivation of time to deterioration.	No	To add more clarity.
		- Section 3.3.6 and 3.3.7: Changes done to clarify that PGIC and PGI-ITT are not about study treatment, but respectively about study	No	To add more clarity.

		involvement and cancer treatment.		
		- Section 3.3.9: Updated definition of received, expected and evaluable questionnaire. Added a caveat for excluding from derivation of overall compliance those PRO where the baseline assessment is not collected as per SoA.	No	To add more clarity.
		- Section 3.5.1: Added clarity on how the drug interruption will be recorded and analyzed.	No	To be consistent with shells and requirements.
		- Section 3.5.2: Rules to derived RDI updated.	No	To correct previous error.
		- Section 3.5.3: Clarified the analysis time window for ILD/pneumonitis. Removed Other significant AEs.	No	To correct previous error in time windowing rules. Deemed as no needed.
		- Section 3.6: Added details on additional PK sample taken for patients with suspected ILD/pneumonitis.	No	For completeness.
		- Section 4.1.4: Changed rules for deriving baseline in case of no time recorded for first dose.	No	To correct previous error.
		- Section 4.1.5: Changed rule to handle end of treatment visit in PROs summary.	No	To align with recent indication from dedicated team.
		- Section 4.1.6: Added other cases of imputation of partial/missing dates.	No	For completeness.
		- Section 4.2.3.1: Added summaries for duration of follow-up in all patients by treatment group. Added more strata about prior lines of endocrine therapy.	No	For completeness.
		- Table 16: Added KM plots where foreseen.	No	For completeness.
		- Section 4.2.6: Removed analyses on EORTC QLQ-C30 scale on financial difficulties; Removed OM option from MMRM model.	No	Deemed as no needed given the other summaries foreseen.
		- Section 4.2.4.2: Added summaries for duration of follow-up in all patients by treatment group and overall.	No	For completeness.
		- Section 4.2.8: Removed summary of long-term tolerability, urinalysis, ECG parameters summary, SpO2 summaries post baseline, ECOG performance status summaries post baseline, ILD/pneumonitis investigation data listing.	No	Deemed as no needed given the other summaries foreseen.
		- Section 4.2.10: Clarified process to identified disallowed concomitant medications.	No	For completeness.

		- Section 4.2.6.1: Data cut off set up with precise value. OM option removed from SAS statement used in analysis.	No	Deemed necessary to specify details for analysis. OM margin removed to obtain estimates in any scenario.
		- Section 2.1 and Section 3.7: Changed definition for ADA evaluable set and ADA/nAb outcomes.	No	Deemed needed to align with regulatory standards.
		- Section 3.7: Updated timing of ADA collection.	Yes	To align with CSP V5.0.
		- Section 2.2: Removed in either treatment group from deviation bias analysis criteria	No	To align with company standard
		- Section 3.2.3: Added NED as possible overall visit response for BICR assessment.	No	To align with section 3.1
		- Abbreviation section: added NED	No	For completeness
		- Section 4.2.3.1: Added details for investigator's choice chemotherapy subgroup analysis	No	To add more clarity
		- Section 3.5.2: removed sentence for consideration of volume before/after infusion	No	Alignment with eCRF instructions

1. STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the analyses in the clinical study protocol (CSP). This SAP is based on version 5.0 of the CSP.

This SAP will apply to the Phase 3 study to determine the efficacy and safety of T-DXd compared with Investigator's choice single agent chemotherapy in HR+, HER2-low breast cancer patients.

1.1 Study Objectives

1.1.1 Primary Objective

The primary objective for this study and the corresponding endpoints/variables are shown in Table 1.

Table 1 Primary Objective and Corresponding Endpoints/Variables

Objective	Endpoints/Variables
To assess the efficacy of T-DXd compared with Investigator's choice chemotherapy in terms of PFS by BICR in the HR+, HER2-low (IHC 2+/ISH- and IHC 1+) population	PFS in the HR+, HER2-low population: Time from date of randomization until the date of objective radiological disease progression by BICR according to RECIST 1.1 or death (by any cause in the absence of progression)

BICR = blinded independent central review; HER2 = human epidermal growth factor receptor 2; HR+ = hormone receptor-positive; IHC = immunohistochemistry; ISH = in situ hybridization; PFS = progression-free survival; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors; T-DXd = trastuzumab deruxtecan.

1.1.2 Secondary Objectives

The secondary objectives for this study and the corresponding endpoints/variables are shown in Table 2.

Table 2 Secondary Objectives and Corresponding Endpoints/Variables

Objectives	Endpoints/Variables
Key Secondary Objectives	
• To assess the efficacy of T-DXd compared with Investigator's choice chemotherapy in terms of OS in the HR+, HER2-low population • To assess the efficacy of T-DXd compared with Investigator's choice chemotherapy in terms of PFS by BICR and OS in the ITT population (HER2 IHC >0 <1+ and HER2-low)	• OS in the HR+, HER2-low population: Time from date of randomization until the date of death by any cause • PFS by BICR according to RECIST 1.1 in the ITT population (HER2 IHC >0 <1+ and HER2-low) • OS in the ITT Population

Table 2 Secondary Objectives and Corresponding Endpoints/Variables

Objectives	Endpoints/Variables
Other Secondary Objectives	
• To further assess the efficacy of T-DXd compared with Investigator's choice chemotherapy in terms of PFS by Investigator assessment, ORR, and DoR by BICR and Investigator assessment in the HR+, HER2-low population • To further assess the efficacy of T-DXd compared with Investigator's choice chemotherapy in terms of ORR and DoR by BICR and Investigator assessment in the ITT population	• ORR in the HR+, HER2-low population: The percentage of patients with at least 1 visit response of CR or PR by BICR and Investigator assessment according to RECIST 1.1 • DoR in the HR+, HER2-low population: Time from date of first detection of objective response until the date of objective radiological disease progression by BICR and Investigator assessment according to RECIST 1.1 or death in the absence of progression • PFS by Investigator assessment according to RECIST 1.1 in the HR+, HER2-low population • ORR and DoR by BICR and by Investigator assessment according to RECIST 1.1 in the ITT population
• To compare the effect of T-DXd with Investigator's choice chemotherapy in terms of PFS2 according to Investigator assessment, time to first subsequent treatment or death (TFST) and time to second subsequent treatment or death (TSST) in the HR+, HER2-low population and the ITT population	• PFS2 in the HR+, HER2-low population and the ITT population: time from randomization until second progression on next-line of treatment, as assessed by Investigator at the local site or death due to any cause • TFST in the HR+, HER2-low population and the ITT population: time from randomization to the start date of subsequent therapy after discontinuation of randomized treatment or death due to any cause • TSST in the HR+, HER2-low population and the ITT population: time from randomization to the start date of second subsequent therapy after discontinuation of randomized treatment or death due to any cause
• To assess the safety and tolerability profile of T-DXd compared with Investigator's choice chemotherapy	• AEs, changes from baseline in laboratory findings, ECHO/MUGA scans, ECGs and vital signs
• To assess the PK of T-DXd	• T-DXd total anti-HER2 antibody and MAAA-1181a concentrations in serum

Table 2 Secondary Objectives and Corresponding Endpoints/Variables

Objectives	Endpoints/Variables
• To assess symptoms, functioning and HRQoL in patients treated with T-DXd compared with Investigator's choice single agent chemotherapy	The PROs include: • Change from baseline in EORTC QLQ-C30 and EORTC QLQ-BR45 scale scores • Time to deterioration in EORTC QLQ-C30 scale scores
• To investigate the immunogenicity of T-DXd	• Number and percentage of patients who develop ADA for T-DXd

ADA = anti-drug antibody; AE = adverse events; BICR = blinded independent central review; CR = complete response; DoR = duration of response; ECG = electrocardiogram; ECHO = echocardiogram; EORTC QLQ-BR45 = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire – Breast Module 45; EORTC QLQ-C30 = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire – Core 30; HER2 = human epidermal growth factor receptor 2; HR+ = hormone receptor-positive; HRQoL = health-related quality of life; IHC = immunohistochemistry; ITT = intent-to-treat; MUGA = multiple gated acquisition scans; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PFS2 = time from randomization to second progression or death; PK = pharmacokinetics; PR = partial response; PRO = patient-reported outcome; RECIST = Response Evaluation Criteria In Solid Tumors; T-DXd = trastuzumab deruxtecan; TFST = time to first subsequent treatment or death; TSST = time to second subsequent treatment or death.

1.1.3 Exploratory Objectives

The exploratory objectives for this study and the corresponding endpoints/variables are shown in Table 3.

Patients randomized into the study in China will be excluded from exploratory objectives requiring the provision of additional tumor or blood samples, with the exception of samples for exploratory safety or clinical benefit analyses to identify candidate markers which may correlate with likelihood of clinical benefit/tolerability. See Section 8.7.1 of the CSP for further details.

The analyses of exploratory biomarkers will be documented in a separate analysis plan.

Table 3 Exploratory Objectives and Corresponding Endpoints/Variables

Objectives	Endpoints/Variables
• To collect blood and tissue samples at pre-treatment, on-treatment and post-treatment for defining biological responses to T-DXd and to investigate predictive markers of response, acquired resistance and other markers that may correlate with likelihood of clinical benefit or tolerability; samples and generated data may be used to support diagnostic development	Biomarkers that include but are not limited to biomarkers of T-DXd sensitivity/resistance and immunological biomarkers are: • Protein expression (IHC and proteomic analysis including, but not limited to ERBB2, related family members and TOPO-1 expression) • Mutational profiling in tissue, blood and ctDNA • Plasma and blood analysis for ctDNA (exploration of genetic alterations in ctDNA and dynamic changes, including ctDNA clearance) • mRNA expression (exploration of gene expression, molecular subtype and gene expression changes following treatment) in tissue and blood
• To explore the impact of treatment and disease state on health utility using the EQ-5D-5L • To assess patient-reported treatment tolerability • To assess the patient's overall impression of the severity of their cancer symptoms, change in condition since starting the study and benefit/risk assessment • To explore the impact of treatment and disease on health care resource use	• The EQ-5D-5L health state utility index will be used to derive health state utility based on patient-reported data • Proportion of patients experiencing treatment-related symptoms as measured by the PRO-CTCAE; patient perceived overall tolerability as measured by the PGI-TT • Proportion of patients reporting different levels of symptom severity as measured by the PGIS, change in condition as measured by the PGIC, and benefit/risk as measured by the PGI-BR • Health care resource use will be captured, including inpatient admissions, intensive care unit admissions, and length of stay in hospital
• To explore and optimize technologies for detection of HER2 protein expression	• Exploration of IHC and non-IHC methods to determine tumoral HER2 expression

ctDNA = circulating tumor DNA; EQ-5D-5L = European Quality of Life 5-Domain 5-Level Scale; ERBB = erythroblastic oncogene B; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; mRNA = messenger RNA; PGI-BR = Patient Global Impression – Benefit/Risk; PGIC = Patient Global Impression – Change; PGIS = Patient Global Impression – Severity; PGI-TT = Patient Global Impression – Treatment Tolerability; PRO-CTCAE = Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; T-DXd = trastuzumab deruxtecan; TOPO = topoisomerase.

1.2 Study Design

The study is an open-label, multi-center, randomized study in HER2-low, HR+ breast cancer patients with disease progression on at least 2 lines of prior endocrine therapy or within 6 months of first line endocrine therapy + CDK4/6 inhibitor use in the metastatic setting. The primary purpose of the study is to determine the efficacy and safety of T-DXd compared with Investigator's choice single agent chemotherapy in the target population. Approximately 850 patients (700 patients with HER2 IHC 1+/2+ [HER2-low] expression and 150 patients with

HER2 IHC >0 <1+ expression) will be randomized 1:1 across approximately 300 centers globally to receive either 5.4 mg/kg T-DXd every 3 weeks (q3w) or Investigator's choice single agent chemotherapy (paclitaxel, nab-paclitaxel or capecitabine) until RECIST 1.1 defined progressive disease (PD), unless there is unacceptable toxicity, withdrawal of main informed consent or another criterion for discontinuation is met.

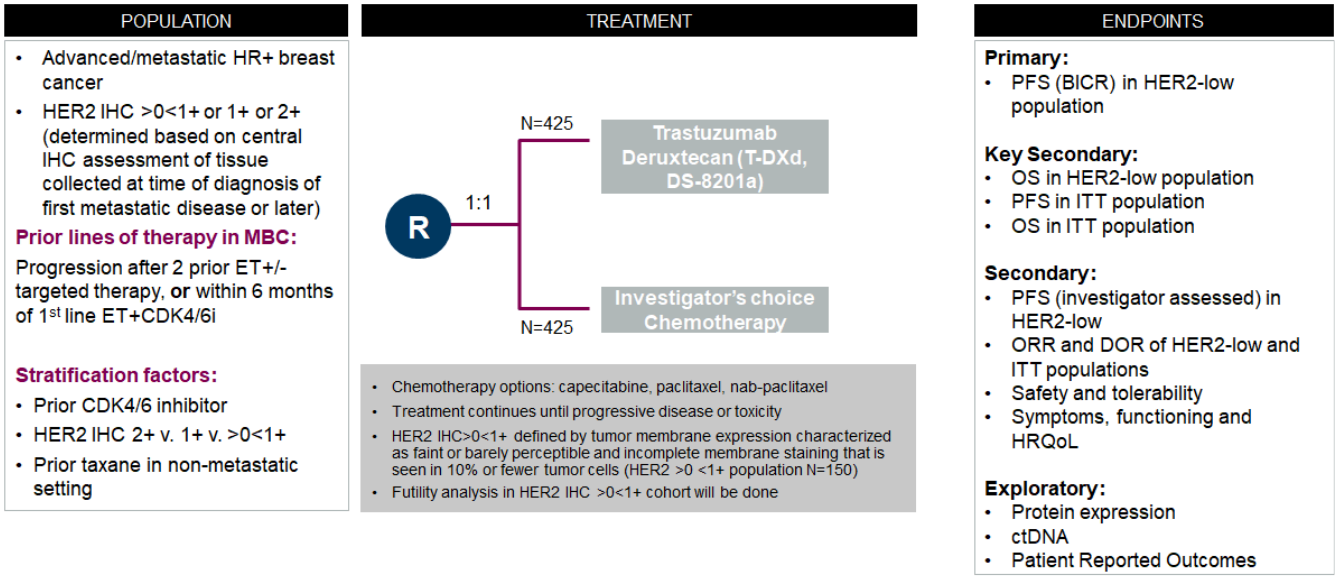
The randomization will be stratified by:

- Prior CDK4/6 inhibitor use (yes vs. no).
- HER2 IHC expression (IHC 2+/ISH- vs. IHC 1+ vs. IHC >0 <1+).
- Prior taxane use in the non-metastatic setting (yes vs. no).

To ensure adequate representation of each subgroup of the HER2-low population, at least 240 patients in each HER2 IHC group (IHC 1+ and IHC 2+/ISH-) across both treatment arms (120 patients per arm) will be randomized. Furthermore, CDK4/6 inhibitors are being increasingly utilized as part of standard of care for patients with HR+ breast cancer. To ensure that majority of patients have received prior CDK4/6 inhibitor therapy in the HER2-low population, no more than 343 patients (49% of 700 patients) who have not received prior CDK4/6 inhibitor therapy will be randomized; a similar proportion of patients who have not received prior CDK4/6 therapy will be randomized in the HER2 IHC >0 <1+ population as well.

The overall study design is summarized in Figure 1.

Figure 1 Study Design



BICR = blinded independent central review; CDK = cyclin-dependent kinase; ctDNA = circulating tumor DNA; DoR = duration of response; ET = endocrine therapy; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; HRQoL = health-related quality of life; IHC = immunohistochemistry; ISH = in situ hybridization; ITT = intent-to-treat; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; T-DXd = trastuzumab deruxtecan

Study Population

The study population will include pathologically documented advanced or metastatic HR+, HER2-low breast cancer patients whose disease has progressed on at least 2 lines of prior endocrine therapy or within 6 months of first line endocrine therapy + CDK4/6 inhibitor use in the metastatic setting. Patients must be ≥18 years of age, RECIST 1.1 evaluable, have an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, and have not received chemotherapy for metastatic disease. In addition to the primary population of HER2-low patients being studied, the study will also randomize approximately 150 patients with HER2 IHC >0 <1+ expression.

Each patient should meet all of the inclusion criteria and none of the exclusion criteria for this study in order to be randomized to a study treatment. Under no circumstances can there be exceptions to this rule.

Study Treatment and Duration of Study Treatment

The details of study treatment and their schedules are provided in Table 4.

Table 4 Study Treatments and Schedule

Compound	Dose	Route	Schedule
T-DXd	5.4 mg/kg	IV	Every 3 weeks
Capecitabine	1000 or 1250 mg/m^2	Oral	Twice daily orally for 2 weeks followed by a 1-week rest period in 3-week cycles
Paclitaxel	80 mg/m^2	IV	Every week (qw) in 3-week cycles
Nab-paclitaxel [a]	100 mg/m^2	IV	Every week (qw) for 3 weeks followed by a one-week rest period in 4-week cycles

IV = intravenous; qw = every week; T-DXd = trastuzumab deruxtecan.
[a] Although nab-paclitaxel is given in 4-week cycles, the schedule of assessments must be followed (see CSP Table 4; e.g., tumor assessment scans every 6 weeks [q6w ± 1 week]).

Unless specific treatment discontinuation criteria are met or the patient withdraws main consent, all patients will continue receiving treatment until disease progression. For patients randomized to the Investigator's choice single agent chemotherapy arm, crossover to T-DXd will not be permitted.

Tumor Assessments

Tumor evaluation scans will be performed at screening (as baseline) with follow-ups every 6 weeks (q6w ± 1 week) from the date of randomization for 48 weeks, and then every 9 weeks (q9w ± 1 week, starting at Week 48) until RECIST 1.1 disease progression.

Following investigator determined RECIST 1.1 progression, it is mandatory to perform an additional imaging assessment (preferably within 4 to 6 weeks after investigator PD) for central review to support the primary endpoint of PFS by BICR. In addition, every attempt should be made to continue to collect and submit subsequent scans (completed per standard practice / as clinically indicated) for central review until the primary PFS analysis DCO, or until the Sponsor notifies the site to discontinue (whichever is earlier) regardless of whether the subject has started another anti-cancer therapy.

Follow-up of Patients post Discontinuation of Study Treatment

After discontinuation of study treatment, all patients will have post-treatment follow-up scheduled at 40 days (+7 days) after their last dose of study treatment. Patients who have discontinued treatment for reasons other than progressive disease will also be followed up with tumor assessments until radiological progression (or death). All patients will be followed up for PFS2 and survival status, unless main consent was withdrawn. Other assessments should also continue to be performed as indicated in CSP.

Survival

All patients randomized should be followed up for survival unless main consent was withdrawn. Long-term/survival follow-up visits will be performed every 3 months (±14 days) from the date of the 40-day (+7 days) follow-up visit until death, withdrawal of main informed consent or study closure, whichever occurs first.

1.3 Number of Subjects

The study will randomize approximately 850 patients (700 HER2 IHC 2+/ISH- and IHC 1+ [HER2-low] patients and 150 HER2 IHC >0 <1+ patients). The HER2-low (IHC 2+/ISH- and IHC 1+) strata will be closed once a total of 700 patients have been randomized in these strata. At this time, a decision will be made by AstraZeneca on whether recruitment into the HER2 IHC >0 <1+ population will be stopped if the 150-patient target has not been met (see Section 5.3 for further details).

Approximately 1417 patients will be screened/enrolled with an approximate screen failure rate of 40% to achieve randomization of approximately 850 patients.

The study provides adequate power to show a statistically significant between-treatment difference in PFS in the HER2-low population. Based on a 2-sided significance level of 5%, a total of 456 PFS events (65% maturity) will provide at least 95% power to detect a hazard ratio of 0.55 (increase in median PFS from 5.5 to 10 months) in the HER2-low population, assuming an exponential distribution for both treatment groups.

If PFS in the HER2-low population is significant, the study also provides sufficient power to demonstrate a statistically significant difference in OS in the HER2-low population. Based on a 2-sided significance level of 3.5% and taking into account 2 interim OS analyses (see Section 4.2.1 for further details), a total of 521 OS events will be required to achieve 80% power to detect a hazard ratio of 0.77 (increase in median OS from 20.5 to 26.6 months) in the HER2-low population, assuming an exponential distribution for both treatment groups. Assuming 74% maturity at the time of the final OS analysis, approximately 700 patients will need to be randomized in the HER2-low subgroup.

2. ANALYSIS SETS

2.1 Definition of Analysis Sets

Table 5 defines the populations upon which the analyses are based.

Table 5 Populations for Analyses

Population/Analysis set	Description
Full analysis set (FAS) (ITT)	The ITT population, also termed as FAS, will include all randomized patients. Treatment arms will be compared on the basis of randomized study treatment, regardless of the study treatment actually received. Patients who were randomized but did not subsequently go on to receive study treatment are included in the analysis in the treatment arm to which they were randomized.
HER2-low	The HER2-low population comprises the subset of patients included in the ITT population with HER2 IHC 2+/ISH- and IHC 1+ as determined per the interactive response technology (IRT) data for the HER2 IHC expression.
HER2 IHC >0 <1+ (interim futility analysis)	The HER2 IHC >0 <1+ population for the interim futility analysis comprises the subset of patients included in the ITT population with HER2 IHC >0 <1+ as determined by central laboratory testing and who were randomized ≥ 24 weeks prior to interim futility DCO. The HER2 result from the most recent evaluable sample prior to randomization will be used to identify the HER2 IHC >0<1+ population.
HER2 IHC >0 <1+	The HER2 IHC >0 <1+ population for the primary analysis comprises the subset of patients included in the ITT population with HER2 IHC >0 <1+ as determined by central laboratory testing. The HER2 result from the most recent evaluable sample prior to randomization will be used to identify the HER2 IHC >0<1+ population.
Safety (SAF)	The SAF will include all patients who received at least one dose of study treatment. Safety data will be summarized according to the treatment received. Erroneously treated patients (e.g., those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received.
HER2-low SAF	The HER2-low SAF comprises the subset of patients included in the SAF with HER2 IHC 2+/ISH- and IHC 1+ as determined per the IRT data for the HER2 IHC expression.
HER2 IHC >0 <1+ SAF	The HER2 IHC >0 <1+ SAF population will include the subset of patients from the HER2 IHC >0 <1+ set who received at least one dose of study treatment.
PK	The PK analysis set will include all patients who receive at least 1 dose of T-DXd per the protocol for whom any post dose data are available.
HER2 IHC >0 <1+ PK	The HER2 IHC >0 <1+ PK analysis set will include the subset of patients from the HER2 IHC >0 <1+ population who receive at least 1

Table 5 Populations for Analyses

Population/Analysis set	Description
	dose of T-DXd per the protocol for whom any post dose data are available.
ADA	All patients who receive at least 1 dose of T-DXd who have a non-missing ADA result at any time. All major ADA analyses will be based on the ADA evaluable set.
HER2 IHC >0 <1+ ADA	The HER2 IHC >0 <1+ ADA analysis set will include the subset of patients from the HER2 IHC >0 <1+ population who receive at least 1 dose of T-DXd who have a non-missing ADA result at any time.

ADA = anti-drug antibody; DCO = data cut-off; FAS = full analysis set; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; ISH = in situ hybridization; ITT = intent-to-treat; PK = pharmacokinetic; SAF = safety analysis set; T-DXd = trastuzumab deruxtecan.

In the SAF and HER2-low SAF, if a patient receives any amount of T-DXd, they will be summarized in the T-DXd group. If a patient only receives Investigator's choice single agent chemotherapy, they will be summarized in the control treatment group.

The analysis populations used in the analysis for each outcome variable are provided in Table 6.

If enrolment in the HER2 IHC >0 <1+ subgroup will be stopped based on the results of the interim futility analysis (see Section 5.2 for further details), efficacy and safety analyses in the ITT population and SAF or in the HER2 IHC >0 <1+ subgroup may be performed for exploratory purposes.

Table 6 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Populations
Efficacy data	
PFS	HER2-low, HER2 IHC >0 <1+ [a,b], and ITT
OS	HER2-low, HER2 IHC >0<1+[b] and ITT
ORR	HER2-low, HER2 IHC >0 <1+ [a,b] and ITT
DoR	HER2-low, HER2 IHC >0 <1+ [a,b] and ITT DoR will be based on the subset of patients who achieved objective tumor response
DCR at 24 weeks	HER2 IHC >0 <1+ [a]
PFS2	HER2-low, HER2 IHC >0 <1+ [b] and ITT
TFST, TSST	HER2-low and ITT
PROs	HER2-low and ITT

Table 6 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Populations
Study population/Demography data	
Patient disposition	All patients, HER2 IHC >0 <1+ [a,b]
Demographic characteristics	HER2-low, HER2 IHC >0 <1+ [a,b] and ITT
Baseline and disease characteristics	HER2-low, HER2 IHC >0 <1+ [a,b] and ITT
Important protocol deviations	HER2-low and ITT
Medical/surgical history	HER2-low, HER2 IHC >0 <1+ [a] and ITT
Previous anti-cancer therapy	HER2-low, HER2 IHC>0<1+ [b] and ITT
Concomitant medications/procedures	HER2-low and ITT
Subsequent anti-cancer therapy	HER2-low and ITT
PK data	
PK data	PK analysis set, HER2 IHC >0 <1+ PK analysis set
Immunogenicity data	
Immunogenicity data	Listings will be based on SAF Summaries will be based on ADA evaluable set and HER2 IHC >0 <1+ ADA evaluable set
Safety data	
Exposure	HER2-low SAF, HER2 IHC >0 <1+ SAF [a, b] and SAF
AEs	HER2-low SAF, HER2 IHC >0 <1+ SAF [a, b] and SAF
Laboratory measurements	HER2-low SAF and SAF, HER2 IHC >0 <1+ SAF[b]
Vital signs	HER2-low SAF and SAF, HER2 IHC >0 <1+ SAF[b]
ECGs	HER2-low SAF and SAF

ADA = anti-drug antibody; AE = adverse event; DCR = disease control rate; DoR = duration of response; ECG = electrocardiogram; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; ITT = intent-to-treat population; KM = Kaplan Meier; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PFS2 = time from randomization to second progression or death; PK = pharmacokinetic; PRO = patient reported outcome; SAF = safety analysis set; TFST = time to first subsequent treatment or death; TSST = time to second subsequent treatment or death

[a] An interim futility analysis will be performed in the HER2 IHC >0 <1+ futility subgroup to compare ORR calculated using investigator data between T-DXd and the Investigator's choice of chemotherapy. As supportive summaries, patient disposition, demographics characteristics, baseline and disease characteristics, medical/surgical history, DoR, PFS KM plot and DCR at 24 weeks by treatment arm will also be provided for the HER2 IHC >0 <1+ futility subgroup. Exposure and safety will also be provided for the HER2 IHC >0 <1+ SAF population at futility analysis.

[b] At PFS analysis, additional summaries/analyses will be conducted on the HER2 IHC >0<1+ subgroup as supportive analyses, including efficacy, PK, immunogenicity, exposure and safety, as well as baseline patient's characteristics.

2.2 Protocol Deviations

For this study, the following general categories will be considered important protocol deviations (IPDs) and will be programmatically identified within the clinical database, and/or identified by Source Data Review, and confirmed by medical review. These will be listed and discussed in the clinical study report (CSR) as appropriate:

- Patients who deviate from key entry criteria per the Clinical Study Protocol (CSP). These are:
    - Inclusion criteria 2, 5, 6, 7 and 9.

      For criterion 2c, violation of local / central HER2 matching requirement in CSP versions 1, 2 & 3 will not be considered an IPD.
    - Exclusion criteria 4, 6, 15 and 16.
- Discontinuation criteria for study treatment met but patient not withdrawn from study treatment:
    - Occurrence of any AE that meets the criteria for permanent discontinuation as defined in Toxicity Management Guidelines for T-DXd (CSP Appendix H).
    - Pregnancy or intent to become pregnant.
    - Patient does not meet the eligibility criteria but is randomized in error, or incorrectly started on treatment, and decision is made to discontinue patient from treatment.
    - Objective progressive disease per criteria set forth in RECIST 1.1 (refer to CSP Appendix G).
    - Initiation of alternative anticancer therapy including another investigational agent.
- Investigational product (IP) deviation:
    - Patients received incorrect study treatment to that which they were randomized.
    - Patients received incorrect dose of IP for at least one cycle.
    - IP non-compliance for oral study treatment (<60% intended dose received for all cycles for reasons other than dose interruption due to AE as instructed by investigator).
    - Use of expired IP.
    - Patients with a known hypersensitivity to IP or any of the excipients of the product.
- Excluded medications taken:

- Received concomitant medication defined as prohibited as per the CSP.

- Deviations related to study procedure:
  - Baseline RECIST scan > 42 days before date of randomization (based upon a 28-day screening period plus 2 weeks allowance, so that only serious violators are identified).
  - No baseline RECIST 1.1 assessment on or before date of randomization.
  - Failure to perform the following mandatory safety assessments in ≥2 consecutive scheduled visits where assessment is expected:
    - RECIST Scans (while on treatment).
    - Pregnancy test.*
    - Hematology and/or clinical chemistry panel.*^
    - ECG.
    - Echo or MUGA.

    * For these assessments 2 consecutive out of window visits would also be considered as an IPD.
    ^ Not applicable for C1D8 and C1D15 safety assessments.
  - No baseline bone scan available within 12 weeks prior to randomization.

- Other important deviations:
  - Lack of provision of informed consent prior to study related procedures.
  - Missing PI electronic case report form (eCRF) signature (identified by remote data check).
  - Any deviation considered important that was not predicted or prespecified.

Patients who receive the wrong treatment at any time will be included in the SAF as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regard to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

The important protocol deviations will be listed and summarized for all patients in the HER2-low and ITT analysis sets (unless otherwise specified) by randomized treatment group. Patients randomized that did not receive study treatment will be excluded from the SAF, the HER2-low SAF and HER2 IHC >0 <1+ SAF. None of the other deviations will lead to patients being excluded from the analysis sets described in Section 2.1.

A per-protocol analysis excluding patients with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed on the PFS by BICR endpoint excluding patients with deviations that may affect the efficacy of the trial therapy if >10% of patients, which include following deviations:

- Did not have the intended disease or indication (which includes inclusion criteria 2*, 5, 6, 7, 9, and exclusion criteria 15, 16)

  * Given that in CSP v4 inclusion criterion #2c was modified to remove the requirement for matched results between historical/local HER2 results and central test results, patients with local vs central HER2 result mismatch will still be considered to have the intended disease.

The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock and will be documented prior to the primary analysis being conducted.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivations of RECIST Visit Responses

For all patients, the RECIST tumor response data will be used to determine each patient's visit response according to RECIST version 1.1 (see further details in Appendix G in the CSP). It will also be used to determine if and when a patient has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumor assessments are to be performed no more than 28 days before randomization and ideally as close as possible to the start of study treatment. Tumor assessments are then performed q6w ± 1 week from the date of randomization for 48 weeks, and then every q9w ± 1 week (starting at Week 48) until RECIST 1.1 disease progression.

Patients who permanently discontinue study treatment for reasons other than objective RECIST 1.1 disease progression, withdrawal of main informed consent, closure of study or death (regardless of whether subsequent anticancer therapy was started) should continue to have RECIST 1.1 scans performed according to the schedule above defined until RECIST 1.1 disease progression.

If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule is to be followed in order to minimize any unintentional bias caused by some patients being assessed at a different frequency than other patients.

Each patient's visit response according to RECIST version 1.1 will be determined:

- From the BICR of imaging scans (see Section 3.1.4 for further details).
- Programmatically, from the Investigator's review of the imaging scans.

At each visit, patients will be assigned a RECIST 1.1 visit response of CR, PR, SD or PD, using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a patient has no evidence of disease (NED) at baseline, then the patient will be assigned a visit response of NED if there is still no evidence of disease or PD if there is evidence of progression. Due to study eligibility criteria, NED at baseline is not expected in investigator data, but may occur in BICR data. If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE) (unless there is evidence of progression in which case the response will be assigned as PD).

Please refer to Section 3.1.3 for the definitions of CR, PR, SD and PD.

RECIST outcomes (i.e., PFS, ORR, etc.) will be calculated programmatically using the BICR and site Investigator data from the overall visit responses. The primary endpoint of PFS in the HER2-low population and the key secondary endpoint of PFS in the ITT population will be based on BICR.

3.1.1 Target Lesions (TLs) – Site Investigator Data

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (LD) (except lymph nodes which must have short axis ≥ 15 mm), with computed tomography (CT) or magnetic resonance imaging (MRI) and which is suitable for accurate repeated measurements. A patient can have a maximum of five measurable lesions recorded at baseline with a maximum of two measurable lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomization will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement. In which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as non-target lesions (NTLs) at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Note: For patients who do not have measurable disease at entry (i.e., no TLs) but have non-measurable disease, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.1.3 for further details). If a

patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Table 7 TL Visit Responses

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs since baseline. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10 mm.
Partial response (PR)	At least a 30% decrease in the sum of the diameters of TL, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progression of disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5 mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
Not evaluable (NE)	Only relevant in certain situations (i.e., if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). **Note:** If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

TL = target lesion.

Rounding of TL Data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to one decimal place before assigning a TL response. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL Data

For a visit to be evaluable then all target lesion measurements should be recorded. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded.
- A NTL visit response of PD is recorded.
- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5mm, from nadir even assuming the non-recorded TLs have disappeared.

Note: the nadir can only be taken from assessments where all the TLs had a lesion diameter recorded.

Lymph nodes

For lymph nodes, if the size reduces to < 10mm then these are considered non-pathological. However, a size will still be given and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are <10mm and all other TLs are 0mm then although the sum may be >0mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

Only CR, PD or NE can follow a CR. If a CR has occurred, then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e., 0 mm or <10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met (i.e., if a lymph node LD increases by 20% and/or the absolute increase is >= 5 mm from nadir but remains <10 mm).
- Step 2: If all lesions are missing, or some lesion measurements are missing but all other lesions meet the CR criteria (i.e., 0 mm or <10 mm for lymph nodes) then response will be set to NE irrespective of whether the other criteria for PD of TL is also met i.e., if a lymph node short axis diameter increases by 20% and/or the absolute increase is >= 5 mm from nadir but remains < 10mm.
- Step 3: If not all lesions are missing and those that are non-missing do not meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis ≥ 10mm, it increases by 20% and the absolute increase is >= 5 mm from nadir) or a new lesion appears, then response will be set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined, the response will be set to remain as CR.

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('x') above which it cannot be accurately measured should be recorded. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment. It is expected that a visit response of PD will remain in the vast majority of cases.

TL too small to measure

If a TL becomes too small to measure, then this will be indicated as such on the CRF and a value of 5mm will be entered into the database and used in TL calculations. However, a smaller value may be used if the radiologist has not indicated 'too small to measure' on the CRF and has

entered a smaller value that can be reliably measured. If a TL response of PD results (at a subsequent visit) then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e., lesions irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment.

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization), should be handled in the following way. Once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumors:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if ≤1/3 of the TLs have missing measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD, then the patient would be assigned a TL response of PD.
- Step 3: If, after both steps, PD has not been assigned, then, if appropriate (i.e., if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention have a value of 0 (or <10mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements, then the visit response will be set as NE.

At subsequent visits, the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up where appropriate (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are treated as missing (because of intervention) then the TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e., if using a value of 0 for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥5 mm from nadir).

If ≤ 1/3 of the TL measurements are treated as missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters) and

this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non-missing lesions to the nadir sum of diameters excluding the lesions with missing measurements.

Example of scaling

Lesion 5 is missing at the follow-up visit; the nadir TL sum including lesions 1-5 was 74 mm. The sum of lesions 1-4 at the follow-up is 68 mm. The sum of the corresponding lesions at the nadir visit is 62 mm.

Scale up as follows to give an estimated TL sum of 81 mm:

68 x 74 / 62 = 81 mm

In general, CR will not be allowed as a TL response for visits where there is missing data. Only PR, SD, or PD (or NE) could be assigned as the TL visit response in these cases. Exceptions to this rule are above described in Step 3 for handling TLs that have had intervention during the study. However, for visits with ≤ 1/3 lesion assessments not recorded, the scaled-up sum of TLs diameters will be included when defining the nadir value for the assessment of progression.

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0mm.

Change in method of assessment of TLs

CT and MRI are the only methods of assessment of TLs that can be used within this trial. If a change in method of assessment occurs between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

3.1.2 Non-target Lesions (NTLs) and New Lesions – Site Investigator Data

At each visit, the Investigator should record an overall assessment of the NTL response. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the Investigator's overall assessment of NTLs as follows:

Table 8 NTL Visit Responses

Visit Responses	Description
Complete response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progressive disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases, the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the Investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. **Note:** For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

NTL = non-target lesion.

To achieve ‘unequivocal progression’ on the basis of NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumor burden has increased sufficiently to merit a determination of disease progression. A modest ‘increase’ in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal, i.e., not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumor.

New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If a new lesion is equivocal and it is subsequently assessed as unequivocal, then all the overall visit responses from the visit when the equivocal lesion was first detected onwards will be PD.

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present but should not overtly affect the derivation.

Symptomatic progression is not a descriptor for progression of NTLs: it is a reason for stopping study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic progression' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumor assessments where possible until objective disease progression is observed.

3.1.3 Overall Visit Response – Site Investigator Data

Table 9 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 9 Overall Visit Responses

Target	Non-target	New Lesions	Overall Visit Response
CR	CR or NA	No (or NE)	**CR**
CR	Non-CR/Non-PD or NE	No (or NE)	**PR**
PR	Non-PD or NE or NA	No (or NE)	**PR**
SD	Non-PD or NE or NA	No (or NE)	**SD**
PD	Any	Any	**PD**
Any	PD	Any	**PD**
Any	Any	Yes	**PD**
NE	Non-PD or NE or NA	No (or NE)	**NE**
NA	CR	No (or NE)	**CR**
NA	Non-CR/Non-PD	No (or NE)	**SD**
NA	NE	No (or NE)	**NE**
NA	NA	No (or NE)	**NED**[a]

CR = complete response; NA = not applicable; NE = not evaluable; NED = No evidence of disease; PD = progressive disease; PR = partial response; SD = stable disease.

[a] Due to study eligibility criteria, NED at baseline is not expected in investigator data, but may be observed in the programmatically derived investigator's visit response.

3.1.4 Independent Review

A planned BICR of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans for all patients (including those at unscheduled visits or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organization for central analysis. The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and will be adjudicated, if required (i.e., two reviewers' review the scans and adjudication is performed by a separate reviewer in case of a disagreement). For each patient, the BICR will define the overall visit response (i.e., the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary (for patients with TLs at baseline: CR, PR, SD, PD, NE; for patients with NTLs only: CR, SD, PD, NE; for patients with no disease identified at baseline the possible response can be: PD, no evidence of disease [NED], NE). If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. Endpoints (of ORR, PFS, DoR, etc.) will be derived programmatically from this information.

Results of this independent review will not be communicated to Investigators and the management of patients will be based solely upon the results of the RECIST 1.1 assessment conducted by the Investigator.

A BICR of all patients will be performed for the database lock for final PFS analysis, which will cover all of the scans up to the DCO.

The primary endpoint of PFS in the HER2-low population and the key secondary endpoint of PFS in the ITT population will be based on BICR.

Further details of the BICR will be documented in the BICR Charter.

3.2 Efficacy Variables

The primary endpoint of the study is PFS by BICR according to RECIST 1.1 in the HER2-low population. The key secondary endpoints include OS in the HER2-low population, PFS by BICR according to RECIST 1.1 in the ITT population and OS in the ITT population.

3.2.1 Progression-free Survival (PFS)

PFS according to RECIST 1.1 by BICR and Investigator assessment will be derived programmatically.

PFS is defined as the time from the date of randomization until the date of disease progression, as defined by RECIST 1.1, or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomized therapy or receives another anticancer therapy prior to progression (i.e., date of PFS event or censoring – date of randomization + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment.

However, if the patient progresses or dies immediately after two or more consecutive missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two missed visits (**note:** NE visit is not considered as missed visit). Given the scheduled visit assessment scheme (i.e., q6w for the first 48 weeks, then q9w thereafter), the definition of 2 missed visits (allowing for early and late visits) will change over time, as described in Table 10.

Table 10 PFS: Definition of 2 Missed Visits

Timing of Previous RECIST Assessment	Definition of 2 Missed Visits
Baseline	13 weeks since randomization (i.e., 2 x 6 weeks + 1 week for a late assessment = 13 weeks)
Post-baseline, before study day 288 (i.e., week 41)	14 weeks since the previous RECIST assessment (i.e., 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks)
From study day 288 to study day 330 (i.e., week 41 to week 47)	17 weeks since the previous RECIST assessment (i.e., 6 weeks + 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 17 weeks)
After study day 330 (i.e., week 47)	20 weeks since the previous RECIST assessment (i.e., 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks)

RECIST = Response Evaluation Criteria in Solid Tumors.

Table 10 PFS: Definition of 2 Missed Visits

Timing of Previous RECIST Assessment	Definition of 2 Missed Visits

Study day is calculated as (assessment date – date of randomization + 1).

If the patients have no evaluable visits or do not have baseline data, they will be censored at Day 1 (randomization date) unless they die within two visits of baseline (12 weeks plus 1 week allowing for a late assessment within the visit window).

The PFS time will always be derived based on scan/assessment dates, not visit dates.

RECIST 1.1 assessments/scans contributing toward a particular visit may be performed on different dates. The following rules will be applied:

- For BICR assessments, the date of progression will be determined based on the **earliest** scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication.
- For Investigator assessments:
    - The date of progression will be determined based on the **earliest** of the RECIST 1.1 assessment/scan dates of the component that indicates progression.
    - If a new lesion is equivocal and it is subsequently assessed as unequivocal, then the progression date will be determined using the date of the initial scan when the equivocal lesion was first detected.
- When censoring a patient for PFS, the patient will be censored at the **latest** of the scan dates contributing to a particular overall visit assessment.

Note: For TLs, only the latest scan date within an imaging visit window is recorded in the RECIST 1.1 CRF out of all scans performed at that assessment for the TLs, and similarly for NTLs, only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.

3.2.2 Overall Survival (OS)

OS is defined as the time from the date of randomization until death due to any cause regardless of whether the patient withdraws from randomized therapy or receives another anticancer therapy (i.e., date of death or censoring – date of randomization + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the CRF).

Note: Survival calls will be made for both the PFS analysis and subsequent OS analyses in the week following the date of data cut-off (DCO) for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly available death registries. If the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note that for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed). The last date for each individual patient is defined as the latest among the following dates recorded on the CRFs:

- AE start and stop dates, AE grade change dates.
- Admission and discharge dates of hospitalization (from HOSPAD and SERAE modules of CRF).
- Study treatment date.
- End of treatment date.
- Overdose date.
- Date of drug dispensed.
- Date of laboratory test, vital signs, ECG, ejection fraction measurement, pulmonary function test, ophthalmologic assessment, physical examination, examinations for interstitial lung disease (ILD)/pneumonitis (ILDIS module of CRF and ADJAEILD from Adjudication Committee), ECOG performance status assessment.
- Dates of questionnaires.
- Date of sampling for PK analysis and T-DXd ADA assessment.
- Dates from pregnancy report.
- Dates from substance abuse report.
- Disease assessment dates on RECIST CRF.
- BICR scan dates.Date of second time to progression.

- Start and stop dates of concomitant medications, concomitant procedures, palliative, and post-treatment radiotherapy and subsequent cancer therapies.
- Date of informed consent withdrawal.
- Date last known alive on survival status CRF.
- End of study date.

For the handling of incomplete and partial dates of death, see Section 4.1.6.

3.2.3 Objective Response Rate (ORR)

ORR according to RECIST 1.1 by BICR, based upon all patients in the analysis set, will be derived programmatically.

ORR is defined as the percentage of patients with at least one visit response of complete or partial response (using RECIST 1.1).

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. However, patients who receive subsequent anticancer therapy (note that for this analysis radiotherapy is not considered a subsequent anticancer therapy) after discontinuing study treatment without progression and then respond will not be included as responders in the ORR.

As a sensitivity analysis, confirmed ORR, defined as the percentage of patients with a confirmed response of CR or PR (according to RECIST 1.1) will also be calculated. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of confirmed ORR. However, patients who receive subsequent anticancer therapy (note that for this analysis radiotherapy is not considered a subsequent anticancer therapy) after discontinuing study treatment without progression and then respond will not be included as responders in the confirmed ORR (i.e., both visits contributing to a response must be prior to subsequent therapy for the patient to be considered as a responder).

In the case where a patient has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the patient will be defined as a responder. Similarly, if a patient has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR will be assigned.

The following variables will be derived for sensitivity purposes:

- Confirmed ORR by BICR based upon all patients in the analysis set.
- ORR and confirmed ORR by Investigator assessment based upon all patients in the analysis set. Of note, the interim futility analysis will be based on confirmed ORR by Investigator assessment (see Section 5.2 for further details).
- ORR and confirmed ORR by BICR based upon the subgroup of patients in the analysis set with measurable disease at baseline (per BICR).
- ORR and confirmed ORR by Investigator assessment based upon the subgroup of patients in the analysis set with measurable disease at baseline (per Investigator assessment).

Best Objective Response (BoR)

As a supportive endpoint for ORR and confirmed ORR, BoR according to RECIST 1.1 by BICR and Investigator assessment will be derived programmatically. The denominators for each case will be consistent with those used in the ORR analysis.

BoR is calculated based on the overall visit responses from each RECIST assessment. It is the best response a patient has had following randomization, but prior to starting any subsequent cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anticancer therapy) and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST assessments prior to death.

Categorization of BoR will be based on RECIST using the following response categories: CR, PR, SD, NED, PD and NE.

Confirmation of CR or PR will not be required in the derivation of BoR as supportive endpoint for ORR, while CR or PR must be confirmed in the derivation of BoR as supportive endpoint for confirmed ORR.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 6 weeks minus 1 week, i.e., at least 35 days (to allow for an early assessment within the assessment window), after randomization. For CR/PR, the initial overall visit assessment that showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤ 13 weeks (i.e., 2 x 6 weeks + 1 week to allow for a late assessment within the assessment window) after randomization, then BoR will be assigned to the progression (PD) category. For patients who

die with no evaluable RECIST assessments, if the death occurs > 13 weeks after randomization then BoR will be assigned to the NE category.

3.2.4 Duration of Response (DoR)

DoR according to RECIST 1.1 by BICR and Investigator assessment will be derived programmatically.

For patients who achieve at least one visit response of CR or PR per RECIST 1.1, DoR is defined as the time from the date of first documented response until date of documented progression (using RECIST 1.1) or death in the absence of disease progression (i.e., date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of PR or CR.

As a sensitivity analysis, DoR based on confirmed response will also be calculated. For patients who achieve confirmed complete or partial response per RECIST 1.1, DoR is defined as the time from the date of first documented response (which is subsequently confirmed) until date of documented progression (using RECIST 1.1) or death in the absence of disease progression (i.e., date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first that was CR or PR that was subsequently confirmed.

If a patient does not progress following a response, then their DoR will use the PFS censoring time.

DoR will not be defined for those patients who do not have documented response.

3.2.5 Disease Control Rate (DCR) at 24 Weeks

DCR at 24 weeks according to RECIST 1.1 by Investigator assessment will be derived programmatically. This endpoint will only be reported at the interim futility analysis using the HER2 IHC >0 <1+ analysis set, to support the analysis of confirmed ORR by Investigator assessment (see Section 5.3 for further details).

DCR at 24 weeks is defined as the percentage of patients who have a BoR of CR or PR (with no confirmation required) or who have SD (without subsequent cancer therapy, note that for this analysis radiotherapy is not considered a subsequent anticancer therapy) for at least 23 weeks after randomization (to allow for an early assessment within the assessment window).

3.2.6 Time from Randomization to Second Progression or Death (PFS2)

PFS2 is defined as time from randomization to second progression (the earliest of the progression event subsequent to first subsequent therapy) or death (i.e., date of PFS2 event or censoring – date of randomization + 1). The date of second progression will be recorded by the Investigator in the CRF and defined according to local standard clinical practice and may include objective radiological imaging, symptomatic progression, or death.

Second progression status will be reviewed (40 days [+7 days] after the last dose of study treatment and then every 3 months [±14 days] thereafter) following the progression event used for the PFS by Investigator assessment (the first progression) and status recorded.

Patients alive and for whom a second disease progression has not been observed should be censored at the earliest of: date of study termination, date last known alive, DCO or, if a patient has not had a first subsequent therapy, the date last known not to have received a first subsequent therapy (TFST censoring date).

3.2.7 Time to First Subsequent Therapy or Death (TFST)

As a supportive summary to PFS, TFST is defined as time from randomization to the start date of the first subsequent anti-cancer therapy after discontinuation of randomized treatment or death due to any cause (i.e., date of first subsequent anti-cancer therapy/death or censoring – date of randomization + 1).

Patients not receiving randomized treatment would have TFST calculated as time from date of randomization to the initial therapy or death.

The use of subsequent anti-cancer therapies will be assessed at the 40-day (+7 days) follow-up visit, followed by visits every 3 months (±14 days) thereafter.

Patients alive and not known to have had a first subsequent anti-cancer therapy will be censored at the earliest of: date of study termination, date last known alive, DCO or the last date that the patient was known not to have received a first subsequent anti-cancer therapy. Patients not receiving randomized treatment would have TFST censored at Day 1 if the patient has no subsequent therapy data.

3.2.8 Time to Second Subsequent Therapy or Death (TSST)

As a supportive summary to PFS2, TSST is defined as time from randomization to the start date of the second subsequent anti-cancer therapy after discontinuation of randomized treatment or death due to any cause (i.e., date of second subsequent anti-cancer therapy/death or censoring – date of randomization + 1).

Patients not receiving randomized treatment would have TSST calculated in the same way, i.e., time from date of randomization to the second subsequent therapy or death.

The use of subsequent anti-cancer therapies will be assessed at the 40-day (+7 days) follow-up visit, followed by visits every 3 months (±14 days) thereafter.

Patients alive and not known to have had a second subsequent anti-cancer therapy will be censored at the earliest of: date of study termination, date last known alive, DCO or the last date that the patient was known not to have received a second subsequent anti-cancer therapy. Patients not receiving randomized treatment would have TSST censored at Day 1 if the patient has no subsequent therapy data.

3.3 Patient-Reported Outcome (PRO) Variables

The following PRO questionnaires will be used to assess the patient experience of treatment, including symptoms, tolerability, functioning and HRQoL: EORTC QLQ-C30, EORTC QLQ-BR45, EQ-5D-5L, PRO-CTCAE, PGIS, PGIC, PGI-TT and PGI-BR. Each PRO questionnaire is described in the sections below. The PRO questionnaires will be assessed throughout the study according to the schedule outlined in the CSP Section 1.3. All PRO analyses will be based on the HER2-low and ITT populations, unless stated otherwise.

Secondary PRO endpoints include change from baseline in EORTC QLQ-C30 and EORTC QLQ-BR45 scale scores, and time to deterioration in all the scales of the EORTC QLQ-C30, except scale on financial difficulties. The remaining PRO endpoints are exploratory.

The individual questionnaires are provided in the CSP Appendix J.

3.3.1 EORTC QLQ-C30

The EORTC QLQ-C30 is a 30-item self-administered questionnaire for all cancer types. Questions are grouped into 5 multi-item functional scales (physical, role, emotional, cognitive, and social), 3 multi-item symptom scales (fatigue, pain, and nausea/vomiting), a 2-item global health status/QoL scale, 5 single items assessing additional symptoms commonly reported by cancer patients (dyspnea, loss of appetite, insomnia, constipation, and diarrhoea), and 1 item on the financial impact of the disease. All but 2 questions have 4-point scales: “Not at All”, “A Little”, “Quite a Bit”, and “Very Much”. The 2 questions concerning global health status/QoL have 7-point scales with ratings ranging from “Very poor” to “Excellent” (Aaronson et al 1993).

For each of the 15 scales, final scores are transformed according to the EORTC QLQ-C30 Scoring Manual (Fayers et al 2001) such that they range from 0 to 100, where higher scores indicate better functioning/QoL or greater level of symptoms/problems (i.e., higher scores have opposite interpretations for functioning/QoL and symptoms/problems).

For each scale, if at least 50% of the items from the scale have been answered, assume that the missing items have values equal to the average of those items which are present for that respondent (Fayers et al 2001). If >50% of the items of a scale are missing, then that scale will be treated as missing. For example, role functioning and cognitive functioning each contain 2 items, and so these scales can be estimated whenever one of their constituent items is present; physical functioning contains 5 items, and so at least 3 need to have been completed. Using this method, none of the single-item measures can be imputed.

Definition of Clinically Meaningful Change

Changes in score compared to baseline will be evaluated. A minimum clinically meaningful change is defined as a change in the score from baseline of ≥10 for scales/items from the EORTC QLQ-C30 (Osoba et al 1998). For example, a clinically meaningful improvement in physical function (as assessed by EORTC QLQ-C30) is defined as an increase in the score from baseline of ≥10, whereas a clinically meaningful deterioration is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, change in symptoms/functioning from baseline will be categorized as improvement, deterioration or no change as shown in Table 11.

Table 11 Clinically Meaningful Changes in EORTC QLQ-C30 Scale Scores

Score	Change from Baseline	Visit Response
Symptom scales/items [a]	≥ +10	Deterioration
	≤ -10	Improvement
	Otherwise	No change
Functional scales and global health status/QoL	≥ +10	Improvement
	≤ -10	Deterioration
	Otherwise	No change

EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; QoL = quality of life.
[a] Including the item on the financial impact of the disease.

For the visit level summaries of improvement/deterioration/no change, all patients with a baseline and post-baseline score will be included, thus the denominator may differ from the time to deterioration endpoints described below.

Time to Deterioration

For each EORTC QLQ-C30 scale, time to deterioration will be defined as the time from randomization until the date of the first clinically meaningful deterioration (as defined in Table 11) that is confirmed at next available assessment, at least 14 days apart, regardless of whether the patient withdraws from study treatment or receives another anti-cancer therapy prior to

deterioration (i.e., date of first deterioration event or censoring – date of randomization + 1). Patients with a single deterioration and no further assessments will be treated as deteriorated in the analysis.

Patients with no clinically meaningful deterioration will be censored at the time of their last PRO assessment where the score could be evaluated. Also, in case of first confirmed deterioration after 2 or more consecutive missed PRO assessment visits (i.e., date from last PRO assessment prior to 2 or more consecutive missed PRO assessment visits and date of first confirmed deterioration ≥ 48 days) the patient will be censored at the time of the last PRO assessment where the score could be evaluated prior to 2 or more consecutive missed visits. If the patient has no evaluable visits, does not have baseline data, does not have post-baseline data, or has a baseline score that will not allow for a 10-point deterioration (baseline score > 90 for symptom scale and < 10 for functional and global health status/QoL scales), they will be censored at Day 1 (randomization date).

In this analysis, RECIST 1.1 progression will not be considered as deterioration and data will not be affected by RECIST 1.1 progression.

A summary of rules to be applied is reported in Table 12.

Table 12 Time to Deterioration Event or Censored

Situation	Event or Censored	Event Date/ Censored Date
No evaluable visits or no baseline data or no post-baseline data or baseline score does not allow for a 10-point deterioration (baseline score > 90 for symptom scale and < 10 for functional and global health status/QoL scales)	Censored	Day 1 (randomization date)
Confirmed deterioration after 2 or more consecutive missed PRO assessment visits (i.e., 48 days)	Censored	Latest evaluable PRO assessment date prior to the two or more missed visits
Confirmed deterioration without 2 or more consecutive missed PRO assessment visits before that (i.e., 48 days)	Event	The first deterioration date
One single deterioration with 2 or more consecutive missed PRO assessment visits before that (i.e., 48 days) and no further assessments	Censored	Latest evaluable PRO assessment date prior to the two or more missed visits
One single deterioration and no further assessments	Event	Last PRO assessment date (the single deterioration date)
No confirmed clinically meaningful deterioration	Censored	Latest evaluable PRO assessment date

Best Overall Response

A patient's best overall response in symptoms, function, and global health status/QoL will be derived as the best response the patient achieved, based on evaluable PRO data collected during the study period. The criteria in Table 13 will be used to assign a best response in symptoms, functioning and global health status/QoL. Summary tables of the best overall response will be provided.

Table 13 Best Overall Response Criteria for EORTC QLQ-C30

Best Overall Response	Criteria

Table 13 Best Overall Response Criteria for EORTC QLQ-C30

Improved	Patient meets one of the following criteria: 1. Has 2 consecutive visit response of 'improvement' at least 14 days apart[a]. 2. Has 1 visit response of 'improvement' and no further assessments.
No Change	Patient does not qualify for an overall score response of 'improved' and meets 1 of the following criteria: 1. Has 2 consecutive visit responses of 'no change' at least 14 days apart[a]. 2. Has 1 visit response of 'no change' and no further assessments.
Deterioration	Patient does not qualify for an overall score response of 'improved' or 'no change' and meets 1 of the following criteria: 1. Has 2 consecutive visit responses of 'deterioration' at least 14 days apart[a]. 2. Has 1 visit response of 'deterioration' and no further assessments.
Non-evaluable	Patient meets one of the following criteria: 1. Does not qualify for one of the above. 2. Has either no baseline or no post-baseline evaluable PRO assessment.
Missing	Patient has no baseline and no post-baseline evaluable PRO assessments.

EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; PRO = patient reported outcome.

[a] The criterion still applies in case of missing visits in between.

3.3.2 EORTC QLQ-BR45

The EORTC QLQ-BR45 is an updated version of the BR23, a validated breast cancer-specific module used in conjunction with the core QLQ-C30 to assess breast cancer-specific HRQoL (Bjelic-Radisic et al, 2020; Sprangers et al, 1996). New breast cancer treatments and diagnostics prompted the update of the QLQ-BR23 to include an additional 22 items. The self-administered instrument includes the original 23-items yielding 5 multi-item scores (body image, sexual functioning, arm symptoms, breast symptoms, and systemic therapy side effects). The additional 22 items yield four additional multi-item scales (breast satisfaction, endocrine therapy symptoms, skin mucosis symptoms, and endocrine sexual symptoms). In addition, single items assess sexual enjoyment, future perspective and being upset by hair loss. Items are scored on a 4-point verbal rating scale: "Not at All," "A Little," "Quite a Bit," and "Very Much". Scores are transformed to a 0 to 100 scale, where higher scores for functioning scales or items indicate better functioning, whereas higher scores for symptom scales or items represent a higher level of symptoms. The free text item in the EORTC QLQ-BR45 instrument is not included in the study, as the utility of this information and the analysis method have not been established.

Definition of Clinically Meaningful Change

Changes in score compared to baseline will be evaluated. Clinically meaningful change and improvement, no change and worsening will be defined as described for the EORTC QLQ-C30 in Section 3.3.1 and Table 11. For the visit level summaries of improvement/deterioration/no change, all patients with a baseline and post-baseline score will be included in the denominator.

Best Overall Response

Best overall response will be defined as described for the EORTC QLQ-C30 in Section 3.3.1 and Table 13. Summary tables of the best overall response will be provided.

3.3.3 EQ-5D-5L

The EQ-5D-5L will be used to explore the impact of treatment and disease state on health state utility.

The EQ-5D-5L, developed by the EuroQol Group, is a generic questionnaire that provides a simple descriptive profile of health and a single index value for health status for economic appraisal (Herdman et al 2011). The EQ-5D-5L questionnaire is comprised of six questions that cover five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, respondents select which statement best describes their health on that day from a possible five options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems, and unable to / extreme problems). A unique EQ-5D health state, termed the EQ-5D-5L profile, is reported as a five-digit code with 3125 possible health states. For example, state 11111 indicates no problems on any of the five dimensions. Respondents also assess their health today using the EQ-VAS (visual analogue scale), which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

The EQ-5D profile will be converted into a weighted health state utility value, termed the EQ-5D index, by applying a country-specific equation to the EQ-5D-5L profile that represents the comparative value of health states. This equation is based on national valuation sets elicited from the general population and the base case will be the UK perspective. Where a valuation set has not been published, the EQ-5D-5L profile will be converted to the EQ-5D index using a crosswalk algorithm (van Hout et al 2012). The EQ-VAS is reported separately.

3.3.4 PRO-CTCAE

PRO-CTCAE is an item library of symptoms experienced by patients while undergoing treatment of their cancer. Items capture the presence, frequency, severity and/or interference with usual activities, depending on the symptom. For each question, patients select the value that best describes their experience over the past week. The items pre-selected for this study are based on a review of the treatment-related symptoms of T-DXd, capecitabine, paclitaxel and nab-paclitaxel and in consideration of symptoms that are already captured in the other PRO

instruments with a view to minimize burden. The free text item in the PRO-CTCAE instrument is not included in the study, as the utility of this information and the analysis method have not been established.

The PRO-CTCAE items are not currently scored.

The PRO-CTCAE will only be administered in the languages where a linguistically validated version is already available.

3.3.5 PGIS

The PGIS item is included to assess how a patient perceives the overall severity of cancer symptoms over the past week. This is a single-item questionnaire and patients will be asked to choose the response that best describes the severity of their overall cancer symptoms with options ranging from “No Symptoms” to “Very Severe”.

The PGIS item is not currently scored.

3.3.6 PGIC

The PGIC item is included to assess how a patient perceives their overall change in health status since randomization. This is a single-item questionnaire and patients will choose from response options ranging from “Much Better” to “Much Worse”.

The PGIC item is not currently scored.

3.3.7 PGI-TT

The PGI-TT item is included to assess how a patient perceives the overall tolerability of his/her cancer treatment. This is a single-item questionnaire and patients will rate the bother associated with any treatment-related symptoms using response options ranging from “Not at all” to “Very much”.

The PGI-TT item is not currently scored.

3.3.8 PGI-BR

The PGI-BR is a 5-item questionnaire assessing the patient’s perception of the overall benefits and risks of treatment. The 5 items assess: overall trial experience, efficacy, side effects, convenience and overall assessment of the benefits and harms of treatment. Items are rated on 5- or 6-point verbal rating or Likert-type scales.

The PGI-BR items are not currently scored.

3.3.9 Compliance

Summary measures of overall compliance and compliance over time by treatment group will be derived for each PRO, where applicable (i.e., for PGIC and PGI-BR baseline and overall compliance will not be derived). These will be based upon:

- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time (e.g., a questionnaire from a patient who has not withdrawn from the study and has not experienced a PFS2 event at the scheduled assessment time but excluding patients in countries with no available translation). Date of study discontinuation and PFS2 event date (whichever occurs earliest) will be mapped to the nearest visit date to define the number of expected forms. If the date of study discontinuation/PFS2 event date (whichever occurs earliest) falls before the end of the visit window, then that visit is only considered expected if they have a received form. If they have not received a form, then this visit is not considered expected as they have not had the full opportunity to complete the questionnaire within the window.
- Received questionnaire = a questionnaire that is expected, has been received and has a completion date and at least one individual item completed.
- Evaluable questionnaire = a questionnaire that is expected and has a completion date and at least one scale that is non-missing (i.e., with at least one score that can be derived from the questionnaire).

Based on the above definitions, the following summary measures will be derived:

- Overall patient compliance rate is defined as: total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.
- Compliance over time (calculated separately for each visit, including baseline) is defined as: number of patients with an evaluable questionnaire at the time point (per definition above), divided by number of patients still expected to complete questionnaires.
- Evaluability rate over time (calculated separately for each visit, including baseline) is defined as: number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.

3.4 Health Care Resource Use Variables

At each scheduled visit, the site should review clinical notes for any non-study related hospital admissions and visits that have occurred. Where any visits have occurred, the site should complete the HOSPAD. This review should be done at every scheduled clinic visit up to and including the post-study treatment discontinuation follow-up visit. If a patient discontinues study

treatment for reasons other than RECIST 1.1 progression, the HOSPAD form should continue to be administered until progression has been confirmed. Study mandated visits should not be included in the HOSPAD.

To investigate the impact of treatment and disease on health care resource, the following variables will be captured:

- Planned and unplanned hospital attendances beyond trial protocol mandated visits (including physician visits, emergency room visits, day cases and admissions).
- Primary sign or symptom the patient presents with.
- Length of hospital stay.
- Length of any time spent in an intensive care unit (ICU) or high dependency unit (HDU).

3.5 Safety Variables

3.5.1 Exposure and Dose Interruptions

Total (or Intended) Exposure

The total (or intended) exposure (i.e., duration of treatment) of a patient to a drug is calculated using the start and stop dates of the drug and the intended dosing interval (or intervals if dosing is not regular). For a dosing period of the drug, the total (or intended) exposure is calculated as the number of days from date A to date B (i.e., B-A+1) where

- A is the date of first dose of the study drug in the dosing period.
- B is the earliest of:
    - the date of death
    - the date of DCO, and
    - the date when the last non-zero dose of the study drug was received (e.g., >0 mg of T-DXd) plus C, where C is equal to the scheduled number of days between doses minus one. Values for C for the study treatments of this study are shown in Table 14.

Table 14 Values Based on Dosing Interval (C) Used in the Calculation of Treatment Exposure

Treatment	C	Notes
T-DXd	20	
Capecitabine	0 or 6	• C = 0 if the last dose was scheduled Day 1 to 13 of a cycle. • C = 6 if the last dose was scheduled Day 14 of a cycle.
Paclitaxel	6	
Nab-paclitaxel	6 or 13	• C = 6 if the last dose was the first (planned on Day 1) or second dose (planned on Day 8) of a cycle. • C = 13 if the last dose was the third one of a cycle (planned on Day 15).

T-DXd = trastuzumab deruxtecan.

Actual Exposure

Actual exposure is defined as the actual treatment duration (days) and will be calculated only for T-DXd as follows:

Actual exposure = intended exposure (days) – total duration of dose delays (days), where intended exposure will be calculated as above.

Since patients will receive T-DXd via IV infusion q3w (±2 days), the total duration of dose delays (for deriving actual exposure) will be calculated as follows:

Total duration of dose delays = sum for all dosing dates of positive values of [date of the dose – date of the previous dose – (21+2) days].

Dose modifications are permitted, and the calculation of actual treatment duration makes no adjustment for any dose modifications that may have occurred.

Number of Treatment Cycles Received

Exposure will also be measured by the number of cycles received. A cycle corresponds to a period of 21 days for T-DXd, capecitabine and paclitaxel and of 28 days for nab-paclitaxel. If a cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Drug interruptions

Two or more consecutive missed and forgotten capecitabine doses should be recorded on the EX1 module as a dose interruption with the reason recorded as "Subject forgot to take dose". Single non-consecutive missed doses will not be recorded on the EX1 module as a dose

interruption, will not be considered to be drug interruptions and therefore will not be included in the summary tables, but the information will appear in the listing for dosing.

Patients who Permanently Discontinue During a Dose Interruption

If a patient permanently discontinues study treatment during a dose interruption, then the date of last administration of study medication recorded on EX will be used in the programming.

Safety Follow-up

- Total safety follow-up = min((last dose date + 47), date of withdrawal of main informed consent, date of death, date of DCO) – first dose date +1.

3.5.2 Dose Intensity

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. It is calculated separately for each drug. RDI will be defined as follows:

RDI = 100% * d/D, where:

- d is the actual cumulative dose delivered up to the actual last day of dosing, and
- D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule, up to discontinuation of that study drug.

Actual cumulative dose will be calculated by summing ((End date of study drug administration - Start date of study drug administration + 1) x Dose) * intended dosing frequency, for each period of study drug administration recorded on the study drug exposure form up to min(date of last dose date where dose > 0, date of death, date of DCO).

Dose modifications should also be considered in the derivation of actual dose administered.

The end of actual dosing period is calculated based on the last non-zero dose, even with skipped doses intended in the schedule.

Intended cumulative dose will be calculated by summing the individual doses that should have been received up to and including the actual last day of treatment according to the protocol planned dose and schedule.

For capecitabine, in case of dose escalation (from 1000mg/m^2 to 1250mg/m^2), if a patient didn't have overdose records or a patient had overdose records, but overdose start date doesn't equal to the date of the earliest cycle where the dose had escalated, then the intended dose will be

calculated separately for the period before the escalation and for the period after the escalation, and will then be combined for the intended cumulative dose across all cycles.

3.5.3 Adverse Events

AEs and serious adverse events (SAEs) will be collected throughout the study, from date of the patient signing the main informed consent form (ICF) until the follow-up period is completed (40 days [+7 days] after the last dose of study treatment). All SAEs, non-serious AEs, and adverse events of special interest (AESIs) will be followed until resolution, stabilization, the event is otherwise explained, or the patient is lost to follow-up. For patients who sign the pre-screening ICF, only SAEs directly related to tissue screening procedure (i.e., if a patient undergoes a tumor biopsy) will be reported during pre-screening period. If an event that starts post the defined safety follow-up period noted above is considered to be due to a late onset toxicity to study treatment, then it should be reported as an AE or SAE as applicable.

A treatment emergent adverse event (TEAE) is defined as an AE that occurs, having been absent before the first dose of study drug, or has worsened in severity or seriousness after initiating the study drug until 47 days after the last dose of study drug.

The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. The grading scales found in the revised NCI CTCAE v5.0 will be utilized for all AEs with an assigned CTCAE grading. For those events without assigned CTCAE grades, the recommendation in the CTCAE criteria that converts mild, moderate, and severe events into CTCAE grades should be used.

Missing start and stop dates for AEs and missing causality will be handled using the rules described in Section 4.1.6.

Adverse Events of Special Interest

An AESI is an AE of scientific and medical interest specific to understanding of the study treatment and may require close monitoring. An AESI may be serious or non-serious.

Based on the available pre-clinical and clinical data, review of the cumulative literature, reported toxicities for the same class of agents and biological plausibility, ILD/pneumonitis and LVEF decrease (re-labelled as ‘Left ventricular dysfunction’ as the undesirable outcome of LVEF reductions, in accordance with the Revision 2 of the European Medicines Agency guidelines on Good Pharmacovigilance Practice) are considered to be AESIs.

AESIs will be identified based on MedDRA preferred terms. Preferred terms used to identify AESIs will be listed before DBL and documented in the Trial Master File.

An ILD adjudication committee and charter will also be established to review all cases of potential ILD/pneumonitis. To ensure adequate and relevant evaluation, additional data may be collected to fully characterize medical history (e.g., smoking, radiation, and pulmonary history), diagnostic evaluation, treatment and outcome of the event. All cases of potential ILD/pneumonitis will be reported, including those outside the safety follow-up period.

3.5.4 Physical Examinations

Physical examinations will be performed from screening until the 40-day (+7 days) follow-up visit, as described in Sections 1.3 and 8.2.2 of the CSP. Full physical examinations will include assessments of the head, eyes, ears, nose, and throat and the respiratory, cardiovascular, gastrointestinal, urogenital, musculoskeletal, neurological, dermatological, hematologic/lymphatic, and endocrine systems. Height will be measured at screening only.

Any new or aggravated clinically relevant abnormal medical finding at a physical examination as compared with the baseline assessment will be reported as an AE unless unequivocally related to the disease under study.

3.5.5 Vital Signs

Vital signs (blood pressure [BP], pulse rate, temperature, and respiratory rate) will be evaluated from screening until the 40-day (+7 days) follow-up visit. Body weight is recorded at Day 1 of each cycle, at end of treatment (EOT) and at the 40-day (+7 days) follow-up visit. Further details are provided in Sections 1.3 and 8.2.3 of the CSP.

Change from baseline in vital sign variables will be calculated for each scheduled post-baseline visit. For the derivation of post-baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 4.1.5 will be used.

3.5.6 ECGs

At screening, ECGs will be obtained in triplicate. Subsequent ECGs will be performed in triplicate only if an abnormality is noted. Twelve-lead ECGs will be performed throughout the study until EOT, and the following ECG variables will be collected: heart rate, QTcF, PR, QT and RR intervals, QRS duration, and an overall evaluation by the Investigator. Further details are provided in Sections 1.3 and 8.2.4 of the CSP.

The overall evaluation of ECG will either be “normal”, “abnormal” or “borderline”, with abnormalities categorized as either “clinically significant” or “not clinically significant”.

Triplicate ECGs will be handled according to the following approach:

- Triplicate ECGs will be identified as 3 ECG assessments performed on the same date within 5 minutes.

- The mean of the 3 ECG assessments will be used to determine the values of ECG parameters at that time point.
- The worst overall evaluation will be used to determine the overall ECG evaluation at that time point.
- For the derivation of baseline (see Section 4.1.4) and of post-baseline visit values considering visit window (see Section 4.1.5), the triplicate ECG will be considered as a single assessment with date/time equal to the earliest date/time of the 3 ECGs.

The following relationship between QTcF, QT and RR (with QTcF and QT expressed in milliseconds and RR in seconds) will be used to derive programmatically the missing parameter in case only two of these variables are available:

$$\mathrm{QTcF} = \frac{\mathrm{QT}}{\sqrt[3]{\mathrm{RR}}}$$

The following relationship between RR and heart rate (with RR expressed in seconds and heart rate in bpm) will be used to derive programmatically the missing parameter in case only one of these variables is available:

$$\mathrm{RR} = \frac{60}{\text{heart rate}}$$

Change from baseline in ECG variables will be calculated for each scheduled post-baseline visit. For the derivation of post-baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 4.1.5 will be used.

3.5.7 Laboratory Measurements

Haematology and clinical chemistry tests will be performed from screening until the 40-day (+7 days) follow-up visit. Urinalysis and coagulation tests will be performed at screening and as clinically indicated until EOT and the end of long-term follow-up, respectively. Blood samples for troponin will be collected at screening, at EOT and if at any time a patient reports signs or symptoms suggesting congestive heart failure, myocardial infarction, or other causes of myocyte necrosis during the treatment period. Further details are provided in Sections 1.3 and 8.2.5 of the CSP.

Change from baseline in laboratory variables will be calculated for each scheduled post-baseline visit. For the derivation of post-baseline visit values considering visit windows and to handle multiple records, derivation rules as described in Section 4.1.5 will be used.

CTCAE grades will be defined at each visit according to the CTCAE grade criteria using project ranges, after conversion of lab result to corresponding project-wide preferred units. High and low

CTCAE grades will be calculated for the following parameters with CTCAE grades defined for both high and low values: potassium, sodium, magnesium, and corrected calcium.

In case of differential white blood cell count (neutrophils, lymphocytes, monocytes, eosinophils, basophils) available only as percentages, the absolute counts will be derived based on the total white blood cell count (leucocytes).

Corrected calcium will be derived during creation of the reporting database using the following formula:

Corrected calcium (mmol/L) = total calcium (mmol/L) + ([40 – albumin (g/L)] x 0.02).

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range), and high (above range).

The maximum or minimum value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post-dose value at any time.

Local reference ranges will be used for the primary interpretation of laboratory data at the local laboratory. Project reference ranges will be used for reporting purposes.

3.5.8 SpO2 and Pulmonary Function Tests

Pulse oximetry (SpO2) and pulmonary function tests will be performed as described in Sections 1.3 and 8.2.6 of the CSP.

3.5.9 LVEF

LVEF will be assessed as described in Sections 1.3 and 8.2.7 of the CSP.

3.5.10 ECOG Performance Status

ECOG performance status will be assessed as described in Sections 1.3 and 8.2.8 of the CSP.

3.5.11 Ophthalmologic Assessments

Ophthalmologic assessments will be performed as described in Sections 1.3 and 8.2.9 of the CSP.

3.5.12 Other Safety Assessments

3.5.12.1 ILD/Pneumonitis Investigation

If new or worsening pulmonary symptoms or radiological abnormality suggestive of ILD/pneumonitis is observed, treatment with study drug should be interrupted and a full

investigation to exclude alternative causes is required, as described in Section 8.2.10.1 and Appendix H of the CSP.

3.6 Pharmacokinetic Variables

For patients who received T-DXd only, blood samples will be obtained at Day 1 of Cycles 1, 2, 4, 6 and 8 for determination of T-DXd, total anti-HER2 antibody and MAAA-1181a concentrations in serum, as described in Sections 1.3 and 8.6.1 of the CSP.

Patients in the T-DXd group who receive chloroquine or hydroxychloroquine during the treatment period should have T-DXd PK blood samples taken also at additional time points.

Patients in the T-DXd group who are suspected of ILD/pneumonitis during the treatment period should have T-DXd PK blood sample taken at time of suspicion.

3.7 Immunogenicity Variables

For patients who received T-DXd only, blood samples for determination of ADA in serum will be collected pre-infusion on Day 1 Cycle 1, Day 1 Cycle 2, and Day 1 Cycle 4, then on Day 1 of every 4 cycles. Further details are provided in Sections 1.3 and 8.6.2 of the CSP.

Samples will be measured for the presence of ADAs and also potentially for ADA-neutralizing antibodies for T-DXd using validated assays.

ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. The neutralizing antibody (nAb) results will be reported as positive or negative.

The following ADA outcomes will be determined:

- ADA positive at any visit (at baseline or post-baseline).
- ADA positive post-baseline and positive at baseline.
- ADA not detected post-baseline and positive at baseline.
- Treatment-induced ADA, defined as ADA positive post-baseline and not detected at baseline (negative or missing).
- Treatment-boosted ADA, defined as a baseline positive ADA titer that was boosted to a 4-fold or higher-level following drug administration.
- Treatment-emergent ADA positive, defined as treatment-induced or treatment-boosted ADA.

- Treatment-emergent persistently positive ADA, defined as being treatment-emergent ADA positive and having at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurement.
- Treatment-emergent transiently positive ADA, defined as being treatment-emergent ADA positive and having at least one post-baseline ADA positive measurement and not fulfilling the conditions for treatment-emergent persistently positive.
- nAb positive at any visit (at baseline or post-baseline), if available.
- Treatment-Emergent nAb positive: treatment-emergent ADA positive patients with any positive nAb assessment post-baseline, if available.

4. ANALYSIS METHODS

4.1 General Principles

4.1.1 Statistical Hypotheses

The null hypothesis for the primary and key secondary efficacy endpoints (PFS and OS) is that there is no difference in terms of PFS or OS distribution between T-DXd and the Investigator's choice chemotherapy. The objective of the study is to demonstrate the superiority of T-DXd over the Investigator's choice chemotherapy, in the HER2-low population and the ITT population.

The approach for adjusting for multiplicity (due to multiple primary/key secondary endpoints and to interim analyses) is described in Section 4.2.1. The strong control of the family-wise Type I error rate at 5% (2-sided) across the primary and key secondary endpoints will be ensured.

4.1.2 Descriptive Statistics

Descriptive statistics will be used for all variables, as appropriate.

Continuous variables will be summarized by the number of observations, mean, standard deviation, median, minimum, and maximum. Upper and lower quartiles will be also presented if appropriate.

Categorical variables will be summarized by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment group. Overall totals will be calculated for baseline summaries only.

4.1.3 Rounding

For continuous data, mean, median and upper and lower quartiles will be rounded to 1 additional decimal place compared to the original data. The same rule will be applied to adjusted means and

their CIs. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

For categorical data, percentages will be rounded to 1 decimal place.

Hazard ratios, odds ratios and their CIs will be rounded to 2 decimal places.

Percentiles of time to event and their CIs estimated from the KM curve will be rounded to 1 decimal place.

p-values will be rounded to 4 decimal places, except for those below 0.00005, which will be displayed as '<0.0001'.

4.1.4 Baseline and Change from Baseline

In general, for efficacy, HER2 expression and PRO endpoints the last observed measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before or on the day of first dose of study treatment then this assessment will be used as baseline. For PRO endpoints, time will be included in the derivation of baseline. Assessments on the day of first dose that are required by the protocol to be conducted before the first dose but where time is not captured (or time of first dose is not captured, for example for Capecitabine doses), will be considered prior to the first dose. In the scenario where there are two nominal pre-dose assessments, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline in case time of first dose is captured. If there are two assessments equally eligible to assess patient status at baseline with assessment time missing or with same time or in case time of first dose is not captured, the average can be taken as a baseline value.

For safety endpoints the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured (or time of first dose is not captured, for example for Capecitabine doses), a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of the first dose where neither time (or time of first dose, for example for Capecitabine doses) nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose. In the scenario where there are two nominal pre-dose assessments on the Day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline in case time of first dose is captured. For safety assessments, if there are two visits equally eligible to assess patient status at baseline with assessment time missing or with same time or in case time of first dose is not captured, the

average can be taken as a baseline value. For non-numeric laboratory tests (e.g., some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as 100 * (post-baseline value - baseline value) / baseline value.

4.1.5 Time Windows for Safety and PRO Assessments

The following conventions will apply:

- Time windows will be defined for any presentations that summarize values by visit.
- The time windows will be exhaustive so that data recorded at any time point has the potential to be summarized. Inclusion within the time window will be based on the actual date and not the intended date of the visit.
- All unscheduled visit data have the potential to be included in the summaries.
- All end of treatment visits will be reported separately in the summaries and will not be included in any time window. This rule will not be applied for PRO assessments for which the end of treatment assessments will be mapped and not displayed separately as per Table 15 along with all the other assessments.
- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls halfway between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits, then the upper limit will be taken as the midpoint value minus 1 day. For assessments planned at specific cycle days, the different duration of cycles between treatments should be taken into account. Table 15, presents three different examples of visit windows, respectively for vital signs, LVEF and PRO assessments.

Table 15 Time Window Examples,

Parameter: Vital Signs				
	T-DXd, capecitabine, paclitaxel		**Nab-paclitaxel**	
Time Point	**Study Day**	**Visit Window**	**Study Day**	**Visit Window**
Cycle 1 Day 8	8	2 – 11	8	2 – 11
Cycle 1 Day 15	15	12 - 18	15	12 - 21
Cycle 2 Day 1	22	19 - 32	29	22 - 42
Cycle 3 Day 1	43	33 - 53	57	43 - 70

Table 15 Time Window Examples,

Parameter: Vital Signs

	T-DXd, capecitabine, paclitaxel		Nab-paclitaxel	
Time Point	**Study Day**	**Visit Window**	**Study Day**	**Visit Window**
Cycle 4 Day 1	64	54 - 74	85	71 - 98
Cycle 5 Day 1	85	…	113	…
…	…	…	…	…

Parameter: LVEF

	T-DXd, capecitabine, paclitaxel		Nab-paclitaxel	
Time Point	**Study Day**	**Visit Window**	**Study Day**	**Visit Window**
Cycle 5 Day 1	85	2 - 126	113	2 - 168
Cycle 9 Day 1	169	127 - 210	225	169 - 280
Cycle 13 Day 1	253	211 - 294	337	281 - 392
Cycle 17 Day 1	337	…	449	…
…	…	…	…	…

Parameter: EORTC QLQ-C30, EORTC QLQ-BR45, EQ-5D-5L, PGIS, PGIC

Any treatment group: T-DXd, capecitabine, paclitaxel and nab-paclitaxel		
Time Point	**Study Day**	**Visit Window**
Week 4 Day 1	22	2 – 32
Week 7 Day 1	43	33 – 53
Week 10 Day 1	64	54 – 74
Week 13 Day 1	85	…
…	…	…

EORTC QLQ = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire. EQ-5D-5L = European Quality of Life 5-Domain 5-Level Scale. LVEF = left ventricular ejection fraction. PGIC = Patient Global Impression–Change. PGIS = Patient Global Impression–Severity. T-DXd = trastuzumab deruxtecan. For safety parameters, study day is calculated as (assessment date – date of Cycle 1 Day 1 + 1). For PRO assessments, study day is calculated as (assessment date – randomization date + 1).

- All values (including those collected at unscheduled visits) will be included when deriving a patient level statistic such as minimum/maximum, regardless of whether or not they fall in an interval and appear in the corresponding visit-based summary.
- For visit based summaries:
    - If there is more than one value per patient within a time window, then the closest value to the scheduled visit date will be used, or the earliest in the event the values are equidistant from the nominal visit date. For PRO assessments, this rule will not be applied for the End of Treatment assessment on or after (within 7 days) the date of decision of discontinuation, that will be selected, whenever available, for the visit window on which it falls, regardless of the distance from the nominal visit date. If there are two values recorded on the same day and the parameter is CTCAE gradable then the record with the highest toxicity grade should be used. Alternatively, if there are two records recorded on the same day and the toxicity grade is the same (or is not calculated for the parameter), then the average of the two records should be used. The listings will highlight the value for the patient that contributed to the summary table, wherever feasible.
    - To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group, visit data will only be summarized if the number of observations is greater than the minimum of 20 and > 1/3 of patients in the analysis population.
- Unlike the other study treatments (administered by infusion), capecitabine will be administered orally. On Day 1 of Cycles 1-3, vital signs and SpO2 will be assessed only once in patients treated with capecitabine (while assessments before infusion and at the end of infusion will be performed in patients receiving the other study treatments). The measurements with capecitabine will be summarized together with the measurements performed before infusion with the other study treatments.
- Listings will display all values contributing to a time point for a patient.

4.1.6 Imputation Rules

Partial date of birth (only year will be available) will have 1st January imputed.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug. Missing CTCAE grades will not be imputed.

AE and Medication Start/Stop Dates

The original incomplete or missing dates will be presented in the listings.

- Adverse events: all AEs will be considered as treatment-emergent unless the opposite can be clearly stated. Imputation will be done only in the context of identifying treatment-emergent adverse events.
- Concomitant medications (including procedures): all medications will be considered as concomitant unless the opposite can be clearly stated.

In practice, for AEs and medications, original incomplete or missing start dates, dates of meeting seriousness criteria and toxicity/severity grade change dates will be imputed as below:

- Missing day: impute the 1st of the month unless month is same as month of first dose of study drug and the end date is on or after the first dose of study drug or ongoing then impute first dose date. If there is no date of first dose then impute the 1st of the month (used to determine the line of prior medication).
- Missing day and month: impute 1st January unless year is the same as first dose date and the end date is on or after the first dose of study drug or ongoing then impute first dose date. If there is no date of first dose then impute 1st January (used to determine the line of prior medication).
- Completely missing: impute first dose date unless the end date suggests it started prior to this in which case impute the date the patient was enrolled. If there is no date of first dose then impute the date the patient was enrolled (used to determine the line of prior medication).

When imputing a start date, ensure that the new imputed date is sensible (i.e., is prior to the end date of the AE or medication).

Original incomplete or missing stop dates for adverse events and medications/therapies will be imputed as below:

- Missing day: Impute the last day of the month unless month is the same as month of study discontinuation, then impute as study discontinuation date. For prior anti-cancer medications impute date of informed consent if month is the same as month informed consent was provided.
- Missing day and month: Impute 31st December unless year is the same as year of study discontinuation then impute study discontinuation date. For prior anti-cancer medications impute date of informed consent if year is the same as year informed consent was provided.

- Completely missing:
    - AE: since there is no ongoing flag recorded in CRF, then assume that AE is still present (i.e., do not impute a date).
    - Medication: if the ongoing flag is missing then assume that medication is still being taken (i.e., do not impute a date). If the medication has stopped and flagged as taken before study treatment and start date of medication is after informed consent date but prior to first dose date then impute the first dose date, if start date of medication is before informed consent date then impute stop date as informed consent date. If the medication has stopped and started on or after first dose date then impute as study discontinuation date. For prior anti-cancer therapies, if the imputed stop date is after or on the same date as randomization, then it will be imputed to the day before randomization.

When imputing a stop date, ensure that the new imputed date is sensible (i.e., is after the start date of the AE or medication).

Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

Date of Death

If a patient is known to have died where only a partial death date is available, then the date of death will be imputed as the latest of the last date known to be alive +1 from the database and the death date using the available information provided:

- For missing day only: using the 1st of the month.
- For missing day and month: using the 1st of January.

If there is evidence of death but the date is entirely missing, it will be treated as missing (i.e., the patient will be censored at the last known alive date).

Date of subsequent anti-cancer therapy

For partial or completely missing subsequent anti-cancer therapy dates, the following rules will be applied for missing start dates:

- Missing day: if the month is the same as treatment end date, then impute to the day after treatment, otherwise first day of the month.
- Missing day and month: if year is the same as treatment end date then impute to the day after treatment, otherwise 1st January of the same year as anti-cancer therapy date.

- Completely missing: impute to the day after treatment.

4.1.7 Analysis of Time-to-event Endpoints

Stratified log-rank tests will be performed using SAS® PROC LIFETEST with a TEST statement.

CIs for median time to event will be estimated from the KM curve based on the Brookmeyer-Crowley method (Brookmeyer and Crowley 1982).

Cox proportional hazards model will be estimated using SAS® PROC PHREG. The Efron method will be used for handling ties and the profile-likelihood approach will be used for the calculation of the confidence interval (CI) for hazard ratio. In case of stratified Cox proportional hazards model, the stratification variables will be included in the STRATA statement.

4.1.8 Other General Principles

The below mentioned general principles will be also followed throughout the study:

- Data will be presented in listings by subject number. All summaries will be presented by treatment group, unless otherwise specified. Data will be summarized and analyzed on the analyses sets as described in Table 5.
- A month is operationally defined to be 30.4375 days. All time-to-event and exposure variables will be expressed in months.
- The stratification factors considered for randomization (prior CDK4/6 inhibitor use, HER2 IHC expression, prior taxane use in the non-metastatic setting) will be used as stratification factors or covariates in the statistical analyses based on the values entered in IRT, even if it is subsequently discovered that these values were incorrect.
- Results of all statistical analyses will be presented using 2-sided 95% CIs and 2-sided p-values, unless otherwise stated.
- SAS® version 9.4 or higher will be used for all analyses.

4.2 Analysis Methods

Table 16 details which efficacy endpoints are to be subjected to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as primary for that endpoint.

All the analyses presented in the table will be performed for the HER2- low and ITT populations, unless otherwise specified. Of note, if enrolment in the HER2 IHC >0 <1+ subgroup will be stopped based on the results of the interim futility analysis (see Section 5.2 for further details),

efficacy analyses in the ITT population or in the HER2 IHC >0 <1+ subgroup may be performed for exploratory purposes.

Table 16 Formal Statistical Analyses to be Conducted and Pre-planned Sensitivity/Supportive Analyses of Efficacy Endpoints and PROs

Endpoint	Notes

PFS	**HER2-low** population (primary endpoint):
	• **Primary analysis**[a] using BICR data: stratified log-rank test
	• Sensitivity analyses using BICR data (stratified log-rank test unless otherwise specified):
	  ○ Analysis considering the midpoint between time of progression and previous RECIST assessment – evaluation-time bias
	  ○ Analysis using alternative censoring rules – attrition bias
	  ○ Analysis excluding patients with deviations affecting the efficacy of trial therapy – deviation bias[b]
	• Sensitivity analysis using Investigator data: stratified log-rank test and Kaplan-Meier estimates and 95% CI – ascertainment bias
	• Subgroup analyses and test for consistency of treatment effect between subgroups using BICR data: Cox regression
	ITT population (key secondary endpoint):
	• **Primary analysis**[a] using BICR data: stratified log-rank test and Kaplan-Meier estimates and 95% CI
	• Sensitivity and subgroup analyses as above described for HER2-low population
	HER2 IHC >0 <1+ population (interim futility analysis):
	• Kaplan-Meier estimates and 95% CI using Investigator data
	HER2 IHC >0 <1+ population:
	• Unstratified Cox regression using BICR and Investigator data
	• Kaplan-Meier estimates and 95% CI using BICR data and Investigator data

OS	**HER2-low** population (key secondary endpoint): • **Primary analysis**[a]: stratified log-rank test and Kaplan-Meier estimates and 95% CI • Sensitivity analysis with censoring indicator reversed – attrition bias: Kaplan-Meier estimates and 95% CI • Subgroup analyses and test for consistency of treatment effect between subgroups: Cox regression **ITT** population (key secondary endpoint):

Table 16 Formal Statistical Analyses to be Conducted and Pre-planned Sensitivity/Supportive Analyses of Efficacy Endpoints and PROs

Endpoint	Notes
	• **Primary analysis**[a]: stratified log-rank test and Kaplan-Meier estimates and 95% CI • Sensitivity and subgroup analyses as above described for HER2-low population **HER2 IHC >0 <1+** population: • Unstratified Cox regression • Kaplan-Meier estimates and 95% CI
ORR	**HER2-low** population, **ITT** population, **HER2 IHC >0 <1+** population: • Logistic regression on objective response by BICR • Same analysis on objective response by Investigator assessment and confirmed objective response by BICR and Investigator assessment, for sensitivity purposes **HER2 IHC >0 <1+** population (interim futility analysis): • Estimation of the difference between T-DXd and Investigator's choice chemotherapy in confirmed ORR by Investigator assessment • Supportive summaries of ORR by Investigator assessment
DoR	**HER2-low** population and **ITT** population: • KM plots and estimates of DoR by BICR • KM plots and estimates of DoR by Investigator assessment, for sensitivity purposes • Estimates of DoR (confirmed response) by BICR and Investigator assessment **HER2 IHC >0 <1+** population (interim futility analysis): • KM plots and estimates of DoR by Investigator assessment **HER2 IHC >0 <1+** population: • KM plots and estimates of DoR by BICR
DCR at 24 weeks	**HER2 IHC >0 <1+** population (interim futility analysis): • Descriptive statistics
PFS2	**HER2-low** population, **ITT** population: • Stratified log-rank test and KM plots **HER2 IHC >0 <1+** population: • Unstratified Cox regression test and KM plots
TFST and TSST	**HER2-low** population and **ITT** population: • Stratified log-rank test and KM plots

Table 16 Formal Statistical Analyses to be Conducted and Pre-planned Sensitivity/Supportive Analyses of Efficacy Endpoints and PROs

Endpoint	Notes
Change from baseline in EORTC QLQ-C30 and EORTC QLQ-BR45 scale scores	**HER2-low** population and **ITT** population: • MMRM
Time to deterioration in EORTC QLQ-C30 scores	**HER2-low** population and **ITT** population: • Stratified log-rank test and KM plots

BICR = blinded independent central review; DCR = disease control rate; DoR = duration of response; EORTC QLQ-BR45 = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire – Breast Module 45; EORTC QLQ-C30 = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire – Core 30; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; ITT = intent-to-treat; KM = Kaplan-Meier; MMRM = mixed model for repeated measures; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PFS2 = time from randomization to second progression or death; RECIST = Response Evaluation Criteria In Solid Tumors; T-DXd = trastuzumab deruxtecan; TFST = time to first subsequent treatment or death; TSST = time to second subsequent treatment or death.
[a] Primary and key secondary endpoints tested according the gatekeeping procedure defined in Section 4.2.1.
[b] To be performed only in the cases described in Section 2.2.

In order to further clarify which endpoints based on ORR, DoR and DCR at 24 weeks will be summarized or analyzed, a summary is provided in Table 17.

Table 17 Endpoints Based on ORR, DoR and DCR to be Summarized or Analyzed

Endpoint	HER2-low and ITT populations	HER2 IHC >0 <1+ population (interim futility analysis)	HER2 IHC >0 <1+ population
ORR	BICR, Investigator assessment	Investigator assessment	BICR, Investigator assessment
BoR	BICR, Investigator assessment	Investigator assessment	BICR, Investigator assessment
Confirmed ORR	BICR, Investigator assessment	Investigator assessment	BICR, Investigator assessment
Confirmed BoR	-[a]	Investigator assessment	-
DoR	BICR, Investigator assessment	Investigator assessment	BICR
DoR based on confirmed response	-	Investigator assessment	-

Table 17 Endpoints Based on ORR, DoR and DCR to be Summarized or Analyzed

Endpoint	HER2-low and ITT populations	HER2 IHC >0 <1+ population (interim futility analysis)	HER2 IHC >0 <1+ population
DCR at 24 weeks	-	Investigator assessment	-

BICR = blinded independent central review; BoR = best objective response; DCR = disease control rate; DoR = duration of response; HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; ITT = intent-to-treat; ORR = objective response rate.
[a] Confirmed BoR will be calculated to derive confirmed ORR, but it will only be presented in listings.

4.2.1 Multiplicity

Timing of PFS Testing

PFS in the HER2-low population and (depending on the results in the previous steps of the gatekeeping procedure and the interim futility analysis, see below for further details) in the ITT population will be tested once, when PFS reaches approximately 65% maturity (456 events) in the HER2-low population. This is estimated to occur 29 months after the first patient is randomized (4 months after randomization is completed) assuming a non-uniform accrual of patients with a duration of 25 months. At this time, it is expected that at least 553 events (65% maturity) will have been observed in the ITT population.

Timing of OS Testing

Depending on the results in the previous steps of the gatekeeping procedure and the interim futility analysis (see below for further details), OS in the HER2-low population and ITT population will be tested at two interim and one final analysis as described below:

- The first interim OS analysis will be performed at the time of the final PFS analysis. It is expected that 216 OS events (31% maturity or 41% information fraction) will have been observed in the HER2-low population and 263 OS events will have been observed in the ITT population.
- The second interim OS analysis will occur when approximately 392 OS events have been observed in the HER2-low population (56% maturity or 75% information fraction). This is anticipated to occur approximately 44 months after the first patient is randomized. It is expected that 477 OS events have been observed in the ITT population at this time.
- The final OS analysis will be performed when approximately 521 OS events have been observed in the HER2-low population (74% maturity), which is expected to occur

approximately 63 months after the first patient is randomized. At this time, it is estimated that 632 OS events will have been observed in the ITT population.

Multiple Testing Procedure

To strongly control the family-wise Type I error rate at 5% (2-sided) in terms of the primary and key secondary endpoints, a multiple testing procedure with the following gatekeeping strategy will be employed:

Step 1: Test PFS in the HER2-low population at a 5% alpha level. If significance is achieved, go to Step 2.

Step 2: Test PFS in the ITT population at a 1.5% alpha level and OS in the HER2-low population at the 3.5% alpha level. If PFS in the ITT population is significant, the 1.5% alpha will be recycled to OS in the HER2-low population. Similarly, if OS in the HER2-low population is significant at either the interim or final analysis, the 3.5% alpha will be recycled to PFS in the ITT population (i.e., PFS in the ITT population at the PFS analysis data cut off (DCO) will be tested at 5% alpha). If both PFS in the ITT population and OS in the HER2-low population are significant, go to Step 3.

Step 3: Test OS in the ITT population at a 5% alpha level.

The 3.5% initial alpha allocated to OS in the HER2-low population will be distributed between the two interim and final analyses using the Lan DeMets spending function that approximates the O'Brien Fleming alpha-spending approach (Lan and DeMets 1983). Under this procedure, the adjusted significance levels at the interim and final analyses are determined by the information fraction available at the time of analysis (i.e., exact number events observed), giving greater weight to analyses performed at the end of the study than those performed earlier. If the study continues to final analysis, the adjusted significance level at the final OS analysis will be based on the actual number of the events at the interim and final analysis, and the alpha already spent at the interim analyses.

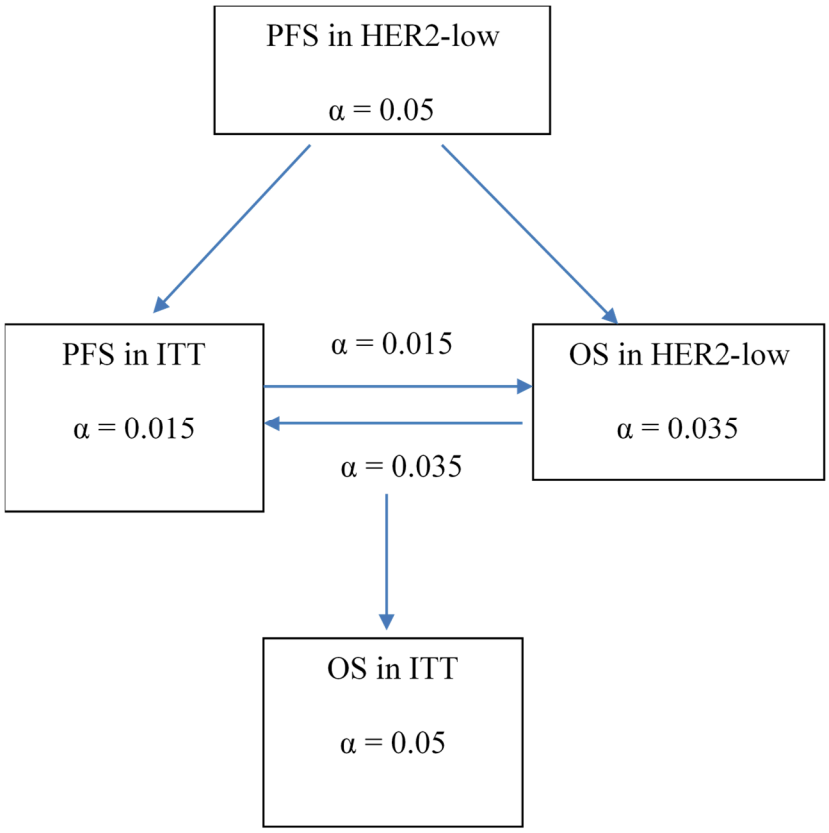
If PFS in the ITT population is significant at 1.5% alpha (Step 2 above) such that this alpha is reallocated to OS in the HER2-low population, the adjusted significance levels at each of the analysis time points for OS in the HER2-low population will be updated per the Group sequential Holm variable procedure (Ye et al 2013). The Lan DeMets spending function that approximates the O'Brien-Fleming alpha-spending approach will be used to derive the updated significance levels. The same significance levels will be used for the interim OS analyses in the HER2-low population and ITT population. The significance level at the final OS analysis will be derived separately for each population and will be based on the actual number of events at the

interim and final analysis, and the alpha already spent at the interim analyses of OS for a given population.

The gatekeeping procedure is also shown in Figure 2

Further details on the interim efficacy analyses are provided in Section 5.3.

Figure 2 Multiple Testing Procedure



HER2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; ITT = intent-to-treat; OS = overall survival; PFS = progression-free survival

Note: Tests on the ITT population will not be performed if the interim futility analysis in the HER2 IHC>0 <1+ population results in the cessation of enrollment of HER2 IHC>0 <1+ patients in this study.

Interim Futility Analysis

An interim futility analysis will be performed in the HER2 IHC >0 <1+ subgroup after the first 70 HER2 IHC >0 <1+ patients are randomized and have had at least 24 weeks of follow-up from the point of randomization, have discontinued from study treatment or withdrawn from the study. If it is decided to limit further recruitment only to the HER2-low population based on the results of the interim futility analysis, neither PFS nor OS will be tested in the ITT population, i.e., testing will only be performed in the HER2-low population. See Section 5.2 for further details.

4.2.2 Pooling Strategy

Stratified log-rank tests and Cox proportional hazards model will adjust for the following factors:

- CDK4/6 inhibitor use: yes vs. no.
- HER2 IHC expression: IHC 1+ vs. IHC 2+/ISH- in the HER2-low population, with the additional level IHC >0 <1+ to be considered in the ITT population.
- Prior taxane use in the non-metastatic setting: yes vs. no.

It is sensible to ensure there are at least 5 events per treatment group in each stratum (Silcocks 2012) (where a stratum is defined as strata1 * strata2 * strata3; so with 3 stratification factors each of 2 levels we have 2*2*2 = 8 strata).

The number of events in each stratum will be evaluated for the primary analysis of PFS by BICR in the HER2-low population (see Section 4.2.3.1 for further details). Stratification factors will be removed from the model in the following order, until at least 5 events per treatment group in each stratum are available:

- Prior taxane use in the non-metastatic setting.
- CDK4/6 inhibitor use.
- HER2 IHC expression.

All analyses will then be conducted in accordance with the pooling strategy defined for the primary analysis.

If there are secondary endpoints that still will not conform to the requirement on number of events above defined, unstratified log-rank tests and Cox proportional hazards model will be used for the analysis of the secondary endpoint. This will be supported by unstratified sensitivity analyses of the primary endpoint. Additional sensitivity analyses may also be required in this situation.

4.2.3 Primary Efficacy Endpoint

4.2.3.1 PFS by BICR in the HER2-low Population

The primary endpoint of the study is PFS by BICR according to RECIST 1.1 in the HER2-low population.

PFS distribution will be compared between T-DXd and Investigator's choice chemotherapy using a stratified log-rank test adjusting for prior CDK4/6 inhibitor use (yes vs. no), HER2 IHC expression (IHC 1+ vs. IHC 2+/ISH-), and prior taxane use in the non-metastatic setting (yes vs. no).

The HR (T-DXd vs. Investigator's choice chemotherapy) and its CI will be estimated from a stratified Cox proportional hazards model, based on the same stratification variables as for the log-rank test.

KM plots of PFS will be presented by treatment group. Summaries of the number and percentage of patients experiencing a PFS event, the type of event (RECIST 1.1 progression or death) and the number and percentage of censored patients and detailed reason for censoring will be provided along with median PFS and its CI for each treatment.

The treatment status at progression of patients at the time of analysis will be summarized. This will include the number (%) of patients who were on treatment at the time of progression, who discontinued study treatment prior to progression, who have not progressed and were on treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) vs. log (time) and, if these raise concerns, by fitting a time-dependent covariate (adding a treatment-by-time or treatment-by-log(time) interaction term) to assess the extent to which this represents random variation. Under non-proportional hazards, the HR can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found, this may be a result of treatment-by-covariate interactions, which will be investigated. The variation in treatment effect will be described by presenting piecewise HR calculated over distinct time-periods, using a Cox regression model including treatment and time-period factors in the model along with the treatment-by-time-period interaction term, stratified by the same factors as the primary analysis.

The Restricted Mean Survival Time (RMST) may also be performed up to the minimum of the largest observed event time in each of the two arms [or suitable clinically relevant timepoint], using the pseudo-values approach (Andersen et al. 2004), to estimate the RMST for each

treatment group with 95% CIs, and the difference in means between treatment groups, with 95% CIs and p-value.

Sensitivity Analyses

The following sensitivity analyses will be performed.

- Evaluation-time bias

  A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be analyzed using a stratified log-rank test, as described for the primary analysis of PFS. Note that midpoint values resulting in non-integer values should be rounded down. For patients whose death was treated as a PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of patient-level average inter-assessment times will be tabulated for each treatment.

- Attrition bias

  Attrition bias will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following two or more missed tumor assessments (see the definition provided in Table 10) will be included. Within the same sensitivity analysis, patients who take subsequent therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last evaluable RECIST assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

  As an additional supportive analysis, a Kaplan-Meier plot of the time to censoring using the PFS data from the primary analysis and where the censoring indicator of the PFS analysis is reversed will be presented.

- Ascertainment bias

  Ascertainment bias will be assessed by analyzing Investigator-reported data. The stratified log-rank test will be repeated on the programmatically derived PFS using the Investigator-reported data based upon RECIST. The HR and its CI will be presented. Further details are provided in Section 4.2.5.1.

If there is an important discrepancy between the primary analysis using the BICR data and this sensitivity analysis using Investigator data, then the proportion of patients with site but no central confirmation of progression will be summarized; such patients have the potential to induce bias in the central review due to informative censoring. An approach of imputing an event at the next visit in the central review analysis may help inform the most likely HR value (Fleischer et al 2011), but only if an important discrepancy exists.

Disagreements between Investigator and central reviews of RECIST progression will be presented for each treatment group. The summary will include the early discrepancy rate (frequency with which the Investigator declared progression earlier than BICR within each treatment group as a proportion of the total number of Investigator assessed progressions) and the late discrepancy rate (frequency with which Investigator declared progression later than BICR as a proportion of total discrepancies within the treatment group) (Amit et al 2011). A tolerance of 2 weeks will be considered when comparing the progression dates by BICR and Investigator assessment.

- Deviation bias

  In the cases described in Section 2.2, deviation bias will be assessed by repeating the PFS analysis excluding patients with deviations that may affect the efficacy of trial therapy.

A forest plot illustrating the HR and CI will be provided to compare the primary and sensitivity analyses of PFS.

Additional Supportive Summaries

In addition, the number of patients prematurely censored will be summarized by treatment group together with baseline potentially prognostic factors of the prematurely censored patients. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest evaluable scan prior to DCO was more than one scheduled tumor assessment interval plus 2 weeks (i.e., 6+2 weeks if time period between randomization and DCO for that patient is 48 weeks or less; 9+2 weeks otherwise) prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to DCO for all censored patients.

A summary of the duration of follow-up will be provided using median time from randomization to date of censoring (date last known to be non-progressor) in censored (not progressed) patients only, presented by treatment group.

Additionally, summary statistics for the number of weeks between the time of progression and the last evaluable RECIST assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who missed two or more consecutive RECIST assessments will be presented for each treatment group.

All of the collected RECIST 1.1 data by BICR will be listed for all randomized patients. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Subgroup Analyses

Subgroup analyses will be conducted comparing PFS between the treatments in the following subgroups of the HER2-low population:

- Sex (male vs. female).
- Age at randomization (<65 vs. ≥65 years of age).
- Race (White, Black/African-American, Asian, Other [Native Hawaiian/Pacific Islander or American Indian/Alaska Native or Others]).
- Region (Asia, North America, Europe, Rest of the World).
- HER2 IHC expression (IHC 1+ vs. IHC 2+/ISH-; of note, the additional level IHC >0 <1+ should be considered when performing the analysis in the ITT population).
- Prior CDK4/6 inhibitor use (yes vs. no).
- Prior taxane use in the non-metastatic setting (yes vs. no).
- Number of prior lines of endocrine therapy in the metastatic setting:
    - 1 prior line
        - 1 prior line and time to progression from start of ET + CDK4/6 inhibitors <= 6 months
        - Remaining patients with 1 prior line
    - 2 prior lines
    - >=3 prior lines
- Estrogen receptor (ER)/ progesterone receptor (PgR) status, according to two different categorizations:
    - ER+ PgR+, ER+ PgR-, ER- PgR+ (where + corresponds to the results “1-10%”, “>10%” and “Positive” and - corresponds to the results “<1%” and “Negative” recorded in the CRF).

  - ER: 1-10% vs. >10%.

- Endocrine resistance (primary vs. secondary).
- Investigator's choice chemotherapy (predefined prior to randomization in IRT):
  - Capecitabine
  - Paclitaxel
  - Nab-paclitaxel
- Prior anthracycline use (yes vs. no).
- Liver metastases (yes vs. no, based on the recorded site of disease in CRF DISEXT module - Extent of Disease Upon Entry to Study).

All subgroup analysis will be based on CRF data including the stratification factors (prior CDK4/6 inhibitor use and prior taxane use in the non-metastatic setting). Except, 'HER2 IHC expression' will be based on HER2 central lab data (the HER2 result from the most recent evaluable sample prior to randomization) and investigator's choice chemotherapy (predefined prior to randomization) entered in IRT.

Other baseline variables may also be assessed if there is clinical or biological justification, or an imbalance is observed between the treatment groups. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors. If a baseline imbalance is observed between treatment groups, ad-hoc subgroup analysis may be used to investigate any potential for impact on the main results.

No adjustment to the significance level will be made since all subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of PFS.

For each subgroup level, the HR and CI will be calculated from a Cox proportional hazards model (with no stratification) that only contains a term for treatment. The Cox models will be fitted using SAS® PROC PHREG, using a BY statement for the subgroup factor.

These HRs and associated CIs will be summarized and presented on a forest plot, along with the results of the overall primary analysis. Number and percentage of patients experiencing a PFS event, along with median PFS and its CI for each treatment will be also provided. The analysis on HER2 IHC >0 <1+ population will be performed as part of the subgroup analysis. For this subgroup, KM curves will also be provided.

If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events across both treatment groups in a subgroup), the HR and CI will not be produced for that subgroup. In this case, only number and percentage of patients with events within each treatment group will be provided.

Test for Consistency of Treatment Effect Between Subgroups

Interactions between treatment and stratification factors included in the analysis of the primary endpoint will also be tested to rule out any qualitative interaction using the approach of Gail and Simon (Gail and Simon 1985).

4.2.4 Key Secondary Efficacy Endpoints

4.2.4.1 PFS by BICR in the ITT Population

PFS by BICR according to RECIST 1.1 in the ITT Population will be analyzed as described for the primary endpoint in Section 4.2.3.1, except for the additional level IHC >0 <1+ for the stratification factor HER2 IHC expression.

The same sensitivity and subgroup analyses (including an assessment of consistency of treatment effect between subgroups) and additional supportive summaries as described for the primary endpoint in Section 4.2.3.1 will also be presented.

4.2.4.2 OS in the HER2-low Population

OS in the HER2-low population will be analyzed as described for the primary endpoint in Section 4.2.3.1. The 2-sided 95% CI and the (1-adjusted alpha) % CI (see Section 5.3 for further details on the adjusted significance level) of HR will be provided.

KM plots of OS will be presented by treatment group. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn main consent will be provided along with the median OS and its CI for each treatment.

The assumption of proportionality of hazards will be assessed as described for the primary endpoint in Section 4.2.3.1.

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias with regard to the primary treatment comparisons, achieved by a KM plot of time to censoring where the censoring indicator of OS is reversed.

The number of patients prematurely censored will be summarized by treatment group. A patient would be defined as prematurely censored if their survival status was not defined at the DCO.

In addition, duration of follow-up will be summarized using median time from randomization to the date of death (i.e., overall survival) or to the date of censoring (date last known to be alive) for censored patients by treatment group, and in all patients by treatment group and overall. .

The percentages of patients alive at 12 and 18 months, with 95% CI, will be reported.

The same subgroup analyses (including an assessment of consistency of treatment effect between subgroups) as described for the primary endpoint in Section 4.2.3.1 will also be presented.

Treatment Switching

Exploratory analyses of overall survival adjusting for the impact of treatment switching will be performed to inform decision-makers, including payers. Methods such as the rank preserving structural failure time (RPSFT), inverse probability of censoring weighting (IPCW), and two steps methods (Latimer et al 2014) will be explored. The final choice of methods will be based on numerous factors including, but not limited to, the completeness of data, the degree of treatment switching, maturity of data, whether switching occurs very early or later in the trial, and the plausibility of the underlying assumptions including the constant treatment effect for the rank preserving method and the exchangeability assumption for IPCW and two step methods. If the described methods are deemed to be insufficient to describe the disease and treatment process, other methods may need to be explored. Further details will be provided in the Payer Analysis Plan.

4.2.4.3 OS in the ITT Population

OS in the ITT Population will be analyzed as described for OS in the HER2-low population in Section 4.2.4.1, except for the additional level IHC >0 <1+ for the stratification factor HER2 IHC expression.

The same sensitivity and subgroup analyses (including an assessment of consistency of treatment effect between subgroups) and additional supportive summaries as described for OS in the HER2-low population in Section 4.2.4.1 will also be presented.

OS will also be analyzed in the HER2 IHC >0 <1+ population as part of the subgroup analysis of OS. KM curves will also be provided.

4.2.5 Secondary Efficacy Endpoints

4.2.5.1 PFS by Investigator Assessment

PFS by Investigator Assessment in the HER2-low and ITT Populations

PFS by Investigator assessment according to RECIST 1.1 in the HER2-low and ITT populations will be analyzed for sensitivity purposes (in order to assess ascertainment bias), as described in Section 4.2.3.1.

PFS by investigator assessment according to RECIST 1.1 will also be analyzed on the HER2 IHC >0 <1+ population as a supportive analysis. The HR and CI will be calculated from a Cox proportional hazards model with no stratification. No p-value will be displayed. KM curves will also be provided.

PFS by Investigator Assessment in the HER2 IHC >0 <1+ Population (Interim Futility Analysis)

The following summaries will be provided at the interim futility analysis for supportive purposes.

KM plots of PFS by Investigator assessment according to RECIST 1.1 in the HER2 IHC >0 <1+ population will be presented by treatment group.

4.2.5.2 ORR by BICR and Investigator Assessment

ORR by BICR and Investigator Assessment in the HER2-low and ITT Populations

ORR by BICR according to RECIST 1.1 based upon all patients in the HER2-low population will be compared between T-DXd and Investigator's choice chemotherapy using a logistic regression model adjusting for the same stratification factors as for the primary endpoint as covariates in the model. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favor T-DXd) together with its associated profile likelihood CI (i.e., using the option LRCI in SAS® PROC GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The same applies to ORR by BICR according to RECIST 1.1 in the ITT population, except for the additional level IHC >0 <1+ for the covariate HER2 IHC expression.

If there are not enough responses (i.e., at least 10 responses per variable, Peduzzi et al 1996) for a meaningful analysis using logistic regression, then a CMH test will be presented.

For sensitivity purposes, the above analyses will be replicated considering only patients in the population with measurable disease at baseline per BICR.

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR) for each treatment group along with the exact Clopper-Pearson 95% CIs based upon:

- the total number of patients in the population and
- the number of patients in the population with measurable disease at baseline per BICR (sensitivity analysis).

For each treatment group, BoR will be summarized by n (%) for each category (CR, PR, SD, PD and NE). No formal statistical analyses are planned for BoR.

For sensitivity purposes, the above analyses will be replicated by considering:

- ORR along with the exact Clopper-Pearson 95% CIs and BoR by Investigator assessment;
- confirmed ORR along with the exact Clopper-Pearson 95% CIs by BICR;
- confirmed ORR along with the exact Clopper-Pearson 95% CIs by Investigator assessment.

Of note, in the analyses based on Investigator assessment patients with measurable disease at baseline will be identified based on Investigator assessment.

Overall visit response data will be listed for all patients (i.e., the ITT population).

The analysis on ORR, confirmed ORR and BoR by BICR and Investigator assessment performed for the HER2-low and ITT population will be also performed (without any stratification factor) on the HER2 IHC >0 <1+ population pertains to ORR primary analysis for supportive purposes. No p-value will be displayed.

ORR by Investigator Assessment in the HER2 IHC >0 <1+ Population (Interim Futility Analysis)

Confirmed ORR along with the exact Clopper-Pearson 95% CIs by Investigator assessment according to RECIST 1.1 will be summarized by treatment group using the HER2 IHC >0 <1+ population. The percentage of patients with confirmed ORR will be based on the total number of patients in the HER2 IHC >0 <1+ population. The difference between T-DXd and Investigator's choice chemotherapy and the corresponding 95% confidence interval using the Miettinen-Nurminen approach (Miettinen 1985) will be calculated.

The observed difference in confirmed ORR will guide the decision of whether to limit recruitment to the HER2-low subgroup only. Further details are provided in Section 5.2.

Summaries will be produced that present the number and percentage of patients with a confirmed tumor response (CR/PR) based upon:

- the total number of patients in the HER2 IHC >0 <1+ population and
- the number of patients in the HER2 IHC >0 <1+ population with measurable disease at baseline per Investigator assessment.

For each treatment group, confirmed BoR will be summarized by n (%) for each category (CR, PR, SD, PD and NE).

For supportive purposes, the above summaries will be replicated by considering ORR and BoR (with no confirmation required) by Investigator assessment.

4.2.5.3 DoR by BICR and Investigator Assessment

DoR by BICR and Investigator Assessment in the HER2-low and ITT Populations

Descriptive data will be provided for the DoR (with no confirmation of response required) and the confirmed DoR, by BICR and, for sensitivity purposes, by Investigator assessment according to RECIST 1.1 in responding patients in the HER2-low and ITT populations, including the associated KM curves and estimates.

The summary for DoR by BICR presented for the HER2-low and ITT population will also be presented in the HER2 IHC >0 <1+ population for supportive purposes.

DoR by Investigator Assessment in the HER2 IHC >0 <1+ Population (Interim Futility Analysis)

In the HER2 IHC >0 <1+ population, the same summaries as in the HER2-low and ITT populations will be provided for DoR based on response (with no confirmation required) and confirmed response by Investigator assessment at the interim futility analysis for supportive purposes.

Of note, all the summaries of DoR will be based on all patients in the population (HER2-low, ITT or HER2 IHC >0 <1+), irrespective of disease measurability at baseline.

4.2.5.4 DCR at 24 Weeks by Investigator Assessment

DCR at 24 Weeks by Investigator assessment will be summarized using descriptive statistics by treatment group in the HER2 IHC >0 <1+ population at the interim futility analysis for supportive purposes.

Of note, all the summaries of DCR at 24 Weeks will be based on all patients in the HER2 IHC >0 <1+ population, irrespective of disease measurability at baseline.

4.2.5.5 PFS2

Time from randomization to second progression or death (PFS2) will be analyzed in the HER2-low and ITT populations using identical methods as outlined for the primary endpoint in Section 4.2.3.1 and adjusting for the same set of covariates, but no subgroup analysis will be performed. The HR for the treatment effect together with its CI will be presented. Medians and their CIs and KM plots will be presented to support the analysis.

The number and percentage of patients experiencing a PFS2 event and the type of progression (objective radiological progression, symptomatic progression in absence of objective radiological progression or other) will also be summarized by treatment group, as well as summaries of deaths in the absence of second progression, and categories of PFS2 censoring. Time from randomization to second progression will be summarized by treatment group.

The analysis will be also performed (without any stratification factor) on the HER2 IHC >0 <1+ population as supportive. The p-value will not be generated.

4.2.5.6 TFST and TSST

For supportive purposes, the time to the start of first subsequent therapy or death (TFST) and to the start of the second subsequent therapy or death (TSST) will be analyzed in the HER2-low and ITT populations using the same methodology and model as that used for the analysis of PFS and described in Section 4.2.3.1. The HR for the treatment effect together with its CI will be presented. In addition, medians and their CIs and a KM plot of the time to the start of subsequent therapy will be presented by treatment group and the time between progression and starting subsequent therapy will be assessed. This will be summarized by treatment group, but no formal comparisons will be made.

4.2.6 Patient Reported Outcomes

PROs are not part of the main multiple testing procedure and will be analyzed as supportive endpoints. Secondary PRO endpoints include change from baseline in EORTC QLQ-C30 and EORTC QLQ-BR45 scale scores, and time to deterioration in all the scales of the EORTC QLQ-C30. Only EORTC QLQ-C30 scale on financial difficulties will not be included in any of the summaries. The outcomes from EQ-5D-5L, PRO-CTCAE, PGIS, PGIC, PGI-TT and PGI-BR are exploratory endpoints.

All the PROs will be analyzed in the HER2-low and ITT populations. Of note, if enrolment in the HER2 IHC >0 <1+ subgroup will be stopped based on the results of the interim futility analysis (see Section 5.2 for further details), analyses of PROs in the ITT population or in the HER2 IHC >0 <1+ subgroup may be performed for exploratory purposes.

4.2.6.1 EORTC QLQ-C30

Summaries of compliance over time by time point and overall will be reported.

Descriptive statistics of EORTC QLQ-C30 scale scores, change from baseline and response (improvement / no change / deterioration) at each scheduled time point will be provided by treatment group. Graphical plots may also be produced, as appropriate.

Change from Baseline in Scale Scores

Change from baseline in EORTC QLQ-C30 scale scores will be analyzed using a mixed model for repeated measures (MMRM). The analysis will make use of all data from baseline up to a selected timepoint cutoff in months. Data cut off will be the date Investigator records PD (as per data collected on SUPPRS) or 7 months after randomization whichever is earlier.. The analysis will be to compare the average treatment effect over the analyzed period, excluding time points with excessive missing data (defined as >75% missing data).

A generic visit variable will be derived for each patient in order that the average treatment effect can be analyzed using the above method. Each visit will be assigned a sequential number. Assessments will be assigned to the visits according to the approach defined in Section 4.1.5. The time from randomization to each of these will be derived to select only those visits occurring within the timepoint cutoff.

The MMRM model will include treatment, visit, treatment-by-visit interaction, the stratification factors as fixed effects, baseline score as a covariate and the baseline-by-visit interaction. Restricted maximum likelihood estimation will be used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each visit equal weight. Adjusted mean estimates by treatment group, by visit and overall, and corresponding CIs will be presented along with an estimate of the treatment difference, CI and p-value. Adjusted mean estimates by treatment group and by visit, and corresponding CIs will be presented also graphically.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: unstructured correlations, Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry.

Time to Deterioration

For each EORTC QLQ-C30 scale, time to deterioration will be analyzed as described for the primary endpoint in Section 4.2.3.1, except for the additional level IHC >0 <1+ for the stratification factor HER2 IHC expression to be considered in the analysis on the ITT population.

For each scale, time to deterioration will be presented using a KM plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration and the median time to deterioration will also be provided by treatment group.

The HR and CI for each scale will be presented graphically on a forest plot.

Best Overall Response

Summary tables of the best overall response will be provided by treatment group.

4.2.6.2 EORTC QLQ-BR45

EORTC QLQ-BR45 will be analyzed as described for EORTC QLQ-C30 in Section 4.2.6.1, except for the analysis of time to deterioration (not performed).

4.2.6.3 EQ-5D-5L

Summaries of compliance over time by time point and overall will be reported.

Descriptive statistics will be reported for health state utility values and the visual analogue scale and their changes from baseline by time point, for each treatment group.

To support future economic evaluations, additional appropriate analyses may be undertaken (e.g., mean health state utility pre- and post-treatment, and pre- and post-progression) and will be outlined in the Payer Analysis Plan.

4.2.6.4 PRO-CTCAE, PGIS, PGIC, PGI-TT and PGI-BR

Summaries of compliance over time by time point and overall will be reported for each PRO, where applicable.

Graphical representation of PRO-CTCAE responses will be provided, presenting percentages of subjects reporting symptom by attribute and level, time point, and treatment group. In addition, percentages of subjects with maximum (worst) level symptom and attribute by treatment group will also be presented. Descriptive statistics will be reported for PGI-TT response by time point and treatment group.

PGIS, PGIC and PGI-BR will be listed using the ITT population, while PRO-CTCAE and PGI-TT will be summarized and listed on the SAF.

4.2.7 Health Care Resource Use

Data on health care resource use will be listed. To support submissions to payers, additional analyses may be undertaken, and these will be outlined in a separate Payer Analysis Plan.

4.2.8 Safety

Data from all cycles of treatment will be combined in the presentation of safety data.

Safety summaries will be provided using the SAF. Selected safety summaries may be provided using HER2-low SAF and HER2 IHC >0<1+ SAF.

Of note, if enrolment in the HER2 IHC >0 <1+ subgroup will be stopped based on the results of the interim futility analysis (see Section 5.2 for further details), safety summaries in the SAF or in the HER2 IHC >0 <1+ SAF subgroup may be provided for exploratory purposes.

At the time of PFS analysis, summaries of AEs and exposure will be provided in the HER2 IHC >0 <1+ population. Additional summaries of laboratory parameters, vital signs, and ECG may be provided in this subgroup.

Safety summaries will be descriptive only. No formal statistical analyses will be performed on the safety variables.

The following sections describe the planned safety summaries. However, additional safety summaries may need to be produced to aid interpretation of the safety data.

4.2.8.1 Adverse Events

AEs occurring prior to start of study treatment will be included in the AE listings, but not in the summary tables.

TEAEs (as defined in Section 3.5.3) that occurred until 47 days after the last dose of the study treatment and before the initiation of the first subsequent cancer therapy will be used for reporting in all AE summary tables, unless otherwise specified. In this definition, palliative radiotherapy is not considered a subsequent cancer therapy. This will more accurately depict safety findings attributable to study treatment only as opposed to presenting all TEAEs, as they may include AEs likely to be attributable to subsequent cancer therapy.

All other TEAEs and post-treatment AEs will be included in the AE listings, but not in the summary tables. AEs occurring 48 days or more after the last dose of study treatment and/or after a patient has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

However, to assess the longer-term toxicity profile, some of the AE summaries may also be produced including AEs observed up until 47 days after following last dose of study treatment, irrespective of initiation of subsequent cancer therapies.

Adverse Events

All AEs, both in terms of MedDRA preferred term and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment group.

All reported AEs will be listed along with the date of onset (including study day), date of resolution (if AE is resolved), Investigator's assessment of CTCAE grade and relationship to study drug.

Summary information (number and percentage of patients reporting at least one event by system organ class and preferred term separated by treatment group) will be tabulated for:

- All AEs (*).
- Most common AEs.
- AEs causally related to any study treatment (as determined by the reporting Investigator) (*).
- AEs by maximum reported CTCAE grade.
- AEs with CTCAE grade 3 or higher (*).
- Most common AEs with CTCAE grade 3 or higher.
- AEs with CTCAE grade 3 or higher, causally related to any study treatment (as determined by the reporting Investigator) (*).
- AEs with outcome of death (*).
- AEs with outcome of death, causally related to any study treatment (as determined by the reporting Investigator) (*).
- SAEs (*).
- SAEs causally related to any study treatment (as determined by the reporting Investigator) (*).
- AEs leading to discontinuation of study medication (*).
- AEs leading to discontinuation of study medication, causally related to any study treatment (as determined by the reporting Investigator) (*).
- AEs leading to interruption of study medication (*).

- AEs leading to interruption of study medication, causally related to any study treatment (as determined by the reporting Investigator) (*).
- AEs leading to dose reduction of study medication (*).
- AEs leading to dose reduction of study medication, causally related to any study treatment (as determined by the reporting Investigator) (*).

For selected categories (*), overall summaries of the number and percentage of patients reporting at least one event in each category and of the number of episodes in each category will be presented.

For the AE tables of most common AEs, all events that occur more than 2% of patients in any treatment group will be summarized by preferred term, by decreasing total frequency (the total column will not be displayed in the AE tables). This cut-off may be modified after review of the data. When applying a cut-off (e.g., 2%), the raw percentage should be compared to the cut-off, no rounding should be applied first.

To adjust for the difference in exposure between the 2 treatment arms, the exposure-adjusted incidence rate (EAIR) will be calculated and summarized for AEs, ≥Grade 3 AEs, SAEs, and AESIs. EAIR is defined as the number of patients with at least one event divided by the total time at risk. Time at risk will be calculated as the number of days of exposure to drug, computed as per Section 3.5.1, summed over all patients.

Deaths

A summary of all deaths (**note:** all deaths on/after date of first dose of the study drug will be considered) will be provided with number and percentage of patients, categorized as:

- Total number of deaths (regardless of date of death).
- Death related to disease under investigation only as determined by the Investigator.
- AE with outcome of death only.
- AE with outcome of death only (AE start date falling after 47-day follow up period) (*).
- Number of patients with death related to disease under investigation (as determined by the Investigator) and an AE with outcome of death.
- Other deaths.

This summary will be repeated for all deaths on-treatment or within 47 days of last dose of study medication. The category marked (*) will not appear in this summary.

Adverse Events of Special Interest

ILD/pneumonitis and left ventricular dysfunction are considered to be AESIs (see Section 3.5.3 for further details).

Summaries of ILD/pneumonitis events will be based on adjudicated drug related ILD/pneumonitis events from the ILD adjudication committee.

Summaries of adjudicated ILD/pneumonitis and left ventricular dysfunction events will include number (%) of patients who have:

- At least one AESI.
- At least one AESI causally related to any study treatment (as determined by the reporting Investigator for left ventricular dysfunction and by the ILD adjudication committee for adjudicated ILD/pneumonitis events).
- At least one AESI by maximum reported CTCAE grade (*).
- At least one AESI by outcome (*).
- At least one serious AESI (*).
- At least one AESI leading to discontinuation of study medication (*).
- At least one AESI leading to interruption of study medication (*).
- At least one AESI leading to dose reduction of study medication (*).

An overall AESI summary will be presented, including the number and percentage of patients in each of these categories. When summarizing adjudicated ILD/pneumonitis events, the categories marked (*) will only be presented for adjudicated drug-related events.

A summary table presenting number (%) of patients who had potential ILD/pneumonitis sent for adjudication will be presented. This will include summaries by outcome and maximum reported CTCAE grade.Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI grouping. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

Time to the first treatment-emergent AESI will be summarized for patients with at least one AESI using descriptive statistics (mean, standard deviation, median, minimum, maximum).

When summarizing time to the first treatment-emergent AESI for adjudicated ILD/pneumonitis events, only adjudicated drug-related events will be considered.

Infusion-Related Reactions

A listing of infusion-related reaction (IRR) will be provided. The definition of IRR includes any AEs occurred on the same or next day of infusion considered by the investigator to be related to study drug and with one of the following PTs: administration related reaction, anaphylactic reaction, hypersensitivity, infusion related hypersensitivity reaction, and infusion related reaction.

4.2.8.2 Laboratory Parameters

Summaries of laboratory parameters will be based on data collected from date/time of the first dose of study treatment up to 47 days following last dose of study treatment and before the initiation of the first subsequent cancer therapy (palliative radiotherapy is not considered a subsequent cancer therapy), whichever occurs first. The same rationale provided in Section 4.2.8.1 for the selection of AEs to be included in summary tables applies.

Data summaries will be provided in preferred units.

Continuous laboratory parameters will be summarized at baseline and over time in terms of absolute values and changes from baseline at each scheduled measurement by treatment group. The minimum and maximum absolute value and change from baseline will be also summarized by treatment group.

Shift tables for laboratory values by worst CTCAE grade will be produced, and for specific parameters separate shift tables indicating directionality of change (i.e., change to low or high) will be produced. The laboratory parameters for which CTCAE grade shift outputs will be produced are:

- **Hematology:** haemoglobin (low), platelets (low), leucocytes (low), lymphocytes (absolute count, low), neutrophils (absolute count, low).
- **Clinical chemistry:** albumin (low), total bilirubin (high), corrected calcium (low and high), creatinine (high), magnesium (low and high), potassium (low and high), sodium (low and high), alkaline phosphatase (high), alanine aminotransferase (ALT) (high), aspartate aminotransferase (AST) (high).
- **Coagulation:** activated partial thromboplastin time (high).

For the parameters with no CTCAE grading that are listed in the CSP Tables 12 and 13, shift tables from baseline to minimum and maximum values will be provided.

Denominators used in laboratory summaries will only include evaluable patients. For example, if a CTCAE criterion involves a change from baseline, evaluable patients are those who have both

a pre-dose and at least 1 post-dose value recorded. If a CTCAE criterion does not consider changes from baseline, evaluable patients are those who have at least 1 post-dose value recorded.

The following figures may be produced for certain parameters if warranted after data review:

- Scatter plots (shift plots) of baseline to minimum/maximum value (as appropriate).
- Box-plots of absolute values by scheduled time point, and box-plots of change from baseline by scheduled time point, may be presented for certain parameters if warranted after data review.

Supportive laboratory listings will cover observed values and changes from baseline for each individual patient as well as abnormalities.

Liver Enzyme Elevations and Hy's Law

Summaries will be provided including the number (%) of patients who have elevated ALT, AST, and total bilirubin during the study:

- **ALT:** ≥3x – ≤5x, >5x – ≤8x, >8x - ≤10x, >10x - ≤20x, and >20x ULN.
- **AST:** ≥3x – ≤5x, >5x – ≤8x, >8x - ≤10x, >10x - ≤20x, and >20x ULN.
- **Total bilirubin:** >1.5x - <2x, ≥2x - ≤3x, >3x - ≤5x, and >5x ULN.
- **ALT or AST:** ≥3x - ≤5x, >5x - ≤8x, >8x - ≤10x, >10x - ≤20x, and >20x ULN.
- **Potential Hy's law:** ALT or AST ≥3x ULN and total bilirubin ≥2x ULN. The onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.

Liver biochemistry test results over time for patients with elevated ALT or AST (i.e., ≥ 3x ULN), and elevated total bilirubin (i.e., ≥ 2x ULN) (at any time) will be plotted. Individual patient data where ALT or AST (i.e., ≥ 3x ULN) plus total bilirubin (i.e., ≥ 2x ULN) are elevated at any time will be listed also.

4.2.8.3 Vital Signs

Summaries of vital signs will be based on data selected as defined in Section 4.2.8.2 for laboratory parameters.

Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, temperature, respiratory rate and weight) will be summarized at baseline and over time in terms of absolute values and changes from baseline at each scheduled measurement by treatment group. The minimum and maximum absolute value and change from baseline will be also summarized by treatment group.

4.2.8.4 ECG

Summaries of ECG will be based on data selected as defined in Section 4.2.8.2 for laboratory parameters.

The number and percentage of patients with normal, borderline, and abnormal (not clinically significant and clinically significant) ECG results will be presented as a shift table from baseline to worst evaluation during the study by treatment group.

4.2.8.5 SpO2

Summaries of SpO2 will be based on data selected as defined in Section 4.2.8.2 for laboratory parameters.

SpO2 will be summarized at baseline only.

4.2.8.6 LVEF

Summaries of LVEF will be based on data selected as defined in Section 4.2.8.2 for laboratory parameters.

LVEF will be summarized by treatment group in terms of:

- Absolute values and number and percentages of patients with LVEF 40-49%, 20-39% and <20% at baseline and at end of treatment, and as minimum and maximum post-baseline value.
- Change from baseline and number and percentages of patients with 10%-19% decrease, >=20% decrease, 10%-19% increase and >=20% increase at end of treatment, and as minimum and maximum post-baseline value.
- Number and percentages of patients with worst LVEF CTCAE grade post-baseline, where per CTCAE:
    - Grade 2: Resting LVEF ≥40% to <50%; 10% to <20% decrease from baseline.
    - Grade 3: Resting LVEF ≥20% to <40%; ≥20% decrease from baseline.
    - Grade 4: Resting LVEF <20%.

Decreased is defined as a drop in absolute value versus baseline. All data collected will be listed.

4.2.8.7 ECOG Performance Status

Summaries of ECOG performance status will be based on data selected as defined in Section 4.2.8.2 for laboratory parameters.

ECOG performance status will be summarized at baseline only.

4.2.9 Demographics and Baseline Characteristics

The following will be summarized for all patients in the HER2-low and ITT populations (unless otherwise specified) by treatment group:

- Patient disposition (including screening failures) (all patients).
- Important protocol deviations.
- Inclusion in analysis sets.
- Demographics (age at randomization, age group [<65, ≥65 years], sex, race and ethnicity).
- Patient characteristics at baseline (height, weight, BMI).
    - BMI will be calculated as: weight (kg)/height $(m)^2$.
- Patient recruitment by region (Asia, North America, Europe, Rest of the World), country and center.
- Previous disease-related treatment modalities.
- Disease characteristics at time of diagnosis (primary tumor location, tumor laterality, histology type, tumor grade and American Joint Committee on Cancer (AJCC) stage).
- Disease characteristics at baseline (ER, PgR and HER expression, measurable disease [yes, no], bone-only disease [yes, no], endocrine resistance, ECOG performance status).
- Extent of disease at baseline.
- Time from diagnosis, from most recent progression and from completion of previous anticancer therapy to randomization.
- Stratification factors as per IRT and CRF data.
- Pulmonary function test at baseline.
- Medical history (past and current).
- Relevant surgical history.
- Disallowed concomitant medications.
- Allowed concomitant medications.
- Post-discontinuation cancer therapy.

Similarly, patient disposition, demographics characteristics, baseline and disease characteristics will also be summarized for patients in the HER2 IHC >0 <1+ population at the time of PFS analysis.

4.2.10 Concomitant and Other Treatments and Procedures

Concomitant and other treatments will be summarized in the HER2-low and ITT populations (unless otherwise specified) by treatment group.

Information on any treatment within the 28 days prior to initiation of study drug and all concomitant treatments given up to 40 days (+7 days) after discontinuation of study treatment, with reasons for the treatment, will be recorded in the CRF. Thereafter, only subsequent regimens of anti-cancer therapy will be recorded in CRF.

Other anti-cancer therapies and investigational agents should not be given while the patient is on study drug. Palliative radiotherapy is permitted, except for palliative radiotherapy to the chest area in patients treated with T-DXd.

Medications received prior to, concomitantly or post-treatment will be coded using the AstraZeneca Drug Dictionary Anatomical Therapeutic Chemical (ATC) Classification codes. Concomitant medications will be summarized by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant medication or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 4.1.6.

Prior medications, concomitant and post-randomized treatment medications are defined based on imputed start and stop dates as follows:

- Prior medications are those taken prior to study treatment with a stop date prior to the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment (and could have started prior to or during treatment).
- Post-treatment medications are those with a start date after the last dose date of study treatment.

In addition, all post-treatment anti-cancer medications and surgical procedures will be summarized.

The following summaries will be produced:

- Summary of allowed and disallowed concomitant medications (based on medical review on collected medications data before unblinding).
- Summary of post study treatment cancer therapies.

Prior, concomitant and post study treatment radiotherapies will be summarized as well.

All concomitant and other treatment and procedure data will be listed.

Missing coding terms should be listed and summarized as "Not coded".

4.2.11 Exposure

Treatment exposure will be summarized for the HER2-low SAF and SAF (unless otherwise specified) by treatment group. The following summaries will be produced:

- Summary of total exposure.
- Summary of actual exposure (T-DXd).
- Summary of number of cycles received.
- Summary of number of infusions received (T-DXd, paclitaxel, nab-paclitaxel).
- Summary of dose interruptions (including duration for capecitabine only) and reductions and cycle delays (for SAF only).
- Summary of RDI (for SAF only).

For patients on study treatment at the time of the interim analysis, the DCO date will be used to calculate exposure.

4.2.12 Pharmacokinetic Data

PK data will be summarized for PK analysis set. The same summaries will be presented for the HER2 IHC >0 <1+ PK set for supportive purposes.

If any deviation is considered to have impact upon PK, a patient or particular data for a patient may be excluded from the PK analysis set.

Serum concentrations of T-DXd, total anti-HER2 antibody and MAAA-1181a will be summarized by nominal sample time using standard summary statistics for PK concentrations (geometric mean, geometric standard deviation, geometric coefficient of variation, arithmetic mean, standard deviation, median, minimum, maximum and n).

All plasma concentrations will be listed.

4.2.13 Immunogenicity Data

Immunogenicity data will be summarized for ADA evaluable set, while listings will be based on SAF. Additional summaries will be presented for the HER2 IHC >0 <1+ ADA evaluable set for supportive purposes.

Immunogenicity results will be listed by patient, and a summary, inclusive of minimum, maximum, and median of maximum post-baseline titers, will be provided by the number and percentage of patients who develop ADA for T-DXd, based on the categories defined in Section 3.7.. The immunogenicity titer and neutralizing ADA data will be listed for samples confirmed positive for the presence of T-DXd antibodies.

The effect of immunogenicity as well as the effect of its neutralizing properties on PK, efficacy, and safety may be evaluated, if the data allow.

4.2.14 Impact of COVID-19

A listing will be provided for patients with reported issues in the Clinical Trial Management System due to coronavirus disease 2019 (COVID-19) pandemic.

A summary table reporting the number of patients in any AE category will be provided for those with confirmed/suspected COVID-19 infection, defined as any AE occurring during the pandemic timeframe (from March 11th 2020, to last known alive date) with PT within the AE search criteria developed by the latest MedDRA Maintenance and Support Services Organization (MSSO) guideline for COVID-19.

Additional analyses may be performed to explore the impact of COVID-19 and implemented contingency measures (e.g., patients discontinued from study treatment and/or study, alternative procedures used to collect critical safety and/or efficacy data, protocol deviations related to COVID-19) on the safety and efficacy results reported for the study. The following may be explored:

- sensitivity analyses for key safety and efficacy endpoints by excluding data from patients from sites that are closed out due to COVID-19;
- safety and laboratory data analysis separately for patients affected by COVID-19 and deaths or adverse events summary attributed to COVID-19 pandemic.

5. INTERIM ANALYSES

5.1 IDMC

An independent data monitoring committee (IDMC) comprised of independent experts will be convened and will meet approximately 6 months after the study has started or after the first 60 patients have been randomized, whichever occurs first. The IDMC will review unblinded safety data and make recommendations to continue, amend or stop the study based on safety findings. The committee will meet approximately every 6 months thereafter and at each meeting make recommendations to continue, amend or stop the study based on safety findings.

In addition, the IDMC will be asked to review:

- Interim futility analysis for HER2 IHC >0 <1+ subgroup (see Sections 4.2.1 and 5.2 for further details).

As AZ study team will be fully unblinded at the time of final PFS analysis, the IDMC will not be required to address specific questions regarding final PFS or interim OS data. However, final PFS and interim OS data will be shared with IDMC.

Full details of the IDMC procedures, processes, and interim analyses can be found in the IDMC Charter.

5.2 Interim Futility Analysis

An interim futility analysis will be carried out after 70 patients have been randomized in the HER2 IHC >0 <1+ subgroup (approximately 35 per treatment group) and have had at least 24 weeks of follow-up from the point of randomization, have discontinued from study treatment or withdrawn from the study. This is expected to occur approximately 20 months after the first patient is randomized.

A non-binding stopping boundary will be applied to guide the decision of whether to limit recruitment to the HER2-low subgroup only. Confirmed ORR by Investigator assessments according to RECIST 1.1 in the HER2 IHC >0 <1+ analysis set will be used in the interim futility analysis. If the observed ORR difference (T-DXd vs. Investigator's choice chemotherapy) in the HER2 IHC >0 <1+ analysis set is ≤ -7%, the recommendation will be to stop enrolment in the HER2 IHC >0 <1+ subgroup.

The stopping boundary was selected such that the probability of crossing the futility boundary is at most 10% if the true ORR difference (T-DXd vs. Investigator's choice chemotherapy) is at least 8%, assuming a 30% ORR in the chemotherapy arm. Furthermore, the probability of crossing the futility boundary is 7% if the true ORR difference is 10%, and 61%, if the true ORR difference is -10%, assuming a 30% ORR in the chemotherapy arm.

As supportive summaries, ORR, DoR, DCR at 24 weeks and PFS KM plot per Investigator assessment will also be provided using the HER2 IHC >0 <1+ analysis set. Patient disposition, demographics characteristics, baseline and disease characteristics, medical/surgical history will also be provided for the HER2 IHC >0 <1+. Exposure and safety will also be provided for the HER2 IHC >0 <1+ SAF analysis set.

The IDMC will make a recommendation on whether or not to stop recruitment in the HER2 IHC >0 <1+ population based on the results of the futility analysis. If recruitment in this subgroup is permanently stopped, patients already dosed may be allowed to continue treatment based on a case-by-case assessment at the discretion of the Investigator, and if the Investigator deems that a patient continues to derive clinical benefit from treatment.

5.3 Interim Efficacy Analyses

OS in the HER2-low population and ITT population will be tested at two interim analyses as described in Section 4.2.1.

The adjusted significance level for the OS analyses will be determined according to the following principles:

- **Interim analyses**:

  - Using the Lan DeMets spending function that approximates the O'Brien Fleming alpha-spending approach, the significance levels for the interim OS analyses in the HER2-low population will be determined based on the information fraction available. This is calculated as the number of OS events observed at the interim look divided by the expected total number of events at the time of the final analysis (i.e., 521 in the HER2-low population).

  - If PFS in ITT population is not significant, the significance levels for the interim OS analyses in the HER2-low population will be calculated from an overall Type I error rate of 3.5%. OS will not be tested in the ITT population.

  - If PFS in ITT population is significant, the significance levels for the interim OS analyses in the HER2-low population will be calculated from an overall Type I error rate of 5%. The same significance levels will be used for the interim OS analyses in the ITT population.

- **Final analysis**: The adjusted significance level will be derived separately for the HER2-low and ITT population. This will be obtained based on the actual number of events at the interim and final analyses, and the alpha spent at interim analyses, in order to control the cumulative Type I error rate at 3.5% or 5% (2-sided) in the HER2-low population depending on the PFS result in the ITT population, and at 5% (2-sided) in the ITT population.

An example for the adjusted significance level to be considered at the interim and final OS analyses (based on the expected number of events in the HER2-low population) is provided in Table 18.

Table 18 Example: Significance Level for Interim and Final OS Analyses

Analysis	Allocated Alpha	Number of Events	Information Fraction	2-sided Adjusted Significance Level
First interim	3.5%	216	41%	0.0004
Second interim		392	75%	0.0122
Final		521	100%	0.0312
First interim	5%	216	41%	0.0010 [a]
Second interim		392	75%	0.0192 [a]
Final		521	100%	0.0441

[a] If PFS in ITT population is significant, the same significance levels will be used for the interim OS analyses in the ITT population.

6. CHANGES OF ANALYSIS FROM PROTOCOL

- The wording “best objective response” according to RECIST 1.1 is used instead of “best overall response”.
- Analyses of ECG parameters, SpO2, LVEF and ECOG performance status have been specified.
- Additional analyses assessing the impact of COVID-19 to be potentially performed have been specified.
- The same significance levels will be used for the interim OS analyses in the HER2-low population and ITT population. In the protocol, the significance levels for the two populations will be derived separately, both using the Lan DeMets spending function that approximates the O'Brien Fleming alpha-spending approach.
- Additional analyses will be performed for interim futility analysis using HER2 IHC >0 <1+ on patient disposition, demographics characteristics, baseline and disease characteristics, medical/surgical history. New defined analysis set HER2 IHC >0 <1+ SAF will be performed on exposure and safety.
- Additional supportive analyses will be performed at the time of the PFS primary analysis on the HER2 IHC >0 <1+ population on patient disposition, demographics characteristics, baseline, and disease characteristics, PFS, OS, ORR, BoR, DoR, PFS2, exposure, AEs, PK, and ADA.

- The definition of ADA evaluable set has been changed from “All patients who receive at least 1 dose of T-DXd per the protocol who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA results.” to “All patients who receive at least 1 dose of T-DXd who have a non-missing ADA result at any time.”.

7. COUNTRY SPECIFIC ANALYSES

Any country specific analyses will be documented in a separate analysis plan.

8. REFERENCES

Aaronson et al 1993
Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365-76.

Andersen et al 2004
Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime Data Anal. 2004 Dec;10(4):335-50.

Amit et al 2011
Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J, Burger HU. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer. 2011;47(12):1772-8.

Bjelic-Radisic et al 2020
Bjelic-Radisic V, Cardoso F, Cameron D, Brain E, Kuljanic K, da Costa RA et al. An international update of the EORTC questionnaire for assessing quality of life in breast cancer patients: EORTC QLQ-BR45. Ann Oncol. 2020;31(2):283-8.

Brookmeyer and Crowley 1982
Brookmeyer F, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982 Mar;38(1):29-41

Fayers et al 2001
Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). European Organisation for Research and Treatment of Cancer, Brussels 2001.

Fleischer et al 2011
Fleischer F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: a quantitative approach. Stat Med 2011 Dec 20;30(29):3373-86.

Gail and Simon 1985
Gail M, Simon R. Tests for qualitative interactions between treatment effects and patient subsets. Biometrics 1985 Jun;41(2):361–72.

Herdman et al 2011
Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Darkin P et al. Development and Preliminary Testing of the New Five-Level Version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-36.

Lan and DeMets 1983
Lan K, DeMets D. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-63.

Latimer et al 2014
Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP et al. Adjusting survival time estimates to account for treatment switching in randomised controlled trials – an economic evaluation context: Methods, limitations and recommendations. Med Decis Making. 2014; 34(3):387-402.

Osoba et al 1998
Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998 Jan;16(1):139-44.

Peduzzi et al 1996
Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis. J Clin Epidemiol 1996 Dec;49(12):1373-9.

Pintilie 2006
Pintilie M. Competing risks: A practical perspective. Wiley 2006.

Silcocks 2012
Silcocks P. How many strata in an RCT? A flexible approach. Br J Cancer. 2012 Mar 27;106(7):1259-61.

Sprangers et al 1996
Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol. 1996;14(10):2756-68.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's Method with the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Res 2010;2(1):97-108.

van Hout et al 2012
van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708-15.

Wassmer and Brannath 2016
Wassmer G, Werner Brannath W. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. Springer 2016.

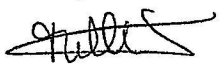
Ye et al 2013
Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints. Stat Med. 2013 Mar 30;32(7):1112-24.

SIGNATURE(S)

A Phase 3, Randomized, Multi-center, Open-label Study of Trastuzumab Deruxtecan (T-DXd) Versus Investigator's Choice Chemotherapy in HER2-Low, Hormone Receptor Positive Breast Cancer Patients whose Disease has Progressed on Endocrine Therapy in the Metastatic Setting (DESTINY-Breast 06)

I approve this document.

IQVIA representative



Delphine Millet
Data Sciences, Safety and Medical
Tour D2, 17bis Place des Reflets
92400 Courbevoie - FRANCE

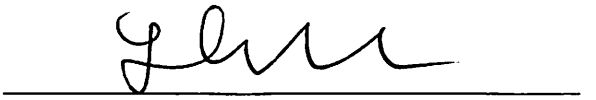
20 Mar 2024
Date
(Day Month Year)

SIGNATURE(S)

A Phase 3, Randomized, Multi-center, Open-label Study of Trastuzumab Deruxtecan (T-DXd) Versus Investigator's Choice Chemotherapy in HER2-Low, Hormone Receptor Positive Breast Cancer Patients whose Disease has Progressed on Endocrine Therapy in the Metastatic Setting (DESTINY-Breast 06)

I approve this document.

Astrazeneca representative



Xiaoling Wu, *PhD, MS, BS*
AstraZeneca US, 1800 Concord Pike,
Wilmington, DE 19803, USA

20MAR2024
Date
(Day Month Year)