Clinical Study Report Appendix 16.1.9

Drug Substance	Datopotamab deruxtecan (Dato-DXd, DS-1062a)
Study Code	D9268C00001

Appendix 16.1.9
Documentation of Statistical Methods and Supporting Statistical Analysis

Statistical Analysis Plan Number	Date
Statistical Analysis Plan Edition 1.0	22-Dec-2021
Statistical Analysis Plan Edition 2.0	21-Jun-2023
Statistical Analysis Plan Edition 3.0	17-Aug-2023

STATISTICAL ANALYSIS PLAN	
Study Code	D9268C00001
Edition Number	1.0
Date	22-Dec-2021

A Phase 3, Open-label, Randomised Study of Dato-DXd Versus Investigator's Choice of Chemotherapy in Participants With Inoperable or Metastatic Hormone Receptor-Positive, HER2-Negative Breast Cancer Who Have Been Treated With One or Two Prior Lines of Systemic Chemotherapy (TROPION-Breast01)

TABLE OF CONTENTS

TITLE PAGE 1
TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 7
AMENDMENT HISTORY 12
1 INTRODUCTION 13
2 CHANGES TO PROTOCOL PLANNED ANALYSES 13
3 DATA ANALYSIS CONSIDERATIONS 13
3.1 Timing of Analyses 13
3.1.1 Ophthalmologic Data Review (DCO1) 13
3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2) 14
3.1.3 Second OS Interim Analysis (DCO3) 14
3.1.4 Primary Analysis of OS (DCO4) 14
3.2 Analysis Populations 14
3.2.1 Enrolled set 14
3.2.2 Full analysis set 14
3.2.3 Safety analysis set 14
3.2.4 Ophthalmologic analysis set 15
3.2.5 Pharmacokinetic analysis set 15
3.2.6 ADA evaluable set 15
3.2.7 Summary of outcome variables and analysis sets 15
3.3 General Considerations 16
3.3.1 Sample Size Determination 17
3.3.2 Investigational Product 19
3.3.3 Baseline 19
3.3.4 On Treatment 19
3.3.5 Visit Window 19
3.3.6 Handling of Unscheduled Visits 20
3.3.7 Missing Data 20
3.3.8 Derivations of RECIST Visit Responses 22
3.3.8.1 Target lesions (TLs) 23
3.3.8.2 Non-target lesions (NTLs) and new lesions 27
3.3.8.3 Overall visit response – site investigator data 28
3.3.8.4 Independent review 29
3.3.9 Multiplicity/Multiple Comparisons 30
3.3.10 Handling of Protocol Deviations in Study Analysis 31
4 STATISTICAL ANALYSIS 33
4.1 Study Population 33
4.1.1 Participant Disposition and Completion Status 33
4.1.1.1 Presentations 33
4.1.2 Analysis Sets 33
4.1.3 Protocol Deviations 34

4.1.4 Demographics 34
4.1.4.1 Definitions and Derivations 34
4.1.4.2 Presentation 34
4.1.5 Baseline Characteristics 34
4.1.5.1 Definitions and Derivations 34
4.1.5.2 Presentation 35
4.1.6 Disease Characteristics 35
4.1.6.1 Definitions and Derivations 35
4.1.6.2 Presentation 35
4.1.7 Medical History and Concomitant Disease 36
4.1.7.1 Definitions and Derivations 36
4.1.7.2 Presentation 36
4.1.8 Prior and Concomitant Medications 36
4.1.8.1 Definitions and Derivations 36
4.1.8.2 Presentation 37
4.1.9 Study Drug Compliance 38
4.2 Endpoint Analyses 38
4.2.1 Primary Endpoint - Progression Free Survival by BICR 45
4.2.1.1 Definition 45
4.2.1.2 Derivations and Censoring Rules 46
4.2.1.3 Handling of Dropouts and Missing Data 48
4.2.1.4 Primary Analysis of Progression Free Survival 48
4.2.1.5 Sensitivity Analyses of Progression Free Survival 50
4.2.1.6 Subgroup Analyses 51
4.2.2 Primary Endpoint - Overall Survival 53
4.2.2.1 Definition 53
4.2.2.2 Derivations and Censoring Rules 54
4.2.2.3 Handling of Dropouts and Missing Data 54
4.2.2.4 Primary Analysis of Overall Survival 55
4.2.2.5 Sensitivity Analyses of Overall Survival 55
4.2.2.6 Subgroup Analyses 56
4.2.3 Secondary Endpoint - Objective Response Rate 56
4.2.3.1 Definition 56
4.2.3.2 Derivations 56
4.2.3.3 Primary Analysis of Objective Response Rate 57
4.2.4 Secondary Endpoint - Best Objective Response 57
4.2.4.1 Definition 57
4.2.4.2 Derivations 58
4.2.4.3 Primary Analysis of Best Objective Response 58
4.2.5 Secondary Endpoint - Duration of Response 58
4.2.5.1 Definition 58
4.2.5.2 Derivations and Censoring Rules 59
4.2.5.3 Primary Analysis of Duration of Response 59
4.2.5.4 Sensitivity Analyses of Duration of Response 59
4.2.5.5 Supplementary Analyses of Duration of Response 59
4.2.6 Secondary Endpoint – PFS by Investigator Assessment 60

4.2.6.1 Definition 60
4.2.6.2 Derivations and Censoring Rules 60
4.2.6.3 Handling of Dropouts and Missing Data 60
4.2.6.4 Primary Analysis of Progression Free Survival 60
4.2.6.5 Sensitivity Analyses of Progression Free Survival 60
4.2.6.6 Subgroup Analyses 60
4.2.7 Secondary Endpoint –Disease Control Rate 60
4.2.7.1 Definition 60
4.2.7.2 Derivations 61
4.2.7.3 Primary Analysis of Disease Control Rate 61
4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30 61
4.2.8.1 Definition 61
4.2.8.2 Derivations and Censoring Rules 62
4.2.8.3 Handling of Dropouts and Missing Data 65
4.2.8.4 Primary Analysis of EORTC QLQ-C30 65
4.2.8.5 Sensitivity Analysis of EORTC QLQ-C30 66
4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death 66
4.2.9.1 Definition 66
4.2.9.2 Derivations and Censoring Rules 66
4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death 66
4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death 67
4.2.10.1 Definition 67
4.2.10.2 Derivations and Censoring Rules 67
4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death 67
4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death 68
4.2.11.1 Definition 68
4.2.11.2 Derivations and Censoring Rules 68
4.2.11.3 Handling of Dropouts and Missing Data 68
4.2.11.4 Primary Analysis of PFS2 69
4.2.12 Secondary Endpoint - Pharmacokinetics 69
4.2.12.1 Derivations 69
4.2.12.2 Primary Analyses of Pharmacokinetics 69
4.2.13 Secondary Endpoint - Immunogenicity 70
4.2.13.1 Derivations 70
4.2.13.2 Primary Analysis of Immunogenicity 71
4.2.14 Exploratory Endpoint – Patient Reported AEs and Treatment Tolerability 71
4.2.14.1 Definition 71
4.2.14.2 Derivations and Censoring rules 72
4.2.14.3 Handling of Dropouts and Missing data 73
4.2.14.4 Primary Analyses of PRO-CTCAE 73
4.2.14.5 Primary Analyses of EORTC IL117 74
4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability 74
4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS) 74

4.2.15.1 Definition 74
4.2.15.2 Derivations and Censoring rules 74
4.2.15.3 Analysis of Patient Global Impression of Severity 74
4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC) 75
4.2.16.1 Definition 75
4.2.16.2 Derivations and Censoring rules 75
4.2.16.3 Analysis of Patient Global Impression of Change 75
4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL 75
4.2.17.1 Definition 75
4.2.17.2 Derivation and Censoring Rules 75
4.2.17.3 Handling of Dropouts and Missing Data 75
4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL 75
4.2.18 Exploratory Endpoint – Breast and Arm Symptoms 76
4.2.18.1 Definition 76
4.2.18.2 Derivations and Censoring Rules 77
4.2.18.3 Handling of Dropouts and Missing Data 77
4.2.18.4 Primary Analysis of Breast and Arm Symptoms 77
4.2.19 Exploratory Endpoint - EQ-5D-5L 78
4.2.19.1 Definition 78
4.2.19.2 Derivations 78
4.2.19.3 Primary Analysis of EQ-5D-5L 78
4.2.19.4 Supplementary Analyses of EQ-5D-5L 79
4.2.20 Exploratory Endpoint - Biomarkers 79
4.2.21 Exploratory Endpoint - Health Care Resource Use 79
4.2.21.1 Definition 79
4.2.21.2 Derivations 79
4.2.21.3 Primary Analysis of Health Care Resource Use 80
4.3 Pharmacodynamic Endpoint(s) 80
4.4 Safety Analyses 80
4.4.1 Exposure 80
4.4.1.1 Definitions and Derivations 80
4.4.1.2 Presentation 85
4.4.2 Adverse Events 85
4.4.2.1 Definitions and Derivations 85
4.4.2.2 Presentation 86
4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest 89
4.4.3 Clinical Laboratory, Blood Sample 89
4.4.3.1 Definitions and Derivations 89
4.4.3.2 Presentations 90
4.4.4 Clinical Laboratory, Urinalysis 92
4.4.4.1 Definitions and Derivations 92
4.4.4.2 Presentations 92
4.4.5 Other Laboratory Evaluations 93
4.4.5.1 Definitions and Derivations 93

4.4.5.2 Presentations 93
4.4.6 Vital Signs 93
4.4.6.1 Definitions and Derivations 93
4.4.6.2 Presentations 93
4.4.7 Electrocardiogram 93
4.4.7.1 Definitions and Derivations 93
4.4.7.2 Presentations 94
4.4.8 Echocardiogram/Multigated Acquisition Scan 95
4.4.8.1 Definitions and Derivations 95
4.4.8.2 Presentations 95
4.4.9 Eastern Cooperative Oncology Group Performance Status 95
4.4.9.1 Definitions and Derivations 95
4.4.9.2 Presentations 96
4.4.10 Physical Examination 96
4.4.10.1 Definitions and Derivations 96
4.4.10.2 Presentations 96
4.4.11 Ophthalmologic Assessments 96
4.4.11.1 Definition and Derivations 96
4.4.11.2 Presentations 97
4.4.12 Pulmonary Function Test 98
4.4.12.1 Definitions and Derivations 98
4.4.12.2 Presentations 98
4.4.13 Impact of COVID-19 99
5 INTERIM ANALYSIS 99
6 REFERENCES 102
7 APPENDIX 102

LIST OF TABLES

Table 1 Summary of outcome variables and analysis sets 15
Table 2 TL visit responses (RECIST 1.1) 23
Table 3 NTL visit responses 27
Table 4 Overall visit response 28
Table 5 Endpoint analyses 39
Table 6 Outcome and date of event for PFS analysis 46
Table 7 Definition of two missed RECIST visits 47
Table 8 EORTC QLQ-C30 scales and scores 61
Table 9 Definition of two missed PRO visits 65
Table 10 EORTC IL116 scales and scores 77
Table 11 Best Corrected Visual Acuity conversion table 97
Table 12 Summary of planned timings of the interim and final OS analyses . 100

LIST OF FIGURES

Figure 1 Multiple Testing Procedure..... 30

LIST OF ABBREVIATIONS

Abbreviation or Specialised Term	Definition
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
AJCC	American joint committee on cancer
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AZ	AstraZeneca
BCVA	Best corrected visual acuity
BICR	Blinded independent central review
BID	Twice daily
BLQ	Below limit of quantification
BMI	Body mass index
BoR	Best overall response
C1D1	Cycle 1, day 1
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
COVID-19	Coronavirus 2019 disease
CRF	Case Report Form
CR	Complete response
CRO	Clinical research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DBL	Data base lock
DCO	Data cut-off
DCR	Disease control rate

Abbreviation or Specialised Term	Definition
DFI	Disease free interval
DoR	Duration of Response
DOSDISC	Treatment discontinuation form
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EORTC	European Organisation for Research and Treatment of Cancer
EQ-5D-5L	EuroQoL 5-dimension, 5-level health state utility index
EQ-VAS	EuroQoL-Visual analogue scale
FAS	Full Analysis Set
FEV1	Forced expiratory volume – 1 second
FVC	Forced vital capacity
GHS	Global health status
HDU	High dependency unit
HER2	Human epidermal growth factor receptor 2
HCRU	Health care resource use
HIV	Human immunodeficiency virus
HR	Hazard ratio
HR-positive	Hormone receptor-positive
HRQoL	Health related quality of life
IA	Interim analysis
ICC	Investigator's Choice Chemotherapy
ICF	Informed consent form
ICR	Independent central review
ICU	Intensive care unit
IDMC	Independent data monitoring committee
ILD	Interstitial lung disease
IP	Investigational product
IPD	Important protocol deviation
IRC	Independent Review Charter
IRT	Interactive Response Technology
ITT	Intention to treat
IV	Intravenous
KM	Kaplan-Meier

Abbreviation or Specialised Term	Definition
LD	Longest diameter
LLOQ	Lower Limit of Quantification
Ln	Natural logarithm or logarithm to the base e
LSCD	Limbal stem cell deficiency
LSmean	Least squares mean
LVEF	Left ventricular ejection fraction
MCID	Minimal clinically important difference
MedDRA	Medical Dictionary for Regulatory Activities
MM	Millimetre
MMRM	Mixed model for repeated measures
MRI	Magnetic Resonance Imaging
MSSO	Maintenance and Support Services Organization
MTP	Multiple testing procedure
MUGA	Multigated acquisition
NA	Not applicable
nAb	Neutralizing antibody
NC	Not calculable
NCI	National Cancer Institute
NE	Not evaluable
NL	New lesion
NQ	Not quantifiable
NR	Not Reportable
NS	No Sample
NTL	Non-target lesion
OAS	Ophthalmologic Analysis Set
ORR	Objective response rate
OS	Overall survival
PAS	Pharmacokinetic analysis set
PAP	Payer Analysis Plan
PD	Progression of disease
PFS	Progression-free survival
PFS2	Second progression-free survival
PGIC	Patients' global impression of change
PGIS	Patients' global impression of severity
PGI-TT	Patients' global impression of treatment tolerability

Abbreviation or Specialised Term	Definition
PID	Percentage intended dose
PK	Pharmacokinetics
PR	Partial response
PRO	Patient-reported outcome
PS	Performance Status
PT	Preferred term
Q3W	Every 3 weeks
Q6W	Every 6 weeks
Q9W	Every 9 weeks
QLQ-C30	EORTC 30-item core quality of life questionnaire
QTcF	QT interval corrected by Fridericia's formula
RDI	Relative dose intensity
RECIST 1.1	Response Evaluation Criteria in Solid Tumours, Version 1.1
RS	Raw score
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Stable disease
SoA	Schedule of activities
SOC	System organ class
TEAE	Treatment emergent adverse event
TELC	Treatment emergent laboratory change
TFBT	Tear film breakup time
TFI	Treatment Free Interval
TFL	Table, figures and listings
TFST	Time to first subsequent therapy or death
TL	Target lesion
TNM	Tumour node metastasis
TROP2	Trophoblast cell surface antigen 2
TSST	Time to second subsequent therapy or death
TTD	Time to deterioration
ULN	Upper limit of normal
US	United States
VAS	Visual analogue scale

Abbreviation or Specialised Term	Definition
VCV	Veeva Clinical Vault
WHO	World Health Organization Drug dictionary

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
N/A	22-Dec-2021	Initial approved SAP	N/A	N/A

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D9268C00001 supporting the clinical study report (CSR). The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection. This statistical analysis plan (SAP) is based on version 2.0 of the CSP dated 27 August 2021. In the event of future amendments to the protocol, this SAP may be modified to account for changes relevant to the statistical analysis.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

For the primary analysis of EuroQoL 5-dimension, 5-level health state utility index (EQ-5D-5L) the CSP specifies analysis on all randomised participants while the SAP specifies analysis on the Full Analysis Set (FAS) who have a baseline EQ-5D-5L assessment. The protocol states that Patient-reported outcome (PROs) will be analysed using the FAS population, however some PROs will be analysed using the Safety Population. See SAP Table 1 for more details. For the sensitivity analysis of attrition bias for progression free survival SAP states two, or more, missed tumour assessments while the CSP states two, or more, non-evaluable tumour assessments. The list of adverse events of special interest (AESI) specified in Section 4.4.2.1 differs from those specified in the CSP. In the SAP geographic region strata is stated as United States, Canada, Europe vs Rest of World while in the CSP it is stated as United States, Europe vs Rest of World.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

As the study is event driven, the accrual of the predetermined number of events included in the study endpoints determines the duration of the data collection phase of the study. There are 4 planned data cut-offs (DCOs) for this study consisting of an ophthalmologic data review (DCO1), primary analysis of progression free survival (PFS)/first overall survival (OS) interim analysis (DCO2), second OS interim analysis (DCO3) and primary analysis of OS (DCO4). These interim analyses and additional safety reviews will be conducted as described in Section 5.

3.1.1 Ophthalmologic Data Review (DCO1)

The DCO for the ophthalmologic data review (DCO1) is planned to occur after completion of the last ophthalmologic assessment and a minimum of 2 assessments per participant has been completed, for the first approximately 100 randomised participants (DCO1).

3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2)

The DCO for the primary analysis of PFS/first OS interim analysis (DCO2) is planned to occur when approximately 419 PFS Blinded independent central review (BICR) events have been observed in the FAS. Based on enrolment assumptions, it is expected that this will occur approximately 21 months after randomisation of the first participant. For the primary analysis of PFS this provides approximately 60% maturity. For the first OS interim analysis this corresponds to approximately 25% maturity and 40% of the information expected at the primary analysis.

3.1.3 Second OS Interim Analysis (DCO3)

The DCO for the interim analysis for superiority in OS (DCO3) is planned to occur when approximately 355 OS events have been observed in the FAS population. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis. Based on enrolment assumptions, it is expected that this will occur approximately 34 months after randomisation of the first participant.

3.1.4 Primary Analysis of OS (DCO4)

The DCO for the primary analysis of OS (DCO4) is planned to occur when approximately 444 OS events have been observed in the FAS population, approximately 44 months after the first participant is randomised (63% maturity). At this time the clinical database will close to new data.

3.2 Analysis Populations

There are six analysis sets defined for this study.

3.2.1 Enrolled set

The Enrolled Set consists all participants who provided informed consent irrespective of whether they received the study treatment.

3.2.2 Full analysis set

The FAS consists of all randomised participants with treatment groups assigned in accordance with the randomisation, regardless of the treatment actually received. Participants who were randomised but did not subsequently receive treatment are included in the FAS. The analysis of data using the FAS therefore follows the principles of intention to treat (ITT).

3.2.3 Safety analysis set

The safety analysis set (SAF) will consist of all randomised participants who received any amount of study treatment (Dato-DXd or Investigator's Choice Chemotherapy (ICC)). Safety data will not be formally analysed but summarised using the SAF, according to the

actual treatment received. If a participant receives any amount of Dato-DXd, they will be summarised in the Dato-DXd treatment group. If a participant only receives ICC, they will be summarised in the ICC treatment group.

3.2.4 Ophthalmologic analysis set

The ophthalmologic analysis set (OAS) will consist of approximately the first 100 randomised participants (approximately 50 per arm, Dato-DXd and ICC). This will consist of all participants who have had ophthalmologic assessments recorded.

3.2.5 Pharmacokinetic analysis set

The pharmacokinetic analysis set (PAS) consists of all participants randomly assigned to study intervention who received at least 1 dose of Dato-DXd for whom there is at least one reportable post-dose pharmacokinetic (PK) concentration. Participants who violate or deviate from the protocol in ways that would significantly affect the PK analyses should not be included in the PK analysis set.

3.2.6 ADA evaluable set

The anti-drug antibody (ADA) evaluable set will consist of participants in the safety analysis set with a non-missing baseline ADA Dato-DXd result and at least 1 post-baseline ADA Dato-DXd result.

3.2.7 Summary of outcome variables and analysis sets

The analysis sets to be used for each outcome are provided in Table 1.

Table 1 Summary of outcome variables and analysis sets

Outcome variable	*Analysis set*
Efficacy Data	
PFS, PFS2, OS, ORR*, DoR*, DCR, TFST, TSST	FAS
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	FAS
Baseline and disease characteristics	FAS
Important protocol deviations	FAS
Medical/surgical history	FAS
Previous anti-cancer therapy	FAS
Concomitant medications/procedures	FAS
Subsequent anti-cancer therapy	FAS
Study drug compliance	FAS
PK/Immunogenicity Data	

Outcome variable	*Analysis set*
PK data	PAS
Immunogenicity	ADA evaluable set
Safety data	
Exposure	SAF
Adverse events	SAF
Laboratory measurements	SAF
Vital signs	SAF
Physical examination	SAF
ECGs	SAF
ECOG PS	SAF
ECHO/MUGA	SAF
Ophthalmologic assessments	OAS
Patient-reported outcomes	
EORTC QLQ-C30, EORTC IL116, EQ-5D-5L, PGI-S, PGI-C	FAS
PGI-TT, PRO-CTCAE, EORTC IL117	SAF
Health care resource use	
HcRU	SAF

*Participants who are evaluable for the analysis of ORR are those with measurable disease at baseline. Participants who are evaluable for the analysis of DoR are those who responded in the ORR analysis.
ADA=antidrug antibody; BoR=best objective response; CTCAE=Common Terminology Criteria for Adverse Events; DCR=disease control rate; DoR= duration of response; ECG=electrocardiogram; ECHO=echocardiogram; ECOG=Eastern Cooperative Oncology Group; EORTC=European Organisation for Research and Treatment of Cancer; EORTC QLQ-C30=EORTC 30-item core quality of life questionnaire; EQ-5D-5L=EuroQoL 5-dimension, 5-level health state utility index; FAS=Full analysis set; HcRU=Healthcare resource use; MUGA=multigated acquisition; ORR=objective response rate; OS=overall survival; PAS=pharmacokinetic analysis set; PGIC=Patients' Global Impression of Change; PGIS=Patients' Global Impression of Severity; PGI-TT= Patient's Global Impression of Treatment Tolerability; PFS=progression-free survival; PFS2=time from randomisation to second progression or death; PK=pharmacokinetic; PRO=patient-reported outcome; PS=Performance Status; SAF=safety analysis set; TFST=time to first subsequent therapy or death; TSST=time to second subsequent therapy or death.

3.3 General Considerations

The below mentioned general principles are followed throughout the study:

- Summary tables are produced by treatment group (Dato-DXd and ICC). Total columns are produced only for tables of disposition, demography, baseline characteristics and EQ-5D-5L data.

- Descriptive statistics are used for all variables, as appropriate. Continuous variables are summarised by the number of observations, mean, standard deviation, median, upper and lower quartiles where indicated, minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables are summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages are calculated out of the population total for the corresponding treatment group.
- For continuous data, the mean and median are rounded to 1 additional decimal place compared to the original data. The standard deviation is rounded to 2 additional decimal places compared to the original data. Minimum and maximum are displayed with the same accuracy as the original data.
- For categorical data, percentages are rounded to 1 decimal place with the exception of 100% which is presented as a whole number.
- Results of all statistical analyses will be presented using a 95% confidence interval (CI) and 2-sided p-value, unless otherwise stated.
- CIs and ratios (including hazard ratios) will be rounded to 2 decimal places. The p-values will be rounded to 4 decimal places, except for those below 0.0001, which will be displayed as '<0.0001'.
- SAS® version 9.4 (or higher) is used for all analyses.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days. One year is defined to be 365.25 days.

Where analysis models are stratified by the randomisation stratification factors, the data from the Interactive Response Technology (IRT) will be used, not the values recorded in the electronic case report form (eCRF).

3.3.1 Sample Size Determination

Approximately 1000 participants will be enrolled to achieve approximately 700 randomly assigned to study intervention.

"Enrolled" means a participant's, or their legally acceptable representative's, agreement to participate in a clinical study following completion of the informed consent process. Potential participants who are screened for the purpose of determining eligibility for the study but are not randomly assigned/assigned in the study, are considered "screen failures", unless otherwise specified by the protocol.

The study is sized for dual primary endpoints to characterise the PFS and OS benefit of Dato-DXd versus ICC in the participants with HR-positive, HER2-negative breast cancer

who have been treated with one or two prior lines of systemic chemotherapy in the inoperable/metastatic setting. The study will be considered positive (a success) if either the PFS analysis results and/or the OS analysis results are statistically significant.

For the primary analysis of PFS (See Section 3.1.2 for timing of analysis) assuming the true PFS treatment effect under the alternative hypothesis is a hazard ratio of 0.55 for Dato-DXd versus ICC, and the median PFS times of 4.7 months and 8.5 months in ICC and Dato-DXd, 419 PFS events from the FAS population (60% maturity) will provide greater than 99% power to demonstrate statistical significance at the 2-sided alpha level of 1.0%. This also assume the median PFS times in both groups are exponentially distributed. The smallest treatment difference that is statistically significant will be a hazard ratio of 0.775. Assuming a recruitment period of 19 months, this analysis is anticipated to be approximately 21 months after the first participant has been randomised.

The primary analysis of OS will be performed when approximately 444 OS events from the FAS have occurred across the Dato-DXd and ICC treatment groups (63% maturity). Assuming the true OS hazard ratio is 0.75 for Dato-DXd versus ICC, and the median OS in ICC is 19.0 months, the study will have 85% power to demonstrate statistical significance at the 5.0% level (using a 2-sided test). This assumes the PFS primary analysis crosses the efficacy threshold, and allowing 2 interim analyses to be conducted at information fractions of approximately 40% and 80% of the target events, respectively (per the O'Brien and Fleming approach (Lan & DeMets, 1983)). The smallest treatment difference that could be statistically significant at the primary OS analysis is a hazard ratio of 0.824.

If the PFS primary analysis does not cross the efficacy threshold, the OS analysis will have 83% power to demonstrate statistical significance at the 4.0% level (using a 2-sided test). The smallest treatment difference that could be statistically significant at the primary analysis is a hazard ratio of 0.817. All OS calculations assume median OS times of 19.0 months and 25.3 months in ICC and Dato-DXd, respectively when the survival times are exponentially distributed.

With a recruitment period of approximately 19 months, it is anticipated that the primary OS analysis will occur approximately 44 months after the first participant has been randomised. The study may continue monitoring participants for OS up to the scheduled primary analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival. Further details of the interim analyses are presented in Section 5.

A nonuniform accrual of participants (with k = 1.5) is assumed when estimating the analysis times. The total proportion of participants randomised at time t [$t \leq 19$ months] following the start of the study is assumed to be $(t/19)^k$.

3.3.2 Investigational Product

Investigational product (IP) refers to Dato-DXd and the ICC. The first and last dates of IP refer to the earliest and latest of this treatment respectively.

3.3.3 Baseline

In general, for efficacy endpoints the last observed measurement prior to randomisation is considered the baseline measurement. For safety endpoints the last evaluable observation before the first dose of IP is considered the baseline measurement unless otherwise specified. For PRO endpoints Cycle 1 Day 1 will be used unless there are multiple assessments before the first dose of IP in which case the last evaluable observation before the first dose of IP will be used.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured are considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

If two visits are equally eligible to assess participant status at baseline (e.g. two assessments both on the same date with no time recorded), the average is used as the baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking the average is not possible, the best value (value closest to none/normal/negative) is used as baseline as this is most conservative. In the scenario where there are two assessments recorded on the same day, one with time recorded and the other without time recorded, the one with the time recorded is selected as baseline. Where safety data are summarised over time, time on study is calculated in relation to date of first IP administration.

In all summaries change from baseline variables are calculated as the post-treatment value minus the value at baseline. The percentage change from baseline is calculated as (post-baseline value - baseline value) / (baseline value) x 100.

3.3.4 On Treatment

For the purposes of summarising safety data assessed at visits, in addition to baseline data, only on treatment data are included in the summary tables. On treatment data is defined as data after the first dose of IP and with assessment date up to and including the date of last IP + 35 days or prior to the start of any subsequent cancer therapy, whichever occurs earlier.

3.3.5 Visit Window

Time windows are defined for all presentations of safety data and PRO data that summarise values by visit according to the following conventions:

- For safety data study day references date of first dose of IP as Day 1, for PK the reference is the time of Dato-DXd administration on the day PK blood samples are taken, for efficacy data study day references date of randomisation as Day 1. All windows for PRO (including those reported on FAS or SAF) will have windows calculated from date of first dose of IP as Day 1. However, time to deterioration will be calculated as described in Section 4.2.8.2.
- The time windows are exhaustive so that data recorded at any time point (scheduled or unscheduled) has the potential to be summarised. Inclusion within the time window is based on the actual date and not the intended date of the visit.
- The window for visits following baseline are 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 is Day 2). If an even number of days exist between two consecutive visits, then the upper limit is taken as the midpoint value minus 1 day.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment (as defined in Section 3.3.4) is used (regardless of where it falls in an interval).
- Listings display all values contributing to a time point for a participant.
- For visit-based summaries, if there is more than one value per participant within a time window then the closest value to the scheduled visit date is summarised. If the values are equidistant from the nominal visit date, then the earlier value is used. Data listings highlight the values used in the summary table, wherever feasible.

Visit data are only included in summary tables if the number of observations is ≥20 in at least one treatment group.

3.3.6 Handling of Unscheduled Visits

Unscheduled visits are included in the method of assigning data to scheduled visits described in the rules in Section 3.3.5 above. Unscheduled visits are not included as a separate visit in the summary tables.

For summaries at participant level, such as of extreme values, all post-baseline values collected are used to derive a participant level statistic including those collected at unscheduled visits and regardless of whether they appear in the corresponding visit-based summary.

3.3.7 Missing Data

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

For missing start dates for adverse events (AEs) and concomitant medications/procedures, the following rules are applied:

- Missing day: Impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.
- Missing day and month: Impute 1st January unless year is the same as first dose date then impute first dose date.
- Completely missing date: Impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date.

An imputed start date of an AE must be prior to the end date of the AE.

For missing stop dates of AEs or concomitant medications/procedures, the following rules are applied:

- Missing day - Impute the last day of the month unless month is the same as month of the last dose of study drug then impute last dose date.
- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date
- Completely missing: If an AE/medication has a completely missing end date then it is treated as ongoing.

The imputation of dates for AEs and concomitant medications is used to determine if an AE is treatment emergent and whether a medication is concomitant. Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations are not calculated.

If a participant is known to have died where only a partial death date is available, then the date of death is 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:

- Missing day only: Using the 1st of the month.
- Missing day and month: Using the 1st January.

For partial subsequent anti-cancer therapy dates, the following rules are applied:

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

If a participant has a partial date of birth, (i.e. for those cases where year of birth only is given) the 1st of the month is imputed if only day is missing, and 1st January is imputed if the day and month are missing.

Other rules for handling missing data are described under the derivation rules for that particular variable.

3.3.8 Derivations of RECIST Visit Responses

For all participants, the Response Evaluation Criteria in Solid Tumours (RECIST) tumour response data will be used to determine each participant's visit response according to RECIST version 1.1. It will also be used to determine if and when a participant has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumour assessments are to be performed no more than 28 days before randomisation and ideally as close as possible to and prior to randomisation. Tumour assessments are then performed every 6 weeks following randomisation for 48 weeks, then every 9 weeks thereafter until disease progression. Following disease progression, 1 additional follow-up scan should be performed as per imaging schedule (i.e. either 6 weeks or 9 weeks later).

If an unscheduled assessment is performed, and the participant 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 minimise any unintentional bias caused by some participants being assessed at a different frequency than other participants.

From the investigator's review of the imaging scans, the RECIST tumour response data are used to determine each participant's visit response according to RECIST version 1.1. At each visit, participants are programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD) or progression of disease (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 participant has had a tumour assessment that cannot be evaluated then the participant is assigned a visit response of not evaluable (NE), (unless there is evidence of progression in which case the response is assigned as PD).

Please refer to Section 3.3.8.3 for the definitions of CR, PR, SD and PD.

RECIST outcomes (i.e. PFS, ORR etc.) are calculated programmatically for the site investigator data (as described in the relevant subsection in Section 4.2) from the overall visit responses.

3.3.8.1 Target lesions (TLs)

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 participant can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as TLs. Lymph nodes are considered to be one organ; therefore, at most 2 target nodal lesions can be selected. If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomisation are 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 target lesion (TL) should be identified as NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

TL visit responses are described in Table 2 below.

Table 2 TL visit responses (RECIST 1.1)

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 <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Stable disease (SD)	Neither sufficient decrease in sum of diameters to qualify for PR nor sufficient increase to qualify for PD.
Progression of disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Not evaluable (NE)	Only relevant if any of the TLs at follow-up were not assessed or not evaluable (e.g. missing anatomy) or had a lesion intervention at this visit. Note: If the sum of diameters

Visit responses	Description
	meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.

TL target lesion.

Rounding of TL data

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

Missing TL data

If the sum of **available** 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, the TL visit response is PD. Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If there is at least one TL measurement missing and a **TL** visit response of PD cannot be assigned, the **TL** visit response is not evaluable (NE).

If all TL measurements are missing, then the TL visit response is NE. Overall visit response is also NE, unless there is a progression of non-TLs or new lesions, in which case the response is PD.

Lymph nodes

For lymph nodes, if the size reduces to <10mm then these are considered non-pathological. However, a size is still given, and this size is still 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 is 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 are applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is 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% but remains <10mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis >10mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response is set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined the response remains as CR.

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure, then this is indicated as such on the CRF and a value of 5mm is 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 is reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) are recorded.

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

- Step 1: the diameters of the TLs (including the lesions that have had intervention) are summed and the calculation are performed in the usual manner. If the visit response is PD, this remains as a valid response category.
- Step 2: If there is no evidence of progression after step 1, the lesion diameter (for those lesions with intervention) are treated 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 participant are 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 is used, and PR or SD assigned as the visit response. Participants 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 participant 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 is set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

If >1/3 of TL measurements are missing (because of intervention) then the TL response is 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 ≥5mm from nadir).

If ≤1/3 of the TL measurements are missing (because of intervention) then the results are scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters) and this is 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 74mm.

The sum of lesions 1-4 at the follow-up is 68mm. The sum of the corresponding lesions at the nadir visit is 62mm.

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

68 x 74 / 62 = 81mm

CR is not allowed as a TL response for visits where there is missing data. Only PR, SD, or PD (or NE) can be assigned as the TL visit response in these cases. However, for visits with ≤1/3 lesion assessments not recorded, the scaled-up sum of TLs diameters is 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 are 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 is recorded for one of the TL sizes and the other TL size is recorded as 0cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs, between CT and MRI is considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions are treated as missing.

3.3.8.2 Non-target lesions (NTLs) and new lesions

At each visit the investigator records an overall assessment of the non-target lesion (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 is derived based on the investigator's overall assessment of NTLs as described in Table 3:

Table 3 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).
Non-CR/non-PD	Persistence of one or more NTLs with no evidence of progression.
Progression (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.
Not evaluable (NE)	Only relevant when one or some of the NTLs are not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit.

Visit responses	Description
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

NTL non-target lesion; TL 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 tumour 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 are also 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 indicates 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 tumour.

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

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

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 that the new lesion response should be treated as NE in the derivation of the overall visit response.

3.3.8.3 Overall visit response – site investigator data

Table 4 defines how the previously defined TL and NTL visit responses are combined with new lesion information to give an overall visit response.

Table 4 Overall visit response

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**

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
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**

CR complete response; PR partial response; SD stable disease; NE not evaluable; NA not applicable; PD progressive disease.

3.3.8.4 Independent review

It is planned that a blinded independent central review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans for all participants (including those at unscheduled visits, or outside visit windows) are collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans are reviewed by two independent radiologists using RECIST 1.1 and are 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 participant, the BICR defines the overall visit response (CR, PR, SD, PD, NE) (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. If a participant has had a tumour assessment that cannot be evaluated, then the participant is assigned a visit response of NE (unless there is evidence of progression in which case the response is 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 are 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 is selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision is used). In the absence of adjudication, the records for all visits for a single reviewer is used. The reviewer selected in the absence of adjudication is the reviewer who read the baseline scan first. The records from the single selected reviewer are used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in TL dimensions. Endpoints (of ORR, PFS, DoR and DCR) are derived programmatically from this information.

Results of this independent review are not communicated to investigators and the management of participants is based solely upon the results of the RECIST 1.1 assessment conducted by the investigator.

An independent central review (ICR) of all participants will be performed for the primary analysis of PFS, which will cover all the scans up to the DCO.

Further details of the BICR are documented in the Independent Review Charter (IRC).

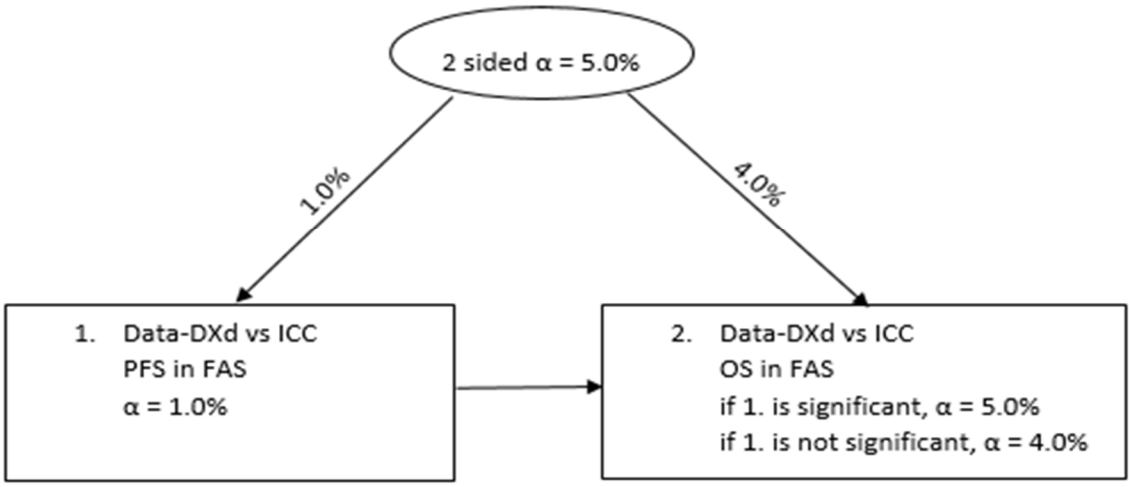
3.3.9 Multiplicity/Multiple Comparisons

The hypotheses of interest with regards to the efficacy for the dual primary endpoints are:

- H0: No differences between Dato-DXd and ICC for PFS and OS.
- H1: Differences between Dato-DXd and ICC for PFS and/or OS.

To preserve the overall type 1 error (familywise error rate) at 5% in the strong sense, a multiple testing procedure (MTP) for the dual primary endpoints of PFS and OS is implemented at DCO2, DCO3 and DCO4. An overview of the MTP with an alpha-splitting and exhaustive recycling strategy (Burman, Sonesson, & Guilbaud, 2009) is provided in Figure 1.

Figure 1 Multiple Testing Procedure



An alpha level of 1.0% will be allocated to the PFS primary analysis and the remaining 4.0% alpha level will be allocated to the OS analyses. If the PFS primary analysis meets statistical significance, the 1.0% type 1 error allocated to PFS endpoint will be reallocated (Burman, Sonesson, & Guilbaud, 2009) to the OS endpoint for a total 2-sided type 1 error of 5.0%. If the PFS primary analysis does not meet statistical significance, the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Alpha spending functions are applied for the OS endpoint in order to preserve the overall 2-sided type 1 error (familywise error rate) in the strong sense across the three planned analyses of OS.

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including 2 interim analyses for superiority of OS. Details are provided in the Interim Analysis section (Section 5).

The significance level alpha for OS across the three analysis times is dependent on the OS information fraction (number of OS events at interim/number of OS events at primary). The significance levels are calculated at the time of the analyses based on the number of events observed.

No multiplicity adjustment is applied for other endpoints as other endpoints are considered supportive endpoints.

3.3.10 Handling of Protocol Deviations in Study Analysis

Protocol deviations are collected, reviewed and reconciled throughout the study. Important protocol deviations (IPDs) are identified from the complete set of protocol deviations. IPDs are those which may significantly impact the reliability of the study data or that may significantly affect a participant's rights, safety, or wellbeing.

A set of pre-determined IPDs are listed in the protocol deviations plan. The protocol deviations plan also indicates which IPDs are identified by programmatic checks.

The IPDs are grouped into the following important protocol deviation (IPD) categories, where full details of the individual IPDs within each IPD category are provided in the protocol deviations plan:

- Inclusion criteria deviations.
- Exclusion criteria deviations.
- Discontinuation criteria for study product met but participant not withdrawn from study treatment.
- Discontinuation Criteria for overall study withdrawal met but participant not withdrawn from study
- IP deviation.
- Excluded medications taken.
- Deviations to study procedure.
- Other important deviations.

The following general categories will be considered important protocol deviations (IPDs) and will be programmatically derived from Veeva Clinical Vault (VCV) data. These will be listed and summarised by randomised treatment group and discussed in the CSR as appropriate. Refer to the CSP for full details of the inclusion/exclusion criteria.

- Inclusion criteria deviations (Deviation 1).
  - Lack of provision of informed consent prior to any study-related procedures.
  - Inclusion criteria 2, 3, 6, 7, 8, 9
- Exclusion criteria deviations (Deviation 2).
  - Exclusion criteria 1, 4, 5, 6, 9, 10-14, 22
- Discontinuation criteria for study product met but participant not withdrawn from study treatment (Deviation 3).
- Discontinuation criteria for overall study withdrawal met but participant not withdrawn from study (Deviation 4).
- Investigational product deviation (Deviation 5).
- Excluded medications taken (Deviation 6).
- Deviations related to study procedure (Deviation 7)
- Other important deviations (Deviation 8)

Participants who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 3.2.3. 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.

None of the deviations will lead to participants being excluded from the analysis sets described in Section 3.2 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding participants with specific important protocol deviations is not planned; however, a ‘deviation bias’ sensitivity analysis may be performed by repeating the PFS analysis excluding participants with deviations that may affect the efficacy of trial therapy. 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.

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Study Population

The domain study population covers participant disposition, analysis sets, protocol deviations, demographics, baseline characteristics, disease characteristics, medical history, prior and concomitant medication.

Study population data is summarised and listed using the FAS unless otherwise stated.

4.1.1 Participant Disposition and Completion Status

4.1.1.1 Presentations

Participant disposition is summarised by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine, presenting the number of participants screened, the number of screen failures, the number and percentage of participants: randomised, who were randomised and did not receive study treatment, who started IP (where participants starting Dato-DXd or ICC are presented according to the randomisation), who are ongoing in the study at the DCO, and who completed and discontinued the study along with reasons for study discontinuation.

The denominator for the percentages in these summaries is based on the number of randomised participants for the treatment group.

The number and percentage of participants ongoing on IP at the time of the DCO and who discontinued each study treatment along with the reasons for treatment discontinuation is also summarised, where the denominator for the percentages is based on the number of participants that started the corresponding study treatment.

Additionally, a summary of recruitment by region, country and site is produced for the enrolled set, the FAS and the safety analysis set. The denominator for the percentages in this summary is based on the number of participants for the analysis set for the treatment group.

Listings presenting details of discontinuations by individual participant are produced for those participants discontinuing treatment and discontinuing the study.

4.1.2 Analysis Sets

The number of participants in each analysis set and reasons for exclusion from each analysis set are summarised.

A listing of individual participants not included in each analysis set is provided.

4.1.3 Protocol Deviations

Refer to Section 3.3.10 for details regarding the definition and derivation of protocol deviations.

A summary table is produced showing the number and percentage of participants with any IPD and by category of IPD, which includes the individual IPDs as detailed in the protocol deviations plan.

The individual participant data for IPDs is also listed.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Demographics are comprised of age, age group, sex, race and ethnicity.

Age is calculated as age at last birthday in whole years using the date of randomisation and date of birth. Where a partial date of birth has been collected, it is imputed as described in Section 3.3.7 to calculate the participant's age for use in listings and summaries tables presenting age and/or age group and subgroup analyses based on age. Age is split into the following categories: <65 and ≥65 years.

Date of birth is listed as it has been collected on the eCRF.

4.1.4.2 Presentation

A summary table of demographic data specified in Section 4.1.4.1 is produced, and demographic data is listed.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Baseline characteristics include height (cm), weight (kg), weight group, body mass index (BMI) (kg/m^2) and ECOG performance status at baseline.

Weight (kg) is categorised into weight groups of:

- <65
- ≥65 and ≤90
- >90

The body mass index (BMI) is calculated as

BMI (kg/m^2) = Weight (kg) / {Height (m)}^{2}

Stratification factors (number of previous lines of chemotherapy, geographic region (with countries within region), and prior use of CDK4/6 inhibitor) are derived from the CRF data as well as recorded by IRT.

4.1.5.2 Presentation

Baseline characteristics specified in Section 4.1.5.1 are summarised and listed.

A summary of the stratification factors recorded both by IRT and CRF is also provided presenting the number and percentage of participants in each distinct stratum.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Disease characteristics include the disease-free interval, time from most recent disease progression to randomisation, overall disease classification (metastatic/locally advanced) and sites of locally advanced and metastatic disease.

Disease-free interval (DFI) in years is defined as:

(date of first relapse – date of diagnosis of primary breast cancer +1)/365.25

Where DFI is calculated only in participants who received adjuvant therapy (i.e. not metastatic at the time of diagnosis).

DFI is categorised into groups of >0 - ≤2 years, >2 - ≤5 years, >5 - ≤10 years and >10 years.

Time from most recent disease progression to randomisation (days) is defined as:

(Date of randomisation - date of most recent disease progression)

Time from most recent disease progression to randomisation is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

Time from diagnosis to randomisation in years is defined as:

(Date of randomisation – original diagnosis date) / 365.25

4.1.6.2 Presentation

The following breast cancer characteristics at diagnosis are summarised: American joint committee on cancer (AJCC) stage, histology type, tumour grade, tumour node metastasis (TNM) classification (primary tumour, regional lymph nodes and distant metastases) and time from diagnosis to randomisation.

In addition, the disease characteristics (described in Section 4.1.6.1) at study entry are summarised.

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Medical history and relevant surgical history are coded using the medical dictionary for regulatory activities (MedDRA) [using the latest or current MedDRA version].

Any medical history which is ongoing at time of informed consent is considered an ongoing condition, otherwise it is considered past medical history.

4.1.7.2 Presentation

Summary tables of past medical history, ongoing conditions and surgical history are presented by MedDRA system organ class (SOC) and preferred term (PT). Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

All medical history, surgical history and ongoing conditions are listed.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

Information on any treatment used from the time of informed consent up to 28 (+7) days after discontinuation of study treatment, with reasons for the treatment, will be recorded in the eCRF. Thereafter, only subsequent regimens of anti-cancer therapy will be recorded in eCRF.

Other anti-cancer therapies, investigational agents, and radiotherapy (except palliative to areas other than the chest) should not be given while the participant is on study drug.

Medications received prior to, concomitantly, or post-treatment are coded using the Anatomical Therapeutic Chemical (ATC) Classification codes.

Medications will be coded using the latest World Health Organization Drug dictionary (WHO Drug) version. The version used will be indicated in the data summaries and listings.

For inclusion in the prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.3.7.

Prior medications, concomitant medications and post-study treatment medications are defined as follows:

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

Missing coding terms are listed and summarised as "Not coded".

Treatment free interval (TFI) (in months) is derived from the CRF data as

(date of most recent progression – date of last dose of anti-cancer therapy prior to the most recent progression + 1)

Treatment free interval is categorised as 0=de novo, ≥12 months and <12 months.

Time from completion of prior anti-cancer therapy to randomisation is defined as

(date for randomisation – last anti-cancer therapy stop date + 1)

Time from completion of prior anti-cancer therapy is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

4.1.8.2 Presentation

The following summaries will be produced:

- Summary of prior medications
- Summary of concomitant medications
- Summary of disallowed concomitant medications
- Summary of post study treatment medications
- Summary of prior disease-related treatment modalities
- Summary of prior cancer therapies
- Summary of non-study cancer therapies whilst on study treatment
- Summary of post study treatment cancer therapies
- Summary of prior radiotherapy
- Summary of on study palliative radiotherapy
- Summary of post study treatment radiotherapy
- Summary of TFI
- Summary of time from completion of prior anti-cancer therapy to randomisation

Prior medications (excluding prior cancer therapies), concomitant medications (including both allowed and disallowed concomitant medications), disallowed concomitant medications and post-study treatment medications (excluding post-study treatment cancer therapies) are presented by ATC classification and generic term, sorted by descending frequency of ATC group and generic term. Participants taking the same concomitant medication/procedure multiple times are counted once per ATC classification and generic term.

A summary of number and percentage of participants receiving prior disease-related treatment modalities are summarised by modality.

Prior cancer therapies, non-study cancer therapies whilst on study treatment, and post-study treatment cancer therapies are summarised by therapy class and ATC group.

A separate summary of number and percentage of participants receiving prior radiotherapy is produced and repeated for on study palliative radiotherapy and post study treatment discontinuation radiotherapy.

TFI and Time from completion of prior anti-cancer therapy to randomisation will be presented using summary statistics (e.g. n, mean, median, standard deviation, min, max). In addition, a separate summary of the number and percentage of participants in each category (as described in Section 4.1.8.1) is produced.

All concomitant, prior and post study treatment discontinuation medication and therapy data are listed for all participants.

4.1.9 Study Drug Compliance

This is covered in Section 4.4.1

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses.

Table 5 Endpoint analyses

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Dual Primary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting, per BICR					
Primary	PFS - BICR Assessments	FAS	Participants who have not progressed or died are censored at latest evaluable RECIST 1.1 assessment. Participants who progress or die after 2 missed visits are censored at last evaluable RECIST 1.1 assessment prior to the two missed visits.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.4
Sensitivity 1 (Evaluation time bias)	PFS - BICR Assessments	FAS	As for primary analysis, but the midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be used as the event time.	Stratified log-rank test.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity 2 (Attrition bias)	PFS - BICR Assessments	FAS	As for primary analysis, but using the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments will be included. Participants 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.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. Reverse Kaplan-Meier plot	4.2.1.5
Sensitivity 3 (Ascertainment bias)	PFS - BICR Assessments	FAS	As for primary analysis but using the site Investigator data which is a secondary efficacy variable.	Discrepancy between primary analysis using BICR assessments and the secondary analysis using Investigator assessment.	4.2.1.5
Sensitivity 4 (Subsequent anti-cancer therapy)	PFS - BICR Assessments	FAS	As for primary analysis, but participants who receive subsequent anti-cancer therapy prior to progression or death are censored at latest evaluable RECIST assessment prior to subsequent anti-cancer therapy.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Dual Primary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of OS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting					
Primary	OS	FAS	Participant not known to have died will be censored at last date participant was known to be alive.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.2.4
Sensitivity 1 (Attrition bias)	OS	FAS	As for primary analysis but censoring indicator of OS is reversed.	Reverse Kaplan-Meier plot Median	4.2.2.5
Secondary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of ORR					
Secondary	ORR	FAS (for participants who have measurable disease at baseline)	Participants without a response included as non-responders. If participants discontinue treatment without response or progression, receive a subsequent anti-cancer therapy and then respond are not included as responders.	Odds ratio from logistic regression model ORR with Clopper-Pearson CI Difference in ORR with Miettinen-Nurminen CI	4.2.3.3
Secondary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DoR					
Secondary	DoR	FAS (for participants who have confirmed CR or PR)	For participants who do not progress or die following a response, the censoring rules follow the rules for PFS censoring for the primary analysis.	Summary statistics. KM plots.	4.2.5.3
Supplementary	DoR	FAS (for participants who have confirmed CR or PR)	As for primary analysis but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.	Summary statistics. KM plots.	4.2.5.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity	DoR	FAS (for participants who have confirmed CR or PR)	For participants who do not progress or die following a response, censor at latest evaluable RECIST assessment prior to subsequent anti-cancer therapy, or at last one prior to two missed visits if death or progression occurs immediately after 2 missed visits.	Median DoR from KM estimates.	4.2.5.4
Secondary Objective 3: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS, per investigator assessment					
Secondary	PFS – Investigator Assessments	FAS	As described for primary PFS	As described for primary PFS	4.2.6.4
Sensitivity	PFS – Investigator Assessments	FAS	As described for sensitivity of PFS for BICR	As described for sensitivity of PFS for BICR	4.2.6.5
Secondary Objective 4: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DCR					
Secondary	DCR at 12 weeks	FAS (for participants who have measurable disease at baseline)	Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.	Odds ratio from logistic regression model	4.2.7.3
Secondary Objective 5: To assess pain in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in pain	FAS	Participants who have not deteriorated or died will be censored at last evaluable assessment. Participants who deteriorate or die after 2 missed visits are censored at last evaluable assessment prior to the 2 missed visits.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.8.4

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity	TTD in pain	FAS	TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.	As described for secondary TTD in pain.	4.2.8.5
Secondary Objective 6: To assess physical functioning in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4
Sensitivity	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5
Secondary Objective 7: To assess global health status/quality of life (GHS/QoL) in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4
Sensitivity	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary Objective 8: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TFST					
Secondary	TFST	FAS	Participants not known to have had a first subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a first subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a first subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots. Median. Time between progression and starting subsequent therapy.	4.2.9.3
Secondary Objective 9: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TSST					
Secondary	TSST	FAS	Participants not known to have had a second subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a second subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a second subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	As described for TFST.	4.2.10.3
Secondary Objective 10: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS2					
Secondary	PFS2	FAS	As described for primary PFS	As described for primary PFS	4.2.11.4

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary Objective 11: To assess the PK of Dato-DXd 6mg/kg IV Q3W					
Secondary	Plasma concentrations of Dato-DXd, total anti-TROP2 antibody, and MAAA-1181a (payload)	PAS	N/A	Summary statistics	4.2.12.2
Secondary Objective 12: To investigate the immunogenicity of Dato-DXd 6mg/kg IV Q3W					
Secondary	Presence of ADA	ADA evaluable set	N/A	Summary statistics	4.2.13.2
Safety: To assess the safety and tolerability profile of Dato-DXd relative to ICC					
Safety	Type, incidence and severity (graded by NCI CTCAE v5.0), seriousness and relationship to study medication of AEs	SAF	N/A	Descriptive	4.4.2
Safety	ECOG PS	SAF	N/A	Descriptive	4.4.9
Safety	Vital signs	SAF	N/A	Descriptive	4.4.6
Safety	Physical examination	SAF	N/A	Descriptive	4.4.10
Safety	Clinical laboratory tests	SAF	N/A	Descriptive	4.4.3, 4.4.4,4.4.5
Safety	ECGs	SAF	N/A	Descriptive	4.4.7
Safety	Echocardiogram/ multigated acquisition	SAF	N/A	Descriptive	4.4.8
Safety	Ophthalmologic assessments	SAF	N/A	Descriptive	4.4.11

4.2.1 Primary Endpoint - Progression Free Survival by BICR

4.2.1.1 Definition

Progression-free survival is defined as the time from the date of randomisation until the date of objective disease progression, as defined by RECIST 1.1, or death (by any cause in the absence of progression) regardless of whether the participant withdraws from

randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1). PFS is assessed by BICR assessment. A secondary endpoint analysis of PFS by investigator assessment is reported.

4.2.1.2 Derivations and Censoring Rules

Participants who have not progressed or died at the time of analysis are censored at the time of the latest date of assessment from their last evaluable RECIST assessment. However, if the participant progresses or dies immediately after two or more consecutive missed visits, the participant is 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).

If the participant has no evaluable RECIST assessment or does not have baseline data, they will be censored at the date of randomisation, unless they die within 2 scheduled scans of baseline (12 weeks + 1 week allowing for a late assessment within the visit window) in which case they are treated as an event with date of death as the event date. PFS censoring rules are described in Table 6.

Table 6 Outcome and date of event for PFS analysis

Scenario	Date of PD/ Death event or Censoring	PFS Outcome
Progression documented between scheduled visits after at most 1 missed assessment	Date of assessment of progression	Event
Death between assessment visits after at most 1 missed assessment	Date of death	Event
No baseline or evaluable RECIST assessment and death within 2 RECIST visits after the date of randomisation	Date of death	Event
No baseline or evaluable RECIST assessment and no death within 2 RECIST visits after the date of randomisation	Day 1 (Date of randomisation)	Censored
No PD or death at time of data cut-off	Date of last evaluable RECIST assessment*	Censored
Death or progression after two or more missed RECIST visits	Date of last evaluable RECIST assessment* prior to the 2 missed visits	Censored

*: if there are no evaluable post-baseline assessments prior to PD or death or data cut-off, participants will be censored at the date of randomisation.

Given the scheduled visit assessment scheme (i.e. every six weeks for the first 48 weeks then every nine weeks thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to nine-weekly this will equate to 17 weeks (i.e. take the average of 6 and 9 weeks which gives 7.5 weeks and then apply same rationale, hence 2 x 7.5 weeks + 1 week for an early assessment + 1 week for a late assessment = 17 weeks). The time period for the previous RECIST assessment will be from study days 288 to 344 (i.e. week 41 to week 49). From week 49 onwards (when the scheduling changes to nine-weekly assessments), two missing visits will equate to 20 weeks (i.e. 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks).

The following is also summarised in Table 7:

Table 7 Definition of two missed RECIST visits

Scheduled Assessment	Previous RECIST assessment	Two missed RECIST visits window
Q6W	No evaluable RECIST visits or no baseline RECIST scan	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W	Day 1	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W up to Week 48	>Day 1 – Day 287 (up to Week 41)	2 x 6 weeks + 2 weeks = 14 weeks (98 days)
	>Day 287 – Day 343 (Week 41 – Week 49) (change period from Q6W to Q9W)	2 x [(6 weeks+9 weeks)/2] + 2 weeks = 17 weeks (119 days)
Q9W thereafter	>Day 343 onwards	2 x 9 weeks + 2 weeks = 20 weeks (140 days)

The PFS time is always derived based on scan/assessment dates and not on visit dates.

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

- For investigator assessments, the date of progression is determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression.

- For BICR assessments, the date of progression is determined based on the earliest of the 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 BICR data.
- For both BICR and investigational assessments when censoring a participant for PFS, the participant is censored at the latest of the scan dates contributing to a particular overall visit assessment.

Note: for TLs only the latest scan date is recorded 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.

4.2.1.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Sections 4.2.1.2 and 4.2.1.5.

4.2.1.4 Primary Analysis of Progression Free Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of PFS in the FAS. The primary analysis of PFS is based on the BICR assessment of PD by RECIST 1.1.

The null hypothesis for the dual primary time to event endpoint of PFS is that there is no difference between Dato-DXd and ICC in the probability of a progression event in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for PFS.

H1: Differences between Dato-DXd and ICC for PFS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

PFS is analysed using a stratified log-rank test adjusting for the stratification factors of number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor.

The stratification variables are based on the values entered into IRT at randomisation, even if it is subsequently discovered that these values were incorrect.

If there are less than 10 events in total for a unique stratum or less than 2 events in either treatment group for a unique stratum then the strata are combined in the following order. The CDK4/6 strata (Yes vs No) are pooled first, followed by the number of previous lines of chemotherapy strata (1 vs 2) and then finally by the geographic region strata (United States, Canada, Europe vs Rest of World).

The hazard ratio (HR) and its confidence interval (CI) (95% and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9) and p-value are presented. The HR and CI are estimated from a stratified Cox Proportional Hazards model (with ties = Efron and stratification variables number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor) and the CI calculated using a profile likelihood approach. A HR less than 1 favours Dato-DXd.

Estimates and 95% CI for PFS rates at 3 months intervals and median PFS for each treatment group are presented.

Proportionality assumption

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of log(-log(survival probability)) versus log (time) and, if these raise concerns, by fitting a time dependent covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods for example 0-6m, 6-12m etc. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over the observed extent of follow-up unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

Summaries

In addition to the analyses described above, the following supportive summaries are produced.

Kaplan-Meier (KM) plots of PFS are presented by treatment group. Summaries of the number and percentage of participants experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided for each treatment.

The treatment status at progression of participants at the time of analysis is summarised. This includes the number (%) of participants who were on treatment at the time of progression, the number (%) of participants who discontinued IP prior to progression, the number (%) of participants who have not progressed and were on IP or discontinued IP.

The number of participants censored may be summarised by treatment group together with baseline prognostic factors of the censored participants. This number and percentage of prematurely censored participants is summarised. A participant will be defined as prematurely censored if they did not progress (or die in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval (+ 2 weeks) prior to the DCO date.

Additionally, summary statistics are given for the number of days from censoring to DCO for all censored participants.

The duration of follow-up is summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) participants only, presented by treatment group.

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

Summaries of the number and percentage of participants who miss two or more consecutive RECIST assessments is presented for each treatment group.

All of the collected RECIST 1.1 data is listed for all randomised participants. In addition, a summary of new lesions (i.e. sites of new lesions) is produced.

4.2.1.5 Sensitivity Analyses of Progression Free Survival

Sensitivity Analysis 1 - 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 analysed 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 participants 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 & Chen, 2010). To support this analysis, the mean of participant-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR RECIST assessments.

Sensitivity Analysis 2 - Attrition bias

Attrition bias is assessed by repeating the primary PFS analysis except that the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments

are included. In addition, and within the same sensitivity analysis, participants 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 are censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis is supported by a KM 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.

Sensitivity Analysis 3 - Ascertainment bias

Ascertainment bias is assessed by analysing the site investigator data which is a secondary efficacy endpoint (analysis methods presented in Section 4.2.6). The stratified log rank test is repeated on PFS using the site investigator data based upon RECIST. The HR and CI are presented.

If there is an important discrepancy between the primary analysis using the BICR data and this sensitivity analysis using site investigator data a summary table is produced showing the number and proportion of participants with site but no central confirmation of progression and with progression determined by central review but not at site. Such participants 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, Gaschler-Markefski, & Bluhmki, 2001), but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group.

Sensitivity Analysis 4 – Subsequent Anti-cancer Therapy

An additional sensitivity analysis is produced which is a repeat of the primary analysis for PFS, but the censoring rule is modified so that participants who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death are censored at their last evaluable assessment prior to taking the subsequent anti-cancer therapy.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival.

Deviation Bias Sensitivity Analysis

A ‘deviation bias’ sensitivity analysis may be performed as discussed in Section 3.3.10.

4.2.1.6 Subgroup Analyses

Subgroup analyses are conducted comparing PFS between the treatments for the following subgroup of the FAS:

1. Stratification factors at randomisation:

    - Number of previous lines of chemotherapy: 1, 2
    - Geographic region: Region 1 [US, Canada, Europe], Region 2 [Rest of World]
    - Prior use of CDK4/6 inhibitor: Yes, No

2. Exploratory factors

- Prior use of taxanes and/or anthracyclines: taxanes alone, anthracyclines alone, both taxanes and anthracyclines, neither taxanes nor anthracyclines
- Age at randomisation: <65, ≥65 years of age
- Race: Asian, non-Asian
- Pre-selected investigator's choice of chemotherapy: Capecitabine, Gemcitabine, Eribulin mesylate, Vinorelbine
- Brain metastases: Yes, No
- Sex: male, female

The subgroup analyses for the stratification factors and the pre-selected investigator's choice of chemotherapy will be based on the values obtained from the IRT system, all other factors will be based on values recorded on the eCRF, or from the third-party vendor data.

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

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

For each subgroup level of a factor, the HR and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties and using a BY statement for the subgroup factor.

These HRs and associated two-sided 95% profile likelihood CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

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 descriptive summaries will be provided.

The presence of quantitative interactions may be assessed by means of an overall global interaction test for strata and possibly subgroups:

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by treatment interaction terms, with one that excludes the interaction terms, and will be assessed at the 2-sided 10% significance level. If there are not more than 10 events per stratum for any covariate (i.e. within each stratum of a treatment*covariate interaction (2 treatments * 2 levels of the covariate = 4 stratum)), a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria, then the covariate-by-treatment interaction term should be omitted from the model. Moreover, if the covariate does not have more than 10 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved, then it will be concluded that overall, the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant, and all excluded interactions are non-significant. Throughout this process all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

Any quantitative interactions identified using this procedure will then be tested to rule out any qualitative interaction using the approach of (Gail & Simon, 1985).

4.2.2 Primary Endpoint - Overall Survival

4.2.2.1 Definition

Overall survival is defined as the time from the date of randomisation until death due to any cause regardless of whether the participant withdraws from randomised therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomisation + 1).

4.2.2.2 Derivations and Censoring Rules

Any participant not known to have died at the time of analysis is censored based on the last recorded date on which the participant was known to be alive.

Note: Survival calls are made in the week following the date of data cut-off (DCO) for the analysis, and if participants are confirmed to be alive or if the death date is after the DCO date, these participants are censored at the date of DCO. This is done at DCO2, DCO3 and DCO4. The status of ongoing, withdrawn (from the study) and "lost to follow-up" participants at the time of the primary OS analysis should be obtained by the site personnel by telephone contact with the participant, participant's family, by contact with the participant's current physician, or local death registries. If the participant has actively withdrawn consent to the processing of their personal data, the vital status of the participant can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For the OS analysis at DCO2 and DCO3, performed prior to the primary 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 participant was known to be alive for those participants still on treatment (since the SURVIVE module is only completed for participants off treatment if a survival sweep is not performed). The last date for each individual participant is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start and stop dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anti-cancer treatment
- Date last known alive on survival status CRF
- End of study date

Duration of follow-up is derived as time from randomisation to the date of death (i.e. overall survival) or to the date of censoring (date last known to be alive).

4.2.2.3 Handling of Dropouts and Missing Data

If a participant is known to have died where only a partial death date is available, then the date of death is imputed following the rules in Section 3.3.7.

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

4.2.2.4 Primary Analysis of Overall Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of OS in the FAS.

The null hypothesis for the dual primary time to event endpoint of OS is that there is no difference between Dato-DXd and ICC in the probability of a death in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for OS.

H1: Differences between Dato-DXd and ICC for OS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

Overall survival will be analysed using a stratified log-rank test, adjusting for the stratification factors at randomisation. The treatment effect of Dato-DXd against ICC will be estimated by the HR together with its 95% CI and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9.

Estimates and 95% CI for OS rates at 6 monthly intervals are presented along with the median OS for each treatment group.

Summaries

Kaplan-Meier (KM) plots of OS are presented by treatment group. Summaries of the number and percentage of participants who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided.

The number of participants prematurely censored will be summarised by treatment arm. A participant would be defined as prematurely censored if their survival status was not defined at the DCO.

In addition, the median duration of follow-up is presented for censored participants by treatment group, and for all participants by treatment group and overall.

4.2.2.5 Sensitivity Analyses of Overall Survival

A sensitivity analysis for OS examining the censoring patterns to rule out attrition bias with regard to the primary treatment comparisons is achieved by a KM plot of time to censoring

where the censoring indicator of OS is reversed. The KM estimates of median follow-up (overall and by treatment group) are also summarised.

4.2.2.6 Subgroup Analyses

Subgroup analyses will be conducted for OS using the same methodology as described for PFS in Section 4.2.1.6.

4.2.3 Secondary Endpoint - Objective Response Rate

4.2.3.1 Definition

Both unconfirmed and confirmed ORR are assessed, where:

- Confirmed ORR is the percentage of participants with an investigator-assessed response of CR or PR 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.
- Unconfirmed ORR is the percentage of participants with at least one investigator-assessed visit response of CR or PR.

The denominator will be defined as subset of all randomised participants with measurable disease at baseline.

ORR will also be defined using the BICR data to define a visit response of CR or PR, with the denominator defined as subset of all randomised participants with measurable disease at baseline per BICR.

4.2.3.2 Derivations

Data obtained up until progression, or last evaluable assessment in the absence of progression, are included in the assessment of ORR, regardless of whether the participant withdraws from therapy. Participants who discontinue randomised treatment without progression, receive a subsequent anti-cancer therapy and then respond are not included as responders in the ORR (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy). For confirmed ORR both visits contributing to a response must be prior to subsequent therapy for the participant to be considered as a responder.

For confirmed ORR, in the case where a participant 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 participant is defined as a responder. Similarly, if a participant 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 is assigned.

ORR is based on all scans regardless of whether they were scheduled or not.

A participant is classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation and confirmed by repeat imaging, prior to RECIST progression and prior to starting any subsequent cancer therapy.

4.2.3.3 Primary Analysis of Objective Response Rate

ORR based on both a confirmed and unconfirmed tumour response are analysed.

A logistic regression model is fitted to tumour response (yes/no) including treatment and the same stratification factors as the primary PFS endpoint as fixed effects. The results of the analysis are presented in terms of an adjusted odds ratio (an odds ratio greater than 1 favour Dato-DXd) together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The adjusted Least Squares (LS) Means response rate (using LSMEANs statement with OM option) from the logistic regression model together with the corresponding 95% CI is presented for each treatment group.

If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran–Mantel–Haenszel (CMH) test is presented. The CMH test is stratified using the same stratification factors as the primary PFS endpoint. The results of the analysis are presented in terms of an odds ratio together with the 95% CI and p-value. The odds ratio, 95% CI and p-value are obtained using SAS PROC FREQ and the CMH test option.

Both unconfirmed and confirmed ORR as assessed by site investigator will be estimated and presented along with the corresponding exact 95% Clopper-Pearson CI for each treatment arm. The difference in ORR between treatment arms will be reported using point estimates and their two-sided 95% CIs by the Miettinen-Nurminen method (Miettinen & Nurminen, 1985). A Summary will be produced that presents the number and percentage of participants with both an unconfirmed and a confirmed tumour response (CR/PR) based upon the number of participants with measurable disease at baseline per the site investigator.

Summaries will also be produced for ORR per BICR, based upon the number of participants with measurable disease at baseline per BICR.

4.2.4 Secondary Endpoint - Best Objective Response

4.2.4.1 Definition

Best objective response (BoR) is a supportive endpoint for ORR. BoR is calculated based on the overall visit responses from each RECIST assessment, described in Section 3.3.8.3. It is the best response a participant has had following randomisation, but prior to starting any subsequent anti-cancer therapy and up to and including RECIST progression or the last

evaluable assessment in the absence of RECIST progression. Categorisation of BoR is based on RECIST using the following response categories: CR, PR, SD, PD and NE.

4.2.4.2 Derivations

BoR is derived using confirmed CR or PR and separately using unconfirmed CR or PR.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment is 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 randomisation. For CR/PR, the initial overall visit assessment that showed a response uses the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST from the overall visit response using all BICR data up until the first progression event. It will also be determined programmatically based on RECIST using all site investigator data up until the first progression event. The denominators for each case will be consistent with those used in the ORR analysis.

BoR is determined based on RECIST using both all site investigator data and separately all BICR data up until the earliest of the first progression event/last evaluable assessment in the absence of RECIST or start of any subsequent cancer therapy. The denominators are consistent with those used in the ORR analysis.

For participants whose progression event is death, BoR is calculated based upon all evaluable RECIST assessments prior to death.

For participants who die with no evaluable RECIST assessments, if the death occurs ≤7 weeks (i.e. 6 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR is assigned to the progression (PD) category. For participants who die with no evaluable RECIST assessments, if the death occurs > 7 weeks after randomisation then BoR is assigned to the NE category. For participants with no evaluable RECIST assessments post randomisation, then BoR is also assigned to the NE category.

4.2.4.3 Primary Analysis of Best Objective Response

For each treatment arm, BoR will be summarised by n (%) for each category (CR, PR, SD, PD, and NE) using confirmed and unconfirmed CR/PR responses.

4.2.5 Secondary Endpoint - Duration of Response

4.2.5.1 Definition

DoR will be defined as the time from the date of first documented confirmed response of CR or PR until date of documented progression per RECIST 1.1 (as assessed by

Investigator assessment) or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1).

DoR will also be defined using the BICR data to define the overall visit response.

4.2.5.2 Derivations and Censoring Rules

The end of response coincides with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of CR or PR as defined in Table 4.

If a participant does not progress following a response, then the PFS censoring time is used.

4.2.5.3 Primary Analysis of Duration of Response

The analysis will include all randomised participants as randomised who have a confirmed response, regardless of whether the participant withdraws from therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression.

Descriptive data will be provided for the DoR, including the associated KM curves (without any formal comparison or p-value attached).

KM estimates for median DoR and 25th and 75th percentiles and corresponding 95% CI for DoR will be summarised. KM estimates for percentage remaining in response at 6 monthly intervals are presented for each treatment group.

Additionally, median, 25th and 75th percentiles for time from randomisation to onset of response will be calculated using standard descriptive statistics.

4.2.5.4 Sensitivity Analyses of Duration of Response

A sensitivity analysis is included using the same definition, but where participants who do not progress or die following a response, censor at latest evaluable RECIST assessment prior to subsequent anti-cancer therapy, or at last one prior to two missed visits if death or progression occurs immediately after 2 missed visits.

Median DoR will be estimated using the same methodology described in Section 4.2.5.3.

4.2.5.5 Supplementary Analyses of Duration of Response

A supplementary analysis is included where the DoR evaluation is repeated but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.

4.2.6 Secondary Endpoint – PFS by Investigator Assessment

4.2.6.1 Definition

PFS by Investigator assessment is defined as the time from the date of randomisation until the date of PD, as defined by RECIST 1.1 (by Investigator assessment) or death (by any cause in the absence of progression) regardless of whether the participant withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1).

4.2.6.2 Derivations and Censoring Rules

This secondary endpoint of PFS based on Investigator assessment will be derived and censored using the same methodology described in Section 4.2.1.2

4.2.6.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.1.3.

4.2.6.4 Primary Analysis of Progression Free Survival

This secondary endpoint of PFS based on Investigator assessment will be analysed using the same methodology described in Section 4.2.1.4.

4.2.6.5 Sensitivity Analyses of Progression Free Survival

Sensitivity analyses for the secondary endpoint of PFS based on Investigator assessment will be analysed using the same methodology described in Section 4.2.1.5.

4.2.6.6 Subgroup Analyses

Unless there is a marked difference between the results of the statistical analyses of the PFS from the BICR data (as described in Section 4.2.1) and that of the site Investigator data, the subgroup analyses described in Section 4.2.1.6 will only be performed on the PFS endpoint using the BICR data.

4.2.7 Secondary Endpoint –Disease Control Rate

4.2.7.1 Definition

Disease control rate at 12 weeks is defined as the percentage of participants who have a confirmed CR or PR or have demonstrated SD for at least 11 weeks (i.e. 12 weeks – 1 week to allow for an early assessment within the assessment window) after randomisation without subsequent cancer therapy per RECIST 1.1, as assessed per Investigator assessment and derived from the raw tumour data.

DCR at 12 weeks will also be defined using the BICR data to define the overall visit response.

4.2.7.2 Derivations

Data obtained from randomisation up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of DCR, regardless of whether the participant withdraws from therapy. Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.

Duration of SD (weeks) is defined as: (date last evaluable assessment of SD in the absence of progression prior to subsequent cancer therapy - randomisation date +1)/7.
Participants without a post-baseline tumour assessment are considered to have no clinical benefit.

DCR is based on all scans regardless of whether they were scheduled or not.

4.2.7.3 Primary Analysis of Disease Control Rate

DCR will be analysed using the same methodology described for ORR in Section 4.2.3.3.

The analysis will be performed on the FAS (for participants who have measurable disease at baseline).

4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30

4.2.8.1 Definition

EORTC QLQ-C30

The European Organisation for Research and Treatment of Cancer (EORTC) 30-item quality of life (QoL) questionnaire (QLQ-C30) consists of 30 questions that are combined to produce 5 multi-item functional scales (physical, role, cognitive, emotional, and social), 3 multi-item symptom scales (fatigue, pain, and nausea/vomiting), a 2-item global health status/QoL scale, 5 individual item symptom scores (appetite loss, dyspnoea, insomnia, constipation, and diarrhoea), and 1 item on the financial impact of the disease.

TTD in pain, physical functioning and GHS/QoL items are secondary endpoints. All other items are part of the exploratory endpoints.

The number of items and item range for each scale/item are displayed in Table 8.

Table 8 EORTC QLQ-C30 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Global health status/ QoL	QL	2	6	29, 30

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Functional scales				
Physical	PF	5	3	1-5
Role	RF	2	3	6, 7
Cognitive	CF	2	3	20, 25
Emotional	EF	4	3	21-24
Social	SF	2	3	26, 27
Symptom scales				
Fatigue	FA	3	3	10, 12, 18
Pain	PA	2	3	9, 19
Nausea/ vomiting	NV	2	3	14, 15
Symptom items				
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17

QoL=Quality of life.

4.2.8.2 Derivations and Censoring Rules

Scoring algorithm

The EORTC QLQ-C30 v3 is scored according to the EORTC QLQ-C30 Scoring Manual (Fayers, et al., 2001). Items are scored on a 4-point verbal rating scale: “Not at all”, “A little”, “Quite a bit”, and “Very much”. Scores are then transformed to give a score from 0 to 100 for each of the symptom scales, functional scales, and the global QoL scale. Higher scores on the global measure of health status/QoL and functional scales indicate better health status/function, but higher scores on symptom scales/scores represent greater symptom severity.

The EORTC QLQ-C30 functional and symptom scales, individual symptom items and global health status are derived as follows:

1. Calculate the average of the items that contribute to the scale or take the value of an individual item, i.e. the raw score (RS):

   $RS = (I1 + I2 + \ldots + In) / n,$

where I1 + I2 + … + In are the items included in a scale and n is the number of items in a scale.

2. Use a linear transformation to standardise the raw score, so that scores range from 0 to 100, where a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

   Functional scales: Score = (1 – [RS – 1] / range) * 100

   Symptom scales/items; global health status: Score = ([RS – 1] / range) * 100,

   where range is the difference between the maximum and the minimum possible value of RS.

Change from baseline

Changes in score from baseline are calculated for each of the functional scales, symptom scales and global health status/QoL scale at each assessment, where baseline is defined and calculated as explained in Section 3.3.3.

Deterioration

Deterioration is defined as change from baseline that reaches a clinically meaningful deterioration threshold. Anchor-based methods using the participant-based anchors PGIS and PGIC will be considered to define thresholds for clinically meaningful within-participant change used in the time to deterioration (TTD) endpoints. Other methods including distribution-based methods, cumulative distribution function, and probability density function curves, and methods using other anchors may also be considered.

Clinically meaningful change thresholds will be estimated for the following outcomes:

- EORTC QLQ-C30: Global health status/QoL, functioning, and select symptom subscales including pain and fatigue
- EORTC QLQ IL116: breast symptoms, arm symptoms (See Section 4.2.18)

The analysis to define clinically meaningful change thresholds in the TTD PRO endpoints will include all randomised participants using the pooled treatment arms data prior to database lock. Further details on methodologies to define these clinically meaningful change thresholds will be provided in a separate PRO SAP.

Improvement, deterioration or no change will be defined based on a minimal clinically important difference (MCID). For each EORTC QLQ-C30 scale/item score not included in

the estimation methods, the MCID will be pre-specified as requiring a 10-point change from baseline.

If the estimation methods have not yet been performed to define the clinically meaningful threshold for improvement, deterioration or no change, then a 10-point change from baseline will be used.

TTD

Time to deterioration (TTD) is defined as time from the date of randomisation to the date of first deterioration, regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

Time to deterioration=date of event or censoring – date of randomisation+1

For the derivation of TTD, the following censoring rules are used:

The date of a participant's death (by any cause) will be used in the absence of a clinically meaningful deterioration, regardless of whether the participant discontinues study treatment or receives another anti-cancer therapy prior to deterioration. Death will be included as an event only if the death occurs within 2 assessments of the last PRO assessment where the change could be evaluated. If death occurs after 2 missed assessments the participant will be censored as described below.

Participants whose symptoms, functioning or global health status/QoL have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptoms, functioning or global health status/QoL could be evaluated. Also, if symptoms, functioning or global health status/QoL deteriorate after 2 or more missed PRO assessments or the participant dies after 2 or more missed PRO assessments, the participant will be censored at the time of the last PRO assessment where the symptoms, functioning or global health status/QoL could be evaluated prior to the 2 missed assessments.

If a participant has no evaluable post-baseline data or does not have baseline data, they will be censored at date of randomisation.

Given the scheduled assessment scheme (i.e. every three weeks for the first 48 weeks then every six-weeks thereafter) the definition of 2 missed assessments will change. If the previous PRO assessment is less than study day 309 (i.e. week 44) then two missing visits will equate to 46 days since the previous PRO assessment allowing for early and late visits (i.e. 2 x 3 weeks + 2 days for an early assessment + 2 days for a late assessment = 46 days). If the two missed assessments occur over the period when the scheduled frequency of PRO assessments changes from three-weekly to six-weekly this will equate to 67 days (i.e. take

the average of 3 and 6 weeks which gives 4.5 weeks and then apply same rationale, hence 2 x 4.5 weeks + 2 days for an early assessment + 2 days for a late assessment = 67 days). The time period for the previous PRO assessment will be from study days 309 to 344 (i.e. week 44 to week 49). From week 49 onwards (when the scheduling changes to six-weekly assessments), two missing assessments will equate to 88 days (i.e. 2 x 6 weeks + 2 days for an early assessment + 2 days for a late assessment = 88 days). Study day will be calculated in line with Section 3.3.5 using first date of dose of IP as Day 1. If participant withdraws treatment prior to week 48 the assessment schedule is assumed to be Q6W relative to C1D1. If participant discontinues treatment after week 48 assessment schedule is the same as above.

The following is also summarised in Table 9:

Table 9 Definition of two missed PRO visits

Scheduled Assessment	Previous PRO assessment	Two missed PRO visits window
Q3W	No evaluable post-baseline PRO data or no baseline PRO data	2 x 3 weeks + 2 days = 44 days
Q3W	Day 1	2 x 3 weeks + 2 days = 44 days
Q3W up to Week 48*	>Day 1 – Day 308 (up to Week 44)	2 x 3 weeks + 4 days = 46 days
	>Day 308 – Day 343 (Week 44 – Week 49) (change period from Q3W to Q6W)	2 x [(3 weeks+6 weeks)/2] + 4 days = 67 days
Q6W thereafter*	>Day 343 onwards	2 x 6 weeks + 4 days = 88 days

*Follow schedule until treatment discontinuation after which a Q6W (relative to C1D1) window will be assumed.

4.2.8.3 Handling of Dropouts and Missing Data

For each subscale, if <50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers, et al., 2001). If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing.

4.2.8.4 Primary Analysis of EORTC QLQ-C30

TTD in the pain scale, physical functioning scale and GHS/QoL scale will be analysed using the same time-to-event analysis methodology described in Section 4.2.1.4.

4.2.8.5 Sensitivity Analysis of EORTC QLQ-C30

A sensitivity analysis for TTD is performed where TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

This will be analysed using the same methodology described in Section 4.2.8.4.

4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death

4.2.9.1 Definition

Time to first subsequent therapy or death (TFST) is defined as the time from the date of randomisation to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomised treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the initial therapy. In this situation, TFST is calculated as time from randomisation to the start of the initial therapy or death.

4.2.9.2 Derivations and Censoring Rules

Any participant not known to have had a first subsequent anti-cancer therapy will be censored at the last date that the participant was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before first subsequent therapy, these participants will be censored at the earliest of their last known to be alive and termination dates.

Any participant without a TTSCAPRX form (not known to have had a first subsequent anti-cancer therapy) and have not died at the time of the analysis is censored at the last date that the participant was known to be alive according to the rules detailed in Section 4.2.2.2, where any participant recorded as alive or to have died after DCO date is censored at the date of DCO.

4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death

Statistical analysis

The time to first subsequent therapy or death (TFST) is analysed using the same methodology as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI is presented. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

Summaries

In addition, medians and a KM plot of the time to the start of subsequent therapy are presented by treatment group and the time between progression and starting subsequent therapy is assessed. This is summarised per treatment group, but no formal comparisons are made.

In participants who received a subsequent anti-cancer therapy, a summary table of first subsequent anti-cancer therapies by therapy class and treatment group is provided, as well as response to first subsequent anti-cancer therapy by treatment group.

A summary of the number of participants prematurely censored is also produced.

4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death

4.2.10.1 Definition

Time to second subsequent therapy or death (TSST) is defined as the time from the date of randomisation to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of first subsequent treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the first subsequent anti-cancer therapy and the second alternative cancer therapy they receive is the second subsequent anti-cancer therapy. In this situation, TSST is calculated as time from randomisation to the start of the second subsequent anti-cancer therapy or death.

4.2.10.2 Derivations and Censoring Rules

Any participant not known to have had a second subsequent anti-cancer therapy or have not died at the time of the analysis is censored at the last date that the participant was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before second subsequent therapy, these participants are censored at the earliest of their last known to be alive and termination dates.

4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death

The time to the second subsequent therapy or death (TSST) is analysed using the same methods as that used for the analysis of TFST (see Section 4.2.9.3). The same statistics and summary tables are produced for TSST as for TFST.

4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death

4.2.11.1 Definition

Time from randomisation to second progression or death (PFS2) is defined as the time from date of randomisation to the earliest progression event following first objective progression subsequent to the first subsequent therapy, or death. The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological progression, symptomatic progression, or death. Second progression status will be reviewed every 3 months following the progression event used for the primary variable PFS (the first progression) and status recorded.

4.2.11.2 Derivations and Censoring Rules

Participants alive and for whom a second disease progression has not been observed are censored at date last known alive and without a second disease progression. Therefore, they are censored at:

- The PFS assessment date if the participant has not had a first progression or death (PFS censoring date).
- The date the participant is last known to not have received a first subsequent therapy if a participant has had a first progression and not started a subsequent therapy (TFST censoring date).
- The latest PFS2 assessment date following first objective progression, if the participant has started a first subsequent therapy and PFS2 event (second progression or death) has not been observed. If a PFS2 assessment has not occurred, then the participant is censored at the day before starting the first subsequent therapy.

However, if the participant experiences a second progression or dies immediately after two or more consecutive missed visits, the participant is censored at the time of the last evaluable PFS2 assessment prior to the two missed visits. Note for deaths prior to 1st progression, but immediately after two or more consecutive missing visits, the participant is censored at the time of the last evaluable PFS1 assessment.

4.2.11.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.11.2.

4.2.11.4 Primary Analysis of PFS2

Statistical Analysis

PFS2 is analysed using identical methods as outlined for PFS (see Section 4.2.1.4) and adjusting for the same stratification factors. The HR for the treatment effect together with its 95% CI are presented. Medians and KM plots are presented to support the analysis.

Summaries

The number and percentage of participants experiencing a PFS2 event and the type of progression (objective progression by RECIST, symptomatic progression or other) are also summarised by treatment group, as well as summaries of deaths in the absence of second progression, and categories of PFS2 censoring. Time from randomisation to second progression will be summarised by treatment arm.

4.2.12 Secondary Endpoint - Pharmacokinetics

4.2.12.1 Derivations

Pharmacokinetic concentration data will be collected according to Section 8.5.1 of the CSP. The schedule of assessment is as per the schedule of activities (SoA) of the CSP. Whole blood samples for determination of plasma concentration of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a, will be obtained for all participants receiving Dato-DXd.

4.2.12.2 Primary Analyses of Pharmacokinetics

All plasma concentrations will be listed for each participant, for each sampling time and each dosing day, regardless of whether they are excluded from summary statistics due to deviation (e.g. as a result of dose interruption, reduction or missing the dose before PK sample collection, or sampling time deviation, etc).

Plasma concentrations of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a will be summarised by visit and nominal sample time using standard summary statistics for PK concentrations (geometric mean, geometric coefficient of variation, geometric mean ± geometric standard deviation, arithmetic mean, standard deviation, minimum, maximum and n) for the Dato-DXd treatment arm.

Individual concentrations below the Lower Limit of Quantification (LLOQ) of the bioanalytical assay will be reported as not quantifiable (NQ) in the listings with the LLOQ defined in the footnotes of the relevant tables, figures and listings (TFLs). Individual plasma concentrations that are Not Reportable will be reported as NR and those that are missing will be reported as NS (No Sample) in the listings. For data below limit of quantification (BLQ), NR or NS the following rules will apply:

- Any values reported as NR or NS will be excluded from the summary tables and corresponding figures.
- If, at a given time point, 50% or less of the plasma concentrations are NQ, the geometric mean, CV%, geometric CV%, mean and SD will be calculated treating the NQ as LLOQ.
- If more than 50%, but not all, of the concentrations are NQ, the geometric mean, CV%, geometric CV%, and SD will be reported as data not calculable (NC). The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.
- If all the concentrations are NQ, the geometric mean, mean, minimum, median and maximum will be reported as NQ and the CV%, geometric CV% and SD as NC.

Participants with protocol deviations seriously impacting PK results are excluded from the summary tables.

Non-compartment PK parameters that can be derived with sparse PK sampling, such as peak and trough concentrations after first dose and at steady state, will be derived directly from concentration summary and reported as data allows.

Population PK, and exploratory exposure response/safety analyses will be performed. This is documented in a separate analysis plan and the results presented separately from the main CSR.

4.2.13 Secondary Endpoint - Immunogenicity

4.2.13.1 Derivations

The presence of ADAs will be assessed in plasma samples taken according to the SoA in the CSP. ADA samples may be further tested for characterisation of the ADA response. ADA results from each sample is reported as either positive or negative. If the sample is positive, the ADA titre is reported as well. In addition, the presence of neutralizing antibody (nAb) will be tested for all ADA-positive samples using a ligand-binding assay. The nAb results is reported as positive or negative.

The ADA categories are defined as follows:

- 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.

- Treatment-boosted ADA, defined as a baseline positive ADA titre that was boosted to a 4-fold or higher-level following drug administration.
- Treatment-emergent ADA, defined as either Treatment-induced or treatment-boosted ADA.
- Persistently positive ADA, defined as having at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurement or an ADA positive result at the last available assessment. The category may include participants meeting these criteria who are ADA positive at baseline.
- Transiently positive ADA, defined as having at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive. The category may include participants meeting these criteria who are ADA positive at baseline.
- nAb positive at any visit (at baseline or post-baseline), if available.

4.2.13.2 Primary Analysis of Immunogenicity

Immunogenicity results will be listed by participant. A summary will be provided of the number and percentage of participants who develop detectable anti-Dato-DXd antibodies by ADA categories (see Section 4.2.13.1) using the ADA evaluable set. Anti-drug antibody titre and neutralising ADA data will be listed for samples confirmed positive for the presence of anti-Dato-DXd antibodies. Immunogenicity data listings will be based on SAF. AEs in ADA positive participants by ADA positive category will be listed.

The effect of ADA on PK, efficacy, and safety will be evaluated, if the data allow.

4.2.14 Exploratory Endpoint – Patient Reported AEs and Treatment Tolerability

4.2.14.1 Definition

PRO-CTCAE

The Patient-Reported Outcomes version of the common criteria for adverse events (PRO-CTCAE), a PRO version of the CTCAE system developed by the National Cancer Institute (NCI), is included to evaluate symptomatic toxicity from the participants' perspective.

PRO-CTCAE is an item library of symptoms experienced by participants while undergoing treatment of their cancer. Symptoms have been converted to participant terms (e.g. CTCAE term "myalgia" converted to "aching muscles"). Items capture the presence, frequency, severity and/or interference with daily activities, depending on the AE. For each question, participants select the value that best describes their experience over the past week, on a 5-point ordinal scale.

The items pre-selected for this study include mouth/throat sores, decreased appetite, nausea, vomiting, constipation, diarrhoea, abdominal pain, shortness of breath, cough, rash, hair loss, hand-foot syndrome, numbness/tingling, and fatigue.

EORTC IL117

The EORTC IL is an online platform comprised of more than 900 individual items from over 60 EORTC questionnaires. The pre-selected items for this study will include dry eyes, mouth pain, and sore mouth (i.e. EORTC IL117).

The recall period is during the past week. Items are scored on a 4-point verbal rating scale: "Not at all", "A little", "Quite a bit", and "Very much".

PGI-TT

The PGI-TT is a single item to assess how a participant perceives the overall tolerability of the IP. The responses indicate how bothered the participant was in the last 7 days by the side effects of their cancer treatment and are scored on a 5-point scale: 1 = Not at all; 2 = A little bit; 3 = Somewhat; 4 = Quite a bit; 5 = Very much.

4.2.14.2 Derivations and Censoring rules

Compliance

Summary measures of compliance over time will be derived for all PRO questionnaires. These will be based upon:

- Expected questionnaire: A questionnaire that is expected to be completed at a scheduled assessment time i.e. a questionnaire from a participant who has not withdrawn from the study at the scheduled assessment time, excluding participants in countries with no available translation.
  - For participants that have progressed or discontinued study treatment, the earliest of date of study treatment discontinuation or date of progression will be used to determine the last on treatment windowed assessment for each participants expected forms using the analysis windows as described in Section 3.3.5. If the date falls before the end of the assessment window, then that assessment will only be considered expected if they have a received form. If they have not received a form, then this assessment is not considered expected as they have not had the full opportunity to complete the questionnaire within the window. For participants who have not discontinued study treatment or progressed, the date of the DCO will be used to determine the last on treatment assessment for their last expected form following the same approach as above.

  - For follow up assessments (EORTC QLQ-C30, EORTC IL116, PGIS and EQ-5D-5L), if a participant has not discontinued study treatment then no follow up forms will be expected. For participants who have discontinued study treatment, and discontinued the study, the date of study discontinuation will be used to determine the last expected assessment that a form should have been completed. For participants who have discontinued study treatment, and not discontinued the study, the date of the DCO will be used to determine whether a form is expected following the same approach as above. For PRO-CTCAE, PGI-TT, EORTC IL117 and PGIC follow up forms will not be expected as these are not collected during post treatment follow-up.

- Received questionnaire: A questionnaire that has been received and has a completion date and at least 1 individual item completed.
- Evaluable questionnaire: A questionnaire with a completion date and at least 1 subscale that is non-missing.

Compliance over time will be calculated separately for each timepoint, including baseline, as the number of participants with an evaluable questionnaire at the time point, divided by number of participants still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculate separately for each timepoint, including baseline, as the number of evaluable questionnaires, divided by the number of received questionnaires. For compliance over time all timepoints (including follow-up time period) with at least 20 participants in one of the treatment arms will be reported.

4.2.14.3 Handling of Dropouts and Missing data

For PRO-CTCAE only participants in countries where a linguistically validated version of the PRO-CTCAE is available for administration are required to complete this questionnaire and thus included the analysis.

4.2.14.4 Primary Analyses of PRO-CTCAE

PRO-CTCAE data will be summarised descriptively for each item by treatment group. The summary will include:

- Number and percentage of participants reporting different levels of responses at each time point
- Number and percentage of the worst response option reported by participants within 12 weeks
- Number and percentage of participants who report the presence of the symptom at baseline
- Number and percentage of participants who report any worsening from baseline at any time within 12 weeks

- Number and percentage of participants who worsen from a score <3 (on a 0-4 scale) at baseline to a score 3 or 4 at any time within 12 weeks

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PRO-CTAE is summarised using the SAF.

4.2.14.5 Primary Analyses of EORTC IL117

The number and percentage of participants with each level of response for each EORTC IL117 item at baseline and over time is summarised by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm. EORTC IL117 is summarised using the SAF.

4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability

Responses for PGI-TT are summarised descriptively as number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

The PGI-TT is summarised using the SAF.

4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS)

4.2.15.1 Definition

The PGIS is a single item to assess how a participant perceives the overall severity of cancer symptoms over the past week. The responses are scored on a 4-point scale: 1 = None; 2 = Mild; 3 = Moderate; 4 = Severe.

4.2.15.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.15.3 Analysis of Patient Global Impression of Severity

Responses for PGIS is summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIS will be based on the FAS.

4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC)

4.2.16.1 Definition

The PGIC is a single item to assess how a participant perceives the overall change in health status since the start of IP. The responses are scored on a 7-point scale: 1 = Much better; 2 = Moderately better; 3 = A little better; 4 = About the same; 5 = A little worse; 6 = Moderately worse; 7 = Much worse.

4.2.16.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.16.3 Analysis of Patient Global Impression of Change

Responses for PGIC are summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

In addition, a stacked horizontal bar chart showing the percentage of participants in each PGIC category by timepoint is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIC will be based on the FAS.

4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL

4.2.17.1 Definition

Refer to Section 4.2.8.1 for the relevant definitions of EORTC QLQ-C30.

4.2.17.2 Derivation and Censoring Rules

Refer to Section 4.2.8.2 for the derivation and censoring rules of scoring algorithm, change from baseline, deterioration and TTD. Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.17.3 Handling of Dropouts and Missing Data

Refer to Section 4.2.8.3 for the handling of missing data for EORTC QLQ-C30.

4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL

TTD of function (including role, cognitive, emotional and social), multi-term symptoms (including fatigue and nausea/vomiting), and single items (dyspnoea, insomnia, appetite loss, constipation and diarrhoea) will be analysed using the same time-to-event analysis methodology described in Section 4.2.1.4.

Change from baseline in subscales of the EORTC QLQ-C30 is analysed using a mixed model for repeated measures (MMRM) of the change from baseline. Participants are included in the mean change from baseline analysis if they have an evaluable baseline assessment and at least one evaluable post-baseline assessment. The model includes treatment, visit, and treatment-by-visit interaction as explanatory variables and the baseline score and the baseline score by visit interaction as covariates.

When less than 20 participants are present at a timepoint in either arm, this timepoint should be excluded from the analysis. An unstructured covariance matrix is used to model the within-participant error and the Kenward-Roger approximation is 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: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, and autoregressive. Adjusted mean change from baseline estimates per treatment group and corresponding 95% CIs are presented along with an overall estimate of the treatment difference, 95% CI, and p-value.

All analyses will have a corresponding graphical plot showing the adjusted mean change from baseline and 95% CI over time. Summary tables of assessment responses (improvement, deterioration, and no change), absolute scores and change from baseline for each EORTC QLQ-C30 scale/item score (global health status/QoL, 5 functions, and all symptoms [fatigue, pain, nausea/vomiting, dyspnoea, insomnia, appetite loss, constipation and diarrhoea]) and for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.18 Exploratory Endpoint – Breast and Arm Symptoms

4.2.18.1 Definition

The EORTC QLQ-BR45 is a breast-cancer-specific module from the EORTC comprising 45 questions to assess breast cancer symptoms. The module includes 5 functional scales/items (body image, future perspective, sexual functioning, sexual enjoyment, breast satisfaction) and 7 symptom scales/items (systemic therapy side effects, upset by hair loss, arm symptoms, breast symptoms, endocrine therapy symptoms, skin mucosis symptoms and endocrine sexual symptoms).

The current study will only include the breast symptoms and arm symptoms scales (7 items) from the BR45, i.e. EORTC IL116.

4.2.18.2 Derivations and Censoring Rules

The EORTC IL116 is scored as described in Table 10 to give a score from 0 to 100 for each of the symptom scales and items.

Table 10 EORTC IL116 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Symptom scales/items				
Arm Symptoms	ARM	3	3	31 - 33
Breast Symptoms	BR	4	3	34 - 37

For each multi-item scale, the average of the corresponding items is calculated. The raw score is then standardised to a 0 - 100 range as for the EORTC QLQ-C30, and missing values are handled in the same way.

The scoring approach for the EORTC IL116 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30. Similarly, to the symptom scales of the EORTC QLQ-C30, higher scores represent greater symptom severity.

The definition of a clinically meaningful change and time to deterioration for EORTC IL116 is the same as that for the EORTC QLQ-C30 described in Section 4.2.8.2. Censoring rules are also the same as described for EORTC QLQ-C30.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.18.3 Handling of Dropouts and Missing Data

Missing data for EORTC IL116 is handled in the same way as EORTC QLQ-C30 (Section 4.2.8.3).

4.2.18.4 Primary Analysis of Breast and Arm Symptoms

TTD in the breast and arm symptom scales will be analysed using the same time-to-event analysis methodology described in Section 4.2.1.4.

Change from baseline in subscales of the EORTC IL116 is analysed using a MMRM of the change from baseline as described in Section 4.2.17.4.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.19 Exploratory Endpoint - EQ-5D-5L

4.2.19.1 Definition

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. The EQ-5D-5L questionnaire comprises 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 a possible 3,125 health states. For example, state 11111 indicates no problems on any of the five dimensions. Respondents also assess their health today using the EuroQoL-Visual analogue scale (EQ-VAS), which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

Minimal clinically important difference

Improvement, deterioration or no change will be defined based on a MCID. For the EQ-VAS score, the MCID will be pre-specified as requiring a 7-point change from baseline.

4.2.19.2 Derivations

The EQ-5D profile is 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 is the UK perspective. Where a valuation set has not been published, the EQ-5D-5L profile is converted to the EQ-5D index using a crosswalk algorithm (Van Hout & Janssen, 2012). The EQ-VAS is reported separately.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.19.3 Primary Analysis of EQ-5D-5L

Descriptive statistics will be calculated for each scheduled time point in the study, for each trial arm and as a total. These will report the number of participants, the number of EQ-5D questionnaires completed at each timepoint, the number and proportion responding to each dimension of the EQ-5D-5L. Additionally, summary statistics (e.g. n, mean, median, standard deviation, min, max) will be reported for the EQ-5D index score and the EQ-VAS score, and the change from baseline for the EQ-5D index score and the EQ-VAS score.

The evaluable population comprises a subset of the FAS who have a baseline EQ-5D-5L assessment.

Graphical plots of the mean EQ-5D index score and EQ-VAS score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study are produced.

For EQ-VAS a summary table of assessment responses (improvement, deterioration, and no change) for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

4.2.19.4 Supplementary Analyses of EQ-5D-5L

To support submissions to payers, additional analyses may be undertaken, and these are outlined in a separate Payer Analysis Plan (PAP).

4.2.20 Exploratory Endpoint - Biomarkers

The relationship of baseline TROP2 expression, tumour mutational profiling and gene expression profiling and, if applicable, of exploratory biomarkers to clinical outcomes (including but not restricted to) of BOR, DoR, PFS and OS may be presented. Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report.

4.2.21 Exploratory Endpoint - Health Care Resource Use

4.2.21.1 Definition

To investigate the impact of treatment and disease on health care resource of non-study protocol related events, the following variables are captured:

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

4.2.21.2 Derivations

Where admitted overnight, the length of hospital stay is calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation + 1).

If there are multiple hospital stays for the same participant, then the length of hospital stay is summed across all hospitalisation admissions for the participant.

Participants with missing discharge dates are calculated as the difference between the last day with available data and the start date of hospitalisation + 1. The length of ICU/HDU stay is calculated using the same method.

4.2.21.3 Primary Analysis of Health Care Resource Use

Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) are provided for each treatment group on the different types of hospital admissions, the length of stay for participants admitted to hospital for at least one overnight stay and length of stay for participants admitted to intensive care / high dependency units, as well as the primary sign or symptom the participant presents with.

Where a participant has admissions for different signs and symptoms, they are included in each category when summarising type of hospital admissions.

This analysis will be done on the SAF.

4.3 Pharmacodynamic Endpoint(s)

Not Applicable.

4.4 Safety Analyses

The domain safety covers exposure, adverse events, clinical laboratory, vital signs, physical examination, ECG, Echocardiogram, ECOG performance score, Ophthalmologic assessments and pulmonary function test.

Tables are provided for the safety set; listings are provided for the safety set.

4.4.1 Exposure

4.4.1.1 Definitions and Derivations

Treatment exposure for Dato-DXd

Dato-DXd is dosed 6 mg/kg intravenously on Day 1 of each 21-day cycle (Q3W). The dose of Dato-DXd may be reduced once to 4 mg/kg intravenous (IV) Q3W and a further second reduction to 3 mg/kg IV Q3W is allowed per participant on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Dato-DXd (months): =

(min (last date where dose > 0 + 20, date of death, date of DCO) – first dose date +1) / (365.25/12)

Actual exposure of Dato-DXd =

total exposure – total duration of dose interruptions,

where the total duration of dose interruption is defined as any length of time when the participant has not taken any of the planned doses. Dose interruptions include missed and delayed doses.

The calculation of actual exposure makes no adjustment for any dose reductions that may have occurred and will only be calculated for Dato-DXd (not for chemotherapy arms).

Treatment exposure for Capecitabine

Capecitabine is scheduled to be dosed 1000 or 1250 mg/m^2 twice daily (BID) on Days 1 to 14 of a 21-day cycle. The choice of dose will be determined by standard institutional practice and the starting dose will be assumed to be the planned dose. The dose of Capecitabine may be reduced by 25% in participants with moderate renal impairment on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Capecitabine (months): =

(min (last Capecitabine dose date where dose > 0, date of death, date of DCO) – first Capecitabine dose date +1) / (365.25/12)

Treatment exposure for Gemcitabine

Gemcitabine is dosed 1000 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Gemcitabine (months): =

(min (last Gemcitabine dose date where dose > 0 + W, date of death, date of DCO) – first Gemcitabine dose date+1) / (365.25/12),

where W=6 if the last Gemcitabine dose was scheduled on Day 1 and W=13 if the last Gemcitabine dose was scheduled on Day 8.

Treatment exposure for Vinorelbine

Vinorelbine is dosed 25 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Vinorelbine (months): =

(min (last Vinorelbine dose date where dose > 0 + W, date of death, date of DCO) – first Gemcitabine dose date +1) / (365.25/12),

where W=6 if the last Vinorelbine dose was scheduled on Day 1 and W=13 if the last Vinorelbine dose was scheduled on Day 8.

Treatment exposure for Eribulin mesylate

Eribulin mesylate is dosed 1.4 mg/m^2 on Days 1 and 8 of a 21-day cycle. A lower starting dose of 1.1 mg/m^2 is recommended for participants with moderate renal impairment. The starting dose will be assumed to be the planned dose. The calculation of exposure is as follows:

Total (or intended) exposure of Eribulin mesylate (months): =

(min (last Eribulin mesylate dose date where dose > 0 + W, date of death, date of DCO) – first Eribulin mesylate dose date +1) / (365.25/12),

where W=6 if the last Eribulin mesylate dose was scheduled on Day 1 and W=13 if the last Eribulin mesylate dose was scheduled on Day 8.

Participants who permanently discontinue during a dose interruption

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

If a participant permanently discontinues study treatment during a dose interruption, then this is not counted as a dose interruption for summary purposes.

Number of treatment cycles received

Exposure is also measured by the number of cycles received. A cycle corresponds to a period of 21 days. If a cycle is prolonged due to toxicity, this is still counted as one cycle. A cycle is counted if any treatment during that cycle is taken.

Safety Follow-up

Total duration of safety follow-up is calculated as:

Total Safety Follow-up (months) = [min (date of safety follow-up assessment, last dose of IP date + 20, date of study discontinuation, date of death, DCO date) – first dose date +1] / (365.25/12)

Dose intensity

Dose intensity is derived for study treatments Dato-DXd, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine.

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. RDI is 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 or the actual last day of dosing. D is the total dose that would be delivered if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 2 days should be added to reflect the protocol allowed window for dosing as shown below.

Dose intensity – Intended cumulative dose

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

The intended dose for Dato-DXd is 6mg/kg on Day 1 (+/-2 days) of each 21-day cycle. The minimum of the participants last dose, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended.

For the calculations below,

DUR = min (date of last dose date where dose > 0, date of death, date of DCO) – first dose date + 1

The intended dose for Dato-DXd is then calculated as:

6 * [integer ((DUR +2)/ 21) + 1]

Similarly, for Eribulin mesylate (dose=1000 mg/m^2), Vinorelbine (dose=1.4 mg/m^2) and Gemcitabine (dose=25 mg/m^2), intended dose will be calculated as:

dose * [integer ((DUR +2)/ 21) + 1]

Note: if the starting dose is reduced by standard institutional practice per protocol, that starting dose will be used for the participant.

For Capecitabine (dose of 1000 or 1250 mg/m^2 BID for 14 days per 21-day cycle, intended dose is given by:

2 * dose * [integer ((DUR+2) / 21)) + 1]

Dose intensity – Actual cumulative dose

For the calculation of actual cumulative dose for Dato-DXd, Gemcitabine, Eribulin mesylate and Vinorelbine, the proportion of volume left after the infusion will be used to calculate how much of the study drug the participant received, i.e.:

- Volume left (proportion) = $\frac{\text{Volume after infusion}}{\text{Volume before infusion}}$
- Actual cumulative dose = sum over all cycles [(1 – Volume left) x dose] (where dose is taken from the exposure CRF page for each cycle)

For the calculation of actual dose for Capecitabine, drug accountability data (of 1000mg tablets and 1250mg tablets) will be used as follows:

Actual cumulative dose = sum dose (mg) dispensed – sum dose (mg) returned

Percentage Intended Dose

Percentage intended dose (PID) is calculated in the same way as RDI, but D is the intended cumulative dose up to the date of progression or study discontinuation instead of date of last dose where dose >0. The minimum of the participant's date of progression or study discontinuation, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended. For the calculations above,

DUR = min (date of date of progression, study discontinuation, date of death, date of DCO) – first dose date + 1.

Similarly, the actual cumulative dose is the cumulative dose taken up to min (date of progression, date of study discontinuation, date of death, date of DCO).

4.4.1.2 Presentation

The following summaries are produced for Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine:

- Total exposure
- Actual exposure (Dato-DXd only)
- RDI and PID.
- Summary of dose interruptions and dose reductions. Number and percentage of participants with any interruptions and number and percentage of participants with any dose reductions are summarised. This is broken down by number of interruptions/dose reductions, and reason for interruption/dose reduction. In addition, the number and percentage of participants with both an interruption and a dose reduction are summarised.

In addition, the number of cycles received, and duration of safety follow-up are also summarised.

4.4.2 Adverse Events

4.4.2.1 Definitions and Derivations

Adverse events (AEs) and serious adverse events (SAEs) are collected from the time of signature of the informed consent form (ICF), throughout the treatment period and until the 28-Day (+7 days) follow-up period is completed. AE causality is as determined by the reporting investigator. For interstitial lung disease (ILD)/pneumonitis, safety follow up will be continued until resolution of ILD/pneumonitis.

Treatment emergent adverse events (TEAEs) are defined as those adverse events (AEs) with onset or that worsen (by investigator report of an increase in CTCAE grade relative to pre-treatment) on or after the first dose of IP and on or before the date of last IP + 28 days (+7 days) and prior to the start of any subsequent anti-cancer therapy.

For the subset of TEAEs, with onset prior to start of study treatment, and which worsened in severity or seriousness after initiating study treatment until 28 days (+7 days) after last dose of study treatment, such worsening should also occur prior to initiation of any first subsequent anti-cancer therapy to be included in the AE summary tables.

Pre-treatment AEs are those which occur before the first dose of IP and do not worsen during the treatment period.

If no onset time is given, and the date of onset of the AE is the same as the date of first dose of IP, then the AE is assumed to have occurred after the first dose of IP.

The medical dictionary for regulatory activities (MedDRA) [using the latest MedDRA version] is used to code AEs. AEs are graded according to the National Cancer Institute (NCI) common terminology criteria for adverse event (CTCAE) version 5.0.

Missing start and stop dates for AEs are handled using the rules described in Section 3.3.7. AEs that have missing causality (after data querying) are assumed to be related to the treatment where causality is missing.

Dose modification describes an AE where action taken is either dose reduced, or drug interrupted.

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to understanding of the AZ DS-1062 (Dato-DXd) safety profile and require close monitoring and rapid communication by the Investigator to the Sponsor.

The AESIs that are collected during this study are: (ILD)/pneumonitis, infusion-related reactions, oral mucositis/stomatitis, mucosal inflammation other than oral mucositis/stomatitis, and ocular surface toxicity.

Preferred terms used to identify AESI will be listed before data base lock (DBL) and documented in the Trial Master File. 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.

The duration of an AESI is calculated as (Stop date of AE - Start date of AE) + 1. The duration of the AESI is not calculated for AESIs which are ongoing. If a participant has multiple AESIs within a category, then the duration is summed for all AESIs within the category but any days where AESIs overlap are only counted once.

Time to first AESI is calculated as:

(Start date of first AESI - Date of IP first dose) + 1.

4.4.2.2 Presentation

All AEs are summarised descriptively by count (n) and percentage (%) for each treatment group.

Unless otherwise stated, only AEs defined as treatment emergent (see Section 4.4.2.1) are included in the summary tables.

All reported AEs (including pre- and post-treatment AEs) are listed including the date of onset, date of resolution (if AE is resolved) and investigator's assessment of CTCAE grade and relationship to IP. An overall summary of the number and percentage of participants in each category below is presented by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine. Each AE category is separately summarised by SOC and PT:

- All AEs
- All AEs possibly related to IP
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to IP
- AEs with outcome of death
- AEs with outcome of death possibly related to IP
- All SAEs
- All SAEs possibly related to IP
- All SAEs with CTCAE grade 3 or higher
- AEs leading to discontinuation of IP
- AEs leading to discontinuation of IP, possibly related to IP
- All SAEs leading to discontinuation of IP
- All SAEs leading to discontinuation of IP possibly related to IP
- AEs leading to interruption of IP
- AEs leading to dose reduction of IP
- AEs leading to dose modification of IP

Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

Additionally, the most common AEs, which are those AEs that occur in at least 5% (where no rounding is applied i.e. an AE with frequency 4.9% does not appear if the cut-off is 5%) of participants in any treatment group, are summarised by PT, by decreasing frequency based on the total number of AEs across treatment groups. This cut-off may be modified after review of the data.

An additional summary of AEs with onset date less than or equal to 28 days after the date of last IP, but after the onset of subsequent cancer therapy is produced by SOC, PT and treatment group.

Further details of summaries by SOC and PT are given below if a participant experienced more than one TEAE:

- The participant will be counted once for each SOC and once for each PT.
- The participant will be counted once for each SOC and once for each PT at the maximum CTCAE grade.
- The participant will be counted once for each SOC and once for each PT using the most related event
- The participant will be counted once for each SOC and once for each PT for related events at the maximum CTCAE grade.

AEs are assigned CTCAE grades and summaries of the number and percentage of participants are provided by maximum reported CTCAE grade, SOC and PT.

Summary tables of AESIs overall and by maximum CTCAE grade are produced, by AESI category. The preferred terms for AESIs are presented in a listing.

Tables are also produced of AESIs by outcome, and AESIs with outcome of recovered/resolved by time of resolution, and by action taken. The time to onset of first AESI and duration of AESI are summarised. In addition, summary tables are produced of number of participants with AESIs possibly related to IP and leading to discontinuation of IP. AESIs are summarised by AESI category rather than SOC and PT.

All AEs and AESIs are listed, and the time to onset of the AE from date of first dose is presented in the listing. Key participant information is provided in 3 separate listings for all SAEs, AEs with an outcome of death and all AEs leading to treatment discontinuation.

Deaths

A separate summary of deaths is provided with number and percentage of participants, categorised as:

- Total number of deaths (regardless of date of death)
- Related to disease under investigation
- AE with outcome of death only and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- AE with outcome of death only and onset date > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first)

- Death with primary or secondary reason related to disease under investigation and AE with outcome of death and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- Death with primary or secondary reason related to disease under investigation and AE with outcome of death > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy, whichever occurs first
- Deaths <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Deaths > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Participants with unknown reason for death
- Other deaths

A corresponding listing is also produced.

4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest

Summaries of ILD/pneumonitis events will be primarily based on adjudicated drug related ILD/pneumonitis events from the ILD adjudication committee. Supportive summaries based on AESI-defined ILD/pneumonitis cases (i.e. identified based on MedDRA preferred terms) will also be provided.

When summarising adjudicated ILD/pneumonitis events, the categories marked (*) will only be presented for adjudicated drug-related events.

When summarising time to and duration of first treatment-emergent AESI for adjudicated ILD/pneumonitis events, only adjudicated drug-related events will be considered.

A listing of ILD/pneumonitis events is produced.

4.4.3 Clinical Laboratory, Blood Sample

4.4.3.1 Definitions and Derivations

Blood samples for determination of clinical chemistry, haematology and coagulation are collected as described in the schedule of activities (SoA) of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records and missing data are followed.

Change from baseline in haematology and clinical chemistry variables are calculated for each post-dose visit on treatment.

CTCAE grades are defined at each visit according to the CTCAE v5.0 grade criteria using local or project ranges as required, after conversion of lab result to corresponding AstraZeneca (AZ) preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, sodium, magnesium and corrected calcium. For these parameters high and low CTCAE grades are calculated.

Corrected calcium product is 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 are compared to the reference range and classified as low (below range), normal (within range or limits of range) and high (above range).

For parameters with no CTCAE grading that are listed in the CSP any increase/decrease/treatment emergent laboratory change (TELC) is derived, where any increase is an increase to a value above the upper local laboratory reference limit at any time on treatment for participants with a value below the upper local laboratory reference limit at baseline, and any decrease is a decrease to any value below the local laboratory reference range limit at any time on treatment for participants with a value above the lower local laboratory reference limit at baseline. A TELC is defined as any on treatment increase or decrease from baseline.

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

Local reference ranges are used for the primary interpretation of laboratory data.

4.4.3.2 Presentations

Only laboratory data that is on treatment as defined in Section 3.3.4 is included in the summary tables.

Data summaries and listings are provided by AZ preferred units.

Laboratory listings will cover observed values and changes from baseline for each individual participant as well as abnormalities. Flags are applied to values falling outside reference ranges and for the CTCAE grade for parameters for which CTCAE grading applies.

For all continuous clinical chemistry and hematology laboratory assessments, absolute value and change from baseline are summarised using descriptive statistics at each scheduled visit.

Shift tables of laboratory values by worst common toxicity criteria (CTCAE) grade on treatment are produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change are produced. Percentages are based on the number of participants with a baseline value and an on-treatment value.

The laboratory parameters for which CTCAE grade shift outputs are produced are:

- Haematology: Haemoglobin, Leukocytes, Lymphocytes (absolute count), Neutrophils (absolute count), Platelets
- Clinical Chemistry: Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Albumin, Alkaline Phosphatase (ALP), Total bilirubin, Magnesium (hypo- and hyper-), Sodium (hypo- and hyper-), Potassium (hypo- and hyper-), Corrected Calcium (hypo- and hyper-), Creatinine

For parameters with no CTCAE grading, the number and percentage of participants with any on treatment increase from baseline, any on treatment decrease from baseline and a TELC is summarised. Percentages are based on the number of participants with a baseline value below/above the local laboratory upper/lower reference limit and an on-treatment value for the any increase/decrease summaries respectively. Percentages for a TELC are based on the number of participants with a baseline value and an on-treatment value.

For parameters with no CTCAE grading, shift tables from baseline to worst value on-treatment are provided.

Hy's law (HL)

A summary table is produced showing the number (%) of participants who have:

- Elevated ALT, AST, and Total bilirubin during the study
- ALT >3x, >5x, >10x and >20x ULN (Upper limit of normal) during the study.
- AST >3x, >5x, >10x and >20x ULN during the study.
- Total bilirubin >1.5x and >2x ULN during the study.
- ALT or AST >3x, >5x, >10x and >20x ULN during the study.
- ALT or AST >3x ULN and total bilirubin >1.5x ULN during the study (potential Hy's law): the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.

- ALT or AST >3x ULN and total bilirubin >2x ULN during the study (potential Hy's law): the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.
- ALT or AST >3x ULN, and no initial ALP ≥ 2x ULN and total bilirubin >2x ULN during the study (potential Hy's law): the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.
- ALP >1.5x and >3x ULN during the study.

Narratives are provided in the CSR for participants with potential Hy's law and in which the elevation in transaminases precede or coincide with (that is, on the same day as) the elevation in total bilirubin.

Liver biochemistry test results over time for participants with elevated ALT (i.e. ≥3x ULN) or AST (i.e. ≥3x ULN), and elevated total bilirubin (i.e. ≥2x ULN) and in which the elevation in transaminases precede or coincide with (that is, on the same day as) the elevation in total bilirubin or participants with ALT/AST ≥ 5x ULN are plotted and listed.

Plots of maximum post-baseline ALT and AST vs. maximum post-baseline total bilirubin, expressed as multiples of ULN, are also produced with reference lines at 3×ULN for ALT and AST, and 2×ULN for Total bilirubin. In each plot, total bilirubin is in the vertical axis.

4.4.4 Clinical Laboratory, Urinalysis

4.4.4.1 Definitions and Derivations

Urine samples for determination of urinalysis are collected as described in the SoA of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.4.2 Presentations

A shift table is produced showing the number and percentage of participants in each category at baseline and the maximum on treatment value for each parameter.

The denominator used only includes participants with a baseline value and at least one on treatment value.

On treatment is defined in Section 3.3.4.

Supportive laboratory listings will cover observed values for each individual participant as well as abnormalities.

4.4.5 Other Laboratory Evaluations

4.4.5.1 Definitions and Derivations

Pregnancy tests (serum at screening and urine at other timepoints) are performed for women of childbearing potential.

In addition, hepatitis B surface antigen, hepatitis C and human immunodeficiency virus (HIV) antibodies is assessed at screening.

4.4.5.2 Presentations

This data is listed only, no summary tables are produced.

4.4.6 Vital Signs

4.4.6.1 Definitions and Derivations

Vital signs are assessed at timelines as specified in the SoA of the CSP. The following vital signs are measured: systolic and diastolic blood pressure, pulse rate, body temperature and respiratory rate. Body weight is also collected. Height is collected at screening only.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.6.2 Presentations

Summaries for vital signs data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values and change from baseline for diastolic and systolic blood pressure, pulse, respiratory rate, temperature and weight are summarised over time at each scheduled visit for each treatment group.

Vital signs data is also listed.

4.4.7 Electrocardiogram

4.4.7.1 Definitions and Derivations

Resting 12-lead electrocardiograms (ECGs) are recorded at timepoints specified in the SoA of the CSP.

The following ECG variables are collected: ECG heart rate, PR duration, QRS duration, QT interval, QTcF interval, RR duration and overall ECG evaluation.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

The overall evaluation of an ECG is either "normal", "borderline" or "abnormal" with abnormalities categorised as either "clinically significant" or "not clinically significant". Any clinically significant ECG abnormalities require triplicate ECG results. Where triplicate ECG results are taken, a single mean value for numeric parameters is used, and the worst case for the overall evaluation is used.

The QT interval corrected for heart rate using Fridericia's correction (QTcF) is calculated as follows (where QT and RR are in seconds):

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

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:

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

4.4.7.2 Presentations

Summaries for ECG data include only on treatment data. On treatment is defined in Section 3.3.4.

A summary table of absolute values and change from baseline for ECG heart rate, PR duration, QRS duration, QT duration, QTcF duration and RR duration is presented over time.

The following summaries for QTcF are also included:

Absolute QTcF interval prolongation at any time on treatment:

- QTcF interval > 450 milliseconds
- QTcF interval > 480 milliseconds
- QTcF interval > 500 milliseconds

Change from baseline in QTcF interval at any time on treatment:

- QTcF interval increases from baseline > 30 milliseconds
- QTcF interval increase from baseline > 60 milliseconds
- QTcF interval >450 milliseconds and change from baseline > 30 milliseconds
- QTcF interval >500 milliseconds and change from baseline > 60 milliseconds

A listing is provided of ECG data.

4.4.8 Echocardiogram/Multigated Acquisition Scan

4.4.8.1 Definitions and Derivations

An echocardiogram (ECHO) or multigated acquisition (MUGA) scan to assess left ventricular ejection fraction (LVEF) is performed at the visits as shown in the SoA of the CSP.

The modality of the cardiac function assessments must be consistent for a given participant, i.e. if an ECHO scan is used for the screening assessment, then ECHO should also be used for subsequent scans. The participants should also be examined using the same machine and operator whenever possible, and quantitative measurements should be taken.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.8.2 Presentations

Summaries for LVEF data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values at baseline and end of treatment visit and change from baseline to end of treatment visit for LVEF results are summarised.

4.4.9 Eastern Cooperative Oncology Group Performance Status

4.4.9.1 Definitions and Derivations

An assessment of ECOG performance status score is performed at the visits as shown in the SoA of the CSP.

The ECOG performance status scores range from 0 to 5, with lower scores indicating greater participant activity:

0. Fully active; able to carry out all usual activities without restrictions

1. Restricted in strenuous activity, but ambulatory and able to carry out light work or work of a sedentary nature (e.g. light housework or office work)

2. Ambulatory and capable of self-care, but unable to carry out any work activities; up and about more than 50% of waking hours

3. Capable of only limited self-care; confined to bed or chair more than 50% of waking hours

4. Completely disabled; unable to carry out any self-care and totally confined to bed or chair

5. Dead

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.9.2 Presentations

Summaries for ECOG data include only on treatment data. On treatment is defined in Section 3.3.4.The number and percentage of participants in each category is summarised at each visit.

4.4.10 Physical Examination

4.4.10.1 Definitions and Derivations

Physical examination, as well as assessment of height and weight, will be performed according to the Schedule of Assessments (SoA) in the CSP.

A full physical examination will be performed at screening which includes assessment of general appearance, respiratory, cardiovascular, abdomen, skin, head and neck (including ears, eyes, nose and throat), oral (mouth), lymph nodes, thyroid, musculoskeletal (including spine and extremities), urogenital, dermatological, gastrointestinal, endocrine, hematologic/lymphatic, and neurological systems. At subsequent visits, targeted physical examinations are to be utilised by the Investigator on the basis of clinical observations and symptomatology. A targeted physical examination includes at a minimum, assessments of the skin, lungs, oral, cardiovascular system, and abdomen (liver and spleen).

4.4.10.2 Presentations

Individual physical examination data will not be summarised.

4.4.11 Ophthalmologic Assessments

4.4.11.1 Definition and Derivations

Ophthalmologic assessments by a licensed eye care provider will be performed as specified in the SoA in the CSP.

The following assessments will be performed for both eyes: daily use of artificial tears, avoidance of contact lenses, eye-related symptoms, best corrected visual acuity (BCVA), corneal sensation, eyelid position, eyelid margins, conjunctiva, slit-lamp examination, anterior chamber cells, fluorescein staining, tear film breakup time (TFBT), limbal stem cell deficiency (LSCD), Oxford grade of punctate epithelial erosions, proparacaine test, intraocular pressure, tear film meniscus, fundoscopy, clinically significant corneal disease

with use of both CTCAE and Corneal Toxicity Severity Grading scales, and other diagnosis and treatment(s) prescribed, if any.

The preferred method for measuring BCVA is the Snellen chart (metric in meter, or imperial in feet). The Snellen chart values will be converted to LogMAR values as illustrated in Table 11. LogMAR values are calculated by taking the $\log_{10}$ of the reciprocal of the Snellen fraction. For example, if the Snellen fraction is 20/50, the LogMAR value is $\log_{10}$ (50/20) = 0.4.

Table 11 Best Corrected Visual Acuity conversion table

Snellen Chart		Snellen Fraction (Decimal)	LogMAR
Feet	Meter		
20/200	6/60	0.10	1.0
20/160	6/48	0.125	0.9
20/125	6/38	0.16	0.8
20/100	6/30	0.20	0.7
20/80	6/24	0.25	0.6
20/63	6/19	0.32	0.5
20/50	6/15	0.40	0.4
20/40	6/12	0.50	0.3
20/32	6/9.5	0.63	0.2
20/25	6/7.5	0.80	0.1
20/20	6/6	1.00	0.0
20/16	6/4.8	1.25	-0.1
20/12.5	6/3.8	1.60	-0.2
20/10	6/3	2.00	-0.3

4.4.11.2 Presentations

Ophthalmologic assessments will be summarised at DCO1 (see Section 3.1.1).

For daily use of artificial tears and BCVA (LogMAR), summary statistics (e.g. n, mean, median, standard deviation, min, max) for the observed value and the change from baseline value will be reported by visit and by treatment group.

For use of contact lenses, fluorescein staining of cornea, tear film breakup time, LSCD with and without fluorescein stain, punctate epithelial erosions in Oxford grade, eye pain and

clinically significant corneal disease by the revised CTCAE grade, summary statistics (e.g. n, mean, median, standard deviation, min, max) for the observed value will be reported by visit and by treatment group.

For intraocular pressure, summary statistics (e.g. n, mean, median, standard deviation, min, max) for the change from baseline value will be reported by visit and by treatment group.

For corneal sensation, eye lid position, eye lid margins abnormalities, tear film meniscus, abnormality results in slit lamp examination with attention to cornea and dilated fundoscopic exam results, shift from baseline will be reported by visit and by treatment group. Percentages will be based on the number of participants with a baseline result and at least one result for the corresponding visit.

For conjunctiva abnormality and anterior chamber (cell) abnormality, baseline to worst result post-baseline shift tables will be reported. Percentages will be based on the number of participants with a baseline result and at least one post baseline result.

The reported diagnosis will be summarised by preferred term (PT). The following PTs will be displayed as separate categories:

1. Keratitis (including Ulcerative keratitis, Corneal perforation)
2. Limbal stem cell deficiency
3. Visual acuity reduced

The prescribed ocular treatments will be summarised by categories of medications (subdivided into corticosteroids, antibiotics, other), adjuvant procedures and surgeries.

Listings will be produced for the following assessments: BCVA, lid margins for abnormalities, slit lamp examination findings, cornea abnormality, LSCD, abnormality findings from dilated fundoscopic exam and clinically significant corneal disease. In addition, listings will be produced for medication used for eye pain, eye exam diagnosis results and eye treatments prescribed.

All analyses will be performed on the OAS.

4.4.12 Pulmonary Function Test

4.4.12.1 Definitions and Derivations

Pulmonary function testing will be performed at screening, including FVC (L), FVC % predicted, FEV1 (L), FEV1 % predicted and FEV1/FVC % as a minimum.

4.4.12.2 Presentations

Absolute values for FVC (L), FVC % predicted, FEV1 (L), FEV1 % predicted and FEV1/FVC % are summarised at screening for each treatment group.

4.4.13 Impact of COVID-19

Depending on the extent of any coronavirus disease 2019 (COVID-19) impact, summaries of data relating to participants diagnosed with COVID-19, and impact of COVID-19 on study conduct (in particular delayed/missed visit, delayed or discontinued IP, discontinuation of study, and COVID-19 related protocol deviations) may be generated, by treatment group, including:

- Disposition (discontinued IP due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruption (visit impact, drug impacted)
- Listing for participants affected by the COVID-19 pandemic
- Listing for participants with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.

Additional analyses may be performed to explore the impact of COVID-19 on key efficacy and safety endpoints, for example repeating the AE summaries separately for participants where events are attributed to COVID-19.

5 INTERIM ANALYSIS

Interim Analysis for Superiority in OS

Two interim analyses for OS are planned.

The first interim analysis will occur at the primary PFS analysis. This corresponds to approximately 178 OS events, 25% maturity and 40% of the information expected at the primary analysis (444 OS events at final, primary).

The second interim analysis will occur when approximately 355 OS events have been observed in the FAS. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis (444 OS events).

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including an interim analysis for superiority. This approach will be used to maintain an overall 2-sided type I error across the three planned analyses of OS.

If the PFS dual primary analysis crosses the efficacy threshold, the 1.0% type I error allocated to the PFS endpoint will be reallocated to the OS endpoint for a total 2-sided type I error of 5.0% (Burman, Sonesson, & Guilbaud, 2009). If the PFS dual primary analysis does not cross the efficacy threshold the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Table 12 Summary of planned timings of the interim and final OS analyses

	Interim Analysis 1		Interim Analysis 2		Primary Analysis	
Projected Timing	21 Months [b]		34 Months		44 Months	
Number of Deaths [a]	178		355		444	
Information Fraction	40%		80%		100%	
Maturity	25%		51%		63%	
Recommendation	**Continue**	**Reject Null Hypothesis**	**Continue**	**Reject Null Hypothesis**	**Do Not Reject Null Hypothesis**	**Reject Null Hypothesis**
At 4.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0005	< 0.0005	≥ 0.0184	< 0.0184	≥ 0.0345	< 0.0345
Estimated hazard ratio	≥ 0.591	< 0.591	≥ 0.777	< 0.777	≥ 0.817	< 0.817
At 5.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0008	< 0.0008	≥ 0.0241	< 0.0241	≥ 0.0427	< 0.0427
Estimated hazard ratio	≥ 0.604	< 0.604	≥ 0.786	< 0.786	≥ 0.824	< 0.824

[a] Estimates based on exponential survival where the median OS is 19.0 months for ICC and 25.3 months for Dato-DXd. The total proportion of participants randomized at time t [t ≤ 19 months] following the start of the study is assumed to be $(t/19)^{1.5}$.
[b] Timing of first IA based on PFS. Number of deaths is an estimate.
[c] Alpha allocated to OS endpoint (4.0% or 5.0%) dependent on statistical significance of PFS.

For a total 2-sided type 1 error of 5.0% this results in a level of significance alpha of approximately 0.0008 for the first interim analysis (IA) and 0.0241 for the second IA. For a total 2-sided type 1 error of 4.0% this results in a level of significance alpha of approximately 0.0005 for the first IA and 0.0184 for the second IA. This is described in more detail in Table 12.

Since the significance level will be dependent on the number of events actually observed, this will be calculated at the time of the analysis.

The interim analyses will be performed by an independent data monitoring committee (IDMC) separate from the study team reporting the final study results so that the study team are kept blinded.

An external IDMC comprised of therapeutic area experts and biostatisticians who are not employed by AstraZeneca and are free from conflict of interest will review the unblinded interim analysis output.

The IDMC will inform the sponsor if superiority has been achieved for OS in the FAS.

Analyses planned to be performed at the interim analysis will include PFS (first IA) and OS, and other key outputs will also be produced. Details of the outputs to be produced for the interim analysis will be specified in the IDMC charter.

If the first interim results do not meet the criterion for declaring superiority for OS in the FAS, then follow-up will continue until the criteria are met for the second OS IA. If the second interim results do not meet the criterion for declaring superiority in the FAS, then follow-up will continue until the criteria is met for the OS primary analysis (approximately 444 OS events in the FAS).

For a total 2-sided type 1 error of 5.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0427 level of significance. For a total 2-sided type 1 error of 4.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0345 level of significance.

The study may continue monitoring participants for OS up to the scheduled final analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival.

IDMC Safety Reviews

This study will use an external IDMC to assess ongoing safety analyses as well as the interim efficacy analysis. The IDMC will meet to review unblinded safety data after the study has started, with an initial early IDMC approximately six months after study start; and then at approximately six-month intervals thereafter. Following each meeting, the IDMC will report to the sponsor and may recommend changes in the conduct of the study.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca and are free from conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will include the recommendation and any potential protocol amendments and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

The safety of all AstraZeneca clinical studies is closely monitored on an ongoing basis by AstraZeneca representatives in consultation with the Participant Safety Department. Issues

identified will be addressed; this could involve, for instance, amendments to the Clinical Study Protocol and letters to investigators.

Full details of the IDMC procedures, processes and the responsibilities of the IDMC will be given in the IDMC charter.

6 REFERENCES

Burman, C., Sonesson, C., & Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med. 28:739-761.

Fayers, P., Aaronson, N., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organization for Research and Treatment of Cancer.

Fleischer, F., Gaschler-Markefski, B., & Bluhmki, E. (2001). How is retrospective independent review.

Gail, M., & Simon, R. (1985). Testing for qualitative interactions between treatment effects and subject subsets. Biometrics. 41:361-72.

Lan, K., & DeMets, D. (1983). Discrete sequential boundaries for clinical trials. Biometrika. 70:659-663.

Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. Stat Med, 4(2). 213-26.

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

Van Hout, B., & Janssen, M. (2012). Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 15(5):708-15.

7 APPENDIX

Not applicable.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d9268c00001-sap-ed-1		
Document Title:	Statistical Analysis Plan Edition 1	
Document ID:	Doc ID-004671764	
Version Label:	1.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
22-Dec-2021 18:52 UTC	PPD	Qualified Person Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.

STATISTICAL ANALYSIS PLAN	
Study Code	D9268C00001
Edition Number	2.0
Date	21-Jun-2023

A Phase 3, Open-label, Randomised Study of Dato-DXd Versus Investigator's Choice of Chemotherapy in Participants With Inoperable or Metastatic Hormone Receptor-Positive, HER2-Negative Breast Cancer Who Have Been Treated With One or Two Prior Lines of Systemic Chemotherapy (TROPION-Breast01)

TABLE OF CONTENTS

TITLE PAGE 1
TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 7
AMENDMENT HISTORY 11
1 INTRODUCTION 20
2 CHANGES TO PROTOCOL PLANNED ANALYSES 20
3 DATA ANALYSIS CONSIDERATIONS 20
3.1 Timing of Analyses 20
3.1.1 Ophthalmologic Data Review (DCO1) 20
3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2) 21
3.1.3 Second OS Interim Analysis (DCO3) 21
3.1.4 Primary Analysis of OS (DCO4) 21
3.2 Analysis Populations 21
3.2.1 Enrolled set 21
3.2.2 Full analysis set 21
3.2.3 Safety analysis set 21
3.2.4 Ophthalmologic analysis set 22
3.2.5 Pharmacokinetic analysis set 22
3.2.6 ADA evaluable set 22
3.2.7 Summary of outcome variables and analysis sets 22
3.3 General Considerations 23
3.3.1 Sample Size Determination 24
3.3.2 Investigational Product 26
3.3.3 Baseline 26
3.3.4 On Treatment 26
3.3.5 Visit Window 27
3.3.6 Handling of Unscheduled Visits 27
3.3.7 Missing Data 28
3.3.8 Derivations of RECIST Visit Responses 29
3.3.8.1 Target lesions (TLs) 30
3.3.8.2 Non-target lesions (NTLs) and new lesions 34
3.3.8.3 Overall visit response – site investigator data 35
3.3.8.4 Independent review 36
3.3.9 Multiplicity/Multiple Comparisons 37
3.3.10 Handling of Protocol Deviations in Study Analysis 39
4 STATISTICAL ANALYSIS 40
4.1 Study Population 40
4.1.1 Participant Disposition and Completion Status 40
4.1.1.1 Presentations 40
4.1.2 Analysis Sets 41
4.1.3 Protocol Deviations 41

4.1.4 Demographics 41
4.1.4.1 Definitions and Derivations 41
4.1.4.2 Presentation 42
4.1.5 Baseline Characteristics 42
4.1.5.1 Definitions and Derivations 42
4.1.5.2 Presentation 42
4.1.6 Disease Characteristics 42
4.1.6.1 Definitions and Derivations 42
4.1.6.2 Presentation 44
4.1.7 Medical History and Concomitant Disease 44
4.1.7.1 Definitions and Derivations 44
4.1.7.2 Presentation 44
4.1.8 Prior and Concomitant Medications 44
4.1.8.1 Definitions and Derivations 44
4.1.8.2 Presentation 45
4.1.9 Study Drug Compliance 46
4.2 Endpoint Analyses 46
4.2.1 Primary Endpoint - Progression Free Survival by BICR 54
4.2.1.1 Definition 54
4.2.1.2 Derivations and Censoring Rules 54
4.2.1.3 Handling of Dropouts and Missing Data 56
4.2.1.4 Primary Analysis of Progression Free Survival 56
4.2.1.5 Sensitivity Analyses of Progression Free Survival 58
4.2.1.6 Subgroup Analyses 60
4.2.2 Primary Endpoint - Overall Survival 62
4.2.2.1 Definition 62
4.2.2.2 Derivations and Censoring Rules 62
4.2.2.3 Handling of Dropouts and Missing Data 63
4.2.2.4 Primary Analysis of Overall Survival 63
4.2.2.5 Sensitivity Analyses of Overall Survival 64
4.2.2.6 Subgroup Analyses 64
4.2.3 Secondary Endpoint - Objective Response Rate 64
4.2.3.1 Definition 64
4.2.3.2 Derivations 65
4.2.3.3 Primary Analysis of Objective Response Rate 65
4.2.4 Secondary Endpoint - Best Objective Response 66
4.2.4.1 Definition 66
4.2.4.2 Derivations 66
4.2.4.3 Primary Analysis of Best Objective Response 67
4.2.5 Secondary Endpoint - Duration of Response 67
4.2.5.1 Definition 67
4.2.5.2 Derivations and Censoring Rules 67
4.2.5.3 Primary Analysis of Duration of Response 68
4.2.5.4 Supplementary Analyses of Duration of Response 68
4.2.6 Secondary Endpoint – PFS by Investigator Assessment 68
4.2.6.1 Definition 68

4.2.6.2 Derivations and Censoring Rules 68
4.2.6.3 Handling of Dropouts and Missing Data 68
4.2.6.4 Primary Analysis of Progression Free Survival 68
4.2.7 Secondary Endpoint –Disease Control Rate 69
4.2.7.1 Definition 69
4.2.7.2 Derivations 69
4.2.7.3 Primary Analysis of Disease Control Rate 70
4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30 70
4.2.8.1 Definition 70
4.2.8.2 Derivations and Censoring Rules 71
4.2.8.3 Handling of Dropouts and Missing Data 74
4.2.8.4 Primary Analysis of TTD Endpoints as Measured by EORTC QLQ-C30 74
4.2.8.5 Sensitivity Analysis of TTD Endpoints as Measured by EORTC QLQ-C30 75
4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death 75
4.2.9.1 Definition 75
4.2.9.2 Derivations and Censoring Rules 75
4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death 75
4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death 76
4.2.10.1 Definition 76
4.2.10.2 Derivations and Censoring Rules 76
4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death 76
4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death 77
4.2.11.1 Definition 77
4.2.11.2 Derivations and Censoring Rules 77
4.2.11.3 Handling of Dropouts and Missing Data 78
4.2.11.4 Primary Analysis of PFS2 78
4.2.12 Secondary Endpoint - Pharmacokinetics 78
4.2.12.1 Derivations 78
4.2.12.2 Primary Analyses of Pharmacokinetics 78
4.2.13 Secondary Endpoint - Immunogenicity 79
4.2.13.1 Derivations 79
4.2.13.2 Primary Analysis of Immunogenicity 80
4.2.14 Exploratory Endpoint – Patient-Reported Symptomatic AEs and Treatment Tolerability 81
4.2.14.1 Definition 81
4.2.14.2 Derivations and Censoring rules 82
4.2.14.3 Handling of Dropouts and Missing data 83
4.2.14.4 Primary Analyses of PRO-CTCAE 84
4.2.14.5 Primary Analyses of EORTC IL117 84
4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability 85
4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS) 85
4.2.15.1 Definition 85

4.2.15.2 Derivations and Censoring rules 85
4.2.15.3 Analysis of Patient Global Impression of Severity 85
4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC) 85
4.2.16.1 Definition 85
4.2.16.2 Derivations and Censoring rules 86
4.2.16.3 Analysis of Patient Global Impression of Change 86
4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL 86
4.2.17.1 Definition 86
4.2.17.2 Derivation and Censoring Rules 86
4.2.17.3 Handling of Dropouts and Missing Data 86
4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL 86
4.2.18 Exploratory Endpoint – Breast and Arm Symptoms 87
4.2.18.1 Definition 87
4.2.18.2 Derivations and Censoring Rules 88
4.2.18.3 Handling of Dropouts and Missing Data 88
4.2.18.4 Primary Analysis of Breast and Arm Symptoms 88
4.2.19 Exploratory Endpoint - EQ-5D-5L 89
4.2.19.1 Definition 89
4.2.19.2 Derivations 89
4.2.19.3 Primary Analysis of EQ-5D-5L 89
4.2.19.4 Supplementary Analyses of EQ-5D-5L 90
4.2.20 Exploratory Endpoint - Biomarkers 90
4.2.21 Exploratory Endpoint - Health Care Resource Use 90
4.2.21.1 Definition 90
4.2.21.2 Derivations 91
4.2.21.3 Primary Analysis of Health Care Resource Use 91
4.3 Pharmacodynamic Endpoint(s) 91
4.4 Safety Analyses 91
4.4.1 Exposure 91
4.4.1.1 Definitions and Derivations 91
4.4.1.2 Presentation 96
4.4.2 Adverse Events 96
4.4.2.1 Definitions and Derivations 96
4.4.2.2 Presentation 97
4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest 101
4.4.3 Clinical Laboratory, Blood Sample 101
4.4.3.1 Definitions and Derivations 101
4.4.3.2 Presentations 102
4.4.4 Clinical Laboratory, Urinalysis 104
4.4.4.1 Definitions and Derivations 104
4.4.4.2 Presentations 104
4.4.5 Other Laboratory Evaluations 105
4.4.5.1 Definitions and Derivations 105
4.4.5.2 Presentations 105

4.4.6 Vital Signs 105
4.4.6.1 Definitions and Derivations 105
4.4.6.2 Presentations 105
4.4.7 Electrocardiogram 105
4.4.7.1 Definitions and Derivations 105
4.4.7.2 Presentations 106
4.4.8 Echocardiogram/Multigated Acquisition Scan 107
4.4.8.1 Definitions and Derivations 107
4.4.8.2 Presentations 107
4.4.9 Eastern Cooperative Oncology Group Performance Status 107
4.4.9.1 Definitions and Derivations 107
4.4.9.2 Presentations 108
4.4.10 Physical Examination 108
4.4.10.1 Definitions and Derivations 108
4.4.10.2 Presentations 108
4.4.11 Ophthalmologic Assessments 108
4.4.11.1 Definition and Derivations 108
4.4.11.2 Presentations 109
4.4.12 Impact of COVID-19 110
5 INTERIM ANALYSIS 111
6 REFERENCES 114
7 APPENDIX 114

LIST OF TABLES

Table 1 Summary of outcome variables and analysis sets 22
Table 2 TL visit responses (RECIST 1.1) 30
Table 3 NTL visit responses 34
Table 4 Overall visit response 36
Table 5 Endpoint analyses 47
Table 6 Outcome and date of event for PFS analysis 54
Table 7 Definition of two missed RECIST visits 55
Table 8 EORTC QLQ-C30 scales and scores 70
Table 9 Definition of two missed PRO visits 74
Table 10 EORTC IL116 scales and scores 88
Table 11 Best Corrected Visual Acuity conversion table 109
Table 12 Summary of planned timings of the interim and final OS analyses 112

LIST OF FIGURES

Figure 1 Multiple Testing Procedure 38

LIST OF ABBREVIATIONS

Abbreviation or Specialised Term	Definition
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
AJCC	American joint committee on cancer
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AZ	AstraZeneca
BCVA	Best corrected visual acuity
BICR	Blinded independent central review
BID	Twice daily
BLQ	Below limit of quantification
BMI	Body mass index
BoR	Best overall response
BSA	Body surface area
C1D1	Cycle 1, day 1
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
COVID-19	Coronavirus 2019 disease
CRF	Case Report Form
CR	Complete response
CRO	Clinical research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DBL	Data base lock
DCO	Data cut-off
DCR	Disease control rate
DoR	Duration of Response

Abbreviation or Specialised Term	Definition
DOSDISC	Treatment discontinuation form
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EORTC	European Organisation for Research and Treatment of Cancer
EQ-5D-5L	EuroQoL 5-dimension, 5-level health state utility index
EQ-VAS	EuroQoL-Visual analogue scale
FAS	Full Analysis Set
FEV1	Forced expiratory volume – 1 second
FVC	Forced vital capacity
GHS	Global health status
HDU	High dependency unit
HER2	Human epidermal growth factor receptor 2
HCRU	Health care resource use
HIV	Human immunodeficiency virus
HR	Hazard ratio
HR-positive	Hormone receptor-positive
HRQoL	Health related quality of life
IA	Interim analysis
ICC	Investigator's Choice Chemotherapy
ICF	Informed consent form
ICR	Independent central review
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IHC	Immunohistochemistry
ILD	Interstitial lung disease
IP	Investigational product
IPD	Important protocol deviation
IRC	Independent Review Charter
IRT	Interactive Response Technology
ITT	Intention to treat
IV	Intravenous
KM	Kaplan-Meier
LD	Longest diameter

Abbreviation or Specialised Term	Definition
LLOQ	Lower Limit of Quantification
Ln	Natural logarithm or logarithm to the base e
LSCD	Limbal stem cell deficiency
LSmean	Least squares mean
LVEF	Left ventricular ejection fraction
MCID	Minimal clinically important difference
MedDRA	Medical Dictionary for Regulatory Activities
MM	Millimetre
MMRM	Mixed model for repeated measures
MRI	Magnetic Resonance Imaging
MSSO	Maintenance and Support Services Organization
MTP	Multiple testing procedure
MUGA	Multigated acquisition
NA	Not applicable
nAb	Neutralizing antibody
NC	Not calculable
NCI	National Cancer Institute
NE	Not evaluable
NL	New lesion
NQ	Not quantifiable
NR	Not Reportable
NS	No Sample
NTL	Non-target lesion
OAS	Ophthalmologic Analysis Set
ORR	Objective response rate
OS	Overall survival
PAS	Pharmacokinetic analysis set
PAP	Payer Analysis Plan
PD	Progression of disease
PFS	Progression-free survival
PFS2	Second progression-free survival
PGIC	Patients' global impression of change
PGIS	Patients' global impression of severity
PGI-TT	Patients' global impression of treatment tolerability
PID	Percentage intended dose

Abbreviation or Specialised Term	Definition
PK	Pharmacokinetics
PR	Partial response
PRO	Patient-reported outcome
PS	Performance Status
PT	Preferred term
Q3W	Every 3 weeks
Q6W	Every 6 weeks
Q9W	Every 9 weeks
QLQ-C30	EORTC 30-item core quality of life questionnaire
QTcF	QT interval corrected by Fridericia's formula
RDI	Relative dose intensity
RECIST 1.1	Response Evaluation Criteria in Solid Tumours, Version 1.1
RS	Raw score
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Stable disease
SoA	Schedule of activities
SOC	System organ class
TEAE	Treatment emergent adverse event
TELC	Treatment emergent laboratory change
TFBT	Tear film breakup time
TFL	Table, figures and listings
TFST	Time to first subsequent therapy or death
TL	Target lesion
TNM	Tumour node metastasis
TROP2	Trophoblast cell surface antigen 2
TSST	Time to second subsequent therapy or death
TTD	Time to deterioration
ULN	Upper limit of normal
US	United States
VAS	Visual analogue scale
VCV	Veeva Clinical Vault
WHO	World Health Organization Drug dictionary

AMENDMENT HISTORY

CATEGORY Change refers to:	**Date**	**Description of change**	**In line with CSP?**	**Rationale**
N/A	22-Dec-2021	Initial approved SAP	N/A	N/A
Two missed visits	21-Jun-2023	Definition of 2 missed visits has been updated for RECIST and PRO assessments.	Yes	Correction to previous error
List of abbreviations	21-Jun-2023	Add BSA and IHC Remove DFI and TFI	N/A	Additional text added to SAP. Text removed from SAP
General	21-Jun-2023	Add additional wording for clarity	Yes	Add clarity
General	21-Jun-2023	Change ORR, DCR and BoR to FAS	Yes	Align with CSP
General	21-Jun-2023	Remove pulmonary function summary	N/A	Change to analysis
Section 2	21-Jun-2023	Add additional wording to clarify difference in PRO analysis	Yes	Explain inconsistency in the CSP
Section 2	21-Jun-2023	Removed text about EQ-5D-5L population	Yes	Changed population to align with CSP
Section 2	21-Jun-2023	Added Hy's law text	No	Hy's law definition is different between CSP and SAP
Section 2	21-Jun-2023	Removed AESI and geographic region	Yes	CSP amendments
ADA evaluable set	21-Jun-2023	Changed population definition	N/A	Changed to analysis population requested
Section 3.2.1	21-Jun-2023	Specify irrespective or randomisation	N/A	Add clarity
Section 3.2.3	21-Jun-2023	No crossover between study treatment arms will be allowed	N/A	Add clarity
Section 3.2.5	21-Jun-2023	Add details on protocol violation	N/A	Add clarity

Section 3.2.6	21-Jun-2023	Change definition to be result at any time	N/A	Change to standards
Section 3.3	21-Jun-2023	Added text regarding precision of summary statistics	N/A	Add clarity to ensure consistency
Section 3.3.2	21-Jun-2023	Update IP definition	N/A	Additional clarity
Section 3.3.3	21-Jun-2023	Add ADA baseline definition	N/A	Additional summary added requiring definition of baseline
Section 3.3.3	21-Jun-2023	Updated definition of PRO endpoints	N/A	To allow participants without treatment to be included in baseline definition
Section 3.3.5	21-Jun-2023	Add ADA baseline definition	N/A	Additional summary added requiring definition of baseline
Section 3.3.5	21-Jun-2023	Updated definition of PRO endpoints	N/A	To allow participants without treatment to be included in baseline definition
Section 3.3.7	21-Jun-2023	Updated missing day imputation rules for stop date	N/A	To ensure imputed stop date will be on or after start date
Table 2 and Table 4	21-Jun-2023	Added NA response for TL	N/A	To account for participants who have no baseline per BICR
Section 4.1.1.1	21-Jun-2023	Remove summary of recruitment by enrolled set and safety analysis set	N/A	Summary removed
Section 4.1.5	21-Jun-2023	Add BSA	N/A	Additional summary
Section 4.1.5.2	21-Jun-2023	Change summary of stratification factors from distinct stratum to category	N/A	Improve interpretation

Section 4.1.6	21-Jun-2023	Add additional disease characteristics	N/A	Additional summaries of interest
Section 4.1.6	21-Jun-2023	Remove DFI and TNM	N/A	Change to summary
Section 4.1.8.1	21-Jun-2023	Remove TFI	N/A	Change to summary
Section 4.1.8.2	21-Jun-2023	Add additional summaries	N/A	Additional summaries of interest
Section 4.1.8.2	21-Jun-2023	Change post-study treatment cancer therapies to summarise by ATC group and generic term	N/A	Change to summary
Section 4.2	21-Jun-2023	Specify BICR and investigator for relevant endpoints	N/A	To provide clarification
Section 4.2.1.2	21-Jun-2023	Reword table 6	No	To provide clarification
Section 4.2.1.4	21-Jun-2023	Specify TEST statement used for log rank test	N/A	To add clarity
Section 4.2.1.5, 4.2.2.5	21-Jun-2023	Add eCRF stratification analysis	N/A	Additional summaries of interest
Section 4.2.1.6	21-Jun-2023	Add HER2 status, ET, ECOG PS, prior use of CDK4/6 inhibitor and early relapse as subgroups	No	Additional subgroups on interest
Section 4.2.1.6	21-Jun-2023	Stratification factors from CRF	N/A	Change to summary
Section 4.2.1.6	21-Jun-2023	Removed qualitive interaction testing	N/A	Does not add any useful supportive information
Section 4.2.2.4	21-Jun-2023	Specify TEST statement used for log rank test	N/A	To add clarity
Section 4.2.2.4, 4.2.9.3	21-Jun-2023	Remove prematurely censored	N/A	Change to standards

Section 4.2.4.2	21-Jun-2023	Define confirmed CR and PR	N/A	To provide clarification
Section 4.2.5	21-Jun-2023	Remove sensitivity analysis	N/A	Change to standards
Section 4.2.5.2	21-Jun-2023	Define confirmed CR and PR	N/A	To provide clarification
Section 4.2.5.4	21-Jun-2023	Specify sensitivity analysis is a repeat of the primary	N/A	Ensure alignment in summary between analyses
Section 4.2.6	21-Jun-2023	Remove sensitivity and subgroup analysis	N/A	Change to summary
Section 4.2.6.4	21-Jun-2023	Specify analysis	N/A	To add clarity
Section 4.2.7.2	21-Jun-2023	Define confirmed CR and PR	N/A	To provide clarification
Section 4.2.8.2	21-Jun-2023	Remove text about MCID when estimation method not performed	N/A	Remove duplicate text
Section 4.2.8.2	21-Jun-2023	Remove death as an event for TTD	N/A	Correction to previous error
Section 4.2.8.2	21-Jun-2023	Specify what evaluable is and changed missed visits to immediately	N/A	To add clarity
Section 4.2.8.2	21-Jun-2023	Specify how participants with baseline score close to maximum/minimum will be censored	N/A	To add clarity
Table 9	21-Jun-2023	Removed first row from table	N/A	As death is no longer an event this row is not needed
Section 4.2.8.3	21-Jun-2023	Change <50% to ≤50%	N/A	To align with scoring manual
Section 4.2.8.4	21-Jun-2023	Specified the analysis being done	N/A	To add clarity

Section 4.2.9.3	21-Jun-2023	Specify which participants are included in time between progression and starting subsequent therapy summary	N/A	To add clarity
Section 4.2.9.3	21-Jun-2023	Remove prematurely censored	N/A	Change to standards
Section 4.2.11.1	21-Jun-2023	Add other as a second progression	N/A	Align with data collection
Section 4.2.11.2	21-Jun-2023	Additional information regarding PFS2 derivation added	N/A	To add clarity
Section 4.2.12.2	21-Jun-2023	Change list of summary statistics	N/A	Change to summary
Section 4.2.12.2	21-Jun-2023	Remove non compartment PK analysis	Yes	Analysis no longer requirement
Section 4.2.13	21-Jun-2023	Additional summary and ADA categories added	N/A	Additional summary of interest
Section 4.2.14.2	21-Jun-2023	Additional information on compliance added	N/A	To add clarity on participants to be included
Section 4.2.14.2	21-Jun-2023	Definition of overall compliance added	N/A	Additional summary of interest
Section 4.2.14.4	21-Jun-2023	Specify 12 weeks excludes baseline record	N/A	To add clarity
Section 4.2.14.4	21-Jun-2023	Change scale from 0-4 to 1-5	N/A	Correction to previous error
Section 4.2.14.4	21-Jun-2023	Add stacked bar chart and pie chart	N/A	Additional summary of interest
Section 4.2.14.4	21-Jun-2023	Specify if summary is on all symptoms and attributes or only 1st attribute	N/A	To add clarity
Section 4.2.14.5	21-Jun-2023	Add additional summaries	N/A	Additional summary of interest. To align with other endpoints

Section 4.2.14.6	21-Jun-2023	Add stacked bar chart and pie chart	N/A	Additional summary of interest
Section 4.2.16.3	21-Jun-2023	Remove stacked bar chart	N/A	Change to analysis
Section 4.2.17.4	21-Jun-2023	Add plot of absolute score and change from baseline	N/A	Additional summary of interest
Section 4.2.17.4	21-Jun-2023	Specified the analysis being done	N/A	To add clarity
Section 4.2.18.4	21-Jun-2023	Add table of assessment response, absolute score and change from baseline. Add plot of absolute score and change from baseline	N/A	Additional summary of interest
Section, 4.2.18.4	21-Jun-2023	Specified the analysis being done	N/A	To add clarity
Section 4.2.19	21-Jun-2023	Add 10-point change from baseline as MCID	N/A	Preference of MCID is different by country
Section 4.2.19.2	21-Jun-2023	Change EQ-5D index derivation	N/A	Change in guidance
Section 4.2.19.3	21-Jun-2023	Change analysis population to FAS	Yes	Align with CSP
Section 4.4.1.1	21-Jun-2023	Update to intended dose calculation	N/A	Correction to previous error
Section 4.4.1.2	21-Jun-2023	Change dose interruption to drug interruption	N/A	To avoid confusion in terminology
Section 4.4.1.2	21-Jun-2023	Add summary of dose delays	N/A	Additional summaries of interest
Section 4.4.2.1	21-Jun-2023	Add information about AESI follow up	Yes	Align with CSP
Section 4.4.2.2	21-Jun-2023	Additional summaries added	N/A	Additional summaries of interest

Section 4.4.2.2	21-Jun-2023	28 days after the date of last IP, but after the onset of subsequent therapy changed to 35 days after the date of last IP and before the onset of subsequent therapy	Yes	To be consistent with safety follow up period
Section 4.4.2.2	21-Jun-2023	Change AEIS summary for AESI category to AESI category and PT	N/A	Additional summaries of interest
Section 4.4.2.3	21-Jun-2023	Remove text regarding categories marked (*)	N/A	Correction to previous error
Section 4.4.3	21-Jun-2023	Specify project ranges used if local range missing	N/A	Change to derivation
Section 4.4.3	21-Jun-2023	Remove coagulation	N/A	Change in data collection
Section 4.4.3.2	21-Jun-2023	Add CTCAE grade to listing	N/A	Additional summary of interest
Section 4.4.3.2	21-Jun-2023	Update to Hy's law summary	N/A	Change to summary
Section 4.4.3.2	21-Jun-2023	Update to Hy's law definition	N/A	Change to standards
Section 4.4.7.1, 4.4.8.1 and 4.4.12.1	21-Jun-2023	Specify summary will be by nominal visits	N/A	Assessments only collected at 2 timepoints and clinically indicated otherwise so cannot apply visit windowing
Section 4.4.7.1	21-Jun-2023	Add additional rules for derived average	N/A	Programming clarification
Section 4.4.7.2	21-Jun-2023	Remove ECG summary table	N/A	Change to analysis
Section 4.4.8.2	21-Jun-2023	Remove EOT summary	N/A	Change in data collection
Section 4.4.11.2	21-Jun-2023	Change to summaries	N/A	Change in data collection

Section 4.4.11.2	21-Jun-2023	Change OAS to SAF	N/A	Correction to previous error
Section 5	21-Jun-2023	Remove duplicate text. Change participant safety department to AstraZeneca global patient safety	N/A	Correction to previous error

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D9268C00001 supporting the clinical study report (CSR). The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection. This statistical analysis plan (SAP) is based on version 2.0 of the CSP dated 27 August 2021. In the event of future amendments to the protocol, this SAP may be modified to account for changes relevant to the statistical analysis.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

Table 9 (population for analysis) in the protocol states that all PROs will be analysed using the FAS population, while the PRO endpoints that measure patient-reported symptomatic adverse events and overall treatment tolerability should be analysed based on the safety population per Table 4 (objectives and endpoints) in the protocol, i.e. among participants who received any amount of study treatment and according to the actual treatment received. See SAP Table 1 for more details and specifying the analysis population to align with Table 4 (objectives and endpoints) in the protocol.

For the sensitivity analysis of attrition bias for progression free survival (PFS), the SAP states two, or more, missed tumour assessments, while the CSP states two, or more, non-evaluable tumour assessments.

In Section 8.3.6 of the CSP Potential Hy's law was clarified and participants with elevated liver enzymes (ALT or AST ≥5x ULN) that have liver metastases present at baseline are now permitted in the study.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

As the study is event driven, the accrual of the predetermined number of events included in the study endpoints determines the duration of the data collection phase of the study. There are 4 planned data cut-offs (DCOs) for this study consisting of an ophthalmologic data review (DCO1), primary analysis of progression free survival (PFS)/first overall survival (OS) interim analysis (DCO2), second OS interim analysis (DCO3) and primary analysis of OS (DCO4). These interim analyses and additional safety reviews will be conducted as described in Section 5.

3.1.1 Ophthalmologic Data Review (DCO1)

The DCO for the ophthalmologic data review (DCO1) is planned to occur after completion of the last ophthalmologic assessment and a minimum of 2 assessments per participant has been completed, for the first approximately 100 randomised participants (DCO1).

3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2)

The DCO for the primary analysis of PFS/first OS interim analysis (DCO2) is planned to occur when approximately 419 PFS Blinded independent central review (BICR) events have been observed in the FAS. Based on enrolment assumptions, it is expected that this will occur approximately 21 months after randomisation of the first participant. For the primary analysis of PFS this provides approximately 60% maturity. For the first OS interim analysis this corresponds to approximately 25% maturity and 40% of the information expected at the primary analysis.

3.1.3 Second OS Interim Analysis (DCO3)

The DCO for the interim analysis for superiority in OS (DCO3) is planned to occur when approximately 355 OS events have been observed in the FAS population. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis. Based on enrolment assumptions, it is expected that this will occur approximately 34 months after randomisation of the first participant.

3.1.4 Primary Analysis of OS (DCO4)

The DCO for the primary analysis of OS (DCO4) is planned to occur when approximately 444 OS events have been observed in the FAS population, approximately 44 months after the first participant is randomised (63% maturity). At this time the clinical database will close to new data.

3.2 Analysis Populations

There are six analysis sets defined for this study.

3.2.1 Enrolled set

The Enrolled Set consists of all participants who provided informed consent irrespective of whether they were randomised or received the study treatment.

3.2.2 Full analysis set

The FAS consists of all randomised participants with treatment groups assigned in accordance with the randomisation, regardless of the treatment actually received. Participants who were randomised but did not subsequently receive treatment are included in the FAS. The analysis of data using the FAS therefore follows the principles of intention to treat (ITT).

3.2.3 Safety analysis set

The safety analysis set (SAF) will consist of all randomised participants who received any amount of study treatment (Dato-DXd or **Investigator's Choice Chemotherapy** (ICC)). Safety data will be summarised using the safety analysis set according to the actual

treatment received. If a participant receives any amount of Dato-DXd, they will be summarised in the Dato-DXd treatment group. If a participant only receives ICC, they will be summarised in the ICC treatment group. No crossover between study treatment arms will be allowed.

3.2.4 Ophthalmologic analysis set

The ophthalmologic analysis set (OAS) will consist of approximately the first 100 randomised participants (approximately 50 per arm, Dato-DXd and ICC). This will consist of all participants who have had ophthalmologic assessments recorded.

3.2.5 Pharmacokinetic analysis set

The pharmacokinetic analysis set (PAS) consists of all participants randomly assigned to study intervention who received at least 1 dose of Dato-DXd for whom there is at least one reportable post-dose pharmacokinetic (PK) concentration.

Participants who violate or deviate from the protocol in ways that would significantly affect the PK analyses should not be included in the PK analysis set. These are participants who have a >10% dose reduction, >10% wrong dose or >10% delay for all timepoints.

3.2.6 ADA evaluable set

The anti-drug antibody (ADA) evaluable set will consist of participants in the safety analysis set with a non-missing ADA Dato-DXd result at any time.

3.2.7 Summary of outcome variables and analysis sets

The analysis sets to be used for each outcome are provided in Table 1.

Table 1 Summary of outcome variables and analysis sets

Outcome variable	*Analysis set*
Efficacy Data	
PFS, PFS2, OS, ORR, DoR, DCR, TFST, TSST	FAS
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	FAS
Baseline and disease characteristics	FAS
Important protocol deviations	FAS
Medical/surgical history	FAS
Previous anti-cancer therapy	FAS
Concomitant medications/procedures	FAS
Subsequent anti-cancer therapy	FAS
Study drug compliance	FAS

Outcome variable	*Analysis set*
PK/Immunogenicity Data	
PK data	PAS
Immunogenicity	ADA evaluable set
Safety data	
Exposure	SAF
Adverse events	SAF
Laboratory measurements	SAF
Vital signs	SAF
Physical examination	SAF
ECGs	SAF
ECOG PS	SAF
ECHO/MUGA	SAF
Ophthalmologic assessments	SAF
Patient-reported outcomes	
EORTC QLQ-C30, EORTC IL116, EQ-5D-5L, PGIS, PGIC	FAS
PGI-TT, PRO-CTCAE, EORTC IL117	SAF
Health care resource use	
HcRU	SAF

Participants who are evaluable for the analysis of DoR are those who responded in the ORR analysis.
ADA=antidrug antibody; BoR=best objective response; CTCAE=Common Terminology Criteria for Adverse Events; DCR=disease control rate; DoR= duration of response; ECG=electrocardiogram; ECHO=echocardiogram; ECOG=Eastern Cooperative Oncology Group; EORTC=European Organisation for Research and Treatment of Cancer; EORTC QLQ-C30=EORTC 30-item core quality of life questionnaire; EQ-5D-5L=EuroQoL 5-dimension, 5-level health state utility index; FAS=Full analysis set; HcRU=Healthcare resource use; MUGA=multigated acquisition; ORR=objective response rate; OS=overall survival; PAS=pharmacokinetic analysis set; PGIC=Patients' Global Impression of Change; PGIS=Patients' Global Impression of Severity; PGI-TT= Patient's Global Impression of Treatment Tolerability; PFS=progression-free survival; PFS2=time from randomisation to second progression or death; PK=pharmacokinetic; PRO=patient-reported outcome; PS=Performance Status; SAF=safety analysis set; TFST=time to first subsequent therapy or death; TSST=time to second subsequent therapy or death.

3.3 General Considerations

The below mentioned general principles are followed throughout the study:

- Summary tables are produced by treatment group (Dato-DXd and ICC). Total columns are produced only for tables of disposition, demography, baseline characteristics and EQ-5D-5L data.

- Descriptive statistics are used for variables, as appropriate. Continuous variables are summarised by the number of observations, mean, standard deviation, median, upper and lower quartiles where indicated, minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables are summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages are calculated out of the population total for the corresponding treatment group.
- For continuous data, the mean and median are rounded to 1 additional decimal place compared to the original data. The standard deviation is rounded to 2 additional decimal places compared to the original data. Minimum and maximum are displayed with the same accuracy as the original data.

  If the number of decimal places in the original data is >3 then display minimum and maximum to 3 decimal places, mean and median to 4 decimal places and the standard deviation to 5 decimal places.
- For categorical data, percentages are rounded to 1 decimal place with the exception of 100% which is presented as a whole number.
- Results of all statistical analyses will be presented using a 95% confidence interval (CI) and 2-sided p-value, unless otherwise stated.
- CIs and ratios (including hazard ratios) will be rounded to 2 decimal places. The p-values will be rounded to 4 decimal places, except for those below 0.0001, which will be displayed as '<0.0001'.
- Survival rates will be rounded to 1 decimal place.
- ORR and DCR will be rounded to 2 decimal places.
- SAS® version 9.4 (or higher) is used for all analyses.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days. One year is defined to be 365.25 days.

Where analysis models are stratified by the randomisation stratification factors, the data from the Interactive Response Technology (IRT) will be used, not the values recorded in the electronic case report form (eCRF), therefore follows the principles of intention to treat (ITT).

3.3.1 Sample Size Determination

Approximately 1000 participants will be enrolled to achieve approximately 700 randomly assigned to study intervention.

"Enrolled" means a participant's, or their legally acceptable representative's, agreement to participate in a clinical study following completion of the informed consent process.

Potential participants who are screened for the purpose of determining eligibility for the study but are not randomly assigned/assigned in the study, are considered “screen failures”, unless otherwise specified by the protocol.

The study is sized for dual primary endpoints to characterise the PFS and OS benefit of Dato-DXd versus ICC in the participants with HR-positive, HER2-negative breast cancer who have been treated with one or two prior lines of systemic chemotherapy in the inoperable/metastatic setting. The study will be considered positive (a success) if either the PFS analysis results and/or the OS analysis results are statistically significant.

For the primary analysis of PFS (See Section 3.1.2 for timing of analysis) assuming the true PFS treatment effect under the alternative hypothesis is a hazard ratio of 0.55 for Dato-DXd versus ICC, and the median PFS times of 4.7 months and 8.5 months in ICC and Dato-DXd, respectively. 419 PFS events from the FAS population (60% maturity) will provide greater than 99% power to demonstrate statistical significance at the 2-sided alpha level of 1.0%. This also assumes the median PFS times in both groups are exponentially distributed. The smallest treatment difference that is statistically significant will be a hazard ratio of 0.775. Assuming a recruitment period of 19 months, this analysis is anticipated to be approximately 21 months after the first participant has been randomised.

The primary analysis of OS will be performed when approximately 444 OS events from the FAS have occurred across the Dato-DXd and ICC treatment groups (63% maturity). Assuming the true OS hazard ratio is 0.75 for Dato-DXd versus ICC, and the median OS in ICC is 19.0 months, the study will have 85% power to demonstrate statistical significance at the 5.0% level (using a 2-sided test). This assumes the PFS primary analysis crosses the efficacy threshold, and allowing 2 interim analyses to be conducted at information fractions of approximately 40% and 80% of the target events, respectively (per the O’Brien and Fleming approach (Lan & DeMets, 1983)). The smallest treatment difference that could be statistically significant at the primary OS analysis is a hazard ratio of 0.824.

If the PFS primary analysis does not cross the efficacy threshold, the OS analysis will have 83% power to demonstrate statistical significance at the 4.0% level (using a 2-sided test). The smallest treatment difference that could be statistically significant at the primary analysis is a hazard ratio of 0.817. All OS calculations assume median OS times of 19.0 months and 25.3 months in ICC and Dato-DXd, respectively when the survival times are exponentially distributed.

With a recruitment period of approximately 19 months, it is anticipated that the primary OS analysis will occur approximately 44 months after the first participant has been randomised. The study may continue monitoring participants for OS up to the scheduled primary analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival. Further details of the interim analyses are presented in Section 5.

A nonuniform accrual of participants (with k=1.5) is assumed when estimating the analysis times. The total proportion of participants randomised at time t [$t \leq 19$ months] following the start of the study is assumed to be $(t/19)^k$.

3.3.2 Investigational Product

Investigational product (IP) refers to Dato-DXd and the ICC. The first and last dates of IP refer to the earliest and last non-zero dose administration of this treatment respectively.

3.3.3 Baseline

In general, for efficacy endpoints the last observed measurement prior to randomisation is considered the baseline measurement. For safety endpoints and ADA the last evaluable observation before the first dose of IP is considered the baseline measurement unless otherwise specified.

For PRO endpoints the last evaluable observation on or before the target date is considered the baseline measurement. Where target date is first dose of IP if present, else Cycle 1 Day 1 visit date if present, else randomisation date.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured are considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

If two visits are equally eligible to assess participant status at baseline (e.g. two assessments both on the same date with no time recorded), the average is used as the baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking the average is not possible, the best value (value closest to none/normal/negative) is used as baseline as this is most conservative. In the scenario where there are two assessments recorded on the same day, one with time recorded and the other without time recorded, the one with the time recorded is selected as baseline. Where safety data are summarised over time, time on study is calculated in relation to date of first IP administration.

In all summaries change from baseline variables are calculated as the post-treatment value minus the value at baseline. The percentage change from baseline is calculated as (post-baseline value - baseline value) / (baseline value) x 100.

3.3.4 On Treatment

For the purposes of summarising safety data assessed at visits, in addition to baseline data, only on treatment data are included in the summary tables. On treatment data is defined as data after the first dose of IP and with assessment date up to and including the date of last IP + 35 days and prior to the start of any subsequent cancer therapy, whichever occurs earlier.

3.3.5 Visit Window

Time windows are defined for all presentations of safety data, ADA data and PRO data that summarise values by visit according to the following conventions:

- For safety and ADA data study day references date of first dose of IP as Day 1, for PK the reference is the time of Dato-DXd administration on the day PK blood samples are taken, for efficacy data study day references date of randomisation as Day 1.
- All windows for PRO (including those reported on FAS or SAF) will have windows calculated from target date. Where target date is first dose of IP if present, else Cycle 1 Day 1 visit date if present, else randomisation date. However, time to deterioration will be calculated as described in Section 4.2.8.2.
- The time windows are exhaustive so that data recorded at any time point (scheduled or unscheduled) has the potential to be summarised. Inclusion within the time window is based on the actual date and not the intended date of the visit.
- The window for visits following baseline is 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 is Day 2). If an even number of days exist between two consecutive visits, then the upper limit is taken as the midpoint value minus 1 day.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment (as defined in Section 3.3.4) is used (regardless of where it falls in an interval).
- Listings display all values contributing to a time point for a participant.
- For visit-based summaries, if there is more than one value per participant within a time window then the closest value to the scheduled visit date is summarised. If the values are equidistant from the nominal visit date, then the earlier value is used. Data listings highlight the values used in the summary table, wherever feasible.

Visit data are only included in summary tables and figures if the number of observations is ≥20 in at least one treatment group.

3.3.6 Handling of Unscheduled Visits

Unscheduled visits are included in the method of assigning data to scheduled visits described in the rules in Section 3.3.5 above. Unscheduled visits are not included as a separate visit in the summary tables.

For summaries at participant level, such as of extreme values, all post-baseline values collected are used to derive a participant level statistic including those collected at unscheduled visits and regardless of whether they appear in the corresponding visit-based summary.

3.3.7 Missing Data

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

For missing start dates for adverse events (AEs) and concomitant medications/procedures, the following rules are applied:

- Missing day: Impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.
- Missing day and month: Impute 1st January unless year is the same as first dose date then impute first dose date.
- Completely missing date: Impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date.

An imputed start date of an AE must be prior to the end date of the AE.

For missing stop dates of AEs or concomitant medications/procedures, the following rules are applied:

- Missing day - If month is same as month of last dose of study drug and start date is after last dose of study drug then impute last day of the month. Otherwise impute the last day of the month unless month is the same as month of the last dose of study drug then impute last dose date.
- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date.
- Completely missing: If an AE/medication has a completely missing end date then it is treated as ongoing.

The imputation of dates for AEs and concomitant medications is used to determine if an AE is treatment emergent and whether a medication is concomitant. Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations are not calculated.

If a participant is known to have died where only a partial death date is available, then the date of death is 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:

- Missing day only: Using the 1st of the month.
- Missing day and month: Using the 1st January.

For partial subsequent anti-cancer therapy dates, the following rules are applied:

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

If a participant has a partial date of birth, (i.e. for those cases where year of birth only is given) the 1st of the month is imputed if only day is missing, and 1st January is imputed if the day and month are missing.

Other rules for handling missing data are described under the derivation rules for that particular variable.

3.3.8 Derivations of RECIST Visit Responses

For all participants, the Response Evaluation Criteria in Solid Tumours (RECIST) tumour response data will be used to determine each participant's visit response according to RECIST version 1.1. It will also be used to determine if and when a participant has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumour assessments are to be performed no more than 28 days before randomisation and ideally as close as possible to and prior to randomisation. Tumour assessments are then performed every 6 weeks following randomisation for 48 weeks, then every 9 weeks thereafter until disease progression. Following disease progression, 1 additional follow-up scan should be performed as per imaging schedule (i.e. either 6 weeks or 9 weeks later).

If an unscheduled assessment is performed, and the participant 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 minimise any unintentional bias caused by some participants being assessed at a different frequency than other participants.

From the investigator's review of the imaging scans, the RECIST tumour response data are used to determine each participant's visit response according to RECIST version 1.1. At each visit, participants are programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD) or progression of disease (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 participant has had a tumour assessment that cannot be evaluated

then the participant is assigned a visit response of not evaluable (NE), (unless there is evidence of progression in which case the response is assigned as PD).

Please refer to Section 3.3.8.3 for the definitions of CR, PR, SD and PD.

RECIST outcomes (i.e. PFS, ORR etc.) are calculated programmatically for the site investigator data (as described in the relevant subsection in Section 4.2) from the overall visit responses.

3.3.8.1 Target lesions (TLs)

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 participant can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as TLs. Lymph nodes are considered to be one organ; therefore, at most 2 target nodal lesions can be selected. If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomisation are 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 target lesion (TL) should be identified as NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

TL visit responses are described in Table 2 below.

Table 2 TL visit responses (RECIST 1.1)

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 <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Stable disease (SD)	Neither sufficient decrease in sum of diameters to qualify for PR nor sufficient increase to qualify for PD.

Visit responses	Description
Progression of disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Not evaluable (NE)	Only relevant if any of the TLs at follow-up were not assessed or not evaluable (e.g. missing anatomy) or had a lesion intervention at this visit. Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not applicable (NA)	Only relevant if 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 is rounded to one decimal place (d.p.) before assigning a TL response. For example, 19.95% is rounded to 20.0% but 19.94% is rounded to 19.9%.

Missing TL data

If the sum of **available** 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, the TL visit response is PD. Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If there is at least one TL measurement missing and a **TL** visit response of PD cannot be assigned, the **TL** visit response is not evaluable (NE).

If all TL measurements are missing, then the TL visit response is NE. Overall visit response is also NE, unless there is a progression of non-TLs or new lesions, in which case the response is PD.

Lymph nodes

For lymph nodes, if the size reduces to <10mm then these are considered non-pathological. However, a size is still given, and this size is still 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 is 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 are applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is 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% but remains <10mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis >10mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response is set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined the response remains as CR.

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure, then this is indicated as such on the CRF and a value of 5mm is 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 is reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) are recorded.

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

- Step 1: the diameters of the TLs (including the lesions that have had intervention) are summed and the calculation are performed in the usual manner. If the visit response is PD, this remains as a valid response category.

- Step 2: If there is no evidence of progression after step 1, the lesion diameter (for those lesions with intervention) are treated 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 participant are 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 is used, and PR or SD assigned as the visit response. Participants 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 participant 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 is set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

If >1/3 of TL measurements are missing (because of intervention) then the TL response is 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 ≥5mm from nadir).

If ≤1/3 of the TL measurements are missing (because of intervention) then the results are scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters) and this is 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 74mm.

The sum of lesions 1-4 at the follow-up is 68mm. The sum of the corresponding lesions at the nadir visit is 62mm.

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

68 x 74 / 62 = 81mm

CR is not allowed as a TL response for visits where there is missing data. Only PR, SD, or PD (or NE) can be assigned as the TL visit response in these cases. However, for visits with

≤1/3 lesion assessments not recorded, the scaled-up sum of TLs diameters is 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 are 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 is recorded for one of the TL sizes and the other TL size is recorded as 0cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs, between CT and MRI is considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions are treated as missing.

3.3.8.2 Non-target lesions (NTLs) and new lesions

At each visit the investigator records an overall assessment of the non-target lesion (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 is derived based on the investigator's overall assessment of NTLs as described in Table 3:

Table 3 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).
Non-CR/non-PD	Persistence of one or more NTLs with no evidence of progression.
Progression (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.

Visit responses	Description
Not evaluable (NE)	Only relevant when one or some of the NTLs are not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

NTL non-target lesion; TL 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 tumour 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 are also 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 indicates 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 tumour.

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

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

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 that the new lesion response should be treated as NE in the derivation of the overall visit response.

3.3.8.3 Overall visit response – site investigator data

Table 4 defines how the previously defined TL and NTL visit responses are combined with new lesion information to give an overall visit response.

Table 4 Overall visit response

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**

CR complete response; PR partial response; SD stable disease; NE not evaluable; NA not applicable; PD progressive disease.

3.3.8.4 Independent review

It is planned that a blinded independent central review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans for all participants (including those at unscheduled visits, or outside visit windows) are collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans are reviewed by two independent radiologists using RECIST 1.1 and are 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 participant, the BICR defines the overall visit response (CR, PR, SD, PD, NE) (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. If a participant has had a tumour assessment that cannot be evaluated, then the participant is assigned a visit response of NE (unless there is evidence of progression in which case the response is 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 are 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 is selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision is used). In the absence of adjudication, the records for all visits for a single reviewer is used. The reviewer selected in the absence of adjudication is the reviewer who read the baseline scan first. The records from the single selected reviewer are used to report

all BICR RECIST information including dates of progression, visit response, censoring and changes in TL dimensions. Endpoints (of ORR, PFS, DoR and DCR) are derived programmatically from this information.

Results of this independent review are not communicated to investigators and the management of participants is based solely upon the results of the RECIST 1.1 assessment conducted by the investigator.

An independent central review (ICR) of all participants will be performed for the primary analysis of PFS, which will cover all the scans up to the DCO.

Further details of the BICR are documented in the Independent Review Charter (IRC).

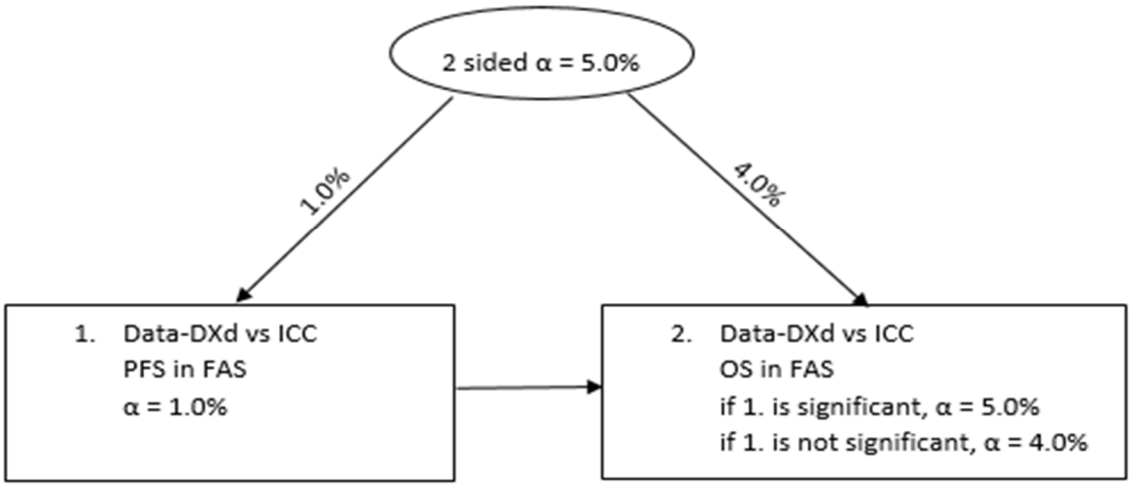
3.3.9 Multiplicity/Multiple Comparisons

The hypotheses of interest with regards to the efficacy for the dual primary endpoints are:

- H0: No differences between Dato-DXd and ICC for PFS and OS.
- H1: Differences between Dato-DXd and ICC for PFS and/or OS.

To preserve the overall type 1 error (familywise error rate) at 5% in the strong sense, a multiple testing procedure (MTP) for the dual primary endpoints of PFS and OS is implemented at DCO2, DCO3 and DCO4. An overview of the MTP with an alpha-splitting and exhaustive recycling strategy (Burman, Sonesson, & Guilbaud, 2009) is provided in Figure 1.

Figure 1 Multiple Testing Procedure



An alpha level of 1.0% will be allocated to the PFS primary analysis and the remaining 4.0% alpha level will be allocated to the OS analyses. If the PFS primary analysis meets statistical significance, the 1.0% type 1 error allocated to PFS endpoint will be reallocated (Burman, Sonesson, & Guilbaud, 2009) to the OS endpoint for a total 2-sided type 1 error of 5.0%. If the PFS primary analysis does not meet statistical significance, the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Alpha spending functions are applied for the OS endpoint in order to preserve the overall 2-sided type 1 error (familywise error rate) in the strong sense across the three planned analyses of OS.

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including 2 interim analyses for superiority of OS. Details are provided in the Interim Analysis section (Section 5).

The significance level alpha for OS across the three analysis times is dependent on the OS information fraction (number of OS events at interim/number of OS events at primary). The significance levels are calculated at the time of the analyses based on the number of events observed.

No multiplicity adjustment is applied for other endpoints as other endpoints are considered supportive endpoints.

3.3.10 Handling of Protocol Deviations in Study Analysis

Protocol deviations are collected, reviewed and reconciled throughout the study. Important protocol deviations (IPDs) are identified from the complete set of protocol deviations. IPDs are those which may significantly impact the reliability of the study data or that may significantly affect a participant's rights, safety, or wellbeing.

A set of pre-determined IPDs are listed in the protocol deviations plan. The protocol deviations plan also indicates which IPDs are identified by programmatic checks.

The IPDs are grouped into the following important protocol deviation (IPD) categories, where full details of the individual IPDs within each IPD category are provided in the protocol deviations plan:

- Inclusion criteria deviations.
- Exclusion criteria deviations.
- Discontinuation criteria for study product met but participant not withdrawn from study treatment.
- Discontinuation Criteria for overall study withdrawal met but participant not withdrawn from study.
- IP deviation.
- Excluded medications taken.
- Deviations to study procedure.
- Other important deviations.

The following general categories will be considered important protocol deviations (IPDs) and will be programmatically derived from Veeva Clinical Vault (VCV) data. These will be listed and summarised by randomised treatment group and discussed in the CSR as appropriate. Refer to the CSP for full details of the inclusion/exclusion criteria.

- Inclusion criteria deviations (Deviation 1).
  - Lack of provision of informed consent prior to any study-related procedures.
  - Inclusion criteria 2, 3, 6, 7, 8, 9
- Exclusion criteria deviations (Deviation 2).
  - Exclusion criteria 1, 4, 5, 6, 9, 10-14, 22
- Discontinuation criteria for study product met but participant not withdrawn from study treatment (Deviation 3).
- Discontinuation criteria for overall study withdrawal met but participant not withdrawn from study (Deviation 4).
- Investigational product deviation (Deviation 5).

- Excluded medications taken (Deviation 6).
- Deviations related to study procedure (Deviation 7)
- Other important deviations (Deviation 8)

Participants who receive incorrect treatment at any time will be included in the safety analysis set as described in Section 3.2.3. 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 global study director and/or statistician.

None of the deviations will lead to participants being excluded from the analysis sets described in Section 3.2 (except for the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding participants with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed by repeating the PFS analysis excluding participants with deviations that may affect the efficacy of trial therapy. 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.

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Study Population

The domain study population covers participant disposition, analysis sets, protocol deviations, demographics, baseline characteristics, disease characteristics, medical history, prior and concomitant medication.

Study population data is summarised and listed using the FAS unless otherwise stated.

4.1.1 Participant Disposition and Completion Status

4.1.1.1 Presentations

Participant disposition is summarised by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine, presenting the number of participants screened, the number of screen failures, the number and percentage of participants: randomised, who were randomised and did not receive study treatment, who started IP (where participants

starting Dato-DXd or ICC are presented according to the randomisation), who are ongoing in the study at the DCO, and who completed and discontinued the study along with reasons for study discontinuation.

The denominator for the percentages in these summaries is based on the number of randomised participants for the treatment group.

The number and percentage of participants ongoing on IP at the time of the DCO and who discontinued each study treatment along with the reasons for treatment discontinuation is also summarised, where the denominator for the percentages is based on the number of participants that started the corresponding study treatment.

Additionally, a summary of recruitment by region, country and site is produced for the FAS. The denominator for the percentages in this summary is based on the number of participants for the analysis set for the treatment group.

Listings presenting details of discontinuations by individual participant are produced for those participants discontinuing treatment and discontinuing the study.

4.1.2 Analysis Sets

The number of participants in each analysis set and reasons for exclusion from each analysis set are summarised.

A listing of individual participants not included in each analysis set is provided.

4.1.3 Protocol Deviations

Refer to Section 3.3.10 for details regarding the definition and derivation of protocol deviations.

A summary table is produced showing the number and percentage of participants with any IPD and by category of IPD, which includes the individual IPDs as detailed in the protocol deviations plan.

The individual participant data for IPDs is also listed.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Demographics are comprised of age, age group, sex, race and ethnicity.

Age is calculated as age at last birthday in whole years using the date of randomisation and date of birth. Where a partial date of birth has been collected, it is imputed as described in Section 3.3.7 to calculate the participant's age for use in listings and summaries tables

presenting age and/or age group and subgroup analyses based on age. Age is split into the following categories: <65 and ≥65 years.

Date of birth is listed as it has been collected on the eCRF.

4.1.4.2 Presentation

A summary table of demographic data specified in Section 4.1.4.1 is produced, and demographic data is listed.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Baseline characteristics include height (cm), weight (kg), weight group, body mass index (BMI) (kg/m^2), body surface area (BSA) (m^2) and ECOG performance status at baseline.

Weight (kg) is categorised into weight groups of:

- <65
- ≥65 and ≤90
- >90

The body mass index (BMI) is calculated as

BMI (kg/m^2) = Weight (kg) / {Height (m)}^{2}

The BSA is calculated as

$$(\text{BSA})\ (\text{m}^2) = \sqrt{\frac{\text{Height (cm)}\ X\ \text{Weight (kg)}}{3600}}$$

Stratification factors (number of previous lines of chemotherapy, geographic region (with countries within region), and prior use of CDK4/6 inhibitor) are derived from the CRF data as well as recorded by IRT.

4.1.5.2 Presentation

Baseline characteristics specified in Section 4.1.5.1 are summarised and listed.

A summary of the stratification factors recorded both by IRT and CRF is also provided presenting the number and percentage of participants in each category.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Disease characteristics include time from most recent disease progression to randomisation, overall disease classification (metastatic/locally advanced), duration of prior breast cancer

CDK4/6 inhibitor therapy (<12 months, >=12 months, missing duration, NA), HER2 status, estrogen receptor status, progesterone receptor status, prior lines of chemotherapy, prior lines of anti-cancer therapy, visceral metastases and de novo metastatic breast cancer. Additionally sites of breast and regional lymph node/other locally advanced sites and all other sites is included.

Time from most recent disease progression to randomisation (days) is defined as:

(Date of randomisation - date of most recent disease progression)

Time from most recent disease progression to randomisation is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

Time from diagnosis to randomisation in years is defined as:

(Date of randomisation – original diagnosis date) / 365.25

Time from diagnosis to randomisation is categorised into groups of ≤5 years and >5 years.

Early relapse is defined as the time between end of (neo) adjuvant chemotherapy and start of first line treatment for relapse to inoperable/metastatic disease being less than one year.

Hepatic function status at baseline is defined as follows:

- Normal
  - Total bilirubin ≤ULN and AST ≤ULN (except for Gilbert syndrome participants)
  - Total bilirubin ≤3x ULN and AST ≤ULN for participants with Gilbert syndrome
- Mild impairment
  - Total bilirubin >ULN, ≤1.5x ULN and any AST except for participants with Gilbert syndrome
  - Total bilirubin >ULN, ≤3x ULN and AST >ULN for participants with Gilbert syndrome
  - Total bilirubin <ULN and AST >ULN regardless of Gilbert syndrome

Renal function status at baseline is defined as follows:

- Normal: serum creatinine clearance ≥90 mL/min
- Mild impairment: serum creatinine clearance ≥60, <90 mL/min
- Moderate impairment: serum creatinine clearance >30, <60 mL/min

4.1.6.2 Presentation

The following breast cancer characteristics at diagnosis are summarised: American joint committee on cancer (AJCC) stage, histology type, tumour grade and time from diagnosis to randomisation.

In addition, the disease characteristics (described in Section 4.1.6.1) are summarised.

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Medical history and relevant surgical history are coded using the medical dictionary for regulatory activities (MedDRA) [using the latest or current MedDRA version].

Any medical history which is ongoing at time of informed consent is considered an ongoing condition, otherwise it is considered past medical history.

4.1.7.2 Presentation

Summary tables of past medical history, ongoing conditions and surgical history are presented by MedDRA system organ class (SOC) and preferred term (PT). Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

All medical history, surgical history and ongoing conditions are listed.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

Information on any treatment used from the time of informed consent up to 28 (+7) days after discontinuation of study treatment, with reasons for the treatment, will be recorded in the eCRF. Thereafter, only subsequent regimens of anti-cancer therapy will be recorded in eCRF.

Other anti-cancer therapies, investigational agents, and radiotherapy (except palliative to areas other than the chest) should not be given while the participant is on study drug.

Medications received prior to, concomitantly, or post-treatment are coded using the Anatomical Therapeutic Chemical (ATC) Classification codes.

Medications will be coded using the latest World Health Organization Drug dictionary (WHO Drug) version. The version used will be indicated in the data summaries and listings.

For inclusion in the prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.3.7.

Prior medications, concomitant medications and post-study treatment medications are defined as follows:

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

Missing coding terms are listed and summarised as "Not coded".

Time from completion of prior anti-cancer therapy to randomisation is defined as

(date for randomisation – last anti-cancer therapy stop date + 1)

Time from completion of prior anti-cancer therapy is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

4.1.8.2 Presentation

The following summaries will be produced:

- Summary of prior medications
- Summary of concomitant medications
- Summary of prohibited concomitant medications
- Summary of post study treatment medications
- Summary of prior disease-related treatment modalities
- Summary of prior cancer therapies
- Summary of time from completion of prior anti-cancer therapy to randomisation
- Summary of non-study cancer therapies whilst on study treatment
- Summary of post study treatment cancer therapies
- Summary of Trastuzumab Deruxtecan and Sacituzumab Govitecan as post study treatment cancer therapies
- Summary of prior radiotherapy
- Summary of on study palliative radiotherapy
- Summary of post study treatment radiotherapy

Prior medications (excluding prior cancer therapies), concomitant medications (including both allowed and prohibited concomitant medications), prohibited concomitant medications

and post-study treatment medications (excluding post-study treatment cancer therapies) are presented by ATC classification and generic term, sorted by descending frequency of ATC group and generic term. Participants taking the same concomitant medication/procedure multiple times are counted once per ATC classification and generic term.

A summary of number and percentage of participants receiving prior disease-related treatment modalities are summarised by modality. Duration of prior cytotoxic chemotherapy (<6 months, >=12 months, missing duration, NA), prior hormonal therapy in the inoperable or metastatic setting (<6 months, >=12 months, missing duration, NA) and prior breast cancer CDK4/6 inhibitor therapy (<12 months, >=12 months, missing duration, NA) are summarised.

Prior cancer therapies and non-study cancer therapies whilst on study treatment are summarised by therapy class and ATC group. Post-study treatment cancer therapies are summarised by ATC group and generic term. In addition, prior cancer therapy in the inoperable or metastatic setting is summarised by therapy class. Prior chemotherapy in the inoperable or metastatic setting is summarised by generic term.

Trastuzumab Deruxtecan and Sacituzumab Govitecan as post study treatment cancer therapies are summarised by generic term.

A separate summary of number and percentage of participants receiving prior radiotherapy is produced and repeated for on study palliative radiotherapy and post study treatment discontinuation radiotherapy.

Time from completion of prior anti-cancer therapy to randomisation will be presented using summary statistics (e.g. n, mean, median, standard deviation, min, max). In addition, a separate summary of the number and percentage of participants in each category (as described in Section 4.1.8.1) is produced.

All concomitant, prior and post study treatment discontinuation medication and therapy data are listed for all participants.

4.1.9 Study Drug Compliance

This is covered in Section 4.4.1

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses.

Table 5 Endpoint analyses

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Dual Primary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting, per BICR					
Primary	PFS - BICR Assessments	FAS	Participants who have not progressed or died are censored at latest evaluable RECIST 1.1 assessment. Participants who progress or die after 2 missed visits are censored at last evaluable RECIST 1.1 assessment prior to the two missed visits.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.4
Sensitivity 1 (Evaluation time bias)	PFS - BICR Assessments	FAS	As for primary analysis, but the midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be used as the event time.	Stratified log-rank test.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity 2 (Attrition bias)	PFS - BICR Assessments	FAS	As for primary analysis, but using the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments will be included. Participants 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.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. Reverse Kaplan-Meier plot	4.2.1.5
Sensitivity 3 (Ascertainment bias)	PFS - BICR Assessments	FAS	As for primary analysis but using the site Investigator data which is a secondary efficacy variable.	Discrepancy between primary analysis using BICR assessments and the secondary analysis using Investigator assessment.	4.2.1.5
Sensitivity 4 (Subsequent anti-cancer therapy)	PFS - BICR Assessments	FAS	As for primary analysis, but participants who receive subsequent anti-cancer therapy prior to progression or death are censored at latest evaluable RECIST assessment prior to subsequent anti-cancer therapy.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity 5 (Stratification according to eCRF)	PFS - BICR Assessments	FAS	As for primary analysis.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test.	4.2.1.5
Dual Primary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of OS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting					
Primary	OS	FAS	Participant not known to have died will be censored at last date participant was known to be alive.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.2.4
Sensitivity 1 (Attrition bias)	OS	FAS	As for primary analysis but censoring indicator of OS is reversed.	Reverse Kaplan-Meier plot Median	4.2.2.5
Sensitivity 2 (Stratification according to eCRF)	OS	FAS	As for primary analysis.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test.	4.2.2.5
Secondary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of ORR per BICR and per investigator assessment					
Secondary	ORR	FAS	Participants without a response included as non-responders. If participants discontinue treatment without response or progression, receive a subsequent anti-cancer therapy and then respond are not included as responders.	Odds ratio from logistic regression model ORR with Clopper-Pearson CI Difference in ORR with Miettinen-Nurminen CI	4.2.3.3
Secondary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DoR per BICR and per investigator assessment					

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary	DoR	FAS (for participants who have confirmed CR or PR)	For participants who do not progress or die following a response, the censoring rules follow the rules for PFS censoring for the primary analysis.	Summary statistics. KM plots.	4.2.5.3
Supplementary	DoR	FAS (for participants who have confirmed CR or PR)	As for primary analysis but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.	Summary statistics. KM plots.	4.2.5.4
Secondary Objective 3: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS, per investigator assessment					
Secondary	PFS – Investigator Assessments	FAS	As described for primary PFS	As described for primary PFS	4.2.6.4
Secondary Objective 4: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DCR per BICR and per investigator assessment					
Secondary	DCR at 12 weeks	FAS	Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.	Odds ratio from logistic regression model	4.2.7.3

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary Objective 5: To assess pain in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in pain	FAS	Participants who have not deteriorated will be censored at last evaluable assessment. Participants who deteriorate after 2 missed visits are censored at last evaluable assessment prior to the 2 missed visits. If a participant has no evaluable post-baseline data or does not have baseline data, they will be censored at date of randomisation.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.8.4
Sensitivity	TTD in pain	FAS	TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.	As described for secondary TTD in pain.	4.2.8.5
Secondary Objective 6: To assess physical functioning in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4
Sensitivity	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5
Secondary Objective 7: To assess global health status/quality of life (GHS/QoL) in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5
Secondary Objective 8: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TFST					
Secondary	TFST	FAS	Participants not known to have had a first subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a first subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a first subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots. Median. Time between progression and starting subsequent therapy.	4.2.9.3
Secondary Objective 9: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TSST					
Secondary	TSST	FAS	Participants not known to have had a second subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a second subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a second subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	As described for TFST.	4.2.10.3
Secondary Objective 10: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS2					

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary	PFS2	FAS	As described for primary PFS	As described for primary PFS	4.2.11.4
Secondary Objective 11: To assess the PK of Dato-DXd 6mg/kg IV Q3W					
Secondary	Plasma concentrations of Dato-DXd, total anti-TROP2 antibody, and MAAA-1181a (payload)	PAS	N/A	Summary statistics	4.2.12.2
Secondary Objective 12: To investigate the immunogenicity of Dato-DXd 6mg/kg IV Q3W					
Secondary	Presence of ADA	ADA evaluable set	N/A	Summary statistics	4.2.13.2
Safety: To assess the safety and tolerability profile of Dato-DXd relative to ICC					
Safety	Type, incidence and severity (graded by NCI CTCAE v5.0), seriousness and relationship to study medication of AEs	SAF	N/A	Descriptive	4.4.2
Safety	ECOG PS	SAF	N/A	Descriptive	4.4.9
Safety	Vital signs	SAF	N/A	Descriptive	4.4.6
Safety	Physical examination	SAF	N/A	Descriptive	4.4.10
Safety	Clinical laboratory tests	SAF	N/A	Descriptive	4.4.3, 4.4.4,4.4.5
Safety	ECGs	SAF	N/A	Descriptive	4.4.7
Safety	Echocardiogram/ multigated acquisition	SAF	N/A	Descriptive	4.4.8
Safety	Ophthalmologic assessments	SAF	N/A	Descriptive	4.4.11

4.2.1 Primary Endpoint - Progression Free Survival by BICR

4.2.1.1 Definition

Progression-free survival is defined as the time from the date of randomisation until the date of objective disease progression, as defined by RECIST 1.1, or death (by any cause in the absence of progression) regardless of whether the participant withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1). PFS is assessed by BICR assessment. A secondary endpoint analysis of PFS by investigator assessment is reported.

4.2.1.2 Derivations and Censoring Rules

Participants who have not progressed or died at the time of analysis are censored at the time of the latest date of assessment from their last evaluable RECIST assessment. However, if the participant progresses or dies immediately after two or more consecutive missed visits, the participant is 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).

If the participant has no evaluable RECIST assessment or does not have baseline data, they will be censored at the date of randomisation, unless they die within 2 scheduled scans of baseline (12 weeks + 1 week allowing for a late assessment within the visit window) in which case they are treated as an event with date of death as the event date. PFS censoring rules are described in Table 6.

Table 6 Outcome and date of event for PFS analysis

Scenario	Date of PD/ Death event or Censoring	PFS Outcome
First progression or death after at most 1 missed visit, regardless of whether the participants withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression	Date of assessment of first progression or death (by any cause in the absence of progression)	Event
No baseline or evaluable RECIST assessment and death within 2 RECIST visits after the date of randomisation	Date of death	Event
No baseline or evaluable RECIST assessment and no death within 2 RECIST visits after the date of randomisation	Day 1 (Date of randomisation)	Censored
No PD or death at time of data cut-off	Date of last evaluable RECIST assessment*	Censored

Scenario	Date of PD/ Death event or Censoring	PFS Outcome
Death or progression after two or more missed RECIST visits	Date of last evaluable RECIST assessment* prior to the 2 missed visits	Censored

*: if there are no evaluable post-baseline assessments prior to PD or death or data cut-off, participants will be censored at the date of randomisation.

Given the scheduled visit assessment scheme (i.e. every six weeks for the first 48 weeks then every nine weeks thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to nine-weekly this will equate to 17 weeks (i.e. take the average of 6 and 9 weeks which gives 7.5 weeks and then apply same rationale, hence 2 x 7.5 weeks + 1 week for an early assessment + 1 week for a late assessment = 17 weeks). The time period for the previous RECIST assessment will be from study days 288 to 329 (i.e. week 41 to week 47). From week 47 onwards (when the scheduling changes to nine-weekly assessments), two missing visits will equate to 20 weeks (i.e. 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks).

The following is also summarised in Table 7:

Table 7 Definition of two missed RECIST visits

Scheduled Assessment	Previous RECIST assessment	Two missed RECIST visits window
Q6W	No evaluable RECIST visits or no baseline RECIST scan	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W	Day 1	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W up to Week 48	>Day 1 – Day 287 (up to Week 41)	2 x 6 weeks + 2 weeks = 14 weeks (98 days)
	>Day 287 – Day 329 (Week 41 – Week 47) (change period from Q6W to Q9W)	2 x [(6 weeks+9 weeks)/2] + 2 weeks = 17 weeks (119 days)
Q9W thereafter	>Day 329 onwards	2 x 9 weeks + 2 weeks = 20 weeks (140 days)

The PFS time is always derived based on scan/assessment dates and not on visit dates.

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

- For investigator assessments, the date of progression is determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression.
- For BICR assessments, the date of progression is determined based on the earliest of the 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 BICR data.
- For both BICR and investigational assessments when censoring a participant for PFS, the participant is censored at the latest of the scan dates contributing to a particular overall visit assessment.

Note: for TLs only the latest scan date is recorded 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.

4.2.1.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Sections 4.2.1.2 and 4.2.1.5.

4.2.1.4 Primary Analysis of Progression Free Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of PFS in the FAS. The primary analysis of PFS is based on the BICR assessment of PD by RECIST 1.1.

The null hypothesis for the dual primary time to event endpoint of PFS is that there is no difference between Dato-DXd and ICC in the probability of a progression event in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for PFS.

H1: Differences between Dato-DXd and ICC for PFS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

PFS is analysed using a stratified log-rank test (using the LIFETEST procedure with a TEST statement) adjusting for the stratification factors of number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor.

The stratification variables are based on the values entered into IRT at randomisation, even if it is subsequently discovered that these values were incorrect.

If there are less than 10 events in total for a unique stratum or less than 2 events in either treatment group for a unique stratum then the strata are combined in the following order. The CDK4/6 strata (Yes vs No) are pooled first, followed by the number of previous lines of chemotherapy strata (1 vs 2) and then finally by the geographic region strata (United States, Canada, Europe vs Rest of World).

The hazard ratio (HR) and its confidence interval (CI) (95% and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9) and p-value are presented. The HR and CI are estimated from a stratified Cox Proportional Hazards model (with ties = Efron and stratification variables number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor) and the CI calculated using a profile likelihood approach. A HR less than 1 favours Dato-DXd.

Estimates and 95% CI for PFS rates at 3 months intervals and median PFS for each treatment group are presented.

Proportionality assumption

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of log(-log(survival probability)) versus log (time) and, if these raise concerns, by fitting a time dependent covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods for example 0-6m, 6-12m etc. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over the observed extent of follow-up unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

Summaries

In addition to the analyses described above, the following supportive summaries are produced.

Kaplan-Meier (KM) plots of PFS are presented by treatment group. Summaries of the number and percentage of participants experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided for each treatment.

The treatment status at progression of participants at the time of analysis is summarised. This includes the number (%) of participants who were on treatment at the time of progression, the number (%) of participants who discontinued IP prior to progression, the number (%) of participants who have not progressed and were on IP or discontinued IP.

The number of participants censored may be summarised by treatment group together with baseline prognostic factors of the censored participants. This number and percentage of prematurely censored participants is summarised. A participant will be defined as prematurely censored if they did not progress (or die in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval (+ 2 weeks) prior to the DCO date.

Additionally, summary statistics are given for the number of days from censoring to DCO for all censored participants.

The duration of follow-up is summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) participants only, presented by treatment group.

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

Summaries of the number and percentage of participants who miss two or more consecutive RECIST assessments is presented for each treatment group.

All of the collected RECIST 1.1 data is listed for all randomised participants. In addition, a summary of new lesions (i.e. sites of new lesions) is produced.

4.2.1.5 Sensitivity Analyses of Progression Free Survival

Sensitivity Analysis 1 - 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 analysed 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 participants 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 & Chen, 2010). To support this analysis, the mean of participant-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR RECIST assessments.

Sensitivity Analysis 2 - Attrition bias

Attrition bias is assessed by repeating the primary PFS analysis except that the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments are included. In addition, and within the same sensitivity analysis, participants 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 are censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis is supported by a KM 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.

Sensitivity Analysis 3 - Ascertainment bias

Ascertainment bias is assessed by analysing the site investigator data which is a secondary efficacy endpoint (analysis methods presented in Section 4.2.6). The stratified log rank test is repeated on PFS using the site investigator data based upon RECIST. The HR and CI are presented.

If there is an important discrepancy between the primary analysis using the BICR data and this sensitivity analysis using site investigator data a summary table is produced showing the number and proportion of participants with site but no central confirmation of progression and with progression determined by central review but not at site. Such participants 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, Gaschler-Markefski, & Bluhmki, 2001), but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group.

Sensitivity Analysis 4 – Subsequent Anti-cancer Therapy

An additional sensitivity analysis is produced which is a repeat of the primary analysis for PFS, but the censoring rule is modified so that participants who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death are censored at their last evaluable assessment prior to taking the subsequent anti-cancer therapy.

Sensitivity Analysis 5 – Stratification according to eCRF

In the event that there are any mis-stratifications during randomisation, the stratified log rank test will be repeated on PFS, where the stratification factors are as recorded according to the eCRF. The HR and CI will also be presented from the Cox proportional hazards analysis.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival.

Deviation Bias Sensitivity Analysis

A 'deviation bias' sensitivity analysis may be performed as discussed in Section 3.3.10.

4.2.1.6 Subgroup Analyses

Subgroup analyses are conducted comparing PFS between the treatments for the following subgroup of the FAS:

1. Stratification factors at randomisation:
    - Number of previous lines of chemotherapy: 1, 2
    - Geographic region: Region 1 [US, Canada, Europe], Region 2 [Rest of World]
    - Prior use of CDK4/6 inhibitor: Yes, No
2. Exploratory factors
    - Prior use of taxanes and/or anthracyclines: taxanes alone, anthracyclines alone, both taxanes and anthracyclines, neither taxanes nor anthracyclines
    - Age at randomisation: <65, ≥65 years of age
    - Race: Asian, non-Asian
    - Pre-selected investigator's choice of chemotherapy: Capecitabine, Gemcitabine, Eribulin, Vinorelbine
    - Brain metastases (at baseline): Yes, No
    - Sex: male, female
    - HER2 status: IHC 0 versus IHC 1+ and [IHC 2+ cases that are FISH-]
    - ECOG performance score: 0, 1
    - Duration of prior use of breast cancer CDK4/6 inhibitor: Yes, ≤12 months, >12 months, Missing duration

- Early relapse: Yes, No
- Duration of prior use of endocrine therapy in the metastatic breast cancer setting: <6 months, ≥6 months

The subgroup analyses for the pre-selected investigator's choice of chemotherapy will be based on the values obtained from the IRT system; all other factors will be based on values recorded on the eCRF, or from the third-party vendor data.

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

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

For each subgroup level of a factor, the HR and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties and using a BY statement for the subgroup factor.

These HRs and associated two-sided 95% profile likelihood CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

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 descriptive summaries will be provided.

The presence of quantitative interactions may be assessed by means of an overall global interaction test for strata and possibly subgroups:

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by treatment interaction terms, with one that excludes the interaction terms, and will be assessed at the 2-sided 10% significance level.

If there are not more than 10 events per stratum for any covariate (i.e. within each stratum of a treatment*covariate interaction (2 treatments * 2 levels of the covariate = 4 stratum)), a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria, then the covariate-by-treatment interaction term should be

omitted from the model. Moreover, if the covariate does not have more than 10 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved, then it will be concluded that overall, the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant, and all excluded interactions are non-significant. Throughout this process all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

4.2.2 Primary Endpoint - Overall Survival

4.2.2.1 Definition

Overall survival is defined as the time from the date of randomisation until death due to any cause regardless of whether the participant withdraws from randomised therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomisation + 1).

4.2.2.2 Derivations and Censoring Rules

Any participant not known to have died at the time of analysis is censored based on the last recorded date on which the participant was known to be alive.

Note: Survival calls are made in the week following the date of data cut-off (DCO) for the analysis, and if participants are confirmed to be alive or if the death date is after the DCO date, these participants are censored at the date of DCO. This is done at DCO2, DCO3 and DCO4. The status of ongoing, withdrawn (from the study) and “lost to follow-up” participants at the time of the primary OS analysis should be obtained by the site personnel by telephone contact with the participant, participant’s family, by contact with the participant’s current physician, or local death registries. If the participant has actively withdrawn consent to the processing of their personal data, the vital status of the participant can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For the OS analysis at DCO2 and DCO3, performed prior to the primary 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 participant was known to be alive for those participants still on treatment (since the SURVIVE module is only completed for participants off treatment if a survival sweep is not performed). The last date for each

individual participant is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start and stop dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anti-cancer treatment
- Date last known alive on survival status CRF
- End of study date

Duration of follow-up is derived as time from randomisation to the date of death (i.e. overall survival) or to the date of censoring (date last known to be alive).

4.2.2.3 Handling of Dropouts and Missing Data

If a participant is known to have died where only a partial death date is available, then the date of death is imputed following the rules in Section 3.3.7.

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

4.2.2.4 Primary Analysis of Overall Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of OS in the FAS.

The null hypothesis for the dual primary time to event endpoint of OS is that there is no difference between Dato-DXd and ICC in the probability of a death in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for OS.

H1: Differences between Dato-DXd and ICC for OS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

Overall survival will be analysed using a stratified log-rank test (using the LIFETEST procedure with a TEST statement), adjusting for the stratification factors at randomisation. The treatment effect of Dato-DXd against ICC will be estimated by the HR together with its 95% CI and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9.

Estimates and 95% CI for OS rates at 6 monthly intervals are presented along with the median OS for each treatment group.

Summaries

Kaplan-Meier (KM) plots of OS are presented by treatment group. Summaries of the number and percentage of participants who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided.

In addition, the median duration of follow-up is presented for censored participants by treatment group, and for all participants by treatment group and overall.

4.2.2.5 Sensitivity Analyses of Overall Survival

Sensitivity Analysis 1 - Attrition bias

A sensitivity analysis for OS examining the censoring patterns to rule out attrition bias with regard to the primary treatment comparisons is achieved by a KM plot of time to censoring where the censoring indicator of OS is reversed. The KM estimates of median follow-up (overall and by treatment group) are also summarised.

Sensitivity Analysis 2 – Stratification according to eCRF

In the event that there are any mis-stratifications during randomisation, the stratified log rank test will be repeated on OS, where the stratification factors are as recorded according to the eCRF. The HR and CI will also be presented from the Cox proportional hazards analysis.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of overall survival.

4.2.2.6 Subgroup Analyses

Subgroup analyses will be conducted for OS using the same methodology as described for PFS in Section 4.2.1.6.

4.2.3 Secondary Endpoint - Objective Response Rate

4.2.3.1 Definition

Both unconfirmed and confirmed ORR are assessed, where:

- Confirmed ORR is the percentage of participants with an investigator-assessed response of CR or PR 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.
- Unconfirmed ORR is the percentage of participants with at least one investigator-assessed visit response of CR or PR.

The denominator will be defined as all randomised participants.

ORR will also be defined using the BICR data to define a visit response of CR or PR, with the denominator defined as all randomised participants.

4.2.3.2 Derivations

Data obtained up until progression, or last evaluable assessment in the absence of progression, are included in the assessment of ORR, regardless of whether the participant withdraws from therapy. Participants who discontinue randomised treatment without progression, receive a subsequent anti-cancer therapy and then respond are not included as responders in the ORR (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy). For confirmed ORR both visits contributing to a response must be prior to subsequent therapy for the participant to be considered as a responder.

For confirmed ORR, in the case where a participant 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 participant is defined as a responder. Similarly, if a participant 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 is assigned.

ORR is based on all scans regardless of whether they were scheduled or not.

A participant is classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation and confirmed by repeat imaging, prior to RECIST progression and prior to starting any subsequent cancer therapy.

4.2.3.3 Primary Analysis of Objective Response Rate

ORR based on both a confirmed and unconfirmed tumour response are analysed.

A logistic regression model is fitted to tumour response (yes/no) including treatment and the same stratification factors as the primary PFS endpoint as fixed effects. The results of the analysis are presented in terms of an adjusted odds ratio (an odds ratio greater than 1 favour Dato-DXd) together with its associated profile likelihood 95% CI (e.g. using the option ‘LRCI’ in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The adjusted Least Squares (LS) Means response rate (using LSMEANs statement with OM option)

from the logistic regression model together with the corresponding 95% CI is presented for each treatment group.

If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran–Mantel–Haenszel (CMH) test is presented. The CMH test is stratified using the same stratification factors as the primary PFS endpoint. The results of the analysis are presented in terms of an odds ratio together with the 95% CI and p-value. The odds ratio, 95% CI and p-value are obtained using SAS PROC FREQ and the CMH test option.

Both unconfirmed and confirmed ORR as assessed by site investigator will be estimated and presented along with the corresponding exact 95% Clopper-Pearson CI for each treatment arm. The difference in ORR between treatment arms will be reported using point estimates and their two-sided 95% CIs by the Miettinen-Nurminen method (Miettinen & Nurminen, 1985). A Summary will be produced that presents the number and percentage of participants with both an unconfirmed and a confirmed tumour response (CR/PR).

Summaries will also be produced for ORR per BICR.

4.2.4 Secondary Endpoint - Best Objective Response

4.2.4.1 Definition

Best objective response (BoR) is a supportive endpoint for ORR. BoR is calculated based on the overall visit responses from each RECIST assessment, described in Section 3.3.8.3. It is the best response a participant has had following randomisation, but prior to starting any subsequent anti-cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorisation of BoR is based on RECIST using the following response categories: CR, PR, SD, PD and NE.

4.2.4.2 Derivations

BoR is derived using confirmed CR or PR and separately using unconfirmed CR or PR.

Confirmed CR or PR coincides with that used for the confirmed ORR endpoint. For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment is 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 randomisation. For CR/PR, the initial overall visit assessment that showed a response uses the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST from the overall visit response using all BICR data up until the first progression event. It will also be determined programmatically based on RECIST using all site investigator data up until the first progression event. The denominators for each case will be consistent with those used in the ORR analysis.

BoR is determined based on RECIST using both all site investigator data and separately all BICR data up until the earliest of the first progression event/last evaluable assessment in the absence of RECIST or start of any subsequent cancer therapy. The denominators are consistent with those used in the ORR analysis.

For participants whose progression event is death, BoR is calculated based upon all evaluable RECIST assessments prior to death.

For participants who die with no evaluable RECIST assessments, if the death occurs ≤7 weeks (i.e. 6 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR is assigned to the progression (PD) category. For participants who die with no evaluable RECIST assessments, if the death occurs > 7 weeks after randomisation then BoR is assigned to the NE category. For participants with no evaluable RECIST assessments post randomisation, then BoR is also assigned to the NE category.

4.2.4.3 Primary Analysis of Best Objective Response

For each treatment arm, BoR will be summarised by n (%) for each category (CR, PR, SD, PD, and NE) using confirmed and unconfirmed CR/PR responses.

4.2.5 Secondary Endpoint - Duration of Response

4.2.5.1 Definition

DoR will be defined as the time from the date of first documented confirmed response of CR or PR until date of documented progression per RECIST 1.1 (as assessed by Investigator assessment) or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1).

DoR will also be defined using the BICR data to define the overall visit response.

4.2.5.2 Derivations and Censoring Rules

The end of response coincides with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of CR or PR as defined in Table 4.

If a participant does not progress following a response, then the PFS censoring time is used.

For confirmed CR and PR both visits contributing to a response must be prior to subsequent therapy. Confirmation assessment must not be 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.

In the case where a participant 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 this is counted as confirmed PR. Similarly, if a participant 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 this is counted as confirmed CR.

4.2.5.3 Primary Analysis of Duration of Response

The analysis will include all randomised participants as randomised who have a confirmed response, regardless of whether the participant withdraws from therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression.

Descriptive data will be provided for the DoR, including the associated KM curves (without any formal comparison or p-value attached).

KM estimates for median DoR and 25th and 75th percentiles and corresponding 95% CI for DoR will be summarised. KM estimates for percentage remaining in response at 6 monthly intervals are presented for each treatment group.

Additionally, median, 25th and 75th percentiles for time from randomisation to onset of response will be calculated using standard descriptive statistics.

4.2.5.4 Supplementary Analyses of Duration of Response

A supplementary analysis is included where the DoR evaluation is repeated but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.

4.2.6 Secondary Endpoint – PFS by Investigator Assessment

4.2.6.1 Definition

PFS by Investigator assessment is defined as the time from the date of randomisation until the date of PD, as defined by RECIST 1.1 (by Investigator assessment) or death (by any cause in the absence of progression) regardless of whether the participant withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1).

4.2.6.2 Derivations and Censoring Rules

This secondary endpoint of PFS based on Investigator assessment will be derived and censored using the same methodology described in Section 4.2.1.2

4.2.6.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.1.3.

4.2.6.4 Primary Analysis of Progression Free Survival

Statistical Analysis

This secondary endpoint of PFS based on Investigator assessment will be analysed using the same methodology described in Section 4.2.1.4. The HR together with its 95% CI and p-value are presented.

Estimates and 95% CI for PFS rates at 3 months intervals and median PFS for each treatment group are presented.

Summaries

Summaries of the number and percentage of participants experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided for each treatment. The duration of follow-up is summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) participants only, presented by treatment group.

In addition a KM plot of PFS are presented by treatment group.

All of the collected RECIST 1.1 data is listed for all randomised participants. In addition, a summary of new lesions (i.e. sites of new lesions) is produced.

4.2.7 Secondary Endpoint –Disease Control Rate

4.2.7.1 Definition

Disease control rate at 12 weeks is defined as the percentage of participants who have a confirmed CR or PR or have demonstrated SD for at least 11 weeks (i.e. 12 weeks – 1 week to allow for an early assessment within the assessment window) after randomisation without subsequent cancer therapy per RECIST 1.1, as assessed per Investigator assessment and derived from the raw tumour data.

DCR at 12 weeks will also be defined using the BICR data to define the overall visit response.

4.2.7.2 Derivations

Data obtained from randomisation up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of DCR, regardless of whether the participant withdraws from therapy. Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.

For confirmed CR and PR both visits contributing to a response must be prior to subsequent therapy. Confirmation assessment must not be 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.

In the case where a participant 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 this is counted as confirmed PR. Similarly, if a participant 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 this is counted as confirmed CR.

Duration of SD (weeks) is defined as: (date last evaluable assessment of SD in the absence of progression prior to subsequent cancer therapy - randomisation date +1)/7.
Participants without a post-baseline tumour assessment are considered to have no clinical benefit.

DCR is based on all scans regardless of whether they were scheduled or not.

4.2.7.3 Primary Analysis of Disease Control Rate

DCR will be analysed using the same methodology described for ORR in Section 4.2.3.3.

The analysis will be performed on the FAS

4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30

4.2.8.1 Definition

EORTC QLQ-C30

The European Organisation for Research and Treatment of Cancer (EORTC) 30-item quality of life (QoL) questionnaire (QLQ-C30) consists of 30 questions that are combined to produce 5 multi-item functional scales (physical, role, cognitive, emotional, and social), 3 multi-item symptom scales (fatigue, pain, and nausea/vomiting), a 2-item global health status/QoL scale, 5 individual item symptom scores (appetite loss, dyspnoea, insomnia, constipation, and diarrhoea), and 1 item on the financial impact of the disease.

TTD in pain, physical functioning and GHS/QoL items are secondary endpoints. All other items are part of the exploratory endpoints.

The number of items and item range for each scale/item are displayed in Table 8.

Table 8 EORTC QLQ-C30 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Global health status/ QoL	QL	2	6	29, 30
Functional scales				

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Physical	PF	5	3	1-5
Role	RF	2	3	6, 7
Cognitive	CF	2	3	20, 25
Emotional	EF	4	3	21-24
Social	SF	2	3	26, 27
Symptom scales				
Fatigue	FA	3	3	10, 12, 18
Pain	PA	2	3	9, 19
Nausea/ vomiting	NV	2	3	14, 15
Symptom items				
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17

QoL=Quality of life.

4.2.8.2 Derivations and Censoring Rules

Scoring algorithm

The EORTC QLQ-C30 v3 is scored according to the EORTC QLQ-C30 Scoring Manual (Fayers, et al., 2001). Items are scored on a 4-point verbal rating scale: “Not at all”, “A little”, “Quite a bit”, and “Very much”. Scores are then transformed to give a score from 0 to 100 for each of the symptom scales, functional scales, and the global QoL scale. Higher scores on the global health status/QoL and functional scales indicate better health status/function, but higher scores on symptom scales/scores represent greater symptom severity.

The EORTC QLQ-C30 functional and symptom scales, individual symptom items and global health status/QoL are derived as follows:

1. Calculate the average of the items that contribute to the scale or take the value of an individual item, i.e. the raw score (RS):

   $RS = (I1 + I2 + \ldots + In) / n,$

   where $I1 + I2 + \ldots + In$ are the items included in a scale and n is the number of items in a scale.

2. Use a linear transformation to standardise the raw score, so that scores range from 0 to 100, where a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

   Functional scales: Score = (1 – [RS – 1] / range) * 100

   Symptom scales/items; global health status/QoL: Score = ([RS – 1] / range) * 100,

   where range is the difference between the maximum and the minimum possible value of RS.

Change from baseline

Changes in score from baseline are calculated for each of the functional scales, symptom scales and global health status/QoL scale at each assessment, where baseline is defined and calculated as explained in Section 3.3.3.

Deterioration

Deterioration is defined as change from baseline that reaches a clinically meaningful deterioration threshold. Anchor-based methods using the participant-based anchors PGIS and PGIC will be considered to define thresholds for clinically meaningful within-participant change used in the time to deterioration (TTD) endpoints. Other methods including distribution-based methods, cumulative distribution function, and probability density function curves, and methods using other anchors may also be considered.

Clinically meaningful change thresholds will be estimated for the following outcomes:

- EORTC QLQ-C30: Global health status/QoL, functioning, and select symptom subscales including pain and fatigue
- EORTC QLQ IL116: breast symptoms, arm symptoms (See Section 4.2.18)

The analysis to define clinically meaningful change thresholds in the TTD PRO endpoints will include all randomised participants using the pooled treatment arms data prior to database lock. Further details on methodologies to define these clinically meaningful change thresholds will be provided in a separate PRO psychometric analysis plan (PAP).

Improvement, deterioration or no change will be defined based on a clinically meaningful change threshold. If the estimation methods have not yet been performed to define the

threshold for improvement, deterioration or no change for a given scale/item score, then a 10-point change from baseline will be used.

TTD

Time to deterioration (TTD) is defined as time from the date of randomisation to the date of first deterioration, regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

Time to deterioration=date of event or censoring – date of randomisation+1

For the derivation of TTD, the following censoring rules are used:

Participants whose symptoms, functioning or global health status/QoL have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom, functioning or global health status/QoL change from baseline could be evaluated. Also, if symptom, functioning or global health status/QoL deteriorate immediately after 2 or more missed PRO assessments, the participant will be censored at the time of the last PRO assessment where the symptom, functioning or global health status/QoL change from baseline could be evaluated prior to the 2 missed assessments.

If a participant has no evaluable post-baseline data or does not have baseline data, they will be censored at date of randomisation.

Participants whose difference between the baseline score and the maximum/minimum subscale score is less than the value of a clinically meaningful change deterioration, which doesn’t allow deterioration to occur, will be censored at date of randomisation.

Given the scheduled assessment scheme (i.e. every three weeks for the first 48 weeks then every six-weeks thereafter) the definition of 2 missed assessments will change. If the previous PRO assessment is less than study day 314 (i.e. week 44) then two missing visits will equate to 46 days since the previous PRO assessment allowing for early and late visits (i.e. 2 x 3 weeks + 2 days for an early assessment + 2 days for a late assessment = 46 days). If the two missed assessments occur over the period when the scheduled frequency of PRO assessments changes from three-weekly to six-weekly this will equate to 67 days (i.e. take the average of 3 and 6 weeks which gives 4.5 weeks and then apply same rationale, hence 2 x 4.5 weeks + 2 days for an early assessment + 2 days for a late assessment = 67 days). The time period for the previous PRO assessment will be from study days 314 to 334 (i.e. week 44 to week 47). From study day 335 (i.e. week 47) onwards (when the scheduling changes to six-weekly assessments), two missing assessments will equate to 88 days (i.e. 2 x 6 weeks + 2 days for an early assessment + 2 days for a late assessment = 88 days). Study

day will be calculated in line with Section 3.3.5. If participant withdraws treatment prior to week 48 the assessment schedule is assumed to be Q6W relative to C1D1. If participant discontinues treatment after week 48 assessment schedule is the same as above.

The following is also summarised in Table 9:

Table 9 Definition of two missed PRO visits

Scheduled Assessment	Previous PRO assessment	Two missed PRO visits window
Q3W	Day 1	2 x 3 weeks + 2 days = 44 days
Q3W up to Week 48*	>Day 1 – Day 313 (i.e. Week 44)	2 x 3 weeks + 4 days = 46 days
	>Day 313 – Day 334 (Week 44 – Week 47) (change period from Q3W to Q6W)	2 x [(3 weeks+6 weeks)/2] + 4 days = 67 days
Q6W thereafter*	>Day 334 onwards	2 x 6 weeks + 4 days = 88 days

*Follow schedule until treatment discontinuation after which a Q6W (relative to C1D1) window will be assumed.

4.2.8.3 Handling of Dropouts and Missing Data

For each subscale, if ≤50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers, et al., 2001). If more than 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing.

4.2.8.4 Primary Analysis of TTD Endpoints as Measured by EORTC QLQ-C30

Statistical analysis

TTD in the pain scale, physical functioning scale and GHS/QoL scale is analysed using the same methodology as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI and p-value are presented. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

Estimates and 95% CI for TTD rates at 3 months intervals and median TTD for each treatment group are presented.

Summaries

Summaries of the number and percentage of participants experiencing a TTD event, and the type of event will be provided for each treatment. The duration of follow-up is summarised

using median time from randomisation to date of censoring in censored participants only, presented by treatment group.

KM plots of the TTD are presented by treatment group.

4.2.8.5 Sensitivity Analysis of TTD Endpoints as Measured by EORTC QLQ-C30

A sensitivity analysis for TTD is performed where TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

This will be analysed using the same methodology described in Section 4.2.8.4.

4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death

4.2.9.1 Definition

Time to first subsequent therapy or death (TFST) is defined as the time from the date of randomisation to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomised treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the initial therapy. In this situation, TFST is calculated as time from randomisation to the start of the initial therapy or death.

4.2.9.2 Derivations and Censoring Rules

Any participant not known to have had a first subsequent anti-cancer therapy will be censored at the last date that the participant was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before first subsequent therapy, these participants will be censored at the earliest of their last known to be alive and termination dates.

Any participant without a TTSCAPRX form (not known to have had a first subsequent anti-cancer therapy) and have not died at the time of the analysis is censored at the last date that the participant was known to be alive according to the rules detailed in Section 4.2.2.2, where any participant recorded as alive or to have died after DCO date is censored at the date of DCO.

4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death

Statistical analysis

The time to first subsequent therapy or death (TFST) is analysed using the same methodology as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI is presented. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

Summaries

In addition, medians and a KM plot of the time to the start of subsequent therapy are presented by treatment group. The time between progression and starting subsequent therapy in participants who have progression per PFS BICR and receive first subsequent therapy is assessed. This is summarised per treatment group, but no formal comparisons are made.

In participants who received a subsequent anti-cancer therapy, a summary table of first subsequent anti-cancer therapies by therapy class and treatment group is provided, as well as response to first subsequent anti-cancer therapy by treatment group.

4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death

4.2.10.1 Definition

Time to second subsequent therapy or death (TSST) is defined as the time from the date of randomisation to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of first subsequent treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the first subsequent anti-cancer therapy and the second alternative cancer therapy they receive is the second subsequent anti-cancer therapy. In this situation, TSST is calculated as time from randomisation to the start of the second subsequent anti-cancer therapy or death.

4.2.10.2 Derivations and Censoring Rules

Any participant not known to have had a second subsequent anti-cancer therapy or have not died at the time of the analysis is censored at the last date that the participant was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before second subsequent therapy, these participants are censored at the earliest of their last known to be alive and termination dates.

4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death

The time to the second subsequent therapy or death (TSST) is analysed using the same methods as that used for the analysis of TFST (see Section 4.2.9.3). The same statistics and summary tables are produced for TSST as for TFST.

4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death

4.2.11.1 Definition

Time from randomisation to second progression or death (PFS2) is defined as the time from date of randomisation to the earliest progression event following first objective progression subsequent to the first subsequent therapy, or death. The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological progression, symptomatic progression, other or death. Second progression status will be reviewed every 3 months following the progression event used for the primary variable PFS (the first progression) and status recorded.

4.2.11.2 Derivations and Censoring Rules

If death occurs within 211 days of first objective progression, or within 211 days of the last evaluable PFS2 assessment, the death will be a PFS2 death event irrespective of whether subsequent therapy has started.

If a participant had a first objective progression which was censored due to 2 missed visits (i.e. was censored for the PFS endpoint), did receive subsequent therapy, and subsequently a second progression was recorded by the investigator, then the participant will be counted as a PFS2 second progression event.

Participants alive and for whom a second disease progression has not been observed are censored at date last known alive and without a second disease progression. Therefore, they are censored at:

- The PFS assessment date if the participant has not had a first progression or death (PFS censoring date).
- The date the participant is last known to not have received a first subsequent therapy if a participant has had a first progression and not started a subsequent therapy (TFST censoring date).
- The latest PFS2 assessment date following first objective progression, if the participant has started a first subsequent therapy and PFS2 event (second progression or death) has not been observed. If a PFS2 assessment has not occurred, then the participant is censored at the day before starting the first subsequent therapy.

Based on two 3-monthly visits plus two allowed 2 week visit windows, a second progression is not evaluable if it was greater than 211 days since last evaluable visit; where the last evaluable visit is the later of the first progression date and any evaluable PFS2 assessment. In addition, if subsequent therapy has not started a PFS2 assessment is not evaluable.

If the participant experiences a second progression that is not evaluable, or dies immediately after two or more consecutive missed visits, the participant is censored at the time of the later of the first progression date and the latest evaluable PFS2 assessment prior to the two missed visits.

For deaths prior to 1st progression, but immediately after two or more consecutive missing visits, the participant is censored at the time of the last evaluable PFS1 assessment.

4.2.11.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.11.2.

4.2.11.4 Primary Analysis of PFS2

Statistical Analysis

PFS2 is analysed using identical methods as outlined for PFS (see Section 4.2.1.4) and adjusting for the same stratification factors. The HR for the treatment effect together with its 95% CI are presented. Medians and KM plots are presented to support the analysis.

Summaries

The number and percentage of participants experiencing a PFS2 event and the type of progression (objective progression by RECIST, symptomatic progression or other) are also summarised by treatment group, as well as summaries of deaths in the absence of second progression, and categories of PFS2 censoring. Time from randomisation to second progression will be summarised by treatment arm.

4.2.12 Secondary Endpoint - Pharmacokinetics

4.2.12.1 Derivations

Pharmacokinetic concentration data will be collected according to Section 8.5.1 of the CSP. The schedule of assessment is as per the schedule of activities (SoA) of the CSP. Whole blood samples for determination of plasma concentration of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a, will be obtained for all participants receiving Dato-DXd.

4.2.12.2 Primary Analyses of Pharmacokinetics

All plasma concentrations will be listed for each participant, for each sampling time and each dosing day, regardless of whether they are excluded from summary statistics due to deviation (e.g. as a result of dose interruption, reduction or missing the dose before PK sample collection, or sampling time deviation, etc).

Plasma concentrations of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a will be summarised by visit and nominal sample time using standard summary statistics for PK

concentrations (geometric mean, geometric coefficient of variation, geometric mean, arithmetic mean, standard deviation, median, minimum, maximum, n and number of concentrations below the Lower Limit of Quantification (LLOQ)) for the Dato-DXd treatment arm.

Individual concentrations below the LLOQ of the bioanalytical assay will be reported as not quantifiable (NQ) in the listings with the LLOQ defined in the footnotes of the relevant tables, figures and listings (TFLs). Individual plasma concentrations that are Not Reportable will be reported as NR and those that are missing will be reported as NS (No Sample) in the listings. For data below limit of quantification (BLQ), NR or NS the following rules will apply:

- Any values reported as NR or NS will be excluded from the summary tables and corresponding figures.
- If, at a given time point, 50% or less of the plasma concentrations are NQ, the geometric mean, CV%, geometric CV%, mean and SD will be calculated treating the NQ as LLOQ.
- If more than 50%, but not all, of the concentrations are NQ, the geometric mean, CV%, geometric CV%, and SD will be reported as data not calculable (NC). The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.
- If all the concentrations are NQ, the geometric mean, mean, minimum, median and maximum will be reported as NQ and the CV%, geometric CV% and SD as NC.

Participants with protocol deviations seriously impacting PK results are excluded from the summary tables.

Population PK, and exploratory exposure response/safety analyses will be performed. This is documented in a separate analysis plan and the results presented separately from the main CSR.

4.2.13 Secondary Endpoint - Immunogenicity

4.2.13.1 Derivations

The presence of ADAs will be assessed in plasma samples taken according to the SoA in the CSP. ADA result from each sample is reported as either positive or negative. If the sample is positive, the ADA titre is reported as well. In addition, the presence of neutralizing antibody (nAb) will be tested for all ADA-positive samples using a ligand-binding assay. The nAb results is reported as positive or negative.

The rules described in Sections 3.3.3, 3.3.5 and 3.3.6 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

Number and percentage of ADA-evaluable participants in the following categories are provided. The number of ADA-evaluable participants in the treatment group will be used as the denominator for percentage calculation.

The ADA categories are defined as follows:

- ADA positive at any visit (at baseline or post-baseline). The percentage of these participants in a population is known as ADA prevalence.
- Treatment-induced ADA positive, defined as ADA positive post-baseline and not detected at baseline (negative or missing).
- Treatment-boosted ADA positive, defined as a baseline positive ADA titre that was boosted to a 4-fold or higher-level following drug administration.
- Treatment-emergent ADA positive (TE-ADA+), defined as either treatment-induced or treatment-boosted ADA. The percentage of these participants in a population is known as ADA incidence.
- Non-treatment-emergent ADA positive (non-TE ADA+), defined as being ADA positive but not fulfilling the conditions for treatment-emergent ADA positive.
- ADA positive post-baseline and positive at baseline.
- ADA not detected post-baseline and positive at baseline.
- Treatment-emergent persistently ADA positive, defined as being TE-ADA+ and having at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurement or an ADA positive result at the last available assessment.
- Those who are classified as treatment-emergent persistently ADA positive based on the second criterion above (last assessment being positive) but not fulfilling the first criterion (based on duration).
- Treatment-emergent transiently ADA positive, defined as being TE-ADA+ and having at least one post-baseline ADA positive measurement but not fulfilling the conditions for persistently positive.
- nAb positive at any visit (at baseline or post-baseline).

4.2.13.2 Primary Analysis of Immunogenicity

A summary will be provided of the number and percentage of participants who develop detectable anti-Dato-DXd antibodies by ADA categories (see Section 4.2.13.1) using the ADA evaluable set. Descriptive statistics (minimum, Q1, median, Q3, and maximum) for the maximum ADA titres of each participant will also be included.

A summary will be provided of the number and percentage of participants who develop detectable anti-Dato-DXd antibodies by visit. Descriptive statistics for ADA titres by visit will also be included.

A summary will be provided of the number and percentage of participants who are ADA positive at a post-baseline assessment for the first time by visit will also be presented. Descriptive statistics for ADA titres by first positive visit will also be included.

Impact of ADA on PK will be explored by presenting plasma Dato-DXd concentration descriptive statistics (Section 4.2.12) in TE-ADA+, non-TE ADA+, ADA-negative, and nAb-positive participants. Spaghetti plots of individual Dato-DXd concentration participants over time profiles of TE ADA+, non-TE ADA+, ADA-negative and nAb-positive participants will be presented.

ADA safety tables will include the number and percentage of participants who had at least 1 AE in any category summarized by ADA status (TE ADA+, non-TE ADA+, ADA negative, and nAb positive).

The effect of ADA on efficacy may be examined by PFS and OS KM plots by ADA status (TE-ADA+ vs non-TE ADA+ vs ADA-negative, and nAb-positive vs nAb-negative), if the data allow (20 events in a subgroup).

Immunogenicity results will be listed for all participants in SAF regardless of ADA-evaluable status. Anti-drug antibody titre and neutralising ADA data will be listed for samples confirmed positive for the presence of anti-Dato-DXd antibodies. AEs in ADA positive participants by ADA positive category will be listed.

4.2.14 Exploratory Endpoint – Patient-Reported Symptomatic AEs and Treatment Tolerability

4.2.14.1 Definition

PRO-CTCAE

The Patient-Reported Outcomes version of the common criteria for adverse events (PRO-CTCAE), a PRO version of the CTCAE system developed by the National Cancer Institute (NCI), is included to evaluate symptomatic toxicity from the participants' perspective.

PRO-CTCAE is an item library of symptoms experienced by participants while undergoing treatment of their cancer. Symptoms have been converted to participant terms (e.g. CTCAE term "myalgia" converted to "aching muscles"). Items capture the presence, frequency, severity and/or interference with daily activities, depending on the AE. For each question,

participants select the value that best describes their experience over the past week, on a 5-point ordinal scale.

The items pre-selected for this study include mouth/throat sores, decreased appetite, nausea, vomiting, constipation, diarrhoea, abdominal pain, shortness of breath, cough, rash, hair loss, hand-foot syndrome, numbness/tingling, and fatigue.

EORTC IL117

The EORTC IL is an online platform comprised of more than 900 individual items from over 60 EORTC questionnaires. The pre-selected items for this study will include dry eyes, mouth pain, and sore mouth (i.e. EORTC IL117).

The recall period is during the past week. Items are scored on a 4-point verbal rating scale: "Not at all", "A little", "Quite a bit", and "Very much".

PGI-TT

The PGI-TT is a single item to assess how a participant perceives the overall tolerability of the IP. The responses indicate how bothered the participant was in the last 7 days by the side effects of their cancer treatment and are scored on a 5-point scale: 1 = Not at all; 2 = A little bit; 3 = Somewhat; 4 = Quite a bit; 5 = Very much.

4.2.14.2 Derivations and Censoring rules

Compliance

Summary measures of compliance over time will be derived for all PRO questionnaires. These will be based upon:

- Expected questionnaire: A questionnaire that is expected to be completed at a scheduled assessment time i.e. a questionnaire from a participant who has not withdrawn from the study at the scheduled assessment time, excluding participants in countries with no available translation and participants who are exempt from PRO completion. Only participants who have started treatment or have a cycle 1 day 1 visit will be included in the compliance summary.
  - For participants that have progressed (RECIST 1.1 progression by Investigator assessment) or discontinued study treatment, the earliest of date of study treatment discontinuation or date of progression will be used to determine the last on treatment windowed assessment for each participants expected forms using the analysis windows as described in Section 3.3.5. If the date falls before the end of the assessment window, then that assessment will only be considered expected if they have a received form. If they have not received a form, then this assessment is

not considered expected as they have not had the full opportunity to complete the questionnaire within the window. For participants who have not discontinued study treatment or progressed, the date of the DCO will be used to determine the last on treatment assessment for their last expected form following the same approach as above.

  - For follow up assessments (EORTC QLQ-C30, EORTC IL116, PGIS and EQ-5D-5L), if a participant has not discontinued study treatment then no follow up forms will be expected. For participants who have discontinued study treatment, and discontinued the study, the earliest of date of study discontinuation or 18 weeks post progression will be used to determine the last expected assessment that a form should have been completed. For participants who have discontinued study treatment, and not discontinued the study, the earliest of date of the DCO or 18 weeks post progression will be used to determine whether a form is expected following the same approach as above. For PRO-CTCAE, PGI-TT, EORTC IL117 and PGIC follow up forms will not be expected as these are not collected during post treatment follow-up.

- Received questionnaire: A questionnaire that has been received and has a completion date and at least 1 individual item completed.
- Evaluable questionnaire: A questionnaire with a completion date and at least 1 subscale that is non-missing.

Compliance over time will be calculated separately for each timepoint, including baseline, as the number of participants with an evaluable questionnaire at the time point, divided by number of participants still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculate separately for each timepoint, including baseline, as the number of evaluable questionnaires, divided by the number of received questionnaires. For compliance over time all timepoints (including follow-up time period) with at least 20 participants in one of the treatment arms will be reported.

Overall compliance will be calculated as the total number of evaluable questionnaires across all timepoint, divided by the total number of questionnaires expected to be received across all timepoints. Similarly, the overall evaluability rate will be calculated as the total number of evaluable questionnaires across all timepoint, divided by the total number of questionnaires received across all timepoints.

4.2.14.3 Handling of Dropouts and Missing data

For PRO-CTCAE only participants in countries where a linguistically validated version of the PRO-CTCAE is available for administration are required to complete this questionnaire and thus included the analysis.

4.2.14.4 Primary Analyses of PRO-CTCAE

PRO-CTCAE data will be summarised descriptively for each symptom by treatment group. The summary will include:

- Number and percentage of participants reporting different levels of responses at each time point for each symptom and attribute
- Number and percentage of the worst response option (defined by the first attribute of each symptom) reported by participants within 12 weeks, excluding baseline
- Number and percentage of participants who report the presence of the symptom (defined by the first attribute of each symptom) at baseline
- Number and percentage of participants who report any worsening from baseline (defined by the first attribute of each symptom) at any time within 12 weeks
- Number and percentage of participants who worsen from a score <4 (on a 1-5 scale) at baseline to a score 4 or 5 (defined by the first attribute of each symptom) at any time within 12 weeks

In addition, a stacked horizontal bar chart showing the percentage of participants with each level of response by timepoint for the first attribute of each PRO-CTCAE symptom is produced. A pie chart showing the percentage of the worst response option (defined by the first attribute of each symptom) reported by participants within 12 weeks, excluding baseline for each PRO-CTCAE item is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm. PRO-CTAE is summarised using the SAF.

4.2.14.5 Primary Analyses of EORTC IL117

EORTC IL117 data will be summarised descriptively for each item by treatment group. The summary will include:

- Number and percentage of the worst response option reported by participants within 12 weeks, excluding baseline
- Number and percentage of participants who report the presence of the symptom at baseline
- Number and percentage of participants who report any worsening from baseline at any time within 12 weeks
- Number and percentage of participants who worsen from a score <3 (on a 1-4 scale) at baseline to a score 3 or 4 at any time within 12 weeks

The number and percentage of participants with each level of response for each EORTC IL117 item at baseline and over time is summarised by treatment group.

In addition, a stacked horizontal bar chart showing the percentage of participants with each level of response by timepoint for each EORTC IL117 item is produced. A pie chart showing the percentage of the worst response option reported by participants within 12 weeks, excluding baseline for each EORTC IL117 item is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm. EORTC IL117 is summarised using the SAF.

4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability

Responses for PGI-TT are summarised descriptively as number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

In addition, a stacked horizontal bar chart showing the percentage of participants in each PGI-TT category by timepoint is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

The PGI-TT is summarised using the SAF.

4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS)

4.2.15.1 Definition

The PGIS is a single item to assess how a participant perceives the overall severity of cancer symptoms over the past week. The responses are scored on a 4-point scale: 1 = None; 2 = Mild; 3 = Moderate; 4 = Severe.

4.2.15.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.15.3 Analysis of Patient Global Impression of Severity

Responses for PGIS is summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group. Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIS will be based on the FAS.

4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC)

4.2.16.1 Definition

The PGIC is a single item to assess how a participant perceives the overall change in health status since the start of IP. The responses are scored on a 7-point scale: 1 = Much better; 2

= Moderately better; 3 = A little better; 4 = About the same; 5 = A little worse; 6 = Moderately worse; 7 = Much worse.

4.2.16.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.16.3 Analysis of Patient Global Impression of Change

Responses for PGIC are summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIC will be based on the FAS.

4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL

4.2.17.1 Definition

Refer to Section 4.2.8.1 for the relevant definitions of EORTC QLQ-C30.

4.2.17.2 Derivation and Censoring Rules

Refer to Section 4.2.8.2 for the derivation and censoring rules of scoring algorithm, change from baseline, deterioration and TTD. Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.17.3 Handling of Dropouts and Missing Data

Refer to Section 4.2.8.3 for the handling of missing data for EORTC QLQ-C30.

4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL

TTD of function (including role, cognitive, emotional and social), multi-term symptoms (including fatigue and nausea/vomiting), and single items (dyspnoea, insomnia, appetite loss, constipation and diarrhoea) will be analysed using the same methods as that used for the analysis of TTD in the pain scale, physical functioning scale and GHS/QoL scale (see Section 4.2.8.4). The same statistics and summary tables are produced.

Change from baseline in subscales of the EORTC QLQ-C30 is analysed using a mixed model for repeated measures (MMRM) of the change from baseline. Participants are included in the mean change from baseline analysis if they have an evaluable baseline assessment and at least one evaluable post-baseline assessment. The model includes

treatment, visit, and treatment-by-visit interaction as explanatory variables and the baseline score and the baseline score by visit interaction as covariates.

When less than 20 participants are present at a timepoint in either arm, this timepoint should be excluded from the analysis. An unstructured covariance matrix is used to model the within-participant error and the Kenward-Roger approximation is 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: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, and autoregressive. Adjusted mean change from baseline estimates per treatment group and corresponding 95% CIs are presented along with an overall estimate of the treatment difference, 95% CI, and p-value.

All analyses will have a corresponding graphical plot showing the adjusted mean change from baseline and 95% CI over time.

Summary tables of assessment responses (improvement, deterioration, and no change), absolute scores and change from baseline for each EORTC QLQ-C30 scale/item score (global health status/QoL, 5 functions, and all symptoms [fatigue, pain, nausea/vomiting, dyspnoea, insomnia, appetite loss, constipation and diarrhoea]) and for each timepoint will be presented by treatment group. Additionally, for each EORTC QLQ-C30 scale/item graphical plots of the mean absolute score and change from baseline along with the associated 95% CI for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

Reason participant did not complete questionnaire will be summarised over time. This summary will not be split by PRO questionnaire.

These analyses will be based on the FAS.

4.2.18 Exploratory Endpoint – Breast and Arm Symptoms

4.2.18.1 Definition

The EORTC QLQ-BR45 is a breast-cancer-specific module from the EORTC comprising 45 questions to assess breast cancer symptoms. The module includes 5 functional scales/items (body image, future perspective, sexual functioning, sexual enjoyment, breast satisfaction) and 7 symptom scales/items (systemic therapy side effects, upset by hair loss, arm symptoms, breast symptoms, endocrine therapy symptoms, skin mucosis symptoms and endocrine sexual symptoms).

The current study will only include the breast symptoms and arm symptoms scales (7 items) from the BR45, i.e. EORTC IL116.

4.2.18.2 Derivations and Censoring Rules

The EORTC IL116 is scored as described in Table 10 to give a score from 0 to 100 for each of the symptom scales and items.

Table 10 EORTC IL116 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Symptom scales/items				
Arm Symptoms	ARM	3	3	31 - 33
Breast Symptoms	BR	4	3	34 - 37

For each multi-item scale, the average of the corresponding items is calculated. The raw score is then standardised to a 0 - 100 range as for the EORTC QLQ-C30, and missing values are handled in the same way.

The scoring approach for the EORTC IL116 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30. Similarly, to the symptom scales of the EORTC QLQ-C30, higher scores represent greater symptom severity.

The definition of a clinically meaningful change and time to deterioration for EORTC IL116 is the same as that for the EORTC QLQ-C30 described in Section 4.2.8.2. Censoring rules are also the same as described for EORTC QLQ-C30.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.18.3 Handling of Dropouts and Missing Data

Missing data for EORTC IL116 is handled in the same way as EORTC QLQ-C30 (Section 4.2.8.3).

4.2.18.4 Primary Analysis of Breast and Arm Symptoms

TTD in the breast and arm symptom scales will be analysed using the same methods as that used for the analysis of TTD in the pain scale, physical functioning scale and GHS/QoL scale (see Section 4.2.8.4). The same statistics and summary tables are produced.

Change from baseline in subscales of the EORTC IL116 is analysed using a MMRM of the change from baseline as described in Section 4.2.17.4.

Summary tables of assessment responses (improvement, deterioration, and no change), absolute scores and change from baseline for each EORTC IL116 scale score (breast symptoms, arm symptoms) and for each timepoint will be presented by treatment group. Additionally, for each EORTC IL116 scale graphical plots of the mean absolute score and change from baseline along with the associated 95% CI for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.19 Exploratory Endpoint - EQ-5D-5L

4.2.19.1 Definition

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. The EQ-5D-5L questionnaire comprises 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 a possible 3,125 health states. For example, state 11111 indicates no problems on any of the five dimensions. Respondents also assess their health today using the EuroQoL-Visual analogue scale (EQ-VAS), which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

Improvement, deterioration or no change of the EQ-VAS score will be defined using a 7-point change from baseline and a 10-point change from baseline.

4.2.19.2 Derivations

The EQ-5D profile is converted into a weighted health state utility value, termed the EQ-5D index using the UK crosswalk algorithm developed by Hernandez Avala (Hernandez Alava, 2020). The UK algorithm is applied to all participants. The EQ-VAS is reported separately.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.19.3 Primary Analysis of EQ-5D-5L

Descriptive statistics will be calculated for each scheduled time point in the study, for each trial arm and as a total. These will report the number of participants, the number of EQ-5D questionnaires completed at each timepoint, the number and proportion responding to each

dimension of the EQ-5D-5L. Additionally, summary statistics (e.g. n, mean, median, standard deviation, min, max) will be reported for the EQ-5D index score and the EQ-VAS score, and the change from baseline for the EQ-5D index score and the EQ-VAS score.

Graphical plots of the mean EQ-5D index score and EQ-VAS score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study are produced.

For EQ-VAS a summary table of assessment responses (improvement, deterioration, and no change) for each timepoint will be presented by treatment group using a 7-point change from baseline and repeated using a 10-point change from baseline.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.19.4 Supplementary Analyses of EQ-5D-5L

To support submissions to payers, additional analyses may be undertaken, and these are outlined in a separate Payer Analysis Plan (PAP).

4.2.20 Exploratory Endpoint - Biomarkers

The relationship of baseline TROP2 expression, tumour mutational profiling and gene expression profiling and, if applicable, of exploratory biomarkers to clinical outcomes (including but not restricted to) of BOR, DoR, PFS and OS may be presented. Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report.

4.2.21 Exploratory Endpoint - Health Care Resource Use

4.2.21.1 Definition

To investigate the impact of treatment and disease on health care resource of non-study protocol related events, the following variables are captured:

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

4.2.21.2 Derivations

Where admitted overnight, the length of hospital stay is calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation + 1).

If there are multiple hospital stays for the same participant, then the length of hospital stay is summed across all hospitalisation admissions for the participant.

Participants with missing discharge dates are calculated as the difference between the last day with available data and the start date of hospitalisation + 1. The length of ICU/HDU stay is calculated using the same method.

4.2.21.3 Primary Analysis of Health Care Resource Use

Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) are provided for each treatment group on the different types of hospital admissions, the length of stay for participants admitted to hospital for at least one overnight stay and length of stay for participants admitted to intensive care / high dependency units, as well as the primary sign or symptom the participant presents with.

Where a participant has admissions for different signs and symptoms, they are included in each category when summarising type of hospital admissions.

This analysis will be done on the SAF.

4.3 Pharmacodynamic Endpoint(s)

Not Applicable.

4.4 Safety Analyses

The domain safety covers exposure, adverse events, clinical laboratory, vital signs, physical examination, ECG, Echocardiogram, ECOG performance score and Ophthalmologic assessments.

Tables are provided for the safety set; listings are provided for the safety set.

4.4.1 Exposure

4.4.1.1 Definitions and Derivations

Treatment exposure for Dato-DXd

Dato-DXd is dosed 6 mg/kg intravenously on Day 1 of each 21-day cycle (Q3W). The dose of Dato-DXd may be reduced once to 4 mg/kg intravenous (IV) Q3W and a further second

reduction to 3 mg/kg IV Q3W is allowed per participant on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Dato-DXd (months): =

(min (last date where dose > 0 + 20, date of death, date of DCO) – first dose date +1) / (365.25/12)

Actual exposure of Dato-DXd =

total exposure – total duration of dose interruptions,

where the total duration of dose interruption is defined as any length of time when the participant has not taken any of the planned doses. Dose interruptions include missed and delayed doses.

The calculation of actual exposure makes no adjustment for any dose reductions that may have occurred and will only be calculated for Dato-DXd (not for chemotherapy arms).

Treatment exposure for Capecitabine

Capecitabine is scheduled to be dosed 1000 or 1250 mg/m^2 twice daily (BID) on Days 1 to 14 of a 21-day cycle. The choice of dose will be determined by standard institutional practice and the starting dose will be assumed to be the planned dose. The dose of Capecitabine may be reduced by 25% in participants with moderate renal impairment on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Capecitabine (months): =

(min (last Capecitabine dose date where dose > 0, date of death, date of DCO) – first Capecitabine dose date +1) / (365.25/12)

Treatment exposure for Gemcitabine

Gemcitabine is dosed 1000 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Gemcitabine (months): =

(min (last Gemcitabine dose date where dose > 0 + W, date of death, date of DCO) – first Gemcitabine dose date+1) / (365.25/12),

where W=6 if the last Gemcitabine dose was scheduled on Day 1 and W=13 if the last Gemcitabine dose was scheduled on Day 8.

Treatment exposure for Vinorelbine

Vinorelbine is dosed 25 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Vinorelbine (months): =

(min (last Vinorelbine dose date where dose > 0 + W, date of death, date of DCO) – first Vinorelbine dose date +1) / (365.25/12),

where W=6 if the last Vinorelbine dose was scheduled on Day 1 and W=13 if the last Vinorelbine dose was scheduled on Day 8.

Treatment exposure for Eribulin mesylate

Eribulin mesylate is dosed 1.4 mg/m^2 on Days 1 and 8 of a 21-day cycle. A lower starting dose of 1.1 mg/m^2 is recommended for participants with moderate renal impairment. The starting dose will be assumed to be the planned dose. The calculation of exposure is as follows:

Total (or intended) exposure of Eribulin mesylate (months): =

(min (last Eribulin mesylate dose date where dose > 0 + W, date of death, date of DCO) – first Eribulin mesylate dose date +1) / (365.25/12),

where W=6 if the last Eribulin mesylate dose was scheduled on Day 1 and W=13 if the last Eribulin mesylate dose was scheduled on Day 8.

Participants who permanently discontinue during a dose interruption

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

If a participant permanently discontinues study treatment during a dose interruption, then this is not counted as a dose interruption for summary purposes.

Number of treatment cycles received

Exposure is also measured by the number of cycles received. A cycle corresponds to a period of 21 days. If a cycle is prolonged due to toxicity, this is still counted as one cycle. A cycle is counted if any treatment during that cycle is taken.

Safety Follow-up

Total duration of safety follow-up is calculated as:

Total Safety Follow-up (months) = [min (date of safety follow-up assessment, last dose of IP date + 20, date of study discontinuation, date of death, DCO date) – first dose date +1] / (365.25/12)

Dose intensity

Dose intensity is derived for study treatments Dato-DXd, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine.

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. RDI is 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 or the actual last day of dosing. D is the total dose that would be delivered if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 2 days should be added to reflect the protocol allowed window for dosing as shown below.

Dose intensity – Intended cumulative dose

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

The intended dose for Dato-DXd is 6mg/kg on Day 1 (+/-2 days) of each 21-day cycle. The minimum of the participants last dose, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended.

For the calculations below,

DUR = min (date of last dose date where dose > 0, date of death, date of DCO) – first dose date + 1

The intended dose for Dato-DXd is then calculated as:

6 * [integer ((DUR +2)/ 21) + 1]

Similarly, for Eribulin mesylate (dose=1.4 mg/m^2), Vinorelbine (dose=25 mg/m^2) and Gemcitabine (dose=1000 mg/m^2), intended dose will be calculated as:

2* dose * [integer ((DUR +2)/ 21) + 1]

Note: if the starting dose is reduced by standard institutional practice per protocol, that starting dose will be used for the participant.

For Capecitabine (dose of 1000 or 1250 mg/m^2 BID for 14 days per 21-day cycle, intended dose is given by:

2 * 14 * dose * [integer ((DUR+2) / 21)) + 1]

Dose intensity – Actual cumulative dose

For the calculation of actual cumulative dose for Dato-DXd, Gemcitabine, Eribulin mesylate and Vinorelbine, the proportion of volume left after the infusion will be used to calculate how much of the study drug the participant received, i.e.:

- Volume left (proportion) = $\frac{\text{Volume after infusion}}{\text{Volume before infusion}}$
- Actual cumulative dose = sum over all cycles [(1 – Volume left) x dose] (where dose is taken from the exposure CRF page for each cycle)

For the calculation of actual dose for Capecitabine, drug accountability data (of 1000mg tablets and 1250mg tablets) will be used as follows:

Actual cumulative dose = sum dose (mg) dispensed – sum dose (mg) returned

Percentage Intended Dose

Percentage intended dose (PID) is calculated in the same way as RDI, but D is the intended cumulative dose up to the date of progression or study discontinuation instead of date of last dose where dose >0. The minimum of the participant's date of progression or study discontinuation, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended. For the calculations above,

DUR = min (date of date of progression, study discontinuation, date of death, date of DCO) – first dose date + 1.

Similarly, the actual cumulative dose is the cumulative dose taken up to min (date of progression, date of study discontinuation, date of death, date of DCO).

4.4.1.2 Presentation

The following summaries are produced for Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine:

- Total exposure
- Actual exposure (Dato-DXd only)
- RDI and PID.
- Summary of drug interruptions, dose delays and dose reductions. Number and percentage of participants with any interruptions, number and percentage of participants with any dose delays and number and percentage of participants with any dose reductions are summarised. This is broken down by number of interruptions/dose delays/dose reductions, and reason for interruption/dose delays/dose reduction. In addition, the number and percentage of participants with both an interruption and a dose reduction are summarised.

In addition, the number of cycles received, and duration of safety follow-up are also summarised.

4.4.2 Adverse Events

4.4.2.1 Definitions and Derivations

Adverse events (AEs) and serious adverse events (SAEs) are collected from the time of signature of the informed consent form (ICF), throughout the treatment period and until the 28-Day (+7 days) follow-up period is completed. AE causality is as determined by the reporting investigator. For interstitial lung disease (ILD)/pneumonitis, safety follow up will be continued until resolution of ILD/pneumonitis.

Treatment emergent adverse events (TEAEs) are defined as those adverse events (AEs) with onset or that worsen (by investigator report of an increase in CTCAE grade relative to pre-treatment) on or after the first dose of IP and on or before the date of last IP + 28 days (+7 days) and prior to the start of any subsequent anti-cancer therapy.

For the subset of TEAEs, with onset prior to start of study treatment, and which worsened in severity or seriousness after initiating study treatment until 28 days (+7 days) after last dose of study treatment, such worsening should also occur prior to initiation of any first subsequent anti-cancer therapy to be included in the AE summary tables.

Pre-treatment AEs are those which occur before the first dose of IP and do not worsen during the treatment period.

If no onset time is given, and the date of onset of the AE is the same as the date of first dose of IP, then the AE is assumed to have occurred after the first dose of IP.

The medical dictionary for regulatory activities (MedDRA) [using the latest MedDRA version] is used to code AEs. AEs are graded according to the National Cancer Institute (NCI) common terminology criteria for adverse event (CTCAE) version 5.0.

Missing start and stop dates for AEs are handled using the rules described in Section 3.3.7. AEs that have missing causality (after data querying) are assumed to be related to the treatment where causality is missing.

Dose modification describes an AE where action taken is either dose reduced, or drug interrupted.

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to understanding of the AZ DS-1062 (Dato-DXd) safety profile and require close monitoring and rapid communication by the Investigator to the Sponsor. All AESIs, regardless of severity or seriousness, must be followed until either event resolution, end of study, trial termination, withdrawal of consent, or participant death.

The AESIs that are collected during this study are: (ILD)/pneumonitis, infusion-related reactions, oral mucositis/stomatitis, mucosal inflammation other than oral mucositis/stomatitis, and ocular surface toxicity.

Preferred terms used to identify AESI will be listed before data base lock (DBL) and documented in the Trial Master File. 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.

The duration of an AESI is calculated as (Stop date of AE - Start date of AE) + 1. The duration of the AESI is not calculated for AESIs which are ongoing. If a participant has multiple AESIs within a category, then the duration is summed for all AESIs within the category but any days where AESIs overlap are only counted once.

Time to first AESI is calculated as:

(Start date of first AESI - Date of IP first dose) + 1.

4.4.2.2 Presentation

All AEs are summarised descriptively by count (n) and percentage (%) for each treatment group.

Unless otherwise stated, only AEs defined as treatment emergent (see Section 4.4.2.1) are included in the summary tables.

All reported AEs (including pre- and post-treatment AEs) are listed including the date of onset, date of resolution (if AE is resolved) and investigator's assessment of CTCAE grade and relationship to IP. An overall summary of the number and percentage of participants in each category below is presented by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine. Each AE category is separately summarised by SOC and PT:

- All AEs
- All AEs possibly related to IP
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to IP
- AEs with outcome of death
- AEs with outcome of death possibly related to IP
- All SAEs
- All SAEs possibly related to IP
- All SAEs with CTCAE grade 3 or higher
- All SAEs with CTCAE grade 3 or higher, possibly related to IP
- AEs leading to discontinuation of IP
- AEs leading to discontinuation of IP, possibly related to IP
- All SAEs leading to discontinuation of IP
- All SAEs leading to discontinuation of IP possibly related to IP
- AEs leading to interruption of IP
- AEs leading to dose reduction of IP
- AEs leading to dose modification of IP

Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

The following AEs that occur in at least 1% of participants in any treatment group, are summarised by PT, by decreasing frequency based on the total number of AEs in the Dato-DXd arm:

- All AEs possibly related to IP
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to IP
- All SAEs

The following AE categories are summarised by PT, by decreasing frequency based on the total number of AEs in the Dato-DXd arm:

- AEs leading to interruption of IP
- AEs leading to dose reduction of IP
- AEs leading to dose modification of IP
- All SAEs possibly related to IP
- All SAEs with CTCAE grade 3 or higher
- AEs leading to discontinuation of IP

Additionally, the most common AEs, which are those AEs that occur in at least 5% (where no rounding is applied i.e. an AE with frequency 4.9% does not appear if the cut-off is 5%) of participants in any treatment group, are summarised by PT, by decreasing frequency based on the total number of AEs across treatment groups. This cut-off may be modified after review of the data. This is repeated by decreasing frequency based on the total number of AEs in the Dato-DXd arm.

An additional summary of AEs with onset date after the date of last IP and less than or equal to 35 days after the date of last IP and before the onset of subsequent cancer therapy is produced by SOC, PT and treatment group.

Further details of summaries by SOC and PT are given below if a participant experienced more than one TEAE:

- The participant will be counted once for each SOC and once for each PT.
- The participant will be counted once for each SOC and once for each PT at the maximum CTCAE grade.
- The participant will be counted once for each SOC and once for each PT using the possibly related to IP (as assessed by the investigator) event
- The participant will be counted once for each SOC and once for each PT for possibly related to IP (as assessed by the investigator) events at the maximum CTCAE grade.

AEs are assigned CTCAE grades and summaries of the number and percentage of participants are provided by maximum reported CTCAE grade, SOC and PT.

An overall summary of the number and percentage of participants in each category below is presented by Dato-DXd and total ICC, by AESI category:

- All AESIs possibly related to IP
- AESIs with CTCAE grade 3 or higher
- All serious AESIs
- AESIs with outcome of recovered/resolved
- AESIs with outcome of not recovered/not resolved
- AESIs leading to interruption of IP
- AESIs leading to dose reduction of IP
- AESIs leading to discontinuation of IP

Summary tables of AESIs overall and by maximum CTCAE grade are produced, by AESI category and PT. The preferred terms for AESIs are presented in a listing.

Tables are also produced of AESIs by outcome and AESIs with outcome of recovered/resolved by time of resolution, and by action taken. The time to onset of first AESI and duration of AESI are summarised. In addition, summary tables are produced of number of participants with AESIs possibly related to IP and leading to discontinuation of IP. AESIs are summarised by AESI category, and PT.

A summary table of AESIs by maximum CTCAE grade is produced, by SOC and PT.

All AEs and AESIs are listed, and the time to onset of the AE from date of first dose is presented in the listing. Key participant information is provided in 4 separate listings for all SAEs, AEs with an outcome of death, all AEs leading to dose modification and all AEs leading to treatment discontinuation.

Deaths

A separate summary of deaths is provided with number and percentage of participants, categorised as:

- Total number of deaths (regardless of date of death)
- Related to disease under investigation

- AE with outcome of death only and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- AE with outcome of death only and onset date > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first)
- Death with primary or secondary reason related to disease under investigation and AE with outcome of death and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- Death with primary or secondary reason related to disease under investigation and AE with outcome of death > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy, whichever occurs first
- Deaths <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Deaths > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Participants with unknown reason for death
- Other deaths

A corresponding listing is also produced.

4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest

Summaries of ILD/pneumonitis events will be primarily based on adjudicated drug related ILD/pneumonitis events from the ILD adjudication committee. Supportive summaries based on AESI-defined ILD/pneumonitis cases (i.e. identified based on pre-defined MedDRA preferred terms) will also be provided.

When summarising time to and duration of first treatment-emergent AESI for adjudicated ILD/pneumonitis events, only adjudicated drug-related events will be considered.

A listing of ILD/pneumonitis events is produced.

4.4.3 Clinical Laboratory, Blood Sample

4.4.3.1 Definitions and Derivations

Blood samples for determination of clinical chemistry and haematology are collected as described in the schedule of activities (SoA) of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records and missing data are followed.

Change from baseline in haematology and clinical chemistry variables are calculated for each post-dose visit on treatment.

CTCAE grades are defined at each visit according to the CTCAE v5.0 grade criteria using local ranges (or project ranges if local ranges are missing), after conversion of lab result to corresponding AstraZeneca (AZ) preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, sodium, magnesium and corrected calcium. For these parameters high and low CTCAE grades are calculated.

Corrected calcium product is 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 are compared to the local reference ranges (or project ranges if local ranges are missing) and classified as low (below range), normal (within range or limits of range) and high (above range).

For parameters with no CTCAE grading that are listed in the CSP any increase/decrease/treatment emergent laboratory change (TELC) is derived, where any increase is an increase to a value above the upper local laboratory reference limit (or project ranges if local ranges are missing) at any time on treatment for participants with a value below the upper local laboratory reference limit (or project ranges if local ranges are missing) at baseline, and any decrease is a decrease to any value below the local laboratory reference range limit (or project ranges if local ranges are missing) at any time on treatment for participants with a value above the lower local laboratory reference limit (or project ranges if local ranges are missing) at baseline. A TELC is defined as any on treatment increase or decrease from baseline.

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

Local reference ranges (or project ranges if local ranges are missing) are used for the primary interpretation of laboratory data.

4.4.3.2 Presentations

Only laboratory data that is on treatment as defined in Section 3.3.4 is included in the summary tables.

Data summaries and listings are provided by AZ preferred units.

Laboratory listings will cover observed values, changes from baseline and CTCAE grade for each individual participant as well as abnormalities. Flags are applied to values falling outside reference ranges and for the CTCAE grade for parameters for which CTCAE grading applies.

For all continuous clinical chemistry and hematology laboratory assessments, absolute value and change from baseline are summarised using descriptive statistics at each scheduled visit.

Shift tables of laboratory values by worst common toxicity criteria (CTCAE) grade on treatment are produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change are produced. Percentages are based on the number of participants with a baseline value and an on-treatment value.

The laboratory parameters for which CTCAE grade shift outputs are produced are:

- Haematology: Haemoglobin, Leukocytes, Lymphocytes (absolute count), Neutrophils (absolute count), Platelets
- Clinical Chemistry: Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Albumin, Alkaline Phosphatase (ALP), Total bilirubin, Magnesium (hypo- and hyper-), Sodium (hypo- and hyper-), Potassium (hypo- and hyper), Corrected Calcium (hypo- and hyper-), Creatinine

For parameters with no CTCAE grading, the number and percentage of participants with any on treatment increase from baseline, any on treatment decrease from baseline and a TELC is summarised. Percentages are based on the number of participants with a baseline value below/above the local laboratory upper/lower reference limit (or project ranges if local ranges are missing) and an on-treatment value for the any increase/decrease summaries respectively. Percentages for a TELC are based on the number of participants with a baseline value and an on-treatment value.

For parameters with no CTCAE grading, shift tables from baseline to worst value on-treatment are provided.

Hy's law (HL)

A summary table is produced showing the number (%) of participants who have:

- Elevated ALT, AST, and Total bilirubin during the study
- ALT ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN (Upper limit of normal) during the study.
- AST ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN during the study.

- Total bilirubin ≥1.5x - ≤2x and >2x ULN during the study.
- ALT or AST ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN during the study.
- ALT or AST ≥3x ULN together with total bilirubin ≥2x ULN, irrespective of ALP, at any point during the study (potential Hy's law) following the start of treatment: the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation*. Exceptions include participants with elevated liver enzymes (ALT or AST ≥5x ULN) that have liver metastases present at baseline. These participants have been permitted in the study as per inclusion criterion number 8 in the CSP.
- ALP ≥1.5x - ≤3x and >3x ULN during the study.

* The ALT or AST elevation occurring less than or equal to 28 days prior to the total bilirubin elevation.

Narratives are provided in the CSR for participants with potential Hy's law.

Liver biochemistry test results over time for participants with potential Hy's law are plotted and listed.

Plots of maximum post-baseline ALT and AST vs. maximum post-baseline total bilirubin, expressed as multiples of ULN, are also produced with reference lines at 3×ULN for ALT and AST, and 2×ULN for Total bilirubin. In each plot, total bilirubin is in the vertical axis.

4.4.4 Clinical Laboratory, Urinalysis

4.4.4.1 Definitions and Derivations

Urine samples for determination of urinalysis are collected as described in the SoA of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.4.2 Presentations

A shift table is produced showing the number and percentage of participants in each category at baseline and the maximum on treatment value for each parameter.

The denominator used only includes participants with a baseline value and at least one on treatment value.

On treatment is defined in Section 3.3.4.

Supportive laboratory listings will cover observed values for each individual participant as well as abnormalities.

4.4.5 Other Laboratory Evaluations

4.4.5.1 Definitions and Derivations

Pregnancy tests (serum at screening and urine at other timepoints) are performed for women of childbearing potential.

In addition, hepatitis B surface antigen, hepatitis C and human immunodeficiency virus (HIV) antibodies is assessed at screening.

4.4.5.2 Presentations

This data is listed only, no summary tables are produced.

4.4.6 Vital Signs

4.4.6.1 Definitions and Derivations

Vital signs are assessed at timelines as specified in the SoA of the CSP. The following vital signs are measured: systolic and diastolic blood pressure, pulse rate, body temperature and respiratory rate. Body weight is also collected. Height is collected at screening only.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.6.2 Presentations

Summaries for vital signs data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values and change from baseline for diastolic and systolic blood pressure, pulse, respiratory rate, temperature and weight are summarised over time at each scheduled visit for each treatment group.

Vital signs data is also listed.

4.4.7 Electrocardiogram

4.4.7.1 Definitions and Derivations

Resting 12-lead electrocardiograms (ECGs) are recorded at timepoints specified in the SoA of the CSP.

The following ECG variables are collected: ECG heart rate, PR duration, QRS duration, QT interval, QTcF interval, RR duration and overall ECG evaluation.

The rules described in Sections 3.3.3, 3.3.6 and 3.3.7 of this document considering definition of baseline and visit windows are followed. ECGs will be presented using nominal visits.

The overall evaluation of an ECG is either "normal", "borderline" or "abnormal" with abnormalities categorised as either "clinically significant" or "not clinically significant". Any clinically significant ECG abnormalities require triplicate ECG results. Where triplicate ECG results are taken, a single mean value for numeric parameters is used, and the worst case for the overall evaluation is used. If there are missing value(s) in triplicates, the other non-missing value(s) in the triplicates will be used to calculate the average. Unscheduled assessments will not be included in the calculation of the average.

The QT interval corrected for heart rate using Fridericia's correction (QTcF) is calculated as follows (where QT and RR are in seconds):

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

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:

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

4.4.7.2 Presentations

Summaries for ECG data include only on treatment data. On treatment is defined in Section 3.3.4.

The following summaries for QTcF are included:

Absolute QTcF interval prolongation at any time on treatment:

- QTcF interval > 450 milliseconds
- QTcF interval > 480 milliseconds
- QTcF interval > 500 milliseconds

Change from baseline in QTcF interval at any time on treatment:

- QTcF interval increases from baseline > 30 milliseconds
- QTcF interval increase from baseline > 60 milliseconds
- QTcF interval >450 milliseconds and change from baseline > 30 milliseconds
- QTcF interval >500 milliseconds and change from baseline > 60 milliseconds

A listing is provided of ECG data.

4.4.8 Echocardiogram/Multigated Acquisition Scan

4.4.8.1 Definitions and Derivations

An echocardiogram (ECHO) or multigated acquisition (MUGA) scan to assess left ventricular ejection fraction (LVEF) is performed at the visits as shown in the SoA of the CSP.

The modality of the cardiac function assessments must be consistent for a given participant, i.e. if an ECHO scan is used for the screening assessment, then ECHO should also be used for subsequent scans. The participants should also be examined using the same machine and operator whenever possible, and quantitative measurements should be taken.

The rules described in Sections 3.3.3 and 3.3.7 of this document considering definition of baseline are followed. LVEF will be presented using nominal visits.

4.4.8.2 Presentations

Summaries for LVEF data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values at baseline for LVEF results are summarised.

4.4.9 Eastern Cooperative Oncology Group Performance Status

4.4.9.1 Definitions and Derivations

An assessment of ECOG performance status score is performed at the visits as shown in the SoA of the CSP.

The ECOG performance status scores range from 0 to 5, with lower scores indicating greater participant activity:

0. Fully active; able to carry out all usual activities without restrictions

1. Restricted in strenuous activity, but ambulatory and able to carry out light work or work of a sedentary nature (e.g. light housework or office work)

2. Ambulatory and capable of self-care, but unable to carry out any work activities; up and about more than 50% of waking hours

3. Capable of only limited self-care; confined to bed or chair more than 50% of waking hours

4. Completely disabled; unable to carry out any self-care and totally confined to bed or chair

5. Dead

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.9.2 Presentations

Summaries for ECOG data include only on treatment data. On treatment is defined in Section 3.3.4.The number and percentage of participants in each category is summarised at each visit.

4.4.10 Physical Examination

4.4.10.1 Definitions and Derivations

Physical examination, as well as assessment of height and weight, will be performed according to the Schedule of Assessments (SoA) in the CSP.

A full physical examination will be performed at screening which includes assessment of general appearance, respiratory, cardiovascular, abdomen, skin, head and neck (including ears, eyes, nose and throat), oral (mouth), lymph nodes, thyroid, musculoskeletal (including spine and extremities), urogenital, dermatological, gastrointestinal, endocrine, hematologic/lymphatic, and neurological systems. At subsequent visits, targeted physical examinations are to be utilised by the Investigator on the basis of clinical observations and symptomatology. A targeted physical examination includes at a minimum, assessments of the skin, lungs, oral, cardiovascular system, and abdomen (liver and spleen).

4.4.10.2 Presentations

Individual physical examination data will not be summarised.

4.4.11 Ophthalmologic Assessments

4.4.11.1 Definition and Derivations

Ophthalmologic assessments by a licensed eye care provider will be performed as specified in the SoA in the CSP.

The following assessments will be performed for both eyes: daily use of artificial tears, avoidance of contact lenses, eye-related symptoms, best corrected visual acuity (BCVA), corneal sensation, eyelid position, eyelid margins, conjunctiva, slit-lamp examination, anterior chamber cells, fluorescein staining, tear film breakup time (TFBT), limbal stem cell deficiency (LSCD), Oxford grade of punctate epithelial erosions, proparacaine test, intraocular pressure, tear film meniscus, fundoscopy, clinically significant corneal disease with use of both CTCAE and Corneal Toxicity Severity Grading scales, and other diagnosis and treatment(s) prescribed, if any.

The preferred method for measuring BCVA is the Snellen chart (metric in meter, or imperial in feet). The Snellen chart values will be converted to LogMAR values as

illustrated in Table 11. LogMAR values are calculated by taking the $\log_{10}$ of the reciprocal of the Snellen fraction. For example, if the Snellen fraction is 20/50, the LogMAR value is $\log_{10}$ (50/20) = 0.4.

Table 11 Best Corrected Visual Acuity conversion table

Snellen Chart		Snellen Fraction (Decimal)	LogMAR
Feet	Meter		
20/200	6/60	0.10	1.0
20/160	6/48	0.125	0.9
20/125	6/38	0.16	0.8
20/100	6/30	0.20	0.7
20/80	6/24	0.25	0.6
20/63	6/19	0.32	0.5
20/50	6/15	0.40	0.4
20/40	6/12	0.50	0.3
20/32	6/9.5	0.63	0.2
20/25	6/7.5	0.80	0.1
20/20	6/6	1.00	0.0
20/16	6/4.8	1.25	-0.1
20/12.5	6/3.8	1.60	-0.2
20/10	6/3	2.00	-0.3

4.4.11.2 Presentations

Ophthalmologic assessments will be summarised at DCO1 (see Section 3.1.1).

For daily use of artificial tears and BCVA (LogMAR), summary statistics (e.g. n, mean, median, standard deviation, min, max) for the observed value and the change from baseline value will be reported by visit and by treatment group.

For tear film breakup time, use of contact lenses, eye pain, fluorescein staining of cornea, abnormality results in slit lamp examination with attention to cornea, punctate epithelial erosions in Oxford grade, clinically significant corneal disease by the revised CTCAE grade and clinically significant corneal disease severity grade, the number and percentage of participants in each category will be reported by visit and treatment group.

For intraocular pressure, summary statistics (e.g. n, mean, median, standard deviation, min, max) for the change from baseline value will be reported by visit and by treatment group.

For corneal sensation, eye lid position, eye lid margins abnormalities, tear film meniscus, LSCD with and without fluorescein stain, and dilated fundoscopic exam results, shift from baseline will be reported by visit and by treatment group. Percentages will be based on the number of participants with a baseline result and at least one result for the corresponding visit.

For conjunctiva abnormality and anterior chamber (cell) abnormality, baseline to worst result post-baseline shift tables will be reported. Percentages will be based on the number of participants with a baseline result and at least one post baseline result.

Eye-related adverse events will be summarised by preferred term (PT). The following PTs will be displayed as separate categories:

1. Keratitis (including Ulcerative keratitis, Corneal perforation)
2. Limbal stem cell deficiency
3. Visual acuity reduced

Eye-related concomitant medications, procedures and surgeries will be summarised by categories of medications (subdivided into corticosteroids, antibiotics, other), ocular procedures and surgeries.

Listings will be produced for the following assessments: BCVA, lid margins for abnormalities, slit lamp examination findings, cornea abnormality, LSCD, abnormality findings from dilated fundoscopic exam and clinically significant corneal disease.

In addition, listings will be produced for:

1. medication used for eye pain
2. eye-related adverse events
3. eye-related concomitant medications, procedures and surgeries

All analyses will be performed on the SAF.

4.4.12 Impact of COVID-19

Depending on the extent of any coronavirus disease 2019 (COVID-19) impact, summaries of data relating to participants diagnosed with COVID-19, and impact of COVID-19 on study conduct (in particular delayed/missed visit, delayed or discontinued IP, discontinuation of study, and COVID-19 related protocol deviations) may be generated, by treatment group, including:

- Disposition (discontinued IP due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruption (visit impact, drug impacted)
- Listing for participants affected by the COVID-19 pandemic
- Listing for participants with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.

Additional analyses may be performed to explore the impact of COVID-19 on key efficacy and safety endpoints, for example repeating the AE summaries separately for participants where events are attributed to COVID-19.

5 INTERIM ANALYSIS

Interim Analysis for Superiority in OS

Two interim analyses for OS are planned.

The first interim analysis will occur at the primary PFS analysis. This corresponds to approximately 178 OS events, 25% maturity and 40% of the information expected at the primary analysis (444 OS events at final, primary).

The second interim analysis will occur when approximately 355 OS events have been observed in the FAS. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis (444 OS events).

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including an interim analysis for superiority. This approach will be used to maintain an overall 2-sided type I error across the three planned analyses of OS.

If the PFS dual primary analysis crosses the efficacy threshold, the 1.0% type I error allocated to the PFS endpoint will be reallocated to the OS endpoint for a total 2-sided type I error of 5.0% (Burman, Sonesson, & Guilbaud, 2009). If the PFS dual primary analysis does not cross the efficacy threshold the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Table 12 Summary of planned timings of the interim and final OS analyses

	Interim Analysis 1		Interim Analysis 2		Primary Analysis	
Projected Timing	21 Months [b]		34 Months		44 Months	
Number of Deaths [a]	178		355		444	
Information Fraction	40%		80%		100%	
Maturity	25%		51%		63%	
Recommendation	**Continue**	**Reject Null Hypothesis**	**Continue**	**Reject Null Hypothesis**	**Do Not Reject Null Hypothesis**	**Reject Null Hypothesis**
At 4.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0005	< 0.0005	≥ 0.0184	< 0.0184	≥ 0.0345	< 0.0345
Estimated hazard ratio	≥ 0.591	< 0.591	≥ 0.777	< 0.777	≥ 0.817	< 0.817
At 5.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0008	< 0.0008	≥ 0.0241	< 0.0241	≥ 0.0427	< 0.0427
Estimated hazard ratio	≥ 0.604	< 0.604	≥ 0.786	< 0.786	≥ 0.824	< 0.824

[a] Estimates based on exponential survival where the median OS is 19.0 months for ICC and 25.3 months for Dato-DXd. The total proportion of participants randomized at time t [t ≤ 19 months] following the start of the study is assumed to be $(t/19)^{1.5}$.
[b] Timing of first IA based on PFS. Number of deaths is an estimate.
[c] Alpha allocated to OS endpoint (4.0% or 5.0%) dependent on statistical significance of PFS.

For a total 2-sided type 1 error of 5.0% this results in a level of significance alpha of approximately 0.0008 for the first interim analysis (IA) and 0.0241 for the second IA. For a total 2-sided type 1 error of 4.0% this results in a level of significance alpha of approximately 0.0005 for the first IA and 0.0184 for the second IA. This is described in more detail in Table 12.

Since the significance level will be dependent on the number of events actually observed, this will be calculated at the time of the analysis.

The interim analyses will be performed by an independent data monitoring committee (IDMC) separate from the study team reporting the final study results so that the study team are kept blinded.

An external IDMC comprised of therapeutic area experts and biostatisticians who are not employed by AstraZeneca and are free from conflict of interest will review the unblinded interim analysis output.

The IDMC will inform the sponsor if superiority has been achieved for OS in the FAS.

Analyses planned to be performed at the interim analysis will include PFS (first IA) and OS, and other key outputs will also be produced. Details of the outputs to be produced for the interim analysis will be specified in the IDMC charter.

If the first interim results do not meet the criterion for declaring superiority for OS in the FAS, then follow-up will continue until the criteria are met for the second OS IA. If the second interim results do not meet the criterion for declaring superiority in the FAS, then follow-up will continue until the criteria is met for the OS primary analysis (approximately 444 OS events in the FAS).

For a total 2-sided type 1 error of 5.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0427 level of significance. For a total 2-sided type 1 error of 4.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0345 level of significance.

The study may continue monitoring participants for OS up to the scheduled final analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival.

IDMC Safety Reviews

This study will use an external IDMC to assess ongoing safety analyses as well as the interim efficacy analyses. The IDMC will meet to review unblinded safety data after the study has started, with an initial early IDMC approximately six months after study start; and then at approximately six-month intervals thereafter.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca and are free from conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will include the recommendation and any potential protocol amendments and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

The safety of all AstraZeneca clinical studies is closely monitored on an ongoing basis by AstraZeneca representatives in consultation with AstraZeneca Global Patient Safety. Issues identified will be addressed; this could involve, for instance, amendments to the Clinical Study Protocol and letters to investigators.

Full details of the IDMC procedures, processes and the responsibilities of the IDMC will be given in the IDMC charter.

6 REFERENCES

Burman, C., Sonesson, C., & Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med. 28:739-761.

Fayers, P., Aaronson, N., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organization for Research and Treatment of Cancer.

Fleischer, F., Gaschler-Markefski, B., & Bluhmki, E. (2001). How is retrospective independent review.

Hernandez Alava, M. P. (2020). Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. *Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York.*, Report 063.

Lan, K., & DeMets, D. (1983). Discrete sequential boundaries for clinical trials. Biometrika. 70:659-663.

Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. Stat Med, 4(2). 213-26.

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

7 APPENDIX

Not applicable.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d9268c00001-sap-ed-2		
Document Title:	Statistical Analysis Plan Edition 2	
Document ID:	Doc ID-004671764	
Version Label:	2.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
25-Jul-2023 17:41 UTC	PPD	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.

STATISTICAL ANALYSIS PLAN	
Study Code	D9268C00001
Edition Number	3.0
Date	17-Aug-2023

A Phase 3, Open-label, Randomised Study of Dato-DXd Versus Investigator's Choice of Chemotherapy in Participants With Inoperable or Metastatic Hormone Receptor-Positive, HER2-Negative Breast Cancer Who Have Been Treated With One or Two Prior Lines of Systemic Chemotherapy (TROPION-Breast01)

TABLE OF CONTENTS

TITLE PAGE 1
TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 7
AMENDMENT HISTORY 12
1 INTRODUCTION 20
2 CHANGES TO PROTOCOL PLANNED ANALYSES 20
3 DATA ANALYSIS CONSIDERATIONS 20
3.1 Timing of Analyses 20
3.1.1 Ophthalmologic Data Review (DCO1) 21
3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2) 21
3.1.3 Second OS Interim Analysis (DCO3) 21
3.1.4 Primary Analysis of OS (DCO4) 21
3.2 Analysis Populations 21
3.2.1 Enrolled set 21
3.2.2 Full analysis set 22
3.2.3 Safety analysis set 22
3.2.4 Ophthalmologic analysis set 22
3.2.5 Pharmacokinetic analysis set 22
3.2.6 ADA evaluable set 22
3.2.7 Summary of outcome variables and analysis sets 22
3.3 General Considerations 24
3.3.1 Sample Size Determination 25
3.3.2 Investigational Product 26
3.3.3 Baseline 26
3.3.4 On Treatment 27
3.3.5 Visit Window 27
3.3.6 Handling of Unscheduled Visits 28
3.3.7 Missing Data 28
3.3.8 Derivations of RECIST Visit Responses 29
3.3.8.1 Target lesions (TLs) 30
3.3.8.2 Non-target lesions (NTLs) and new lesions 34
3.3.8.3 Overall visit response – site investigator data 36
3.3.8.4 Independent review 36
3.3.9 Multiplicity/Multiple Comparisons 37
3.3.10 Handling of Protocol Deviations in Study Analysis 39
4 STATISTICAL ANALYSIS 40
4.1 Study Population 40
4.1.1 Participant Disposition and Completion Status 41
4.1.1.1 Presentations 41
4.1.2 Analysis Sets 41
4.1.3 Protocol Deviations 41

4.1.4 Demographics 42
4.1.4.1 Definitions and Derivations 42
4.1.4.2 Presentation 42
4.1.5 Baseline Characteristics 42
4.1.5.1 Definitions and Derivations 42
4.1.5.2 Presentation 42
4.1.6 Disease Characteristics 43
4.1.6.1 Definitions and Derivations 43
4.1.6.2 Presentation 44
4.1.7 Medical History and Concomitant Disease 44
4.1.7.1 Definitions and Derivations 44
4.1.7.2 Presentation 44
4.1.8 Prior and Concomitant Medications 44
4.1.8.1 Definitions and Derivations 44
4.1.8.2 Presentation 45
4.1.9 Study Drug Compliance 46
4.2 Endpoint Analyses 46
4.2.1 Primary Endpoint - Progression Free Survival by BICR 54
4.2.1.1 Definition 54
4.2.1.2 Derivations and Censoring Rules 54
4.2.1.3 Handling of Dropouts and Missing Data 56
4.2.1.4 Primary Analysis of Progression Free Survival 56
4.2.1.5 Sensitivity Analyses of Progression Free Survival 58
4.2.1.6 Subgroup Analyses 60
4.2.2 Primary Endpoint - Overall Survival 62
4.2.2.1 Definition 62
4.2.2.2 Derivations and Censoring Rules 62
4.2.2.3 Handling of Dropouts and Missing Data 63
4.2.2.4 Primary Analysis of Overall Survival 63
4.2.2.5 Sensitivity Analyses of Overall Survival 64
4.2.2.6 Subgroup Analyses 64
4.2.3 Secondary Endpoint - Objective Response Rate 64
4.2.3.1 Definition 64
4.2.3.2 Derivations 65
4.2.3.3 Primary Analysis of Objective Response Rate 65
4.2.4 Secondary Endpoint - Best Objective Response 66
4.2.4.1 Definition 66
4.2.4.2 Derivations 66
4.2.4.3 Primary Analysis of Best Objective Response 67
4.2.5 Secondary Endpoint - Duration of Response 67
4.2.5.1 Definition 67
4.2.5.2 Derivations and Censoring Rules 67
4.2.5.3 Primary Analysis of Duration of Response 68
4.2.5.4 Supplementary Analyses of Duration of Response 68
4.2.6 Secondary Endpoint – PFS by Investigator Assessment 68
4.2.6.1 Definition 68

4.2.6.2 Derivations and Censoring Rules 68
4.2.6.3 Handling of Dropouts and Missing Data 69
4.2.6.4 Primary Analysis of Progression Free Survival 69
4.2.7 Secondary Endpoint –Disease Control Rate 69
4.2.7.1 Definition 69
4.2.7.2 Derivations 69
4.2.7.3 Primary Analysis of Disease Control Rate 70
4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30 70
4.2.8.1 Definition 70
4.2.8.2 Derivations and Censoring Rules 71
4.2.8.3 Handling of Dropouts and Missing Data 74
4.2.8.4 Primary Analysis of TTD Endpoints as Measured by EORTC QLQ-C30 74
4.2.8.5 Sensitivity Analysis of TTD Endpoints as Measured by EORTC QLQ-C30 75
4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death 75
4.2.9.1 Definition 75
4.2.9.2 Derivations and Censoring Rules 75
4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death 76
4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death 76
4.2.10.1 Definition 76
4.2.10.2 Derivations and Censoring Rules 76
4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death 77
4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death 77
4.2.11.1 Definition 77
4.2.11.2 Derivations and Censoring Rules 77
4.2.11.3 Handling of Dropouts and Missing Data 78
4.2.11.4 Primary Analysis of PFS2 78
4.2.12 Secondary Endpoint - Pharmacokinetics 78
4.2.12.1 Derivations 78
4.2.12.2 Primary Analyses of Pharmacokinetics 79
4.2.13 Secondary Endpoint - Immunogenicity 80
4.2.13.1 Derivations 80
4.2.13.2 Primary Analysis of Immunogenicity 81
4.2.14 Exploratory Endpoint – Patient-Reported Symptomatic AEs and Treatment Tolerability 81
4.2.14.1 Definition 81
4.2.14.2 Derivations and Censoring rules 82
4.2.14.3 Handling of Dropouts and Missing data 84
4.2.14.4 Primary Analyses of PRO-CTCAE 84
4.2.14.5 Primary Analyses of EORTC IL117 84
4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability 85
4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS) 85
4.2.15.1 Definition 85

4.2.15.2 Derivations and Censoring rules 85
4.2.15.3 Analysis of Patient Global Impression of Severity 85
4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC) 86
4.2.16.1 Definition 86
4.2.16.2 Derivations and Censoring rules 86
4.2.16.3 Analysis of Patient Global Impression of Change 86
4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL 86
4.2.17.1 Definition 86
4.2.17.2 Derivation and Censoring Rules 86
4.2.17.3 Handling of Dropouts and Missing Data 86
4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL 86
4.2.18 Exploratory Endpoint – Breast and Arm Symptoms 87
4.2.18.1 Definition 87
4.2.18.2 Derivations and Censoring Rules 88
4.2.18.3 Handling of Dropouts and Missing Data 88
4.2.18.4 Primary Analysis of Breast and Arm Symptoms 88
4.2.19 Exploratory Endpoint - EQ-5D-5L 89
4.2.19.1 Definition 89
4.2.19.2 Derivations 89
4.2.19.3 Primary Analysis of EQ-5D-5L 90
4.2.19.4 Supplementary Analyses of EQ-5D-5L 90
4.2.20 Exploratory Endpoint - Biomarkers 90
4.2.21 Exploratory Endpoint - Health Care Resource Use 90
4.2.21.1 Definition 90
4.2.21.2 Derivations 91
4.2.21.3 Primary Analysis of Health Care Resource Use 91
4.3 Pharmacodynamic Endpoint(s) 91
4.4 Safety Analyses 91
4.4.1 Exposure 91
4.4.1.1 Definitions and Derivations 91
4.4.1.2 Presentation 96
4.4.2 Adverse Events 96
4.4.2.1 Definitions and Derivations 96
4.4.2.2 Presentation 97
4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest 101
4.4.3 Clinical Laboratory, Blood Sample 101
4.4.3.1 Definitions and Derivations 101
4.4.3.2 Presentations 102
4.4.4 Clinical Laboratory, Urinalysis 104
4.4.4.1 Definitions and Derivations 104
4.4.4.2 Presentations 104
4.4.5 Other Laboratory Evaluations 105
4.4.5.1 Definitions and Derivations 105
4.4.5.2 Presentations 105

4.4.6 Vital Signs 105
4.4.6.1 Definitions and Derivations 105
4.4.6.2 Presentations 105
4.4.7 Electrocardiogram 105
4.4.7.1 Definitions and Derivations 105
4.4.7.2 Presentations 106
4.4.8 Echocardiogram/Multigated Acquisition Scan 107
4.4.8.1 Definitions and Derivations 107
4.4.8.2 Presentations 107
4.4.9 Eastern Cooperative Oncology Group Performance Status 107
4.4.9.1 Definitions and Derivations 107
4.4.9.2 Presentations 108
4.4.10 Physical Examination 108
4.4.10.1 Definitions and Derivations 108
4.4.10.2 Presentations 108
4.4.11 Ophthalmologic Assessments 108
4.4.11.1 Definition and Derivations 108
4.4.11.2 Presentations 109
4.4.12 Impact of COVID-19 110
5 INTERIM ANALYSIS 111
6 REFERENCES 114
7 APPENDIX 114

LIST OF TABLES

Table 1 Summary of outcome variables and analysis sets 23
Table 2 TL visit responses (RECIST 1.1) 31
Table 3 NTL visit responses 35
Table 4 Overall visit response 36
Table 5 Endpoint analyses 47
Table 6 Outcome and date of event for PFS analysis 54
Table 7 Definition of two missed RECIST visits 55
Table 8 EORTC QLQ-C30 scales and scores 71
Table 9 Definition of two missed PRO visits 74
Table 10 EORTC IL116 scales and scores 88
Table 11 Best Corrected Visual Acuity conversion table 109
Table 12 Summary of planned timings of the interim and final OS analyses 112

LIST OF FIGURES

Figure 1 Multiple Testing Procedure 38

LIST OF ABBREVIATIONS

Abbreviation or Specialised Term	Definition
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
AJCC	American joint committee on cancer
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AZ	AstraZeneca
BCVA	Best corrected visual acuity
BICR	Blinded independent central review
BID	Twice daily
BLQ	Below limit of quantification
BMI	Body mass index
BoR	Best overall response
BSA	Body surface area
C1D1	Cycle 1, day 1
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
COVID-19	Coronavirus 2019 disease
CRF	Case Report Form
CR	Complete response
CRO	Clinical research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DBL	Data base lock
DCO	Data cut-off
DCR	Disease control rate
DoR	Duration of Response

Abbreviation or Specialised Term	Definition
DOSDISC	Treatment discontinuation form
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EORTC	European Organisation for Research and Treatment of Cancer
EQ-5D-5L	EuroQoL 5-dimension, 5-level health state utility index
EQ-VAS	EuroQoL-Visual analogue scale
FAS	Full Analysis Set
FEV1	Forced expiratory volume – 1 second
FVC	Forced vital capacity
GHS	Global health status
HDU	High dependency unit
HER2	Human epidermal growth factor receptor 2
HCRU	Health care resource use
HIV	Human immunodeficiency virus
HR	Hazard ratio
HR-positive	Hormone receptor-positive
HRQoL	Health related quality of life
IA	Interim analysis
ICC	Investigator's Choice Chemotherapy
ICF	Informed consent form
ICR	Independent central review
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IHC	Immunohistochemistry
ILD	Interstitial lung disease
IP	Investigational product
IPD	Important protocol deviation
IRC	Independent Review Charter
IRT	Interactive Response Technology
ITT	Intention to treat
IV	Intravenous
KM	Kaplan-Meier
LD	Longest diameter

Abbreviation or Specialised Term	Definition
LLOQ	Lower Limit of Quantification
Ln	Natural logarithm or logarithm to the base e
LSCD	Limbal stem cell deficiency
LSmean	Least squares mean
LVEF	Left ventricular ejection fraction
MCID	Minimal clinically important difference
MedDRA	Medical Dictionary for Regulatory Activities
MM	Millimetre
MMRM	Mixed model for repeated measures
MRI	Magnetic Resonance Imaging
MSSO	Maintenance and Support Services Organization
MTP	Multiple testing procedure
MUGA	Multigated acquisition
NA	Not applicable
nAb	Neutralizing antibody
NC	Not calculable
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NL	New lesion
NQ	Not quantifiable
NR	Not Reportable
NS	No Sample
NTL	Non-target lesion
OAS	Ophthalmologic Analysis Set
ORR	Objective response rate
OS	Overall survival
PAS	Pharmacokinetic analysis set
PAP	Payer Analysis Plan
PD	Progression of disease
PFS	Progression-free survival
PFS2	Second progression-free survival
PGIC	Patients' global impression of change
PGIS	Patients' global impression of severity
PGI-TT	Patients' global impression of treatment tolerability

Abbreviation or Specialised Term	Definition
PID	Percentage intended dose
PK	Pharmacokinetics
PR	Partial response
PRO	Patient-reported outcome
PS	Performance Status
PT	Preferred term
Q3W	Every 3 weeks
Q6W	Every 6 weeks
Q9W	Every 9 weeks
QLQ-C30	EORTC 30-item core quality of life questionnaire
QTcF	QT interval corrected by Fridericia's formula
RDI	Relative dose intensity
RECIST 1.1	Response Evaluation Criteria in Solid Tumours, Version 1.1
RS	Raw score
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Stable disease
SoA	Schedule of activities
SOC	System organ class
TEAE	Treatment emergent adverse event
TELC	Treatment emergent laboratory change
TFBT	Tear film breakup time
TFL	Table, figures and listings
TFST	Time to first subsequent therapy or death
TL	Target lesion
TNM	Tumour node metastasis
TROP2	Trophoblast cell surface antigen 2
TSST	Time to second subsequent therapy or death
TTD	Time to deterioration
ULN	Upper limit of normal
US	United States
VAS	Visual analogue scale
VCV	Veeva Clinical Vault

Abbreviation or Specialised Term	Definition
WHO	World Health Organization Drug dictionary

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
N/A	22-Dec-2021	Initial approved SAP	N/A	N/A
Two missed visits	21-Jun-2023	Definition of 2 missed visits has been updated for RECIST and PRO assessments.	Yes	Correction to previous error
List of abbreviations	21-Jun-2023	Added BSA and IHC Removed DFI and TFI	N/A	Additional text added to SAP. Text removed from SAP
General	21-Jun-2023	Added additional wording for clarity	Yes	Added clarity
General	21-Jun-2023	Changed ORR, DCR and BoR to FAS	Yes	Aligned with CSP
General	21-Jun-2023	Removed pulmonary function summary	N/A	Changed analysis
Section 2	21-Jun-2023	Added additional wording to clarify difference in PRO analysis	Yes	Clarified inconsistency in the CSP
Section 2	21-Jun-2023	Removed text about EQ-5D-5L population	Yes	Changed population to align with CSP
Section 2	21-Jun-2023	Added Hy's law text	No	Hy's law definition was different between CSP and SAP
Section 2	21-Jun-2023	Removed AESI and geographic region	Yes	CSP amendments
ADA evaluable set	21-Jun-2023	Changed population definition	N/A	Changed to analysis population requested
Section 3.2.1	21-Jun-2023	Specified irrespective of randomisation	N/A	Added clarity
Section 3.2.3	21-Jun-2023	No crossover between study treatment arms will be allowed	N/A	Added clarity
Section 3.2.5	21-Jun-2023	Added details on protocol violation	N/A	Added clarity
Section 3.2.6	21-Jun-2023	Changed definition to be result at any time	N/A	Standards changed

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 3.3	21-Jun-2023	Added text regarding precision of summary statistics	N/A	Added clarity to ensure consistency
Section 3.3.2	21-Jun-2023	Updated IP definition	N/A	Additional clarity
Section 3.3.3	21-Jun-2023	Added ADA baseline definition	N/A	Additional summary added requiring definition of baseline
Section 3.3.3	21-Jun-2023	Updated definition of PRO endpoints	N/A	To allow participants without treatment to be included in baseline definition
Section 3.3.5	21-Jun-2023	Added ADA baseline definition	N/A	Additional summary added requiring definition of baseline
Section 3.3.5	21-Jun-2023	Updated definition of PRO endpoints	N/A	To allow participants without treatment to be included in baseline definition
Section 3.3.7	21-Jun-2023	Updated missing day imputation rules for stop date	N/A	To ensure imputed stop date will be on or after start date
Table 2 and Table 4	21-Jun-2023	Added NA response for TL	N/A	To account for participants who have no baseline per BICR
Section 4.1.1.1	21-Jun-2023	Removed summary of recruitment by enrolled set and safety analysis set	N/A	Summary removed
Section 4.1.5	21-Jun-2023	Added BSA	N/A	Additional summary
Section 4.1.5.2	21-Jun-2023	Changed summary of stratification factors from distinct stratum to category	N/A	Improved interpretation
Section 4.1.6	21-Jun-2023	Added additional disease characteristics	N/A	Additional summaries of interest

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.1.6	21-Jun-2023	Removed DFI and TNM	N/A	Summary changed
Section 4.1.8.1	21-Jun-2023	Removed TFI	N/A	Summary changed
Section 4.1.8.2	21-Jun-2023	Added additional summaries	N/A	Additional summaries of interest
Section 4.1.8.2	21-Jun-2023	Changed post-study treatment cancer therapies to summarise by ATC group and generic term	N/A	Summary changed
Section 4.2	21-Jun-2023	Specified BICR and investigator for relevant endpoints	N/A	Clarification provided
Section 4.2.1.2	21-Jun-2023	Reworded table 6	No	Clarification provided
Section 4.2.1.4	21-Jun-2023	Specified TEST statement used for log rank test	N/A	Added clarity
Section 4.2.1.5, 4.2.2.5	21-Jun-2023	Added eCRF stratification analysis	N/A	Additional summaries of interest
Section 4.2.1.6	21-Jun-2023	Added HER2 status, ET, ECOG PS, prior use of CDK4/6 inhibitor and early relapse as subgroups	No	Additional subgroups on interest
Section 4.2.1.6	21-Jun-2023	Stratification factors from CRF	N/A	Summary changed
Section 4.2.1.6	21-Jun-2023	Removed qualitive interaction testing	N/A	Does not add any useful supportive information
Section 4.2.2.4	21-Jun-2023	Specified TEST statement used for log rank test	N/A	Added clarity
Section 4.2.2.4, 4.2.9.3	21-Jun-2023	Removed prematurely censored	N/A	Change to standards
Section 4.2.4.2	21-Jun-2023	Defined confirmed CR and PR	N/A	Clarification provided
Section 4.2.5	21-Jun-2023	Removed sensitivity analysis	N/A	Change to standards
Section 4.2.5.2	21-Jun-2023	Defined confirmed CR and PR	N/A	Clarification provided

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.2.5.4	21-Jun-2023	Specified sensitivity analysis is a repeat of the primary	N/A	Ensure alignment in summary between analyses
Section 4.2.6	21-Jun-2023	Removed sensitivity and subgroup analysis	N/A	Summary changed
Section 4.2.6.4	21-Jun-2023	Specified analysis	N/A	Added clarity
Section 4.2.7.2	21-Jun-2023	Defined confirmed CR and PR	N/A	Clarification provided
Section 4.2.8.2	21-Jun-2023	Removed text about MCID when estimation method not performed	N/A	Removed duplicate text
Section 4.2.8.2	21-Jun-2023	Removed death as an event for TTD	N/A	Correction to previous error
Section 4.2.8.2	21-Jun-2023	Specified what evaluable is and changed missed visits to immediately	N/A	Added clarity
Section 4.2.8.2	21-Jun-2023	Specified how participants with baseline score close to maximum/minimum will be censored	N/A	Added clarity
Table 9	21-Jun-2023	Removed first row from table	N/A	As death is no longer an event this row is not needed
Section 4.2.8.3	21-Jun-2023	Changed <50% to ≤50%	N/A	Aligned with scoring manual
Section 4.2.8.4	21-Jun-2023	Specified the analysis being done	N/A	Added clarity
Section 4.2.9.3	21-Jun-2023	Specified which participants are included in time between progression and starting subsequent therapy summary	N/A	Added clarity
Section 4.2.9.3	21-Jun-2023	Removed prematurely censored	N/A	Change to standards
Section 4.2.11.1	21-Jun-2023	Added other as a second progression	N/A	Aligned with data collection
Section 4.2.11.2	21-Jun-2023	Additional information regarding PFS2 derivation added	N/A	Added clarity

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.2.12.2	21-Jun-2023	Changed list of summary statistics	N/A	Summary changed
Section 4.2.12.2	21-Jun-2023	Removed non compartment PK analysis	Yes	Analysis no longer requirement
Section 4.2.13	21-Jun-2023	Additional summary and ADA categories added	N/A	Additional summary of interest
Section 4.2.14.2	21-Jun-2023	Additional information on compliance added	N/A	Added clarity on participants to be included
Section 4.2.14.2	21-Jun-2023	Definition of overall compliance added	N/A	Additional summary of interest
Section 4.2.14.4	21-Jun-2023	Specified 12 weeks excludes baseline record	N/A	Added clarity
Section 4.2.14.4	21-Jun-2023	Changed scale from 0-4 to 1-5	N/A	Correction to previous error
Section 4.2.14.4	21-Jun-2023	Added stacked bar chart and pie chart	N/A	Additional summary of interest
Section 4.2.14.4	21-Jun-2023	Specified if summary is on all symptoms and attributes or only 1st attribute	N/A	Added clarity
Section 4.2.14.5	21-Jun-2023	Added additional summaries	N/A	Additional summary of interest. To align with other endpoints
Section 4.2.14.6	21-Jun-2023	Added stacked bar chart and pie chart	N/A	Additional summary of interest
Section 4.2.16.3	21-Jun-2023	Removed stacked bar chart	N/A	Changed analysis
Section 4.2.17.4	21-Jun-2023	Added plot of absolute score and change from baseline	N/A	Additional summary of interest
Section 4.2.17.4	21-Jun-2023	Specified the analysis being done	N/A	Added clarity
Section 4.2.18.4	21-Jun-2023	Added table of assessment response, absolute score and change from baseline. Added plot of absolute score and change from baseline	N/A	Additional summary of interest
Section, 4.2.18.4	21-Jun-2023	Specified the analysis being done	N/A	Added clarity

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.2.19	21-Jun-2023	Added 10-point change from baseline as MCID	N/A	Preference of MCID is different by country
Section 4.2.19.2	21-Jun-2023	Changed EQ-5D index derivation	N/A	Guidance changed
Section 4.2.19.3	21-Jun-2023	Changed analysis population to FAS	Yes	Aligned with CSP
Section 4.4.1.1	21-Jun-2023	Updated to intended dose calculation	N/A	Correction to previous error
Section 4.4.1.2	21-Jun-2023	Changed dose interruption to drug interruption	N/A	To avoid confusion in terminology
Section 4.4.1.2	21-Jun-2023	Added summary of dose delays	N/A	Additional summaries of interest
Section 4.4.2.1	21-Jun-2023	Added information about AESI follow up	Yes	Aligned with CSP
Section 4.4.2.2	21-Jun-2023	Additional summaries added	N/A	Additional summaries of interest
Section 4.4.2.2	21-Jun-2023	28 days after the date of last IP, but after the onset of subsequent therapy changed to 35 days after the date of last IP and before the onset of subsequent therapy	Yes	To be consistent with safety follow up period
Section 4.4.2.2	21-Jun-2023	Changed AEIS summary for AESI category to AESI category and PT	N/A	Additional summaries of interest
Section 4.4.2.3	21-Jun-2023	Removed text regarding categories marked (*)	N/A	Correction to previous error
Section 4.4.3	21-Jun-2023	Specified project ranges used if local range missing	N/A	Changed derivation
Section 4.4.3	21-Jun-2023	Removed coagulation	N/A	Data collection changed
Section 4.4.3.2	21-Jun-2023	Added CTCAE grade to listing	N/A	Additional summary of interest
Section 4.4.3.2	21-Jun-2023	Updated to Hy's law summary	N/A	Summary changed

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.4.3.2	21-Jun-2023	Updated to Hy's law definition	N/A	Standards changed
Section 4.4.7.1, 4.4.8.1 and 4.4.12.1	21-Jun-2023	Specified summary will be by nominal visits	N/A	Assessments only collected at 2 timepoints and clinically indicated otherwise so cannot apply visit windowing
Section 4.4.7.1	21-Jun-2023	Added additional rules for derived average	N/A	Programming clarification
Section 4.4.7.2	21-Jun-2023	Removed ECG summary table	N/A	Analysis changed
Section 4.4.8.2	21-Jun-2023	Removed EOT summary	N/A	Data collection changed
Section 4.4.11.2	21-Jun-2023	Changed summary	N/A	Data collection changed
Section 5	21-Jun-2023	Removed duplicate text. Changed participant safety department to AstraZeneca global patient safety	N/A	Correction to previous error
General	17-Aug-2023	Updated grammar	Yes	Correction to previous error
Section 3.2.6	17-Aug-2023	Specified ADA set only includes participants receiving Dato-DXd	N/A	Added clarity
Section 3.3.7	17-Aug-2023	Added text for calculating duration when imputing dates	N/A	Added clarity
Section 3.3.10	17-Aug-2023	Changed exclusion criteria to match CSP numbering	Yes	Aligned with CSP
Section 3.3.10	17-Aug-2023	Specified deviation 4 is no longer an IPD	Yes	Protocol deviation plan updated
Section 4.1.1.1	17-Aug-2023	Replaced screen failure summary with participants not randomised	N/A	Aligned with data collection
Section 4.1.8.2	17-Aug-2023	Updated categories for prior disease-related treatment modalities	N/A	Standards changed

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Section 4.4.2.1	17-Aug-2023	Removed AZ DS 1062 from AESI definition	Yes	Aligned with CSP
Section 4.2.1.6	17-Aug-2023	Removed missing duration subgroup from prior use of breast cancer CDK4/6 inhibitor	Yes	Standards changed
Section 4.2.2.4	17-Aug-2023	Specified OS analysed using same methods as primary PFS	Yes	Added clarity to ensure consistency
Section 4.2.9.3	17-Aug-2023	Changed therapy class to generic term	N/A	Aligned with post study treatment cancer therapy summary
Section 4.2.13.1	17-Aug-2023	Treatment-emergent persistently ADA positive definition changed	N/A	Standards changed
Section 4.2.13.2	17-Aug-2023	Added additional wording regarding ADA titre summary	N/A	Added clarity
Section 2, 3.2.4, 3.2.7, 4.4.11	17-Aug-2023	Updated OAS definition. Changed summary from SAF to OAS	No	Made analysis set more appropriate
Section 2, 3.3.8.4, 4.2.4	17-Aug-2023	Added text for NED BICR response	No	Corrected previous error

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D9268C00001 supporting the clinical study report (CSR). The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection. This statistical analysis plan (SAP) is based on version 2.0 of the CSP dated 27 August 2021. In the event of future amendments to the protocol, this SAP may be modified to account for changes relevant to the statistical analysis.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

Table 9 (population for analysis) in the protocol states that all PROs will be analysed using the FAS population, while the PRO endpoints that measure patient-reported symptomatic adverse events and overall treatment tolerability should be analysed based on the safety population per Table 4 (objectives and endpoints) in the protocol, i.e. among participants who received any amount of study treatment and according to the actual treatment received. See SAP Table 1 for more details and specifying the analysis population to align with Table 4 (objectives and endpoints) in the protocol.

For the sensitivity analysis of attrition bias for progression free survival (PFS), the SAP states two, or more, missed tumour assessments, while the CSP states two, or more, non-evaluable tumour assessments.

In Section 8.3.6 of the CSP Potential Hy's law was clarified and participants with elevated liver enzymes (ALT or AST ≥5x ULN) that have liver metastases present at baseline are now permitted in the study.

In Section 3.2.4 of the SAP the ophthalmologic analysis set was defined differently compared to the CSP. The SAP has additional conditions on the number of on-treatment assessments required to be included in the analysis set.

For BICR the SAP mentions a response of NED while the CSP doesn't. The study requires measurable disease at baseline but encountered cases of no measurable disease at baseline in BICR data.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

As the study is event driven, the accrual of the predetermined number of events included in the study endpoints determines the duration of the data collection phase of the study. There are 4 planned data cut-offs (DCOs) for this study consisting of an ophthalmologic data review (DCO1), primary analysis of progression free survival (PFS)/first overall survival

(OS) interim analysis (DCO2), second OS interim analysis (DCO3) and primary analysis of OS (DCO4). These interim analyses and additional safety reviews will be conducted as described in Section 5.

3.1.1 Ophthalmologic Data Review (DCO1)

The DCO for the ophthalmologic data review (DCO1) is planned to occur after completion of the last ophthalmologic assessment and a minimum of 2 assessments per participant has been completed, for the first approximately 100 randomised participants (DCO1).

3.1.2 Primary Analysis of PFS/First OS Interim Analysis (DCO2)

The DCO for the primary analysis of PFS/first OS interim analysis (DCO2) is planned to occur when approximately 419 PFS Blinded independent central review (BICR) events have been observed in the FAS. Based on enrolment assumptions, it is expected that this will occur approximately 21 months after randomisation of the first participant. For the primary analysis of PFS this provides approximately 60% maturity. For the first OS interim analysis this corresponds to approximately 25% maturity and 40% of the information expected at the primary analysis.

3.1.3 Second OS Interim Analysis (DCO3)

The DCO for the interim analysis for superiority in OS (DCO3) is planned to occur when approximately 355 OS events have been observed in the FAS population. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis. Based on enrolment assumptions, it is expected that this will occur approximately 34 months after randomisation of the first participant.

3.1.4 Primary Analysis of OS (DCO4)

The DCO for the primary analysis of OS (DCO4) is planned to occur when approximately 444 OS events have been observed in the FAS population, approximately 44 months after the first participant is randomised (63% maturity). At this time the clinical database will close to new data.

3.2 Analysis Populations

There are six analysis sets defined for this study.

3.2.1 Enrolled set

The Enrolled Set consists of all participants who provided informed consent irrespective of whether they were randomised or received the study treatment.

3.2.2 Full analysis set

The FAS consists of all randomised participants with treatment groups assigned in accordance with the randomisation, regardless of the treatment actually received. Participants who were randomised but did not subsequently receive treatment are included in the FAS. The analysis of data using the FAS therefore follows the principles of intention to treat (ITT).

3.2.3 Safety analysis set

The safety analysis set (SAF) will consist of all randomised participants who received any amount of study treatment (Dato-DXd or **Investigator's Choice Chemotherapy** (ICC)). Safety data will be summarised using the safety analysis set according to the actual treatment received. If a participant receives any amount of Dato-DXd, they will be summarised in the Dato-DXd treatment group. If a participant only receives ICC, they will be summarised in the ICC treatment group. No crossover between study treatment arms will be allowed.

3.2.4 Ophthalmologic analysis set

For DCO1, the ophthalmologic analysis set (OAS) will consist of approximately the first 100 randomised participants (approximately 50 per arm, Dato-DXd and ICC) that have a minimum of two on-treatment assessments per participant. For subsequent DCOs the OAS will be extended to include all participants that have a baseline ophthalmologic assessment and a minimum of two on-treatment assessments per participant.

3.2.5 Pharmacokinetic analysis set

The pharmacokinetic analysis set (PAS) consists of all participants randomly assigned to study intervention who received at least 1 dose of Dato-DXd for whom there is at least one reportable post-dose pharmacokinetic (PK) concentration.

Participants who violate or deviate from the protocol in ways that would significantly affect the PK analyses should not be included in the PK analysis set. These are participants who have a >10% dose reduction, >10% wrong dose or >10% delay for all timepoints.

3.2.6 ADA evaluable set

The anti-drug antibody (ADA) evaluable set will consist of participants in the safety analysis set who received at least 1 dose of Dato-DXd and with a non-missing ADA Dato-DXd result at any time.

3.2.7 Summary of outcome variables and analysis sets

The analysis sets to be used for each outcome are provided in Table 1.

Table 1 Summary of outcome variables and analysis sets

Outcome variable	*Analysis set*
Efficacy Data	
PFS, PFS2, OS, ORR, DoR, DCR, TFST, TSST	FAS
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	FAS
Baseline and disease characteristics	FAS
Important protocol deviations	FAS
Medical/surgical history	FAS
Previous anti-cancer therapy	FAS
Concomitant medications/procedures	FAS
Subsequent anti-cancer therapy	FAS
Study drug compliance	FAS
PK/Immunogenicity Data	
PK data	PAS
Immunogenicity	ADA evaluable set
Safety data	
Exposure	SAF
Adverse events	SAF
Laboratory measurements	SAF
Vital signs	SAF
Physical examination	SAF
ECGs	SAF
ECOG PS	SAF
ECHO/MUGA	SAF
Ophthalmologic assessments	OAS
Patient-reported outcomes	
EORTC QLQ-C30, EORTC IL116, EQ-5D-5L, PGIS, PGIC	FAS
PGI-TT, PRO-CTCAE, EORTC IL117	SAF
Health care resource use	
HcRU	SAF

Participants who are evaluable for the analysis of DoR are those who responded in the ORR analysis.
ADA=antidrug antibody; BoR=best objective response; CTCAE=Common Terminology Criteria for Adverse Events; DCR=disease control rate; DoR= duration of response; ECG=electrocardiogram;

ECHO=echocardiogram; ECOG=Eastern Cooperative Oncology Group; EORTC=European Organisation for Research and Treatment of Cancer; EORTC QLQ-C30=EORTC 30-item core quality of life questionnaire; EQ-5D-5L=EuroQoL 5-dimension, 5-level health state utility index; FAS=Full analysis set; HcRU=Healthcare resource use; MUGA=multigated acquisition; ORR=objective response rate; OS=overall survival; PAS=pharmacokinetic analysis set; PGIC=Patients' Global Impression of Change; PGIS=Patients' Global Impression of Severity; PGI-TT= Patient's Global Impression of Treatment Tolerability; PFS=progression-free survival; PFS2=time from randomisation to second progression or death; PK=pharmacokinetic; PRO=patient-reported outcome; PS=Performance Status; SAF=safety analysis set; TFST=time to first subsequent therapy or death; TSST=time to second subsequent therapy or death.

3.3 General Considerations

The below mentioned general principles are followed throughout the study:

- Summary tables are produced by treatment group (Dato-DXd and ICC). Total columns are produced only for tables of disposition, demography, baseline characteristics and EQ-5D-5L data.
- Descriptive statistics are used for variables, as appropriate. Continuous variables are summarised by the number of observations, mean, standard deviation, median, upper and lower quartiles where indicated, minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables are summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages are calculated out of the population total for the corresponding treatment group.
- For continuous data, the mean and median are rounded to 1 additional decimal place compared to the original data. The standard deviation is rounded to 2 additional decimal places compared to the original data. Minimum and maximum are displayed with the same accuracy as the original data.

  If the number of decimal places in the original data is >3 then display minimum and maximum to 3 decimal places, mean and median to 4 decimal places and the standard deviation to 5 decimal places.
- For categorical data, percentages are rounded to 1 decimal place with the exception of 100% which is presented as a whole number.
- Results of all statistical analyses will be presented using a 95% confidence interval (CI) and 2-sided p-value, unless otherwise stated.
- CIs and ratios (including hazard ratios) will be rounded to 2 decimal places. The p-values will be rounded to 4 decimal places, except for those below 0.0001, which will be displayed as '<0.0001'.
- Survival rates will be rounded to 1 decimal place.
- ORR and DCR will be rounded to 2 decimal places.

- SAS® version 9.4 (or higher) is used for all analyses.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days. One year is defined to be 365.25 days.

Where analysis models are stratified by the randomisation stratification factors, the data from the Interactive Response Technology (IRT) will be used, not the values recorded in the electronic case report form (eCRF), therefore follows the principles of intention to treat (ITT).

3.3.1 Sample Size Determination

Approximately 1000 participants will be enrolled to achieve approximately 700 participants randomly assigned to study intervention.

"Enrolled" means a participant's, or their legally acceptable representative's, agreement to participate in a clinical study following completion of the informed consent process. Potential participants who are screened for the purpose of determining eligibility for the study but are not randomly assigned/assigned in the study, are considered "screen failures", unless otherwise specified by the protocol.

The study is sized for dual primary endpoints to characterise the PFS and OS benefit of Dato-DXd versus ICC in the participants with HR-positive, HER2-negative breast cancer who have been treated with one or two prior lines of systemic chemotherapy in the inoperable/metastatic setting. The study will be considered positive (a success) if either the PFS analysis results and/or the OS analysis results are statistically significant.

For the primary analysis of PFS (See Section 3.1.2 for timing of analysis) assuming the true PFS treatment effect under the alternative hypothesis is a hazard ratio of 0.55 for Dato-DXd versus ICC, and the median PFS times of 4.7 months and 8.5 months in ICC and Dato-DXd, respectively. 419 PFS events from the FAS population (60% maturity) will provide greater than 99% power to demonstrate statistical significance at the 2-sided alpha level of 1.0%. This also assumes the median PFS times in both groups are exponentially distributed. The smallest treatment difference that is statistically significant will be a hazard ratio of 0.775. Assuming a recruitment period of 19 months, this analysis is anticipated to be approximately 21 months after the first participant has been randomised.

The primary analysis of OS will be performed when approximately 444 OS events from the FAS have occurred across the Dato-DXd and ICC treatment groups (63% maturity). Assuming the true OS hazard ratio is 0.75 for Dato-DXd versus ICC, and the median OS in ICC is 19.0 months, the study will have 85% power to demonstrate statistical significance at the 5.0% level (using a 2-sided test). This assumes the PFS primary analysis crosses the efficacy threshold, and allowing 2 interim analyses to be conducted at information fractions

of approximately 40% and 80% of the target events, respectively (per the O'Brien and Fleming approach (Lan & DeMets, 1983)). The smallest treatment difference that could be statistically significant at the primary OS analysis is a hazard ratio of 0.824.

If the PFS primary analysis does not cross the efficacy threshold, the OS analysis will have 83% power to demonstrate statistical significance at the 4.0% level (using a 2-sided test). The smallest treatment difference that could be statistically significant at the primary analysis is a hazard ratio of 0.817. All OS calculations assume median OS times of 19.0 months and 25.3 months in ICC and Dato-DXd, respectively when the survival times are exponentially distributed.

With a recruitment period of approximately 19 months, it is anticipated that the primary OS analysis will occur approximately 44 months after the first participant has been randomised. The study may continue monitoring participants for OS up to the scheduled primary analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival. Further details of the interim analyses are presented in Section 5.

A nonuniform accrual of participants (with k=1.5) is assumed when estimating the analysis times. The total proportion of participants randomised at time t [$t \leq 19$ months] following the start of the study is assumed to be $(t/19)^k$.

3.3.2 Investigational Product

Investigational product (IP) refers to Dato-DXd and the ICC. The first and last dates of IP refer to the earliest and last non-zero dose administration of this treatment respectively.

3.3.3 Baseline

In general, for efficacy endpoints the last observed measurement prior to randomisation is considered the baseline measurement. For safety endpoints and ADA the last evaluable observation before the first dose of IP is considered the baseline measurement unless otherwise specified.

For PRO endpoints the last evaluable observation on or before the target date is considered the baseline measurement. Where target date is first dose of IP if present, else Cycle 1 Day 1 visit date if present, else randomisation date.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured are considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

If two visits are equally eligible to assess participant status at baseline (e.g. two assessments both on the same date with no time recorded), the average is used as the baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking the

average is not possible, the best value (value closest to none/normal/negative) is used as baseline as this is most conservative. In the scenario where there are two assessments recorded on the same day, one with time recorded and the other without time recorded, the one with the time recorded is selected as baseline. Where safety data are summarised over time, time on study is calculated in relation to date of first IP administration.

In all summaries change from baseline variables are calculated as the post-treatment value minus the value at baseline. The percentage change from baseline is calculated as (post-baseline value - baseline value) / (baseline value) x 100.

3.3.4 On Treatment

For the purposes of summarising safety data assessed at visits, in addition to baseline data, only on treatment data are included in the summary tables. On treatment data is defined as data after the first dose of IP and with assessment date up to and including the date of last IP + 35 days and prior to the start of any subsequent cancer therapy, whichever occurs earlier.

3.3.5 Visit Window

Time windows are defined for all presentations of safety data, ADA data and PRO data that summarise values by visit according to the following conventions:

- For safety and ADA data study day references date of first dose of IP as Day 1, for PK the reference is the time of Dato-DXd administration on the day PK blood samples are taken, for efficacy data study day references date of randomisation as Day 1.
- All windows for PRO (including those reported on FAS or SAF) will have windows calculated from target date. Where target date is first dose of IP if present, else Cycle 1 Day 1 visit date if present, else randomisation date. However, time to deterioration will be calculated as described in Section 4.2.8.2.
- The time windows are exhaustive so that data recorded at any time point (scheduled or unscheduled) has the potential to be summarised. Inclusion within the time window is based on the actual date and not the intended date of the visit.
- The window for visits following baseline is 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 is Day 2). If an even number of days exist between two consecutive visits, then the upper limit is taken as the midpoint value minus 1 day.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment (as defined in Section 3.3.4) is used (regardless of where it falls in an interval).
- Listings display all values contributing to a time point for a participant.

- For visit-based summaries, if there is more than one value per participant within a time window then the closest value to the scheduled visit date is summarised. If the values are equidistant from the nominal visit date, then the earlier value is used. Data listings highlight the values used in the summary table, wherever feasible.

Visit data are only included in summary tables and figures if the number of observations is ≥20 in at least one treatment group.

3.3.6 Handling of Unscheduled Visits

Unscheduled visits are included in the method of assigning data to scheduled visits described in the rules in Section 3.3.5 above. Unscheduled visits are not included as a separate visit in the summary tables.

For summaries at participant level, such as of extreme values, all post-baseline values collected are used to derive a participant level statistic including those collected at unscheduled visits and regardless of whether they appear in the corresponding visit-based summary.

3.3.7 Missing Data

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

For missing start dates for adverse events (AEs) and concomitant medications/procedures, the following rules are applied:

- Missing day: Impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.
- Missing day and month: Impute 1st January unless year is the same as first dose date then impute first dose date.
- Completely missing date: Impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date.

An imputed start date of an AE must be prior to the end date of the AE.

For missing stop dates of AEs or concomitant medications/procedures, the following rules are applied:

- Missing day - If month is same as month of last dose of study drug and start date is after last dose of study drug then impute last day of the month. Otherwise impute the

last day of the month unless month is the same as month of the last dose of study drug then impute last dose date.

- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date.
- Completely missing: If an AE/medication has a completely missing end date then it is treated as ongoing.

The imputation of dates for AEs and concomitant medications is used to determine if an AE is treatment emergent and whether a medication is concomitant. Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations are not calculated except for prior medications.

If a participant is known to have died where only a partial death date is available, then the date of death is 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:

- Missing day only: Using the 1st of the month.
- Missing day and month: Using the 1st January.

For partial subsequent anti-cancer therapy dates, the following rules are applied:

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

If a participant has a partial date of birth, (i.e. for those cases where year of birth only is given) the 1st of the month is imputed if only day is missing, and 1st January is imputed if the day and month are missing.

Other rules for handling missing data are described under the derivation rules for that particular variable.

3.3.8 Derivations of RECIST Visit Responses

For all participants, the Response Evaluation Criteria in Solid Tumours (RECIST) tumour response data will be used to determine each participant's visit response according to RECIST version 1.1. It will also be used to determine if and when a participant has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumour assessments are to be performed no more than 28 days before randomisation and ideally as close as possible to and prior to randomisation. Tumour

assessments are then performed every 6 weeks following randomisation for 48 weeks, then every 9 weeks thereafter until disease progression. Following disease progression, 1 additional follow-up scan should be performed as per imaging schedule (i.e. either 6 weeks or 9 weeks later).

If an unscheduled assessment is performed, and the participant 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 minimise any unintentional bias caused by some participants being assessed at a different frequency than other participants.

From the investigator's review of the imaging scans, the RECIST tumour response data are used to determine each participant's visit response according to RECIST version 1.1. At each visit, participants are programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD) or progression of disease (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 participant has had a tumour assessment that cannot be evaluated then the participant is assigned a visit response of not evaluable (NE), (unless there is evidence of progression in which case the response is assigned as PD).

Please refer to Section 3.3.8.3 for the definitions of CR, PR, SD and PD.

RECIST outcomes (i.e. PFS, ORR etc.) are calculated programmatically for the site investigator data (as described in the relevant subsection in Section 4.2) from the overall visit responses.

3.3.8.1 Target lesions (TLs)

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 participant can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as TLs. Lymph nodes are considered to be one organ; therefore, at most 2 target nodal lesions can be selected. If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomisation are 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 target lesion (TL) should be identified as NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

TL visit responses are described in Table 2 below.

Table 2 TL visit responses (RECIST 1.1)

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 <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Stable disease (SD)	Neither sufficient decrease in sum of diameters to qualify for PR nor sufficient increase to qualify for PD.
Progression of disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Not evaluable (NE)	Only relevant if any of the TLs at follow-up were not assessed or not evaluable (e.g. missing anatomy) or had a lesion intervention at this visit. Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not applicable (NA)	Only relevant if 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 is rounded to one decimal place (d.p.) before assigning a TL response. For example, 19.95% is rounded to 20.0% but 19.94% is rounded to 19.9%.

Missing TL data

If the sum of **available** 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, the TL visit response is PD. Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If there is at least one TL measurement missing and a **TL** visit response of PD cannot be assigned, the **TL** visit response is not evaluable (NE).

If all TL measurements are missing, then the TL visit response is NE. Overall visit response is also NE, unless there is a progression of non-TLs or new lesions, in which case the response is PD.

Lymph nodes

For lymph nodes, if the size reduces to <10mm then these are considered non-pathological. However, a size is still given, and this size is still 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 is 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 are applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is 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% but remains <10mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or <10mm for lymph nodes) then response is set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis >10mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response is set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined the response remains as CR.

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure, then this is indicated as such on the CRF and a value of 5mm is 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 is reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) are recorded.

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

- Step 1: the diameters of the TLs (including the lesions that have had intervention) are summed and the calculation are performed in the usual manner. If the visit response is PD, this remains as a valid response category.
- Step 2: If there is no evidence of progression after step 1, the lesion diameter (for those lesions with intervention) are treated 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 participant are 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 is used, and PR or SD assigned as the visit response. Participants 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 participant 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 is set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

If >1/3 of TL measurements are missing (because of intervention) then the TL response is 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 ≥5mm from nadir).

If ≤1/3 of the TL measurements are missing (because of intervention) then the results are scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters) and this is 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 74mm.

The sum of lesions 1-4 at the follow-up is 68mm. The sum of the corresponding lesions at the nadir visit is 62mm.

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

68 x 74 / 62 = 81mm

CR is not allowed as a TL response for visits where there is missing data. Only PR, SD, or PD (or NE) can be assigned as the TL visit response in these cases. However, for visits with ≤1/3 lesion assessments not recorded, the scaled-up sum of TLs diameters is 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 are 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 is recorded for one of the TL sizes and the other TL size is recorded as 0cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs, between CT and MRI is considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions are treated as missing.

3.3.8.2 Non-target lesions (NTLs) and new lesions

At each visit the investigator records an overall assessment of the non-target lesion (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 is derived based on the investigator's overall assessment of NTLs as described in Table 3:

Table 3 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).
Non-CR/non-PD	Persistence of one or more NTLs with no evidence of progression.
Progression (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.
Not evaluable (NE)	Only relevant when one or some of the NTLs are not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

NTL non-target lesion; TL 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 tumour 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 are also 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 indicates 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 tumour.

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

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

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 that the new lesion response should be treated as NE in the derivation of the overall visit response.

3.3.8.3 Overall visit response – site investigator data

Table 4 defines how the previously defined TL and NTL visit responses are combined with new lesion information to give an overall visit response.

Table 4 Overall visit response

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**

CR complete response; PR partial response; SD stable disease; NE not evaluable; NA not applicable; PD progressive disease.

3.3.8.4 Independent review

It is planned that a blinded independent central review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans for all participants (including those at unscheduled visits, or outside visit windows) are collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans are reviewed by two independent radiologists using RECIST 1.1 and are 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 participant, the BICR defines the overall visit response (CR, PR, SD, PD, NED, NE) (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. If a participant has had a tumour assessment that cannot be evaluated, then the participant is assigned a visit response of NE (unless there is evidence of progression in which case the response is 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 are 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 is selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision is used). In the absence of adjudication, the records for all visits for a single reviewer is used. The reviewer selected in the absence of adjudication is the reviewer who read the baseline scan first. The records from the single selected reviewer are used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in TL dimensions. Endpoints (of ORR, PFS, DoR and DCR) are derived programmatically from this information.

Results of this independent review are not communicated to investigators and the management of participants is based solely upon the results of the RECIST 1.1 assessment conducted by the investigator.

An independent central review (ICR) of all participants will be performed for the primary analysis of PFS, which will cover all the scans up to the DCO.

Further details of the BICR are documented in the Independent Review Charter (IRC).

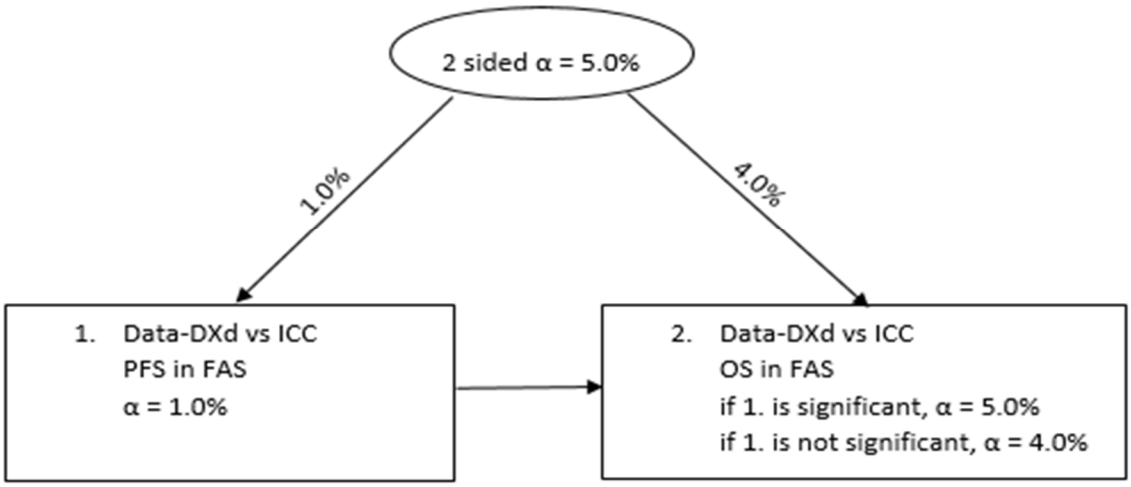
3.3.9 Multiplicity/Multiple Comparisons

The hypotheses of interest with regards to the efficacy for the dual primary endpoints are:

- H0: No differences between Dato-DXd and ICC for PFS and OS.
- H1: Differences between Dato-DXd and ICC for PFS and/or OS.

To preserve the overall type 1 error (familywise error rate) at 5% in the strong sense, a multiple testing procedure (MTP) for the dual primary endpoints of PFS and OS is implemented at DCO2, DCO3 and DCO4. An overview of the MTP with an alpha-splitting and exhaustive recycling strategy (Burman, Sonesson, & Guilbaud, 2009) is provided in Figure 1.

Figure 1 Multiple Testing Procedure



An alpha level of 1.0% will be allocated to the PFS primary analysis and the remaining 4.0% alpha level will be allocated to the OS analyses. If the PFS primary analysis meets statistical significance, the 1.0% type 1 error allocated to PFS endpoint will be reallocated (Burman, Sonesson, & Guilbaud, 2009) to the OS endpoint for a total 2-sided type 1 error of 5.0%. If the PFS primary analysis does not meet statistical significance, the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Alpha spending functions are applied for the OS endpoint in order to preserve the overall 2-sided type 1 error (familywise error rate) in the strong sense across the three planned analyses of OS.

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including 2 interim analyses for superiority of OS. Details are provided in the Interim Analysis section (Section 5).

The significance level alpha for OS across the three analysis times is dependent on the OS information fraction (number of OS events at interim/number of OS events at primary). The significance levels are calculated at the time of the analyses based on the number of events observed.

No multiplicity adjustment is applied for other endpoints as other endpoints are considered supportive endpoints.

3.3.10 Handling of Protocol Deviations in Study Analysis

Protocol deviations are collected, reviewed and reconciled throughout the study. Important protocol deviations (IPDs) are identified from the complete set of protocol deviations. IPDs are those which may significantly impact the reliability of the study data or that may significantly affect a participant's rights, safety, or wellbeing.

A set of pre-determined IPDs are listed in the protocol deviations plan. The protocol deviations plan also indicates which IPDs are identified by programmatic checks.

The IPDs are grouped into the following important protocol deviation (IPD) categories, where full details of the individual IPDs within each IPD category are provided in the protocol deviations plan:

- Inclusion criteria deviations.
- Exclusion criteria deviations.
- Discontinuation criteria for study product met but participant not withdrawn from study treatment.
- Discontinuation Criteria for overall study withdrawal met but participant not withdrawn from study.
- IP deviation.
- Excluded medications taken.
- Deviations to study procedure.
- Other important deviations.

The following general categories will be considered important protocol deviations (IPDs) and will be programmatically derived from Veeva Clinical Vault (VCV) data. These will be listed and summarised by randomised treatment group and discussed in the CSR as appropriate. Refer to the CSP for full details of the inclusion/exclusion criteria.

- Inclusion criteria deviations (Deviation 1).
  - Lack of provision of informed consent prior to any study-related procedures.
  - Inclusion criteria 2, 3, 6, 7, 8, 9
- Exclusion criteria deviations (Deviation 2).
  - Exclusion criteria 19, 22, 23, 24, 27-32, 40
- Discontinuation criteria for study product met but participant not withdrawn from study treatment (Deviation 3).
- Discontinuation criteria for overall study withdrawal met but participant not withdrawn from study (Deviation 4). (All withdrawal criteria were reviewed and not deemed to be

IPDs because they are neither considered affecting participant safety/rights/wellbeing nor affecting the efficacy assessment).

- Investigational product deviation (Deviation 5).
- Excluded medications taken (Deviation 6).
- Deviations related to study procedure (Deviation 7)
- Other important deviations (Deviation 8)

Participants who receive incorrect treatment at any time will be included in the safety analysis set as described in Section 3.2.3. 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 global study director and/or statistician.

None of the deviations will lead to participants being excluded from the analysis sets described in Section 3.2 (except for the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding participants with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed by repeating the PFS analysis excluding participants with deviations that may affect the efficacy of trial therapy. 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.

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Study Population

The domain study population covers participant disposition, analysis sets, protocol deviations, demographics, baseline characteristics, disease characteristics, medical history, prior and concomitant medication.

Study population data is summarised and listed using the FAS unless otherwise stated.

4.1.1 Participant Disposition and Completion Status

4.1.1.1 Presentations

Participant disposition is summarised by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine, presenting the number of participants screened, the number of participants not randomised along with reasons for not randomised, the number and percentage of participants: randomised, who were randomised and did not receive study treatment, who started IP (where participants starting Dato-DXd or ICC are presented according to the randomisation), who are ongoing in the study at the DCO, and who completed and discontinued the study along with reasons for study discontinuation.

The denominator for the percentages in these summaries is based on the number of randomised participants for the treatment group.

The number and percentage of participants ongoing on IP at the time of the DCO and who discontinued each study treatment along with the reasons for treatment discontinuation is also summarised, where the denominator for the percentages is based on the number of participants that started the corresponding study treatment.

Additionally, a summary of recruitment by region, country and site is produced for the FAS. The denominator for the percentages in this summary is based on the number of participants for the analysis set for the treatment group.

Listings presenting details of discontinuations by individual participant are produced for those participants discontinuing treatment and discontinuing the study.

4.1.2 Analysis Sets

The number of participants in each analysis set and reasons for exclusion from each analysis set are summarised.

A listing of individual participants not included in each analysis set is provided.

4.1.3 Protocol Deviations

Refer to Section 3.3.10 for details regarding the definition and derivation of protocol deviations.

A summary table is produced showing the number and percentage of participants with any IPD and by category of IPD, which includes the individual IPDs as detailed in the protocol deviations plan.

The individual participant data for IPDs is also listed.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Demographics are comprised of age, age group, sex, race and ethnicity.

Age is calculated as age at last birthday in whole years using the date of randomisation and date of birth. Where a partial date of birth has been collected, it is imputed as described in Section 3.3.7 to calculate the participant's age for use in listings and summaries tables presenting age and/or age group and subgroup analyses based on age. Age is split into the following categories: <65 and ≥65 years.

Date of birth is listed as it has been collected on the eCRF.

4.1.4.2 Presentation

A summary table of demographic data specified in Section 4.1.4.1 is produced, and demographic data is listed.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Baseline characteristics include height (cm), weight (kg), weight group, body mass index (BMI) (kg/m^2), body surface area (BSA) (m^2) and ECOG performance status at baseline.

Weight (kg) is categorised into weight groups of:

- <65
- ≥65 and ≤90
- >90

The body mass index (BMI) is calculated as

BMI (kg/m^2) = Weight (kg) / {Height (m)}^{2}

The BSA is calculated as

$$\text{(BSA) (m}^2\text{)} = \sqrt{\frac{\text{Height (cm)} \; X \; \text{Weight (kg)}}{3600}}$$

Stratification factors (number of previous lines of chemotherapy, geographic region (with countries within region), and prior use of CDK4/6 inhibitor) are derived from the CRF data as well as recorded by IRT.

4.1.5.2 Presentation

Baseline characteristics specified in Section 4.1.5.1 are summarised and listed.

A summary of the stratification factors recorded both by IRT and CRF is also provided presenting the number and percentage of participants in each category.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Disease characteristics include time from most recent disease progression to randomisation, overall disease classification (metastatic/locally advanced), duration of prior breast cancer CDK4/6 inhibitor therapy (<12 months, >=12 months, missing duration, NA), HER2 status, estrogen receptor status, progesterone receptor status, prior lines of chemotherapy, prior lines of anti-cancer therapy, visceral metastases and de novo metastatic breast cancer. Additionally, sites of breast and regional lymph node/other locally advanced sites and all other sites is included.

Time from most recent disease progression to randomisation (days) is defined as:

(Date of randomisation - date of most recent disease progression)

Time from most recent disease progression to randomisation is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

Time from diagnosis to randomisation in years is defined as:

(Date of randomisation – original diagnosis date) / 365.25

Time from diagnosis to randomisation is categorised into groups of ≤5 years and >5 years.

Early relapse is defined as the time between end of (neo) adjuvant chemotherapy and start of first line treatment for relapse to inoperable/metastatic disease being less than one year.

Hepatic function status at baseline is defined as follows:

- Normal
  - Total bilirubin ≤ULN and AST ≤ULN (except for Gilbert syndrome participants)
  - Total bilirubin ≤3x ULN and AST ≤ULN for participants with Gilbert syndrome
- Mild impairment
  - Total bilirubin >ULN, ≤1.5x ULN and any AST except for participants with Gilbert syndrome
  - Total bilirubin >ULN, ≤3x ULN and AST >ULN for participants with Gilbert syndrome
  - Total bilirubin <ULN and AST >ULN regardless of Gilbert syndrome

Renal function status at baseline is defined as follows:

- Normal: serum creatinine clearance ≥90 mL/min
- Mild impairment: serum creatinine clearance ≥60, <90 mL/min
- Moderate impairment: serum creatinine clearance >30, <60 mL/min

4.1.6.2 Presentation

The following breast cancer characteristics at diagnosis are summarised: American joint committee on cancer (AJCC) stage, histology type, tumour grade and time from diagnosis to randomisation.

In addition, the disease characteristics (described in Section 4.1.6.1) are summarised.

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Medical history and relevant surgical history are coded using the medical dictionary for regulatory activities (MedDRA) [using the latest or current MedDRA version].

Any medical history which is ongoing at time of informed consent is considered an ongoing condition, otherwise it is considered past medical history.

4.1.7.2 Presentation

Summary tables of past medical history, ongoing conditions and surgical history are presented by MedDRA system organ class (SOC) and preferred term (PT). Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

All medical history, surgical history and ongoing conditions are listed.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

Information on any treatment used from the time of informed consent up to 28 (+7) days after discontinuation of study treatment, with reasons for the treatment, will be recorded in the eCRF. Thereafter, only subsequent regimens of anti-cancer therapy will be recorded in eCRF.

Other anti-cancer therapies, investigational agents, and radiotherapy (except palliative to areas other than the chest) should not be given while the participant is on study drug.

Medications received prior to, concomitantly, or post-treatment are coded using the Anatomical Therapeutic Chemical (ATC) Classification codes.

Medications will be coded using the latest World Health Organization Drug dictionary (WHO Drug) version. The version used will be indicated in the data summaries and listings.

For inclusion in the prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.3.7.

Prior medications, concomitant medications and post-study treatment medications are defined as follows:

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

Missing coding terms are listed and summarised as "Not coded".

Time from completion of prior anti-cancer therapy to randomisation is defined as

(date for randomisation – last anti-cancer therapy stop date + 1)

Time from completion of prior anti-cancer therapy is categorised into groups of >0 - <2 weeks, ≥2 weeks - <1 month, ≥1 - <2 months, ≥2 - <3 months and ≥3 months.

4.1.8.2 Presentation

The following summaries will be produced:

- Summary of prior medications
- Summary of concomitant medications
- Summary of prohibited concomitant medications
- Summary of post study treatment medications
- Summary of prior disease-related treatment modalities
- Summary of prior cancer therapies
- Summary of time from completion of prior anti-cancer therapy to randomisation
- Summary of non-study cancer therapies whilst on study treatment
- Summary of post study treatment cancer therapies
- Summary of Trastuzumab Deruxtecan and Sacituzumab Govitecan as post study treatment cancer therapies
- Summary of prior radiotherapy
- Summary of on study palliative radiotherapy

- Summary of post study treatment radiotherapy

Prior medications (excluding prior cancer therapies), concomitant medications (including both allowed and prohibited concomitant medications), prohibited concomitant medications and post-study treatment medications (excluding post-study treatment cancer therapies) are presented by ATC classification and generic term, sorted by descending frequency of ATC group and generic term. Participants taking the same concomitant medication/procedure multiple times are counted once per ATC classification and generic term.

A summary of number and percentage of participants receiving prior disease-related treatment modalities are summarised by modality. Duration of prior cytotoxic chemotherapy (<6 months, >=6 months, NA), prior hormonal therapy in the inoperable or metastatic setting (<6 months, >=6 months, NA) and prior breast cancer CDK4/6 inhibitor therapy (<12 months, >=12 months, NA) are summarised.

Prior cancer therapies and non-study cancer therapies whilst on study treatment are summarised by therapy class and ATC group. Post-study treatment cancer therapies are summarised by ATC group and generic term. In addition, prior cancer therapy in the inoperable or metastatic setting is summarised by therapy class. Prior chemotherapy in the inoperable or metastatic setting is summarised by generic term.

Trastuzumab Deruxtecan and Sacituzumab Govitecan as post study treatment cancer therapies are summarised by generic term.

A separate summary of number and percentage of participants receiving prior radiotherapy is produced and repeated for on study palliative radiotherapy and post study treatment discontinuation radiotherapy.

Time from completion of prior anti-cancer therapy to randomisation will be presented using summary statistics (e.g. n, mean, median, standard deviation, min, max). In addition, a separate summary of the number and percentage of participants in each category (as described in Section 4.1.8.1) is produced.

All concomitant, prior and post study treatment discontinuation medication and therapy data are listed for all participants.

4.1.9 Study Drug Compliance

This is covered in Section 4.4.1

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses.

Table 5 Endpoint analyses

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Dual Primary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting, per BICR					
Primary	PFS - BICR Assessments	FAS	Participants who have not progressed or died are censored at latest evaluable RECIST 1.1 assessment. Participants who progress or die after 2 missed visits are censored at last evaluable RECIST 1.1 assessment prior to the two missed visits.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.4
Sensitivity 1 (Evaluation time bias)	PFS - BICR Assessments	FAS	As for primary analysis, but the midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be used as the event time.	Stratified log-rank test.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity 2 (Attrition bias)	PFS - BICR Assessments	FAS	As for primary analysis, but using the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments will be included. Participants 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.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. Reverse Kaplan-Meier plot	4.2.1.5
Sensitivity 3 (Ascertainment bias)	PFS - BICR Assessments	FAS	As for primary analysis but using the site Investigator data which is a secondary efficacy variable.	Discrepancy between primary analysis using BICR assessments and the secondary analysis using Investigator assessment.	4.2.1.5
Sensitivity 4 (Subsequent anti-cancer therapy)	PFS - BICR Assessments	FAS	As for primary analysis, but participants who receive subsequent anti-cancer therapy prior to progression or death are censored at latest evaluable RECIST assessment prior to subsequent anti-cancer therapy.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.1.5

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity 5 (Stratification according to eCRF)	PFS - BICR Assessments	FAS	As for primary analysis.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test.	4.2.1.5
Dual Primary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of OS in participants with inoperable or metastatic HR-positive, HER2-negative breast cancer, who have been treated with 1 or 2 lines of chemotherapy in the inoperable/metastatic setting					
Primary	OS	FAS	Participant not known to have died will be censored at last date participant was known to be alive.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.2.4
Sensitivity 1 (Attrition bias)	OS	FAS	As for primary analysis but censoring indicator of OS is reversed.	Reverse Kaplan-Meier plot Median	4.2.2.5
Sensitivity 2 (Stratification according to eCRF)	OS	FAS	As for primary analysis.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test.	4.2.2.5
Secondary Objective 1: To demonstrate superiority of Dato-DXd relative to ICC by assessment of ORR per BICR and per investigator assessment					
Secondary	ORR	FAS	Participants without a response included as non-responders. If participants discontinue treatment without response or progression, receive a subsequent anti-cancer therapy and then respond are not included as responders.	Odds ratio from logistic regression model ORR with Clopper-Pearson CI Difference in ORR with Miettinen-Nurminen CI	4.2.3.3
Secondary Objective 2: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DoR per BICR and per investigator assessment					

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary	DoR	FAS (for participants who have confirmed CR or PR)	For participants who do not progress or die following a response, the censoring rules follow the rules for PFS censoring for the primary analysis.	Summary statistics. KM plots.	4.2.5.3
Supplementary	DoR	FAS (for participants who have confirmed CR or PR)	As for primary analysis but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.	Summary statistics. KM plots.	4.2.5.4
Secondary Objective 3: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS, per investigator assessment					
Secondary	PFS – Investigator Assessments	FAS	As described for primary PFS	As described for primary PFS	4.2.6.4
Secondary Objective 4: To demonstrate superiority of Dato-DXd relative to ICC by assessment of DCR per BICR and per investigator assessment					
Secondary	DCR at 12 weeks	FAS	Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.	Odds ratio from logistic regression model	4.2.7.3

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary Objective 5: To assess pain in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in pain	FAS	Participants who have not deteriorated will be censored at last evaluable assessment. Participants who deteriorate after 2 missed visits are censored at last evaluable assessment prior to the 2 missed visits. If a participant has no evaluable post-baseline data or does not have baseline data, they will be censored at date of randomisation.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots.	4.2.8.4
Sensitivity	TTD in pain	FAS	TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.	As described for secondary TTD in pain.	4.2.8.5
Secondary Objective 6: To assess physical functioning in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4
Sensitivity	TTD in physical functioning	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5
Secondary Objective 7: To assess global health status/quality of life (GHS/QoL) in participants treated with Dato-DXd relative to ICC					
Secondary	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.4

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Sensitivity	TTD in GHS/QOL	FAS	As described for TTD in pain.	As described for TTD in pain.	4.2.8.5
Secondary Objective 8: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TFST					
Secondary	TFST	FAS	Participants not known to have had a first subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a first subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a first subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	Hazard Ratio from stratified Cox proportional hazards model. Stratified log-rank test. KM plots. Median. Time between progression and starting subsequent therapy.	4.2.9.3
Secondary Objective 9: To demonstrate superiority of Dato-DXd relative to ICC by assessment of TSST					
Secondary	TSST	FAS	Participants not known to have had a second subsequent anti-cancer therapy and who have not died are censored at the last date that they were known not to have received a second subsequent anti-cancer therapy. Participants who terminate the study for reason other than death prior to receiving a second subsequent anti-cancer therapy are censored at earliest of date last known to be alive and termination date.	As described for TFST.	4.2.10.3
Secondary Objective 10: To demonstrate superiority of Dato-DXd relative to ICC by assessment of PFS2					

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Secondary	PFS2	FAS	As described for primary PFS	As described for primary PFS	4.2.11.4
Secondary Objective 11: To assess the PK of Dato-DXd 6mg/kg IV Q3W					
Secondary	Plasma concentrations of Dato-DXd, total anti-TROP2 antibody, and MAAA-1181a (payload)	PAS	N/A	Summary statistics	4.2.12.2
Secondary Objective 12: To investigate the immunogenicity of Dato-DXd 6mg/kg IV Q3W					
Secondary	Presence of ADA	ADA evaluable set	N/A	Summary statistics	4.2.13.2
Safety: To assess the safety and tolerability profile of Dato-DXd relative to ICC					
Safety	Type, incidence and severity (graded by NCI CTCAE v5.0), seriousness and relationship to study medication of AEs	SAF	N/A	Descriptive	4.4.2
Safety	ECOG PS	SAF	N/A	Descriptive	4.4.9
Safety	Vital signs	SAF	N/A	Descriptive	4.4.6
Safety	Physical examination	SAF	N/A	Descriptive	4.4.10
Safety	Clinical laboratory tests	SAF	N/A	Descriptive	4.4.3, 4.4.4,4.4.5
Safety	ECGs	SAF	N/A	Descriptive	4.4.7
Safety	Echocardiogram/ multigated acquisition	SAF	N/A	Descriptive	4.4.8
Safety	Ophthalmologic assessments	SAF	N/A	Descriptive	4.4.11

4.2.1 Primary Endpoint - Progression Free Survival by BICR

4.2.1.1 Definition

Progression-free survival is defined as the time from the date of randomisation until the date of objective disease progression, as defined by RECIST 1.1, or death (by any cause in the absence of progression) regardless of whether the participant withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1). PFS is assessed by BICR assessment. A secondary endpoint analysis of PFS by investigator assessment is reported.

4.2.1.2 Derivations and Censoring Rules

Participants who have not progressed or died at the time of analysis are censored at the time of the latest date of assessment from their last evaluable RECIST assessment. However, if the participant progresses or dies immediately after two or more consecutive missed visits, the participant is 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).

If the participant has no evaluable RECIST assessment or does not have baseline data, they will be censored at the date of randomisation, unless they die within 2 scheduled scans of baseline (12 weeks + 1 week allowing for a late assessment within the visit window) in which case they are treated as an event with date of death as the event date. PFS censoring rules are described in Table 6.

Table 6 Outcome and date of event for PFS analysis

Scenario	Date of PD/ Death event or Censoring	PFS Outcome
First progression or death after at most 1 missed visit, regardless of whether the participants withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression	Date of assessment of first progression or death (by any cause in the absence of progression)	Event
No baseline or evaluable RECIST assessment and death within 2 RECIST visits after the date of randomisation	Date of death	Event
No baseline or evaluable RECIST assessment and no death within 2 RECIST visits after the date of randomisation	Day 1 (Date of randomisation)	Censored
No PD or death at time of data cut-off	Date of last evaluable RECIST assessment*	Censored

Scenario	Date of PD/ Death event or Censoring	PFS Outcome
Death or progression after two or more missed RECIST visits	Date of last evaluable RECIST assessment* prior to the 2 missed visits	Censored

*: if there are no evaluable post-baseline assessments prior to PD or death or data cut-off, participants will be censored at the date of randomisation.

Given the scheduled visit assessment scheme (i.e. every six weeks for the first 48 weeks then every nine weeks thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to nine-weekly this will equate to 17 weeks (i.e. take the average of 6 and 9 weeks which gives 7.5 weeks and then apply same rationale, hence 2 x 7.5 weeks + 1 week for an early assessment + 1 week for a late assessment = 17 weeks). The time period for the previous RECIST assessment will be from study days 288 to 329 (i.e. week 41 to week 47). From week 47 onwards (when the scheduling changes to nine-weekly assessments), two missing visits will equate to 20 weeks (i.e. 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks).

The following is also summarised in Table 7:

Table 7 Definition of two missed RECIST visits

Scheduled Assessment	Previous RECIST assessment	Two missed RECIST visits window
Q6W	No evaluable RECIST visits or no baseline RECIST scan	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W	Day 1	2 x 6 weeks + 1 week = 13 weeks (91 days)
Q6W up to Week 48	>Day 1 – Day 287 (up to Week 41)	2 x 6 weeks + 2 weeks = 14 weeks (98 days)
	>Day 287 – Day 329 (Week 41 – Week 47) (change period from Q6W to Q9W)	2 x [(6 weeks+9 weeks)/2] + 2 weeks = 17 weeks (119 days)
Q9W thereafter	>Day 329 onwards	2 x 9 weeks + 2 weeks = 20 weeks (140 days)

The PFS time is always derived based on scan/assessment dates and not on visit dates.

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

- For investigator assessments, the date of progression is determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression.
- For BICR assessments, the date of progression is determined based on the earliest of the 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 BICR data.
- For both BICR and investigational assessments when censoring a participant for PFS, the participant is censored at the latest of the scan dates contributing to a particular overall visit assessment.

Note: for TLs only the latest scan date is recorded 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.

4.2.1.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Sections 4.2.1.2 and 4.2.1.5.

4.2.1.4 Primary Analysis of Progression Free Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of PFS in the FAS. The primary analysis of PFS is based on the BICR assessment of PD by RECIST 1.1.

The null hypothesis for the dual primary time to event endpoint of PFS is that there is no difference between Dato-DXd and ICC in the probability of a progression event in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for PFS.

H1: Differences between Dato-DXd and ICC for PFS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

PFS is analysed using a stratified log-rank test (using the LIFETEST procedure with a TEST statement) adjusting for the stratification factors of number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor.

The stratification variables are based on the values entered into IRT at randomisation, even if it is subsequently discovered that these values were incorrect.

If there are less than 10 events in total for a unique stratum or less than 2 events in either treatment group for a unique stratum then the strata are combined in the following order. The CDK4/6 strata (Yes vs No) are pooled first, followed by the number of previous lines of chemotherapy strata (1 vs 2) and then finally by the geographic region strata (United States, Canada, Europe vs Rest of World).

The hazard ratio (HR) and its confidence interval (CI) (95% and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9) and p-value are presented. The HR and CI are estimated from a stratified Cox Proportional Hazards model (with ties = Efron and stratification variables number of previous lines of chemotherapy, geographic region, and prior use of CDK4/6 inhibitor) and the CI calculated using a profile likelihood approach. A HR less than 1 favours Dato-DXd.

Estimates and 95% CI for PFS rates at 3 months intervals and median PFS for each treatment group are presented.

Proportionality assumption

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of log(-log(survival probability)) versus log (time) and, if these raise concerns, by fitting a time dependent covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods for example 0-6m, 6-12m etc. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over the observed extent of follow-up unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

Summaries

In addition to the analyses described above, the following supportive summaries are produced.

Kaplan-Meier (KM) plots of PFS are presented by treatment group. Summaries of the number and percentage of participants experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided for each treatment.

The treatment status at progression of participants at the time of analysis is summarised. This includes the number (%) of participants who were on treatment at the time of progression, the number (%) of participants who discontinued IP prior to progression, the number (%) of participants who have not progressed and were on IP or discontinued IP.

The number of participants censored may be summarised by treatment group together with baseline prognostic factors of the censored participants. This number and percentage of prematurely censored participants is summarised. A participant will be defined as prematurely censored if they did not progress (or die in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval (+ 2 weeks) prior to the DCO date.

Additionally, summary statistics are given for the number of days from censoring to DCO for all censored participants.

The duration of follow-up is summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) participants only, presented by treatment group.

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

Summaries of the number and percentage of participants who miss two or more consecutive RECIST assessments is presented for each treatment group.

All of the collected RECIST 1.1 data is listed for all randomised participants. In addition, a summary of new lesions (i.e. sites of new lesions) is produced.

4.2.1.5 Sensitivity Analyses of Progression Free Survival

Sensitivity Analysis 1 - 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 analysed 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 participants 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 & Chen, 2010). To support this analysis, the mean of participant-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR RECIST assessments.

Sensitivity Analysis 2 - Attrition bias

Attrition bias is assessed by repeating the primary PFS analysis except that the actual PFS event times, rather than the censored times, of participants who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments are included. In addition, and within the same sensitivity analysis, participants 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 are censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis is supported by a KM 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.

Sensitivity Analysis 3 - Ascertainment bias

Ascertainment bias is assessed by analysing the site investigator data which is a secondary efficacy endpoint (analysis methods presented in Section 4.2.6). The stratified log rank test is repeated on PFS using the site investigator data based upon RECIST. The HR and CI are presented.

If there is an important discrepancy between the primary analysis using the BICR data and this sensitivity analysis using site investigator data a summary table is produced showing the number and proportion of participants with site but no central confirmation of progression and with progression determined by central review but not at site. Such participants 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, Gaschler-Markefski, & Bluhmki, 2001), but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group.

Sensitivity Analysis 4 – Subsequent Anti-cancer Therapy

An additional sensitivity analysis is produced which is a repeat of the primary analysis for PFS, but the censoring rule is modified so that participants who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death are censored at their last evaluable assessment prior to taking the subsequent anti-cancer therapy.

Sensitivity Analysis 5 – Stratification according to eCRF

In the event that there are any mis-stratifications during randomisation, the stratified log rank test will be repeated on PFS, where the stratification factors are as recorded according to the eCRF. The HR and CI will also be presented from the Cox proportional hazards analysis.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival.

Deviation Bias Sensitivity Analysis

A 'deviation bias' sensitivity analysis may be performed as discussed in Section 3.3.10.

4.2.1.6 Subgroup Analyses

Subgroup analyses are conducted comparing PFS between the treatments for the following subgroup of the FAS:

1. Stratification factors at randomisation:
    - Number of previous lines of chemotherapy: 1, 2
    - Geographic region: Region 1 [US, Canada, Europe], Region 2 [Rest of World]
    - Prior use of CDK4/6 inhibitor: Yes, No
2. Exploratory factors
    - Prior use of taxanes and/or anthracyclines: taxanes alone, anthracyclines alone, both taxanes and anthracyclines, neither taxanes nor anthracyclines
    - Age at randomisation: <65, ≥65 years of age
    - Race: Asian, non-Asian
    - Pre-selected investigator's choice of chemotherapy: Capecitabine, Gemcitabine, Eribulin, Vinorelbine
    - Brain metastases (at baseline): Yes, No
    - Sex: male, female
    - HER2 status: IHC 0 versus IHC 1+ and [IHC 2+ cases that are FISH-]
    - ECOG performance score: 0, 1
    - Duration of prior use of breast cancer CDK4/6 inhibitor: Yes, ≤12 months, >12 months

- Early relapse: Yes, No
- Duration of prior use of endocrine therapy in the metastatic breast cancer setting: <6 months, ≥6 months

The subgroup analyses for the pre-selected investigator's choice of chemotherapy will be based on the values obtained from the IRT system; all other factors will be based on values recorded on the eCRF, or from the third-party vendor data.

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

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

For each subgroup level of a factor, the HR and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties and using a BY statement for the subgroup factor.

These HRs and associated two-sided 95% profile likelihood CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

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 descriptive summaries will be provided.

The presence of quantitative interactions may be assessed by means of an overall global interaction test for strata and possibly subgroups:

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by treatment interaction terms, with one that excludes the interaction terms, and will be assessed at the 2-sided 10% significance level.

If there are not more than 10 events per stratum for any covariate (i.e. within each stratum of a treatment*covariate interaction (2 treatments * 2 levels of the covariate = 4 stratum)), a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria, then the covariate-by-treatment interaction term should be

omitted from the model. Moreover, if the covariate does not have more than 10 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved, then it will be concluded that overall, the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant, and all excluded interactions are non-significant. Throughout this process all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

4.2.2 Primary Endpoint - Overall Survival

4.2.2.1 Definition

Overall survival is defined as the time from the date of randomisation until death due to any cause regardless of whether the participant withdraws from randomised therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomisation + 1).

4.2.2.2 Derivations and Censoring Rules

Any participant not known to have died at the time of analysis is censored based on the last recorded date on which the participant was known to be alive.

Note: Survival calls are made in the week following the date of data cut-off (DCO) for the analysis, and if participants are confirmed to be alive or if the death date is after the DCO date, these participants are censored at the date of DCO. This is done at DCO2, DCO3 and DCO4. The status of ongoing, withdrawn (from the study) and “lost to follow-up” participants at the time of the primary OS analysis should be obtained by the site personnel by telephone contact with the participant, participant’s family, by contact with the participant’s current physician, or local death registries. If the participant has actively withdrawn consent to the processing of their personal data, the vital status of the participant can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For the OS analysis at DCO2 and DCO3, performed prior to the primary 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 participant was known to be alive for those participants still on treatment (since the SURVIVE module is only completed for participants off treatment if a survival sweep is not performed). The last date for each

individual participant is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start and stop dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anti-cancer treatment
- Date last known alive on survival status CRF
- End of study date

Duration of follow-up is derived as time from randomisation to the date of death (i.e. overall survival) or to the date of censoring (date last known to be alive).

4.2.2.3 Handling of Dropouts and Missing Data

If a participant is known to have died where only a partial death date is available, then the date of death is imputed following the rules in Section 3.3.7.

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

4.2.2.4 Primary Analysis of Overall Survival

Statistical Analysis

One primary objective of the study is to demonstrate the superiority of Dato-DXd relative to ICC by assessment of OS in the FAS.

The null hypothesis for the dual primary time to event endpoint of OS is that there is no difference between Dato-DXd and ICC in the probability of a death in the FAS at any time point. The intention of the study is to demonstrate the superiority of Dato-DXd over ICC.

H0: No differences between Dato-DXd and ICC for OS.

H1: Differences between Dato-DXd and ICC for OS.

NOTE: as there are dual primary endpoints, the significance levels will be determined using the MTP for PFS and OS as described in Section 3.3.9.

Overall survival will be analysed using the same methodology as that used for the primary analysis of PFS. The treatment effect of Dato-DXd against ICC will be estimated by the HR together with its 95% CI and the appropriate CI according to the significance level in the MTP as described in Section 3.3.9.

Estimates and 95% CI for OS rates at 6 monthly intervals are presented along with the median OS for each treatment group.

Summaries

Kaplan-Meier (KM) plots of OS are presented by treatment group. Summaries of the number and percentage of participants who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided.

In addition, the median duration of follow-up is presented for censored participants by treatment group, and for all participants by treatment group and overall.

4.2.2.5 Sensitivity Analyses of Overall Survival

Sensitivity Analysis 1 - Attrition bias

A sensitivity analysis for OS examining the censoring patterns to rule out attrition bias with regard to the primary treatment comparisons is achieved by a KM plot of time to censoring where the censoring indicator of OS is reversed. The KM estimates of median follow-up (overall and by treatment group) are also summarised.

Sensitivity Analysis 2 – Stratification according to eCRF

In the event that there are any mis-stratifications during randomisation, the stratified log rank test will be repeated on OS, where the stratification factors are as recorded according to the eCRF. The HR and CI will also be presented from the Cox proportional hazards analysis.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of overall survival.

4.2.2.6 Subgroup Analyses

Subgroup analyses will be conducted for OS using the same methodology as described for PFS in Section 4.2.1.6.

4.2.3 Secondary Endpoint - Objective Response Rate

4.2.3.1 Definition

Both unconfirmed and confirmed ORR are assessed, where:

- Confirmed ORR is the percentage of participants with an investigator-assessed response of CR or PR 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.
- Unconfirmed ORR is the percentage of participants with at least one investigator-assessed visit response of CR or PR.

The denominator will be defined as all randomised participants.

ORR will also be defined using the BICR data to define a visit response of CR or PR, with the denominator defined as all randomised participants.

4.2.3.2 Derivations

Data obtained up until progression, or last evaluable assessment in the absence of progression, are included in the assessment of ORR, regardless of whether the participant withdraws from therapy. Participants who discontinue randomised treatment without progression, receive a subsequent anti-cancer therapy and then respond are not included as responders in the ORR (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy). For confirmed ORR both visits contributing to a response must be prior to subsequent therapy for the participant to be considered as a responder.

For confirmed ORR, in the case where a participant 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 participant is defined as a responder. Similarly, if a participant 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 is assigned.

ORR is based on all scans regardless of whether they were scheduled or not.

A participant is classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation and confirmed by repeat imaging, prior to RECIST progression and prior to starting any subsequent cancer therapy.

4.2.3.3 Primary Analysis of Objective Response Rate

ORR based on both a confirmed and unconfirmed tumour response are analysed.

A logistic regression model is fitted to tumour response (yes/no) including treatment and the same stratification factors as the primary PFS endpoint as fixed effects. The results of the analysis are presented in terms of an adjusted odds ratio (an odds ratio greater than 1 favour Dato-DXd) together with its associated profile likelihood 95% CI (e.g. using the option ‘LRCI’ in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The adjusted Least Squares (LS) Means response rate (using LSMEANs statement with OM option)

from the logistic regression model together with the corresponding 95% CI is presented for each treatment group.

If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran–Mantel–Haenszel (CMH) test is presented. The CMH test is stratified using the same stratification factors as the primary PFS endpoint. The results of the analysis are presented in terms of an odds ratio together with the 95% CI and p-value. The odds ratio, 95% CI and p-value are obtained using SAS PROC FREQ and the CMH test option.

Both unconfirmed and confirmed ORR as assessed by site investigator will be estimated and presented along with the corresponding exact 95% Clopper-Pearson CI for each treatment arm. The difference in ORR between treatment arms will be reported using point estimates and their two-sided 95% CIs by the Miettinen-Nurminen method (Miettinen & Nurminen, 1985). A Summary will be produced that presents the number and percentage of participants with both an unconfirmed and a confirmed tumour response (CR/PR).

Summaries will also be produced for ORR per BICR.

4.2.4 Secondary Endpoint - Best Objective Response

4.2.4.1 Definition

Best objective response (BoR) is a supportive endpoint for ORR. BoR is calculated based on the overall visit responses from each RECIST assessment, described in Section 3.3.8.3. It is the best response a participant has had following randomisation, but prior to starting any subsequent anti-cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorisation of BoR for investigator data is based on RECIST using the following response categories: CR, PR, SD, PD and NE. Categorisation of BoR for BICR data is based on RECIST using the following response categories: CR, PR, SD, PD, NED and NE.

4.2.4.2 Derivations

BoR is derived using confirmed CR or PR and separately using unconfirmed CR or PR.

Confirmed CR or PR coincides with that used for the confirmed ORR endpoint. For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment is 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 randomisation. For CR/PR, the initial overall visit assessment that showed a response uses the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST from the overall visit response using all BICR data up until the first progression event. It will also be determined programmatically based on RECIST using all site investigator data up until the first

progression event. The denominators for each case will be consistent with those used in the ORR analysis.

BoR is determined based on RECIST using both all site investigator data and separately all BICR data up until the earliest of the first progression event/last evaluable assessment in the absence of RECIST or start of any subsequent cancer therapy. The denominators are consistent with those used in the ORR analysis.

For participants whose progression event is death, BoR is calculated based upon all evaluable RECIST assessments prior to death.

For participants who die with no evaluable RECIST assessments, if the death occurs ≤7 weeks (i.e. 6 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR is assigned to the progression (PD) category. For participants who die with no evaluable RECIST assessments, if the death occurs > 7 weeks after randomisation then BoR is assigned to the NE category. For participants with no evaluable RECIST assessments post randomisation, then BoR is also assigned to the NE category.

4.2.4.3 Primary Analysis of Best Objective Response

For each treatment arm, BoR for investigator data will be summarised by n (%) for each category (CR, PR, SD, PD, and NE) using confirmed and unconfirmed CR/PR responses.

For each treatment arm, BoR for BICR data will be summarised by n (%) for each category (CR, PR, SD, PD, NED, and NE) using confirmed and unconfirmed CR/PR responses.

4.2.5 Secondary Endpoint - Duration of Response

4.2.5.1 Definition

DoR will be defined as the time from the date of first documented confirmed response of CR or PR until date of documented progression per RECIST 1.1 (as assessed by Investigator assessment) or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1).

DoR will also be defined using the BICR data to define the overall visit response.

4.2.5.2 Derivations and Censoring Rules

The end of response coincides with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of CR or PR as defined in Table 4.

If a participant does not progress following a response, then the PFS censoring time is used.

For confirmed CR and PR both visits contributing to a response must be prior to subsequent therapy. Confirmation assessment must not be 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.

In the case where a participant 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 this is counted as confirmed PR. Similarly, if a participant 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 this is counted as confirmed CR.

4.2.5.3 Primary Analysis of Duration of Response

The analysis will include all randomised participants as randomised who have a confirmed response, regardless of whether the participant withdraws from therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression.

Descriptive data will be provided for the DoR, including the associated KM curves (without any formal comparison or p-value attached).

KM estimates for median DoR and 25th and 75th percentiles and corresponding 95% CI for DoR will be summarised. KM estimates for percentage remaining in response at 6 monthly intervals are presented for each treatment group.

Additionally, median, 25th and 75th percentiles for time from randomisation to onset of response will be calculated using standard descriptive statistics.

4.2.5.4 Supplementary Analyses of Duration of Response

A supplementary analysis is included where the DoR evaluation is repeated but participants who receive another anti-cancer therapy prior to progression or death are censored at the time of the latest assessment prior to receiving the new anti-cancer therapy.

4.2.6 Secondary Endpoint – PFS by Investigator Assessment

4.2.6.1 Definition

PFS by Investigator assessment is defined as the time from the date of randomisation until the date of PD, as defined by RECIST 1.1 (by Investigator assessment) or death (by any cause in the absence of progression) regardless of whether the participant withdraws from randomised therapy, receives another anti-cancer therapy or clinically progresses prior to RECIST 1.1 progression (i.e. date of PFS event or censoring – date of randomisation + 1).

4.2.6.2 Derivations and Censoring Rules

This secondary endpoint of PFS based on Investigator assessment will be derived and censored using the same methodology described in Section 4.2.1.2

4.2.6.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.1.3.

4.2.6.4 Primary Analysis of Progression Free Survival

Statistical Analysis

This secondary endpoint of PFS based on Investigator assessment will be analysed using the same methodology described in Section 4.2.1.4. The HR together with its 95% CI and p-value are presented.

Estimates and 95% CI for PFS rates at 3 months intervals and median PFS for each treatment group are presented.

Summaries

Summaries of the number and percentage of participants experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided for each treatment. The duration of follow-up is summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) participants only, presented by treatment group.

In addition a KM plot of PFS are presented by treatment group.

All of the collected RECIST 1.1 data is listed for all randomised participants. In addition, a summary of new lesions (i.e. sites of new lesions) is produced.

4.2.7 Secondary Endpoint –Disease Control Rate

4.2.7.1 Definition

Disease control rate at 12 weeks is defined as the percentage of participants who have a confirmed CR or PR or have demonstrated SD for at least 11 weeks (i.e. 12 weeks – 1 week to allow for an early assessment within the assessment window) after randomisation without subsequent cancer therapy per RECIST 1.1, as assessed per Investigator assessment and derived from the raw tumour data.

DCR at 12 weeks will also be defined using the BICR data to define the overall visit response.

4.2.7.2 Derivations

Data obtained from randomisation up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of DCR, regardless of

whether the participant withdraws from therapy. Participants who receive a subsequent therapy prior to week 11 will not be considered to have disease control in the analysis.

For confirmed CR and PR both visits contributing to a response must be prior to subsequent therapy. Confirmation assessment must not be 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.

In the case where a participant 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 this is counted as confirmed PR. Similarly, if a participant 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 this is counted as confirmed CR.

Duration of SD (weeks) is defined as: (date last evaluable assessment of SD in the absence of progression prior to subsequent cancer therapy - randomisation date +1)/7.
Participants without a post-baseline tumour assessment are considered to have no clinical benefit.

DCR is based on all scans regardless of whether they were scheduled or not.

4.2.7.3 Primary Analysis of Disease Control Rate

DCR will be analysed using the same methodology described for ORR in Section 4.2.3.3.

The analysis will be performed on the FAS

4.2.8 Secondary Endpoint – Time to Deterioration (TTD) in Pain, Physical Functioning, and Global Health Status/Quality of Life (GHS/QoL) as measured by EORTC QLQ-C30

4.2.8.1 Definition

EORTC QLQ-C30

The European Organisation for Research and Treatment of Cancer (EORTC) 30-item quality of life (QoL) questionnaire (QLQ-C30) consists of 30 questions that are combined to produce 5 multi-item functional scales (physical, role, cognitive, emotional, and social), 3 multi-item symptom scales (fatigue, pain, and nausea/vomiting), a 2-item global health status/QoL scale, 5 individual item symptom scores (appetite loss, dyspnoea, insomnia, constipation, and diarrhoea), and 1 item on the financial impact of the disease.

TTD in pain, physical functioning and GHS/QoL items are secondary endpoints. All other items are part of the exploratory endpoints.

The number of items and item range for each scale/item are displayed in Table 8.

Table 8 EORTC QLQ-C30 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Global health status/ QoL	QL	2	6	29, 30
Functional scales				
Physical	PF	5	3	1-5
Role	RF	2	3	6, 7
Cognitive	CF	2	3	20, 25
Emotional	EF	4	3	21-24
Social	SF	2	3	26, 27
Symptom scales				
Fatigue	FA	3	3	10, 12, 18
Pain	PA	2	3	9, 19
Nausea/ vomiting	NV	2	3	14, 15
Symptom items				
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17

QoL=Quality of life.

4.2.8.2 Derivations and Censoring Rules

Scoring algorithm

The EORTC QLQ-C30 v3 is scored according to the EORTC QLQ-C30 Scoring Manual (Fayers, et al., 2001). Items are scored on a 4-point verbal rating scale: “Not at all”, “A little”, “Quite a bit”, and “Very much”. Scores are then transformed to give a score from 0 to 100 for each of the symptom scales, functional scales, and the global QoL scale. Higher scores on the global health status/QoL and functional scales indicate better health status/function, but higher scores on symptom scales/scores represent greater symptom severity.

The EORTC QLQ-C30 functional and symptom scales, individual symptom items and global health status/QoL are derived as follows:

1. Calculate the average of the items that contribute to the scale or take the value of an individual item, i.e. the raw score (RS):

RS = (I1 + I2 + … + In) / n,

where I1 + I2 + … + In are the items included in a scale and n is the number of items in a scale.

2. Use a linear transformation to standardise the raw score, so that scores range from 0 to 100, where a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

   Functional scales: Score = (1 – [RS – 1] / range) * 100

   Symptom scales/items; global health status/QoL: Score = ([RS – 1] / range) * 100,

   where range is the difference between the maximum and the minimum possible value of RS.

Change from baseline

Changes in score from baseline are calculated for each of the functional scales, symptom scales and global health status/QoL scale at each assessment, where baseline is defined and calculated as explained in Section 3.3.3.

Deterioration

Deterioration is defined as change from baseline that reaches a clinically meaningful deterioration threshold. Anchor-based methods using the participant-based anchors PGIS and PGIC will be considered to define thresholds for clinically meaningful within-participant change used in the time to deterioration (TTD) endpoints. Other methods including distribution-based methods, cumulative distribution function, and probability density function curves, and methods using other anchors may also be considered.

Clinically meaningful change thresholds will be estimated for the following outcomes:

- EORTC QLQ-C30: Global health status/QoL, functioning, and select symptom subscales including pain and fatigue
- EORTC QLQ IL116: breast symptoms, arm symptoms (See Section 4.2.18)

The analysis to define clinically meaningful change thresholds in the TTD PRO endpoints will include all randomised participants using the pooled treatment arms data prior to database lock. Further details on methodologies to define these clinically meaningful change thresholds will be provided in a separate PRO psychometric analysis plan (PAP).

Improvement, deterioration or no change will be defined based on a clinically meaningful change threshold. If the estimation methods have not yet been performed to define the threshold for improvement, deterioration or no change for a given scale/item score, then a 10-point change from baseline will be used.

TTD

Time to deterioration (TTD) is defined as time from the date of randomisation to the date of first deterioration, regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

Time to deterioration=date of event or censoring – date of randomisation+1

For the derivation of TTD, the following censoring rules are used:

Participants whose symptoms, functioning or global health status/QoL have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom, functioning or global health status/QoL change from baseline could be evaluated. Also, if symptom, functioning or global health status/QoL deteriorate immediately after 2 or more missed PRO assessments, the participant will be censored at the time of the last PRO assessment where the symptom, functioning or global health status/QoL change from baseline could be evaluated prior to the 2 missed assessments.

If a participant has no evaluable post-baseline data or does not have baseline data, they will be censored at date of randomisation.

Participants whose difference between the baseline score and the maximum/minimum subscale score is less than the value of a clinically meaningful change deterioration, which doesn’t allow deterioration to occur, will be censored at date of randomisation.

Given the scheduled assessment scheme (i.e. every three weeks for the first 48 weeks then every six-weeks thereafter) the definition of 2 missed assessments will change. If the previous PRO assessment is less than study day 314 (i.e. week 44) then two missing visits will equate to 46 days since the previous PRO assessment allowing for early and late visits (i.e. 2 x 3 weeks + 2 days for an early assessment + 2 days for a late assessment = 46 days). If the two missed assessments occur over the period when the scheduled frequency of PRO assessments changes from three-weekly to six-weekly this will equate to 67 days (i.e. take the average of 3 and 6 weeks which gives 4.5 weeks and then apply same rationale, hence 2 x 4.5 weeks + 2 days for an early assessment + 2 days for a late assessment = 67 days). The time period for the previous PRO assessment will be from study days 314 to 334 (i.e. week 44 to week 47). From study day 335 (i.e. week 47) onwards (when the scheduling changes

to six-weekly assessments), two missing assessments will equate to 88 days (i.e. 2 x 6 weeks + 2 days for an early assessment + 2 days for a late assessment = 88 days). Study day will be calculated in line with Section 3.3.5. If participant withdraws treatment prior to week 48 the assessment schedule is assumed to be Q6W relative to C1D1. If participant discontinues treatment after week 48 assessment schedule is the same as above.

The following is also summarised in Table 9:

Table 9 Definition of two missed PRO visits

Scheduled Assessment	Previous PRO assessment	Two missed PRO visits window
Q3W	Day 1	2 x 3 weeks + 2 days = 44 days
Q3W up to Week 48*	>Day 1 – Day 313 (i.e. Week 44)	2 x 3 weeks + 4 days = 46 days
	>Day 313 – Day 334 (Week 44 – Week 47) (change period from Q3W to Q6W)	2 x [(3 weeks+6 weeks)/2] + 4 days = 67 days
Q6W thereafter*	>Day 334 onwards	2 x 6 weeks + 4 days = 88 days

*Follow schedule until treatment discontinuation after which a Q6W (relative to C1D1) window will be assumed.

4.2.8.3 Handling of Dropouts and Missing Data

For each subscale, if ≤50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers, et al., 2001). If more than 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing.

4.2.8.4 Primary Analysis of TTD Endpoints as Measured by EORTC QLQ-C30

Statistical analysis

TTD in the pain scale, physical functioning scale and GHS/QoL scale is analysed using the same methodology as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI and p-value are presented. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

Estimates and 95% CI for TTD rates at 3 months intervals and median TTD for each treatment group are presented.

Summaries

Summaries of the number and percentage of participants experiencing a TTD event, and the type of event will be provided for each treatment. The duration of follow-up is summarised using median time from randomisation to date of censoring in censored participants only, presented by treatment group.

KM plots of the TTD are presented by treatment group.

4.2.8.5 Sensitivity Analysis of TTD Endpoints as Measured by EORTC QLQ-C30

A sensitivity analysis for TTD is performed where TTD is defined as the time from the date of randomisation to the date of first deterioration that is confirmed at a subsequent timepoint (except if the first deterioration is at the participant's last available assessment), regardless of whether the participant discontinues the IP or receives another anti-cancer therapy prior to deterioration.

This will be analysed using the same methodology described in Section 4.2.8.4.

4.2.9 Secondary Endpoint - Time to First Subsequent Therapy or Death

4.2.9.1 Definition

Time to first subsequent therapy or death (TFST) is defined as the time from the date of randomisation to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomised treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the initial therapy. In this situation, TFST is calculated as time from randomisation to the start of the initial therapy or death.

4.2.9.2 Derivations and Censoring Rules

Any participant not known to have had a first subsequent anti-cancer therapy will be censored at the last date that the participant was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before first subsequent therapy, these participants will be censored at the earliest of their last known to be alive and termination dates.

Any participant without a TTSCAPRX form (not known to have had a first subsequent anti-cancer therapy) and have not died at the time of the analysis is censored at the last date that the participant was known to be alive according to the rules detailed in Section 4.2.2.2, where any participant recorded as alive or to have died after DCO date is censored at the date of DCO.

4.2.9.3 Primary Analysis of Time to First Subsequent Therapy or Death

Statistical analysis

The time to first subsequent therapy or death (TFST) is analysed using the same methodology as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI is presented. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

Summaries

In addition, medians and a KM plot of the time to the start of subsequent therapy are presented by treatment group. The time between progression and starting subsequent therapy in participants who have progression per PFS BICR and receive first subsequent therapy is assessed. This is summarised per treatment group, but no formal comparisons are made.

In participants who received a subsequent anti-cancer therapy, a summary table of first subsequent anti-cancer therapies by generic term and treatment group is provided, as well as response to first subsequent anti-cancer therapy by treatment group.

4.2.10 Secondary Endpoint - Time to Second Subsequent Therapy or Death

4.2.10.1 Definition

Time to second subsequent therapy or death (TSST) is defined as the time from the date of randomisation to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of first subsequent treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomisation + 1).

For participants not receiving randomised treatment but remaining in the study, the first alternative cancer therapy they receive is the first subsequent anti-cancer therapy and the second alternative cancer therapy they receive is the second subsequent anti-cancer therapy. In this situation, TSST is calculated as time from randomisation to the start of the second subsequent anti-cancer therapy or death.

4.2.10.2 Derivations and Censoring Rules

Any participant not known to have had a second subsequent anti-cancer therapy or have not died at the time of the analysis is censored at the last date that the participant was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a participant terminated the study for reason other than death before second subsequent therapy, these participants are censored at the earliest of their last known to be alive and termination dates.

4.2.10.3 Primary Analysis of Time to Second Subsequent Therapy or Death

The time to the second subsequent therapy or death (TSST) is analysed using the same methods as that used for the analysis of TFST (see Section 4.2.9.3). The same statistics and summary tables are produced for TSST as for TFST.

4.2.11 Secondary Endpoint - Time from Randomisation to Second Progression or Death

4.2.11.1 Definition

Time from randomisation to second progression or death (PFS2) is defined as the time from date of randomisation to the earliest progression event following first objective progression subsequent to the first subsequent therapy, or death. The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological progression, symptomatic progression, other or death. Second progression status will be reviewed every 3 months following the progression event used for the primary variable PFS (the first progression) and status recorded.

4.2.11.2 Derivations and Censoring Rules

If death occurs within 211 days of first objective progression, or within 211 days of the last evaluable PFS2 assessment, the death will be a PFS2 death event irrespective of whether subsequent therapy has started.

If a participant had a first objective progression which was censored due to 2 missed visits (i.e. was censored for the PFS endpoint), did receive subsequent therapy, and subsequently a second progression was recorded by the investigator, then the participant will be counted as a PFS2 second progression event.

Participants alive and for whom a second disease progression has not been observed are censored at date last known alive and without a second disease progression. Therefore, they are censored at:

- The PFS assessment date if the participant has not had a first progression or death (PFS censoring date).
- The date the participant is last known to not have received a first subsequent therapy if a participant has had a first progression and not started a subsequent therapy (TFST censoring date).
- The latest PFS2 assessment date following first objective progression, if the participant has started a first subsequent therapy and PFS2 event (second progression or death) has not been observed. If a PFS2 assessment has not occurred, then the participant is censored at the day before starting the first subsequent therapy.

Based on two 3-monthly visits plus two allowed 2 week visit windows, a second progression is not evaluable if it was greater than 211 days since last evaluable visit; where the last evaluable visit is the later of the first progression date and any evaluable PFS2 assessment. In addition, if subsequent therapy has not started a PFS2 assessment is not evaluable.

If the participant experiences a second progression that is not evaluable, or dies immediately after two or more consecutive missed visits, the participant is censored at the time of the later of the first progression date and the latest evaluable PFS2 assessment prior to the two missed visits.

For deaths prior to 1st progression, but immediately after two or more consecutive missing visits, the participant is censored at the time of the last evaluable PFS1 assessment.

4.2.11.3 Handling of Dropouts and Missing Data

Dropouts and missing data are handled according to the censoring rules detailed in Section 4.2.11.2.

4.2.11.4 Primary Analysis of PFS2

Statistical Analysis

PFS2 is analysed using identical methods as outlined for PFS (see Section 4.2.1.4) and adjusting for the same stratification factors. The HR for the treatment effect together with its 95% CI are presented. Medians and KM plots are presented to support the analysis.

Summaries

The number and percentage of participants experiencing a PFS2 event and the type of progression (objective progression by RECIST, symptomatic progression or other) are also summarised by treatment group, as well as summaries of deaths in the absence of second progression, and categories of PFS2 censoring. Time from randomisation to second progression will be summarised by treatment arm.

4.2.12 Secondary Endpoint - Pharmacokinetics

4.2.12.1 Derivations

Pharmacokinetic concentration data will be collected according to Section 8.5.1 of the CSP. The schedule of assessment is as per the schedule of activities (SoA) of the CSP. Whole blood samples for determination of plasma concentration of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a, will be obtained for all participants receiving Dato-DXd.

4.2.12.2 Primary Analyses of Pharmacokinetics

All plasma concentrations will be listed for each participant, for each sampling time and each dosing day, regardless of whether they are excluded from summary statistics due to deviation (e.g. as a result of dose interruption, reduction or missing the dose before PK sample collection, or sampling time deviation, etc).

Plasma concentrations of Dato-DXd, total anti-TROP2 antibody and MAAA-1181a will be summarised by visit and nominal sample time using standard summary statistics for PK concentrations (geometric mean, geometric coefficient of variation, geometric mean, arithmetic mean, standard deviation, median, minimum, maximum, n and number of concentrations below the Lower Limit of Quantification (LLOQ)) for the Dato-DXd treatment arm.

Individual concentrations below the LLOQ of the bioanalytical assay will be reported as not quantifiable (NQ) in the listings with the LLOQ defined in the footnotes of the relevant tables, figures and listings (TFLs). Individual plasma concentrations that are Not Reportable will be reported as NR and those that are missing will be reported as NS (No Sample) in the listings. For data below limit of quantification (BLQ), NR or NS the following rules will apply:

- Any values reported as NR or NS will be excluded from the summary tables and corresponding figures.
- If, at a given time point, 50% or less of the plasma concentrations are NQ, the geometric mean, CV%, geometric CV%, mean and SD will be calculated treating the NQ as LLOQ.
- If more than 50%, but not all, of the concentrations are NQ, the geometric mean, CV%, geometric CV%, and SD will be reported as data not calculable (NC). The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.
- If all the concentrations are NQ, the geometric mean, mean, minimum, median and maximum will be reported as NQ and the CV%, geometric CV% and SD as NC.

Participants with protocol deviations seriously impacting PK results are excluded from the summary tables.

Population PK, and exploratory exposure response/safety analyses will be performed. This is documented in a separate analysis plan and the results presented separately from the main CSR.

4.2.13 Secondary Endpoint - Immunogenicity

4.2.13.1 Derivations

The presence of ADAs will be assessed in plasma samples taken according to the SoA in the CSP. ADA result from each sample is reported as either positive or negative. If the sample is positive, the ADA titre is reported as well. In addition, the presence of neutralizing antibody (nAb) will be tested for all ADA-positive samples using a ligand-binding assay. The nAb results is reported as positive or negative.

The rules described in Sections 3.3.3, 3.3.5 and 3.3.6 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

Number and percentage of ADA-evaluable participants in the following categories are provided. The number of ADA-evaluable participants in the treatment group will be used as the denominator for percentage calculation.

The ADA categories are defined as follows:

- ADA positive at any visit (at baseline or post-baseline). The percentage of these participants in a population is known as ADA prevalence.
- Treatment-induced ADA positive, defined as ADA positive post-baseline and not detected at baseline (negative or missing).
- Treatment-boosted ADA positive, defined as a baseline positive ADA titre that was boosted to a 4-fold or higher-level following drug administration.
- Treatment-emergent ADA positive (TE-ADA+), defined as either treatment-induced or treatment-boosted ADA. The percentage of these participants in a population is known as ADA incidence.
- Non-treatment-emergent ADA positive (non-TE ADA+), defined as being ADA positive but not fulfilling the conditions for treatment-emergent ADA positive.
- ADA positive post-baseline and positive at baseline.
- ADA not detected post-baseline and positive at baseline.
- Treatment-emergent persistently ADA positive, defined as being TE-ADA+ 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 ADA positive, defined as being TE-ADA+ and having at least one post-baseline ADA positive measurement but not fulfilling the conditions for persistently positive.
- nAb positive at any visit (at baseline or post-baseline).

4.2.13.2 Primary Analysis of Immunogenicity

A summary will be provided of the number and percentage of participants who develop detectable anti-Dato-DXd antibodies by ADA categories (see Section 4.2.13.1) using the ADA evaluable set. Descriptive statistics (minimum, Q1, median, Q3, and maximum) for the maximum post-baseline ADA titres of participants in each ADA category will also be included.

A summary will be provided of the number and percentage of participants who develop detectable anti-Dato-DXd antibodies by visit. Descriptive statistics for ADA titres by visit will also be included.

A summary will be provided of the number and percentage of participants who are ADA positive at a post-baseline assessment for the first time by visit will also be presented. Descriptive statistics for ADA titres by first positive visit will also be included.

Impact of ADA on PK will be explored by presenting plasma Dato-DXd concentration descriptive statistics (Section 4.2.12) in TE-ADA+, non-TE ADA+, ADA-negative, and nAb-positive participants. Spaghetti plots of individual Dato-DXd concentration participants over time profiles of TE ADA+, non-TE ADA+, ADA-negative and nAb-positive participants will be presented.

ADA safety tables will include the number and percentage of participants who had at least 1 AE in any category summarized by ADA status (TE ADA+, non-TE ADA+, ADA negative, and nAb positive).

The effect of ADA on efficacy may be examined by PFS and OS KM plots by ADA status (TE-ADA+ vs non-TE ADA+ vs ADA-negative, and nAb-positive vs nAb-negative), if the data allow (20 events in a subgroup).

Immunogenicity results will be listed for all participants in SAF regardless of ADA-evaluable status. Anti-drug antibody titre and neutralising ADA data will be listed for samples confirmed positive for the presence of anti-Dato-DXd antibodies. AEs in ADA positive participants by ADA positive category will be listed.

4.2.14 Exploratory Endpoint – Patient-Reported Symptomatic AEs and Treatment Tolerability

4.2.14.1 Definition

PRO-CTCAE

The Patient-Reported Outcomes version of the common criteria for adverse events (PRO-CTCAE), a PRO version of the CTCAE system developed by the National Cancer

Institute (NCI), is included to evaluate symptomatic toxicity from the participants' perspective.

PRO-CTCAE is an item library of symptoms experienced by participants while undergoing treatment of their cancer. Symptoms have been converted to participant terms (e.g. CTCAE term "myalgia" converted to "aching muscles"). Items capture the presence, frequency, severity and/or interference with daily activities, depending on the AE. For each question, participants select the value that best describes their experience over the past week, on a 5-point ordinal scale.

The items pre-selected for this study include mouth/throat sores, decreased appetite, nausea, vomiting, constipation, diarrhoea, abdominal pain, shortness of breath, cough, rash, hair loss, hand-foot syndrome, numbness/tingling, and fatigue.

EORTC IL117

The EORTC IL is an online platform comprised of more than 900 individual items from over 60 EORTC questionnaires. The pre-selected items for this study will include dry eyes, mouth pain, and sore mouth (i.e. EORTC IL117).

The recall period is during the past week. Items are scored on a 4-point verbal rating scale: "Not at all", "A little", "Quite a bit", and "Very much".

PGI-TT

The PGI-TT is a single item to assess how a participant perceives the overall tolerability of the IP. The responses indicate how bothered the participant was in the last 7 days by the side effects of their cancer treatment and are scored on a 5-point scale: 1 = Not at all; 2 = A little bit; 3 = Somewhat; 4 = Quite a bit; 5 = Very much.

4.2.14.2 Derivations and Censoring rules

Compliance

Summary measures of compliance over time will be derived for all PRO questionnaires. These will be based upon:

- Expected questionnaire: A questionnaire that is expected to be completed at a scheduled assessment time i.e. a questionnaire from a participant who has not withdrawn from the study at the scheduled assessment time, excluding participants in countries with no available translation and participants who are exempt from PRO completion. Only participants who have started treatment or have a cycle 1 day 1 visit will be included in the compliance summary.

  - For participants that have progressed (RECIST 1.1 progression by Investigator assessment) or discontinued study treatment, the earliest of date of study treatment discontinuation or date of progression will be used to determine the last on treatment windowed assessment for each participants expected forms using the analysis windows as described in Section 3.3.5. If the date falls before the end of the assessment window, then that assessment will only be considered expected if they have a received form. If they have not received a form, then this assessment is not considered expected as they have not had the full opportunity to complete the questionnaire within the window. For participants who have not discontinued study treatment or progressed, the date of the DCO will be used to determine the last on treatment assessment for their last expected form following the same approach as above.
  - For follow up assessments (EORTC QLQ-C30, EORTC IL116, PGIS and EQ-5D-5L), if a participant has not discontinued study treatment then no follow up forms will be expected. For participants who have discontinued study treatment, and discontinued the study, the earliest of date of study discontinuation or 18 weeks post progression will be used to determine the last expected assessment that a form should have been completed. For participants who have discontinued study treatment, and not discontinued the study, the earliest of date of the DCO or 18 weeks post progression will be used to determine whether a form is expected following the same approach as above. For PRO-CTCAE, PGI-TT, EORTC IL117 and PGIC follow up forms will not be expected as these are not collected during post treatment follow-up.

- Received questionnaire: A questionnaire that has been received and has a completion date and at least 1 individual item completed.
- Evaluable questionnaire: A questionnaire with a completion date and at least 1 subscale that is non-missing.

Compliance over time will be calculated separately for each timepoint, including baseline, as the number of participants with an evaluable questionnaire at the time point, divided by number of participants still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculate separately for each timepoint, including baseline, as the number of evaluable questionnaires, divided by the number of received questionnaires. For compliance over time all timepoints (including follow-up time period) with at least 20 participants in one of the treatment arms will be reported.

Overall compliance will be calculated as the total number of evaluable questionnaires across all timepoint, divided by the total number of questionnaires expected to be received across all timepoints. Similarly, the overall evaluability rate will be calculated as the total number of evaluable questionnaires across all timepoint, divided by the total number of questionnaires received across all timepoints.

4.2.14.3 Handling of Dropouts and Missing data

For PRO-CTCAE only participants in countries where a linguistically validated version of the PRO-CTCAE is available for administration are required to complete this questionnaire and thus included the analysis.

4.2.14.4 Primary Analyses of PRO-CTCAE

PRO-CTCAE data will be summarised descriptively for each symptom by treatment group. The summary will include:

- Number and percentage of participants reporting different levels of responses at each time point for each symptom and attribute
- Number and percentage of the worst response option (defined by the first attribute of each symptom) reported by participants within 12 weeks, excluding baseline
- Number and percentage of participants who report the presence of the symptom (defined by the first attribute of each symptom) at baseline
- Number and percentage of participants who report any worsening from baseline (defined by the first attribute of each symptom) at any time within 12 weeks
- Number and percentage of participants who worsen from a score <4 (on a 1-5 scale) at baseline to a score 4 or 5 (defined by the first attribute of each symptom) at any time within 12 weeks

In addition, a stacked horizontal bar chart showing the percentage of participants with each level of response by timepoint for the first attribute of each PRO-CTCAE symptom is produced. A pie chart showing the percentage of the worst response option (defined by the first attribute of each symptom) reported by participants within 12 weeks, excluding baseline for each PRO-CTCAE item is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm. PRO-CTAE is summarised using the SAF.

4.2.14.5 Primary Analyses of EORTC IL117

EORTC IL117 data will be summarised descriptively for each item by treatment group. The summary will include:

- Number and percentage of the worst response option reported by participants within 12 weeks, excluding baseline
- Number and percentage of participants who report the presence of the symptom at baseline
- Number and percentage of participants who report any worsening from baseline at any time within 12 weeks

- Number and percentage of participants who worsen from a score <3 (on a 1-4 scale) at baseline to a score 3 or 4 at any time within 12 weeks

The number and percentage of participants with each level of response for each EORTC IL117 item at baseline and over time is summarised by treatment group.

In addition, a stacked horizontal bar chart showing the percentage of participants with each level of response by timepoint for each EORTC IL117 item is produced. A pie chart showing the percentage of the worst response option reported by participants within 12 weeks, excluding baseline for each EORTC IL117 item is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm. EORTC IL117 is summarised using the SAF.

4.2.14.6 Primary Analyses of Global Impression of Treatment Tolerability

Responses for PGI-TT are summarised descriptively as number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

In addition, a stacked horizontal bar chart showing the percentage of participants in each PGI-TT category by timepoint is produced.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

The PGI-TT is summarised using the SAF.

4.2.15 Exploratory Endpoint - Patient Global Impression of Severity (PGIS)

4.2.15.1 Definition

The PGIS is a single item to assess how a participant perceives the overall severity of cancer symptoms over the past week. The responses are scored on a 4-point scale: 1 = None; 2 = Mild; 3 = Moderate; 4 = Severe.

4.2.15.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.15.3 Analysis of Patient Global Impression of Severity

Responses for PGIS is summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group. Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIS will be based on the FAS.

4.2.16 Exploratory Endpoint - Patient Global Impression of Change (PGIC)

4.2.16.1 Definition

The PGIC is a single item to assess how a participant perceives the overall change in health status since the start of IP. The responses are scored on a 7-point scale: 1 = Much better; 2 = Moderately better; 3 = A little better; 4 = About the same; 5 = A little worse; 6 = Moderately worse; 7 = Much worse.

4.2.16.2 Derivations and Censoring rules

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.16.3 Analysis of Patient Global Impression of Change

Responses for PGIC are summarised descriptively as the number of participants and corresponding percentage in each category of the questionnaire over time by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

PGIC will be based on the FAS.

4.2.17 Exploratory Endpoint - Patient Reported Symptoms, Functioning and Health Related QoL

4.2.17.1 Definition

Refer to Section 4.2.8.1 for the relevant definitions of EORTC QLQ-C30.

4.2.17.2 Derivation and Censoring Rules

Refer to Section 4.2.8.2 for the derivation and censoring rules of scoring algorithm, change from baseline, deterioration and TTD. Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.17.3 Handling of Dropouts and Missing Data

Refer to Section 4.2.8.3 for the handling of missing data for EORTC QLQ-C30.

4.2.17.4 Primary Analysis of Patient Reported Symptoms, Functioning and Health Related QoL

TTD of function (including role, cognitive, emotional and social), multi-term symptoms (including fatigue and nausea/vomiting), and single items (dyspnoea, insomnia, appetite loss, constipation and diarrhoea) will be analysed using the same methods as that used for the analysis of TTD in the pain scale, physical functioning scale and GHS/QoL scale (see Section 4.2.8.4). The same statistics and summary tables are produced.

Change from baseline in subscales of the EORTC QLQ-C30 is analysed using a mixed model for repeated measures (MMRM) of the change from baseline. Participants are included in the mean change from baseline analysis if they have an evaluable baseline assessment and at least one evaluable post-baseline assessment. The model includes treatment, visit, and treatment-by-visit interaction as explanatory variables and the baseline score and the baseline score by visit interaction as covariates.

When less than 20 participants are present at a timepoint in either arm, this timepoint should be excluded from the analysis. An unstructured covariance matrix is used to model the within-participant error and the Kenward-Roger approximation is 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: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, and autoregressive. Adjusted mean change from baseline estimates per treatment group and corresponding 95% CIs are presented along with an overall estimate of the treatment difference, 95% CI, and p-value.

All analyses will have a corresponding graphical plot showing the adjusted mean change from baseline and 95% CI over time.

Summary tables of assessment responses (improvement, deterioration, and no change), absolute scores and change from baseline for each EORTC QLQ-C30 scale/item score (global health status/QoL, 5 functions, and all symptoms [fatigue, pain, nausea/vomiting, dyspnoea, insomnia, appetite loss, constipation and diarrhoea]) and for each timepoint will be presented by treatment group. Additionally, for each EORTC QLQ-C30 scale/item graphical plots of the mean absolute score and change from baseline along with the associated 95% CI for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

Reason participant did not complete questionnaire will be summarised over time. This summary will not be split by PRO questionnaire.

These analyses will be based on the FAS.

4.2.18 Exploratory Endpoint – Breast and Arm Symptoms

4.2.18.1 Definition

The EORTC QLQ-BR45 is a breast-cancer-specific module from the EORTC comprising 45 questions to assess breast cancer symptoms. The module includes 5 functional scales/items (body image, future perspective, sexual functioning, sexual enjoyment, breast satisfaction) and 7 symptom scales/items (systemic therapy side effects, upset by hair loss,

arm symptoms, breast symptoms, endocrine therapy symptoms, skin mucosis symptoms and endocrine sexual symptoms).

The current study will only include the breast symptoms and arm symptoms scales (7 items) from the BR45, i.e. EORTC IL116.

4.2.18.2 Derivations and Censoring Rules

The EORTC IL116 is scored as described in Table 10 to give a score from 0 to 100 for each of the symptom scales and items.

Table 10 EORTC IL116 scales and scores

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Symptom scales/items				
Arm Symptoms	ARM	3	3	31 - 33
Breast Symptoms	BR	4	3	34 - 37

For each multi-item scale, the average of the corresponding items is calculated. The raw score is then standardised to a 0 - 100 range as for the EORTC QLQ-C30, and missing values are handled in the same way.

The scoring approach for the EORTC IL116 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30. Similarly, to the symptom scales of the EORTC QLQ-C30, higher scores represent greater symptom severity.

The definition of a clinically meaningful change and time to deterioration for EORTC IL116 is the same as that for the EORTC QLQ-C30 described in Section 4.2.8.2. Censoring rules are also the same as described for EORTC QLQ-C30.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.18.3 Handling of Dropouts and Missing Data

Missing data for EORTC IL116 is handled in the same way as EORTC QLQ-C30 (Section 4.2.8.3).

4.2.18.4 Primary Analysis of Breast and Arm Symptoms

TTD in the breast and arm symptom scales will be analysed using the same methods as that used for the analysis of TTD in the pain scale, physical functioning scale and GHS/QoL scale (see Section 4.2.8.4). The same statistics and summary tables are produced.

Change from baseline in subscales of the EORTC IL116 is analysed using a MMRM of the change from baseline as described in Section 4.2.17.4.

Summary tables of assessment responses (improvement, deterioration, and no change), absolute scores and change from baseline for each EORTC IL116 scale score (breast symptoms, arm symptoms) and for each timepoint will be presented by treatment group. Additionally, for each EORTC IL116 scale graphical plots of the mean absolute score and change from baseline along with the associated 95% CI for each timepoint will be presented by treatment group.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.19 Exploratory Endpoint - EQ-5D-5L

4.2.19.1 Definition

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. The EQ-5D-5L questionnaire comprises 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 a possible 3,125 health states. For example, state 11111 indicates no problems on any of the five dimensions. Respondents also assess their health today using the EuroQoL-Visual analogue scale (EQ-VAS), which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

Improvement, deterioration or no change of the EQ-VAS score will be defined using a 7-point change from baseline and a 10-point change from baseline.

4.2.19.2 Derivations

The EQ-5D profile is converted into a weighted health state utility value, termed the EQ-5D index using the UK crosswalk algorithm developed by Hernandez Avala (Hernandez Alava, 2020). The UK algorithm is applied to all participants. The EQ-VAS is reported separately.

Refer to Section 4.2.14.2 for the derivation of compliance.

4.2.19.3 Primary Analysis of EQ-5D-5L

Descriptive statistics will be calculated for each scheduled time point in the study, for each trial arm and as a total. These will report the number of participants, the number of EQ-5D questionnaires completed at each timepoint, the number and proportion responding to each dimension of the EQ-5D-5L. Additionally, summary statistics (e.g. n, mean, median, standard deviation, min, max) will be reported for the EQ-5D index score and the EQ-VAS score, and the change from baseline for the EQ-5D index score and the EQ-VAS score.

Graphical plots of the mean EQ-5D index score and EQ-VAS score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study are produced.

For EQ-VAS a summary table of assessment responses (improvement, deterioration, and no change) for each timepoint will be presented by treatment group using a 7-point change from baseline and repeated using a 10-point change from baseline.

Summary statistics will be used to summarise measures of compliance and evaluability over time and by treatment arm.

These analyses will be based on the FAS.

4.2.19.4 Supplementary Analyses of EQ-5D-5L

To support submissions to payers, additional analyses may be undertaken, and these are outlined in a separate Payer Analysis Plan (PAP).

4.2.20 Exploratory Endpoint - Biomarkers

The relationship of baseline TROP2 expression, tumour mutational profiling and gene expression profiling and, if applicable, of exploratory biomarkers to clinical outcomes (including but not restricted to) of BOR, DoR, PFS and OS may be presented. Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report.

4.2.21 Exploratory Endpoint - Health Care Resource Use

4.2.21.1 Definition

To investigate the impact of treatment and disease on health care resource of non-study protocol related events, the following variables are captured:

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

- Procedures and tests

4.2.21.2 Derivations

Where admitted overnight, the length of hospital stay is calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation + 1).

If there are multiple hospital stays for the same participant, then the length of hospital stay is summed across all hospitalisation admissions for the participant.

Participants with missing discharge dates are calculated as the difference between the last day with available data and the start date of hospitalisation + 1. The length of ICU/HDU stay is calculated using the same method.

4.2.21.3 Primary Analysis of Health Care Resource Use

Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) are provided for each treatment group on the different types of hospital admissions, the length of stay for participants admitted to hospital for at least one overnight stay and length of stay for participants admitted to intensive care / high dependency units, as well as the primary sign or symptom the participant presents with.

Where a participant has admissions for different signs and symptoms, they are included in each category when summarising type of hospital admissions.

This analysis will be done on the SAF.

4.3 Pharmacodynamic Endpoint(s)

Not Applicable.

4.4 Safety Analyses

The domain safety covers exposure, adverse events, clinical laboratory, vital signs, physical examination, ECG, Echocardiogram, ECOG performance score and Ophthalmologic assessments.

Tables are provided for the safety set; listings are provided for the safety set.

4.4.1 Exposure

4.4.1.1 Definitions and Derivations

Treatment exposure for Dato-DXd

Dato-DXd is dosed 6 mg/kg intravenously on Day 1 of each 21-day cycle (Q3W). The dose of Dato-DXd may be reduced once to 4 mg/kg intravenous (IV) Q3W and a further second reduction to 3 mg/kg IV Q3W is allowed per participant on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Dato-DXd (months): =

(min (last date where dose > 0 + 20, date of death, date of DCO) – first dose date +1) / (365.25/12)

Actual exposure of Dato-DXd =

total exposure – total duration of dose interruptions,

where the total duration of dose interruption is defined as any length of time when the participant has not taken any of the planned doses. Dose interruptions include missed and delayed doses.

The calculation of actual exposure makes no adjustment for any dose reductions that may have occurred and will only be calculated for Dato-DXd (not for chemotherapy arms).

Treatment exposure for Capecitabine

Capecitabine is scheduled to be dosed 1000 or 1250 mg/m^2 twice daily (BID) on Days 1 to 14 of a 21-day cycle. The choice of dose will be determined by standard institutional practice and the starting dose will be assumed to be the planned dose. The dose of Capecitabine may be reduced by 25% in participants with moderate renal impairment on the study. The calculation of exposure is as follows:

Total (or intended) exposure of Capecitabine (months): =

(min (last Capecitabine dose date where dose > 0, date of death, date of DCO) – first Capecitabine dose date +1) / (365.25/12)

Treatment exposure for Gemcitabine

Gemcitabine is dosed 1000 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Gemcitabine (months): =

(min (last Gemcitabine dose date where dose > 0 + W, date of death, date of DCO) – first Gemcitabine dose date+1) / (365.25/12),

where W=6 if the last Gemcitabine dose was scheduled on Day 1 and W=13 if the last Gemcitabine dose was scheduled on Day 8.

Treatment exposure for Vinorelbine

Vinorelbine is dosed 25 mg/m^2 on Days 1 and 8 of a 21-day cycle. The calculation of exposure is as follows:

Total (or intended) exposure of Vinorelbine (months): =

(min (last Vinorelbine dose date where dose > 0 + W, date of death, date of DCO) – first Vinorelbine dose date +1) / (365.25/12),

where W=6 if the last Vinorelbine dose was scheduled on Day 1 and W=13 if the last Vinorelbine dose was scheduled on Day 8.

Treatment exposure for Eribulin mesylate

Eribulin mesylate is dosed 1.4 mg/m^2 on Days 1 and 8 of a 21-day cycle. A lower starting dose of 1.1 mg/m^2 is recommended for participants with moderate renal impairment. The starting dose will be assumed to be the planned dose. The calculation of exposure is as follows:

Total (or intended) exposure of Eribulin mesylate (months): =

(min (last Eribulin mesylate dose date where dose > 0 + W, date of death, date of DCO) – first Eribulin mesylate dose date +1) / (365.25/12),

where W=6 if the last Eribulin mesylate dose was scheduled on Day 1 and W=13 if the last Eribulin mesylate dose was scheduled on Day 8.

Participants who permanently discontinue during a dose interruption

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

If a participant permanently discontinues study treatment during a dose interruption, then this is not counted as a dose interruption for summary purposes.

Number of treatment cycles received

Exposure is also measured by the number of cycles received. A cycle corresponds to a period of 21 days. If a cycle is prolonged due to toxicity, this is still counted as one cycle. A cycle is counted if any treatment during that cycle is taken.

Safety Follow-up

Total duration of safety follow-up is calculated as:

Total Safety Follow-up (months) = [min (date of safety follow-up assessment, last dose of IP date + 20, date of study discontinuation, date of death, DCO date) – first dose date +1] / (365.25/12)

Dose intensity

Dose intensity is derived for study treatments Dato-DXd, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine.

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. RDI is 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 or the actual last day of dosing. D is the total dose that would be delivered if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 2 days should be added to reflect the protocol allowed window for dosing as shown below.

Dose intensity – Intended cumulative dose

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

The intended dose for Dato-DXd is 6mg/kg on Day 1 (+/-2 days) of each 21-day cycle. The minimum of the participants last dose, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended.

For the calculations below,

DUR = min (date of last dose date where dose > 0, date of death, date of DCO) – first dose date + 1

The intended dose for Dato-DXd is then calculated as:

6 * [integer ((DUR +2)/ 21) + 1]

Similarly, for Eribulin mesylate (dose=1.4 mg/m^2), Vinorelbine (dose=25 mg/m^2) and Gemcitabine (dose=1000 mg/m^2), intended dose will be calculated as:

2* dose * [integer ((DUR +2)/ 21) + 1]

Note: if the starting dose is reduced by standard institutional practice per protocol, that starting dose will be used for the participant.

For Capecitabine (dose of 1000 or 1250 mg/m^2 BID for 14 days per 21-day cycle, intended dose is given by:

2 * 14 * dose * [integer ((DUR+2) / 21)) + 1]

Dose intensity – Actual cumulative dose

For the calculation of actual cumulative dose for Dato-DXd, Gemcitabine, Eribulin mesylate and Vinorelbine, the proportion of volume left after the infusion will be used to calculate how much of the study drug the participant received, i.e.:

- Volume left (proportion) = $\frac{\text{Volume after infusion}}{\text{Volume before infusion}}$
- Actual cumulative dose = sum over all cycles [(1 – Volume left) x dose] (where dose is taken from the exposure CRF page for each cycle)

For the calculation of actual dose for Capecitabine, drug accountability data (of 1000mg tablets and 1250mg tablets) will be used as follows:

Actual cumulative dose = sum dose (mg) dispensed – sum dose (mg) returned

Percentage Intended Dose

Percentage intended dose (PID) is calculated in the same way as RDI, but D is the intended cumulative dose up to the date of progression or study discontinuation instead of date of last dose where dose >0. The minimum of the participant's date of progression or study discontinuation, date of death, date of DCO will be used to calculate the duration the participant has been on the study with dosing intended. For the calculations above,

DUR = min (date of date of progression, study discontinuation, date of death, date of DCO) – first dose date + 1.

Similarly, the actual cumulative dose is the cumulative dose taken up to min (date of progression, date of study discontinuation, date of death, date of DCO).

4.4.1.2 Presentation

The following summaries are produced for Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine:

- Total exposure
- Actual exposure (Dato-DXd only)
- RDI and PID.
- Summary of drug interruptions, dose delays and dose reductions. Number and percentage of participants with any interruptions, number and percentage of participants with any dose delays and number and percentage of participants with any dose reductions are summarised. This is broken down by number of interruptions/dose delays/dose reductions, and reason for interruption/dose delays/dose reduction. In addition, the number and percentage of participants with both an interruption and a dose reduction are summarised.

In addition, the number of cycles received, and duration of safety follow-up are also summarised.

4.4.2 Adverse Events

4.4.2.1 Definitions and Derivations

Adverse events (AEs) and serious adverse events (SAEs) are collected from the time of signature of the informed consent form (ICF), throughout the treatment period and until the 28-Day (+7 days) follow-up period is completed. AE causality is as determined by the reporting investigator. For interstitial lung disease (ILD)/pneumonitis, safety follow up will be continued until resolution of ILD/pneumonitis.

Treatment emergent adverse events (TEAEs) are defined as those adverse events (AEs) with onset or that worsen (by investigator report of an increase in CTCAE grade relative to pre-treatment) on or after the first dose of IP and on or before the date of last IP + 28 days (+7 days) and prior to the start of any subsequent anti-cancer therapy.

For the subset of TEAEs, with onset prior to start of study treatment, and which worsened in severity or seriousness after initiating study treatment until 28 days (+7 days) after last dose of study treatment, such worsening should also occur prior to initiation of any first subsequent anti-cancer therapy to be included in the AE summary tables.

Pre-treatment AEs are those which occur before the first dose of IP and do not worsen during the treatment period.

If no onset time is given, and the date of onset of the AE is the same as the date of first dose of IP, then the AE is assumed to have occurred after the first dose of IP.

The medical dictionary for regulatory activities (MedDRA) [using the latest MedDRA version] is used to code AEs. AEs are graded according to the National Cancer Institute (NCI) common terminology criteria for adverse event (CTCAE) version 5.0.

Missing start and stop dates for AEs are handled using the rules described in Section 3.3.7. AEs that have missing causality (after data querying) are assumed to be related to the treatment where causality is missing.

Dose modification describes an AE where action taken is either dose reduced, or drug interrupted.

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to understanding of the Dato-DXd safety profile and require close monitoring and rapid communication by the Investigator to the Sponsor. All AESIs, regardless of severity or seriousness, must be followed until either event resolution, end of study, trial termination, withdrawal of consent, or participant death.

The AESIs that are collected during this study are: ILD/pneumonitis, infusion-related reactions, oral mucositis/stomatitis, mucosal inflammation other than oral mucositis/stomatitis, and ocular surface toxicity.

Preferred terms used to identify AESI will be listed before data base lock (DBL) and documented in the Trial Master File. 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.

The duration of an AESI is calculated as (Stop date of AE - Start date of AE) + 1. The duration of the AESI is not calculated for AESIs which are ongoing. If a participant has multiple AESIs within a category, then the duration is summed for all AESIs within the category but any days where AESIs overlap are only counted once.

Time to first AESI is calculated as:

(Start date of first AESI - Date of IP first dose) + 1.

4.4.2.2 Presentation

All AEs are summarised descriptively by count (n) and percentage (%) for each treatment group.

Unless otherwise stated, only AEs defined as treatment emergent (see Section 4.4.2.1) are included in the summary tables.

All reported AEs (including pre- and post-treatment AEs) are listed including the date of onset, date of resolution (if AE is resolved) and investigator's assessment of CTCAE grade and relationship to IP. An overall summary of the number and percentage of participants in each category below is presented by Dato-DXd, total ICC, Capecitabine, Gemcitabine, Eribulin mesylate and Vinorelbine. Each AE category is separately summarised by SOC and PT:

- All AEs
- All AEs possibly related to IP
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to IP
- AEs with outcome of death
- AEs with outcome of death possibly related to IP
- All SAEs
- All SAEs possibly related to IP
- All SAEs with CTCAE grade 3 or higher
- All SAEs with CTCAE grade 3 or higher, possibly related to IP
- AEs leading to discontinuation of IP
- AEs leading to discontinuation of IP, possibly related to IP
- All SAEs leading to discontinuation of IP
- All SAEs leading to discontinuation of IP possibly related to IP
- AEs leading to interruption of IP
- AEs leading to dose reduction of IP
- AEs leading to dose modification of IP

Sorting is by internationally agreed order for SOC, and alphabetically for PT within SOC.

The following AEs that occur in at least 1% of participants in any treatment group, are summarised by PT, by decreasing frequency based on the total number of AEs in the Dato-DXd arm:

- All AEs possibly related to IP
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to IP
- All SAEs

The following AE categories are summarised by PT, by decreasing frequency based on the total number of AEs in the Dato-DXd arm:

- AEs leading to interruption of IP
- AEs leading to dose reduction of IP
- AEs leading to dose modification of IP
- All SAEs possibly related to IP
- All SAEs with CTCAE grade 3 or higher
- AEs leading to discontinuation of IP

Additionally, the most common AEs, which are those AEs that occur in at least 5% (where no rounding is applied i.e. an AE with frequency 4.9% does not appear if the cut-off is 5%) of participants in any treatment group, are summarised by PT, by decreasing frequency based on the total number of AEs across treatment groups. This cut-off may be modified after review of the data. This is repeated by decreasing frequency based on the total number of AEs in the Dato-DXd arm.

An additional summary of AEs with onset date after the date of last IP and less than or equal to 35 days after the date of last IP and before the onset of subsequent cancer therapy is produced by SOC, PT and treatment group.

Further details of summaries by SOC and PT are given below if a participant experienced more than one TEAE:

- The participant will be counted once for each SOC and once for each PT.
- The participant will be counted once for each SOC and once for each PT at the maximum CTCAE grade.
- The participant will be counted once for each SOC and once for each PT using the possibly related to IP (as assessed by the investigator) event.
- The participant will be counted once for each SOC and once for each PT for possibly related to IP (as assessed by the investigator) events at the maximum CTCAE grade.

AEs are assigned CTCAE grades and summaries of the number and percentage of participants are provided by maximum reported CTCAE grade, SOC and PT.

An overall summary of the number and percentage of participants in each category below is presented by Dato-DXd and total ICC, by AESI category:

- All AESIs possibly related to IP
- AESIs with CTCAE grade 3 or higher
- All serious AESIs
- AESIs with outcome of recovered/resolved
- AESIs with outcome of not recovered/not resolved
- AESIs leading to interruption of IP
- AESIs leading to dose reduction of IP
- AESIs leading to discontinuation of IP

Summary tables of AESIs overall and by maximum CTCAE grade are produced, by AESI category and PT. The preferred terms for AESIs are presented in a listing.

Tables are also produced of AESIs by outcome and AESIs with outcome of recovered/resolved by time of resolution, and by action taken. The time to onset of first AESI and duration of AESI are summarised. In addition, summary tables are produced of number of participants with AESIs possibly related to IP and leading to discontinuation of IP. AESIs are summarised by AESI category, and PT.

A summary table of AESIs by maximum CTCAE grade is produced, by SOC and PT.

All AEs and AESIs are listed, and the time to onset of the AE from date of first dose is presented in the listing. Key participant information is provided in 4 separate listings for all SAEs, AEs with an outcome of death, all AEs leading to dose modification and all AEs leading to treatment discontinuation.

Deaths

A separate summary of deaths is provided with number and percentage of participants, categorised as:

- Total number of deaths (regardless of date of death)
- Related to disease under investigation

- AE with outcome of death only and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- AE with outcome of death only and onset date > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first)
- Death with primary or secondary reason related to disease under investigation and AE with outcome of death and onset date <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first)
- Death with primary or secondary reason related to disease under investigation and AE with outcome of death > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy, whichever occurs first
- Deaths <= 28 days (+7 days) after last dose of IP and prior to initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Deaths > 28 days (+7 days) after last dose of IP or on or after initiation of subsequent anti-cancer therapy (whichever occurs first), unrelated to AE or disease under investigation
- Participants with unknown reason for death
- Other deaths

A corresponding listing is also produced.

4.4.2.3 ILD/pneumonitis Adverse Event of Special Interest

Summaries of ILD/pneumonitis events will be primarily based on adjudicated drug related ILD/pneumonitis events from the ILD adjudication committee. Supportive summaries based on AESI-defined ILD/pneumonitis cases (i.e. identified based on pre-defined MedDRA preferred terms) will also be provided.

When summarising time to and duration of first treatment-emergent AESI for adjudicated ILD/pneumonitis events, only adjudicated drug-related events will be considered.

A listing of ILD/pneumonitis events is produced.

4.4.3 Clinical Laboratory, Blood Sample

4.4.3.1 Definitions and Derivations

Blood samples for determination of clinical chemistry and haematology are collected as described in the schedule of activities (SoA) of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records and missing data are followed.

Change from baseline in haematology and clinical chemistry variables are calculated for each post-dose visit on treatment.

CTCAE grades are defined at each visit according to the CTCAE v5.0 grade criteria using local ranges (or project ranges if local ranges are missing), after conversion of lab result to corresponding AstraZeneca (AZ) preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, sodium, magnesium and corrected calcium. For these parameters high and low CTCAE grades are calculated.

Corrected calcium product is 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 are compared to the local reference ranges (or project ranges if local ranges are missing) and classified as low (below range), normal (within range or limits of range) and high (above range).

For parameters with no CTCAE grading that are listed in the CSP any increase/decrease/treatment emergent laboratory change (TELC) is derived, where any increase is an increase to a value above the upper local laboratory reference limit (or project ranges if local ranges are missing) at any time on treatment for participants with a value below the upper local laboratory reference limit (or project ranges if local ranges are missing) at baseline, and any decrease is a decrease to any value below the local laboratory reference range limit (or project ranges if local ranges are missing) at any time on treatment for participants with a value above the lower local laboratory reference limit (or project ranges if local ranges are missing) at baseline. A TELC is defined as any on treatment increase or decrease from baseline.

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

Local reference ranges (or project ranges if local ranges are missing) are used for the primary interpretation of laboratory data.

4.4.3.2 Presentations

Only laboratory data that is on treatment as defined in Section 3.3.4 is included in the summary tables.

Data summaries and listings are provided by AZ preferred units.

Laboratory listings will cover observed values, changes from baseline and CTCAE grade for each individual participant as well as abnormalities. Flags are applied to values falling outside reference ranges and for the CTCAE grade for parameters for which CTCAE grading applies.

For all continuous clinical chemistry and hematology laboratory assessments, absolute value and change from baseline are summarised using descriptive statistics at each scheduled visit.

Shift tables of laboratory values by worst common toxicity criteria (CTCAE) grade on treatment are produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change are produced. Percentages are based on the number of participants with a baseline value and an on-treatment value.

The laboratory parameters for which CTCAE grade shift outputs are produced are:

- Haematology: Haemoglobin, Leukocytes, Lymphocytes (absolute count), Neutrophils (absolute count), Platelets
- Clinical Chemistry: Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Albumin, Alkaline Phosphatase (ALP), Total bilirubin, Magnesium (hypo- and hyper-), Sodium (hypo- and hyper-), Potassium (hypo- and hyper), Corrected Calcium (hypo- and hyper-), Creatinine

For parameters with no CTCAE grading, the number and percentage of participants with any on treatment increase from baseline, any on treatment decrease from baseline and a TELC is summarised. Percentages are based on the number of participants with a baseline value below/above the local laboratory upper/lower reference limit (or project ranges if local ranges are missing) and an on-treatment value for the any increase/decrease summaries respectively. Percentages for a TELC are based on the number of participants with a baseline value and an on-treatment value.

For parameters with no CTCAE grading, shift tables from baseline to worst value on-treatment are provided.

Hy's law (HL)

A summary table is produced showing the number (%) of participants who have:

- Elevated ALT, AST, and Total bilirubin during the study
- ALT ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN (Upper limit of normal) during the study.
- AST ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN during the study.

- Total bilirubin ≥1.5x - ≤2x and >2x ULN during the study.
- ALT or AST ≥3x - ≤5x, >5x - ≤10x, >10x - ≤20x and >20x ULN during the study.
- ALT or AST ≥3x ULN together with total bilirubin ≥2x ULN, irrespective of ALP, at any point during the study (potential Hy's law) following the start of treatment: the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation*. Exceptions include participants with elevated liver enzymes (ALT or AST ≥5x ULN) that have liver metastases present at baseline. These participants have been permitted in the study as per inclusion criterion number 8 in the CSP.
- ALP ≥1.5x - ≤3x and >3x ULN during the study.

* The ALT or AST elevation occurring less than or equal to 28 days prior to the total bilirubin elevation.

Narratives are provided in the CSR for participants with potential Hy's law.

Liver biochemistry test results over time for participants with potential Hy's law are plotted and listed.

Plots of maximum post-baseline ALT and AST vs. maximum post-baseline total bilirubin, expressed as multiples of ULN, are also produced with reference lines at 3×ULN for ALT and AST, and 2×ULN for Total bilirubin. In each plot, total bilirubin is in the vertical axis.

4.4.4 Clinical Laboratory, Urinalysis

4.4.4.1 Definitions and Derivations

Urine samples for determination of urinalysis are collected as described in the SoA of the CSP.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.4.2 Presentations

A shift table is produced showing the number and percentage of participants in each category at baseline and the maximum on treatment value for each parameter.

The denominator used only includes participants with a baseline value and at least one on treatment value.

On treatment is defined in Section 3.3.4.

Supportive laboratory listings will cover observed values for each individual participant as well as abnormalities.

4.4.5 Other Laboratory Evaluations

4.4.5.1 Definitions and Derivations

Pregnancy tests (serum at screening and urine at other timepoints) are performed for women of childbearing potential.

In addition, hepatitis B surface antigen, hepatitis C and human immunodeficiency virus (HIV) antibodies is assessed at screening.

4.4.5.2 Presentations

This data is listed only, no summary tables are produced.

4.4.6 Vital Signs

4.4.6.1 Definitions and Derivations

Vital signs are assessed at timelines as specified in the SoA of the CSP. The following vital signs are measured: systolic and diastolic blood pressure, pulse rate, body temperature and respiratory rate. Body weight is also collected. Height is collected at screening only.

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.6.2 Presentations

Summaries for vital signs data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values and change from baseline for diastolic and systolic blood pressure, pulse, respiratory rate, temperature and weight are summarised over time at each scheduled visit for each treatment group.

Vital signs data is also listed.

4.4.7 Electrocardiogram

4.4.7.1 Definitions and Derivations

Resting 12-lead electrocardiograms (ECGs) are recorded at timepoints specified in the SoA of the CSP.

The following ECG variables are collected: ECG heart rate, PR duration, QRS duration, QT interval, QTcF interval, RR duration and overall ECG evaluation.

The rules described in Sections 3.3.3, 3.3.6 and 3.3.7 of this document considering definition of baseline and visit windows are followed. ECGs will be presented using nominal visits.

The overall evaluation of an ECG is either "normal", "borderline" or "abnormal" with abnormalities categorised as either "clinically significant" or "not clinically significant". Any clinically significant ECG abnormalities require triplicate ECG results. Where triplicate ECG results are taken, a single mean value for numeric parameters is used, and the worst case for the overall evaluation is used. If there are missing value(s) in triplicates, the other non-missing value(s) in the triplicates will be used to calculate the average. Unscheduled assessments will not be included in the calculation of the average.

The QT interval corrected for heart rate using Fridericia's correction (QTcF) is calculated as follows (where QT and RR are in seconds):

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

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:

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

4.4.7.2 Presentations

Summaries for ECG data include only on treatment data. On treatment is defined in Section 3.3.4.

The following summaries for QTcF are included:

Absolute QTcF interval prolongation at any time on treatment:

- QTcF interval > 450 milliseconds
- QTcF interval > 480 milliseconds
- QTcF interval > 500 milliseconds

Change from baseline in QTcF interval at any time on treatment:

- QTcF interval increases from baseline > 30 milliseconds
- QTcF interval increase from baseline > 60 milliseconds
- QTcF interval >450 milliseconds and change from baseline > 30 milliseconds
- QTcF interval >500 milliseconds and change from baseline > 60 milliseconds

A listing is provided of ECG data.

4.4.8 Echocardiogram/Multigated Acquisition Scan

4.4.8.1 Definitions and Derivations

An echocardiogram (ECHO) or multigated acquisition (MUGA) scan to assess left ventricular ejection fraction (LVEF) is performed at the visits as shown in the SoA of the CSP.

The modality of the cardiac function assessments must be consistent for a given participant, i.e. if an ECHO scan is used for the screening assessment, then ECHO should also be used for subsequent scans. The participants should also be examined using the same machine and operator whenever possible, and quantitative measurements should be taken.

The rules described in Sections 3.3.3 and 3.3.7 of this document considering definition of baseline are followed. LVEF will be presented using nominal visits.

4.4.8.2 Presentations

Summaries for LVEF data include only on treatment data. On treatment is defined in Section 3.3.4.

Absolute values at baseline for LVEF results are summarised.

4.4.9 Eastern Cooperative Oncology Group Performance Status

4.4.9.1 Definitions and Derivations

An assessment of ECOG performance status score is performed at the visits as shown in the SoA of the CSP.

The ECOG performance status scores range from 0 to 5, with lower scores indicating greater participant activity:

0. Fully active; able to carry out all usual activities without restrictions

1. Restricted in strenuous activity, but ambulatory and able to carry out light work or work of a sedentary nature (e.g. light housework or office work)

2. Ambulatory and capable of self-care, but unable to carry out any work activities; up and about more than 50% of waking hours

3. Capable of only limited self-care; confined to bed or chair more than 50% of waking hours

4. Completely disabled; unable to carry out any self-care and totally confined to bed or chair

5. Dead

The rules described in Sections 3.3.3, 3.3.5, 3.3.6 and 3.3.7 of this document considering definition of baseline, visit windows and how to handle multiple records are followed.

4.4.9.2 Presentations

Summaries for ECOG data include only on treatment data. On treatment is defined in Section 3.3.4.The number and percentage of participants in each category is summarised at each visit.

4.4.10 Physical Examination

4.4.10.1 Definitions and Derivations

Physical examination, as well as assessment of height and weight, will be performed according to the Schedule of Assessments (SoA) in the CSP.

A full physical examination will be performed at screening which includes assessment of general appearance, respiratory, cardiovascular, abdomen, skin, head and neck (including ears, eyes, nose and throat), oral (mouth), lymph nodes, thyroid, musculoskeletal (including spine and extremities), urogenital, dermatological, gastrointestinal, endocrine, hematologic/lymphatic, and neurological systems. At subsequent visits, targeted physical examinations are to be utilised by the Investigator on the basis of clinical observations and symptomatology. A targeted physical examination includes at a minimum, assessments of the skin, lungs, oral, cardiovascular system, and abdomen (liver and spleen).

4.4.10.2 Presentations

Individual physical examination data will not be summarised.

4.4.11 Ophthalmologic Assessments

4.4.11.1 Definition and Derivations

Ophthalmologic assessments by a licensed eye care provider will be performed as specified in the SoA in the CSP.

The following assessments will be performed for both eyes: daily use of artificial tears, avoidance of contact lenses, eye-related symptoms, best corrected visual acuity (BCVA), corneal sensation, eyelid position, eyelid margins, conjunctiva, slit-lamp examination, anterior chamber cells, fluorescein staining, tear film breakup time (TFBT), limbal stem cell deficiency (LSCD), Oxford grade of punctate epithelial erosions, proparacaine test, intraocular pressure, tear film meniscus, fundoscopy, clinically significant corneal disease with use of both CTCAE and Corneal Toxicity Severity Grading scales, and other diagnosis and treatment(s) prescribed, if any.

The preferred method for measuring BCVA is the Snellen chart (metric in meter, or imperial in feet). The Snellen chart values will be converted to LogMAR values as

illustrated in Table 11. LogMAR values are calculated by taking the $\log_{10}$ of the reciprocal of the Snellen fraction. For example, if the Snellen fraction is 20/50, the LogMAR value is $\log_{10}$ (50/20) = 0.4.

Table 11 Best Corrected Visual Acuity conversion table

Snellen Chart		Snellen Fraction (Decimal)	LogMAR
Feet	Meter		
20/200	6/60	0.10	1.0
20/160	6/48	0.125	0.9
20/125	6/38	0.16	0.8
20/100	6/30	0.20	0.7
20/80	6/24	0.25	0.6
20/63	6/19	0.32	0.5
20/50	6/15	0.40	0.4
20/40	6/12	0.50	0.3
20/32	6/9.5	0.63	0.2
20/25	6/7.5	0.80	0.1
20/20	6/6	1.00	0.0
20/16	6/4.8	1.25	-0.1
20/12.5	6/3.8	1.60	-0.2
20/10	6/3	2.00	-0.3

4.4.11.2 Presentations

Ophthalmologic assessments will be summarised at DCO1 (see Section 3.1.1).

For daily use of artificial tears and BCVA (LogMAR), summary statistics (e.g. n, mean, median, standard deviation, min, max) for the observed value and the change from baseline value will be reported by visit and by treatment group.

For tear film breakup time, use of contact lenses, eye pain, fluorescein staining of cornea, abnormality results in slit lamp examination with attention to cornea, punctate epithelial erosions in Oxford grade, clinically significant corneal disease by the revised CTCAE grade and clinically significant corneal disease severity grade, the number and percentage of participants in each category will be reported by visit and treatment group.

For intraocular pressure, summary statistics (e.g. n, mean, median, standard deviation, min, max) for the change from baseline value will be reported by visit and by treatment group.

For corneal sensation, eye lid position, eye lid margins abnormalities, tear film meniscus, LSCD with and without fluorescein stain, and dilated fundoscopic exam results, shift from baseline will be reported by visit and by treatment group. Percentages will be based on the number of participants with a baseline result and at least one result for the corresponding visit.

For conjunctiva abnormality and anterior chamber (cell) abnormality, baseline to worst result post-baseline shift tables will be reported. Percentages will be based on the number of participants with a baseline result and at least one post baseline result.

Eye-related adverse events will be summarised by preferred term (PT). The following PTs will be displayed as separate categories:

1. Keratitis (including Ulcerative keratitis, Corneal perforation)
2. Limbal stem cell deficiency
3. Visual acuity reduced

Eye-related concomitant medications, procedures and surgeries will be summarised by categories of medications (subdivided into corticosteroids, antibiotics, other), ocular procedures and surgeries.

Listings will be produced for the following assessments: BCVA, lid margins for abnormalities, slit lamp examination findings, cornea abnormality, LSCD, abnormality findings from dilated fundoscopic exam and clinically significant corneal disease.

In addition, listings will be produced for:

1. medication used for eye pain
2. eye-related adverse events
3. eye-related concomitant medications, procedures and surgeries

All analyses will be performed on the OAS.

4.4.12 Impact of COVID-19

Depending on the extent of any coronavirus disease 2019 (COVID-19) impact, summaries of data relating to participants diagnosed with COVID-19, and impact of COVID-19 on study conduct (in particular delayed/missed visit, delayed or discontinued IP, discontinuation of study, and COVID-19 related protocol deviations) may be generated, by treatment group, including:

- Disposition (discontinued IP due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruption (visit impact, drug impacted)
- Listing for participants affected by the COVID-19 pandemic
- Listing for participants with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.

Additional analyses may be performed to explore the impact of COVID-19 on key efficacy and safety endpoints, for example repeating the AE summaries separately for participants where events are attributed to COVID-19.

5 INTERIM ANALYSIS

Interim Analysis for Superiority in OS

Two interim analyses for OS are planned.

The first interim analysis will occur at the primary PFS analysis. This corresponds to approximately 178 OS events, 25% maturity and 40% of the information expected at the primary analysis (444 OS events at final, primary).

The second interim analysis will occur when approximately 355 OS events have been observed in the FAS. This corresponds to approximately 51% maturity and 80% of the information expected at the primary analysis (444 OS events).

The Lan DeMets approach (Lan & DeMets, 1983) that approximates the O'Brien and Fleming spending function will be used to account for multiplicity introduced by including an interim analysis for superiority. This approach will be used to maintain an overall 2-sided type I error across the three planned analyses of OS.

If the PFS dual primary analysis crosses the efficacy threshold, the 1.0% type I error allocated to the PFS endpoint will be reallocated to the OS endpoint for a total 2-sided type I error of 5.0% (Burman, Sonesson, & Guilbaud, 2009). If the PFS dual primary analysis does not cross the efficacy threshold the OS endpoint will have a total 2-sided type 1 error of 4.0%.

Table 12 Summary of planned timings of the interim and final OS analyses

	Interim Analysis 1		Interim Analysis 2		Primary Analysis	
Projected Timing	21 Months [b]		34 Months		44 Months	
Number of Deaths [a]	178		355		444	
Information Fraction	40%		80%		100%	
Maturity	25%		51%		63%	
Recommendation	**Continue**	**Reject Null Hypothesis**	**Continue**	**Reject Null Hypothesis**	**Do Not Reject Null Hypothesis**	**Reject Null Hypothesis**
At 4.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0005	< 0.0005	≥ 0.0184	< 0.0184	≥ 0.0345	< 0.0345
Estimated hazard ratio	≥ 0.591	< 0.591	≥ 0.777	< 0.777	≥ 0.817	< 0.817
At 5.0% 2-sided alpha [c]						
2-sided nominal p-value	≥ 0.0008	< 0.0008	≥ 0.0241	< 0.0241	≥ 0.0427	< 0.0427
Estimated hazard ratio	≥ 0.604	< 0.604	≥ 0.786	< 0.786	≥ 0.824	< 0.824

[a] Estimates based on exponential survival where the median OS is 19.0 months for ICC and 25.3 months for Dato-DXd. The total proportion of participants randomized at time t [t ≤ 19 months] following the start of the study is assumed to be $(t/19)^{1.5}$.
[b] Timing of first IA based on PFS. Number of deaths is an estimate.
[c] Alpha allocated to OS endpoint (4.0% or 5.0%) dependent on statistical significance of PFS.

For a total 2-sided type 1 error of 5.0% this results in a level of significance alpha of approximately 0.0008 for the first interim analysis (IA) and 0.0241 for the second IA. For a total 2-sided type 1 error of 4.0% this results in a level of significance alpha of approximately 0.0005 for the first IA and 0.0184 for the second IA. This is described in more detail in Table 12.

Since the significance level will be dependent on the number of events actually observed, this will be calculated at the time of the analysis.

The interim analyses will be performed by an independent data monitoring committee (IDMC) separate from the study team reporting the final study results so that the study team are kept blinded.

An external IDMC comprised of therapeutic area experts and biostatisticians who are not employed by AstraZeneca and are free from conflict of interest will review the unblinded interim analysis output.

The IDMC will inform the sponsor if superiority has been achieved for OS in the FAS.

Analyses planned to be performed at the interim analysis will include PFS (first IA) and OS, and other key outputs will also be produced. Details of the outputs to be produced for the interim analysis will be specified in the IDMC charter.

If the first interim results do not meet the criterion for declaring superiority for OS in the FAS, then follow-up will continue until the criteria are met for the second OS IA. If the second interim results do not meet the criterion for declaring superiority in the FAS, then follow-up will continue until the criteria is met for the OS primary analysis (approximately 444 OS events in the FAS).

For a total 2-sided type 1 error of 5.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0427 level of significance. For a total 2-sided type 1 error of 4.0% the OS will be tested in the FAS at the primary analysis at the alpha ≈ 0.0345 level of significance.

The study may continue monitoring participants for OS up to the scheduled final analysis, beyond planned interim analyses, to provide more refined estimates of treatment effects for survival.

IDMC Safety Reviews

This study will use an external IDMC to assess ongoing safety analyses as well as the interim efficacy analyses. The IDMC will meet to review unblinded safety data after the study has started, with an initial early IDMC approximately six months after study start; and then at approximately six-month intervals thereafter.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca and are free from conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will include the recommendation and any potential protocol amendments and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

The safety of all AstraZeneca clinical studies is closely monitored on an ongoing basis by AstraZeneca representatives in consultation with AstraZeneca Global Patient Safety. Issues identified will be addressed; this could involve, for instance, amendments to the Clinical Study Protocol and letters to investigators.

Full details of the IDMC procedures, processes and the responsibilities of the IDMC will be given in the IDMC charter.

6 REFERENCES

Burman, C., Sonesson, C., & Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med. 28:739-761.

Fayers, P., Aaronson, N., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organization for Research and Treatment of Cancer.

Fleischer, F., Gaschler-Markefski, B., & Bluhmki, E. (2001). How is retrospective independent review.

Hernandez Alava, M. P. (2020). Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. *Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York.*, Report 063.

Lan, K., & DeMets, D. (1983). Discrete sequential boundaries for clinical trials. Biometrika. 70:659-663.

Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. Stat Med, 4(2). 213-26.

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

7 APPENDIX

Not applicable.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d9268c00001-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	Doc ID-004671764	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
21-Aug-2023 12:31 UTC	PPD	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.