



STATISTICAL ANALYSIS PLAN

Protocol Title: Phase 2 Randomized, Double-Blinded, Controlled Study of Tucatinib vs. Placebo in Combination with Capecitabine and Trastuzumab in Patients with Pretreated Unresectable Locally Advanced or Metastatic HER2+ Breast Carcinoma **(HER2CLIMB)**

Protocol Number: ONT-380-206

Sponsor: Seattle Genetics, Inc.
21823 30th Drive SE
Bothell, WA 98021, USA

Version: 1.0, 7-Aug-2019

SPONSOR APPROVAL PAGE

Document Title: Statistical Analysis Plan

Protocol Number: ONT-380-206

Version: 1.0; 7-Aug-2019

The individuals signing below have reviewed and approve of this statistical analysis plan.

Approval(s)

[Redacted Signature]

8- Aug - 2019
Date

Seattle Genetics, Inc.

[Redacted Signature]

8 Aug 2019
Date

Seattle Genetics, Inc.

[Redacted Signature]

7- Aug - 2019
Date

Seattle Genetics, Inc.

[Redacted Signature]

7- Aug - 2019
Date

Seattle Genetics, Inc.

TABLE OF CONTENTS

SPONSOR APPROVAL PAGE.....	2
LIST OF ABBREVIATIONS	5
1 INTRODUCTION.....	6
2 STUDY OBJECTIVES	6
2.1 Primary Objective.....	6
2.2 Secondary Objectives	6
2.2.1 Key Secondary Objectives	6
2.2.2 Other Secondary Objectives	6
2.2.3 Exploratory Objectives.....	7
3 STUDY DESIGN.....	7
3.1 Description	7
3.2 Method of Assigning Subjects to Treatment Arms.....	7
3.3 Endpoints.....	8
3.3.1 Primary Endpoint	8
3.3.2 Secondary Endpoints.....	8
3.3.3 Exploratory Endpoints.....	9
3.4 Data Monitoring Committee.....	9
3.5 Blinding.....	10
4 GENERAL STATISTICAL CONSIDERATIONS.....	10
4.1 Analysis Sets	10
4.2 Subgroups.....	11
4.3 Handling of Missing Data.....	11
4.3.1 Efficacy	11
4.3.2 Safety.....	12
4.3.3 Pharmacokinetics.....	13
4.4 Multicenter Studies.....	13
4.5 Determination of Sample Size	13
4.6 Timing of Analyses	13
4.7 Multiple Comparison/Multiplicity	14
4.8 Data Conventions, Definitions, and Formulas	16
5 STATISTICAL METHODOLOGY.....	18
5.1 Trial Details	18
5.1.1 Subject Disposition.....	18
5.1.2 Protocol Deviations	18
5.1.3 Baseline Characteristics and Disease History.....	18
5.1.4 Concomitant Therapy	19
5.1.5 Extent of Exposure	19
5.1.6 Subsequent anticancer treatment	20
5.2 Analysis of Efficacy	20
5.2.1 Primary Endpoint	21
5.2.2 Key Secondary Endpoints	25
5.2.3 Other Secondary Endpoints.....	26
5.2.4 Exploratory Efficacy Endpoints	28
5.3 Analysis of Safety.....	30
5.3.1 Adverse Events.....	30
5.3.2 Serious Adverse Events.....	31
5.3.3 Adverse Events of Special Interest.....	31
5.3.4 Clinical Laboratory Results	32
5.3.5 Ejection Fraction	32
5.3.6 Vital Signs	32
5.3.7 Deaths.....	32

5.4	Pharmacokinetic Analyses.....	33
5.5	Health Economics and Outcomes.....	33
5.5.1	Health Care Resource Utilization.....	33
5.5.2	Health-Related Quality of Life.....	33
5.6	Interim Analyses.....	33
5.7	Changes in the Planned Analysis.....	33
6	REFERENCES.....	34

LIST OF IN-TEXT TABLES

Table 1:	Initial LD (OF) boundaries for PFS _{BM} (2 analyses) and OS (3 analyses).....	14
Table 2:	LD (OF) boundaries for PFS _{BM} (2 lo analyses) and OS (3 analyses) at $\alpha=0.05$ level.....	15
Table 3:	Initial LD (OF) boundaries for PFS _{BM} (2 analyses) and OS (2 analyses).....	15
Table 4:	LD (OF) boundaries for PFS _{BM} (2 analyses) and OS (2 analyses) at $\alpha=0.05$ level.....	16
Table 5:	Analysis population and re-randomization procedure for efficacy endpoints	21
Table 6:	Censoring Scheme for Primary Analysis of PFS.....	22
Table 7:	Analysis Data Set Structure based on Alternative Randomizations	23
Table 8:	Censoring Scheme for the Primary Analysis of OS.....	26

LIST OF IN-TEXT FIGURES

Figure 1:	Illustration of Timing of Analyses.....	14
-----------	---	----

LIST OF ABBREVIATIONS

AE	Adverse Event(s)
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BICR	Blinded Independent Central Review
CBR	Clinical Benefit Rate
CDISC	Clinical Data Interchange Standards Consortium
CDS	Clinical Drug Safety
CMH	Cochran Mantel Haenszel
CNS	Central Nervous System
CR	Complete Response
CTCAE	Common Toxicity Criteria for Adverse Events
ctDNA	Circulating tumor DNA
DMC	Data Monitoring Committee
DOR	Duration of response
ECHO	Echocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic Case Report Form
ED	Emergency Department
HER2	Human Epidermal Growth Factor Receptor 2
IRT	Interactive Response Technology
ITT	Intent-To-Treat
LFT	liver Function Test
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
MUGA	multiple-gated acquisition scan
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PFS	Progression Free Survival
PK	Pharmacokinetics
PR	Partial Response
PT	Preferred Term
RANO-BM	Response Assessment in Neuro-Oncology Brain Metastases
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SD	Stable Disease
SMQ	Standardized MedDRA Queries
SOC	System Organ Class
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TESAE	Treatment Emergent Serious Adverse Event
T-DMI	ado-trastuzumab emtansine or trastuzumab emtansine
ULN	Upper Limit of Normal
WHODRUG	World Health Organization Drug Dictionary

1 INTRODUCTION

This document outlines the statistical methods to be implemented within the scope of Protocol ONT-380-206, entitled “Phase 2 Randomized, Double-Blinded, Controlled Study of Tucatinib vs. Placebo in Combination with Capecitabine and Trastuzumab in Patients with Pretreated Unresectable Locally Advanced or Metastatic HER2+ Breast Carcinoma (HER2CLIMB)”. Results of the proposed analyses will become the basis of the clinical study report for this protocol.

The purpose of this plan is to provide specific guidelines from which the analyses will proceed. All planned analyses specified in this document will be performed. Any changes will either be reflected in amendments to this plan before the database lock or specifically documented in the clinical study report. Any changes to this plan, in the form of “post hoc” or “data driven” analyses will be identified as such in the clinical study report.

2 STUDY OBJECTIVES

2.1 Primary Objective

- To assess the effect of tucatinib vs. placebo in combination with capecitabine and trastuzumab on progression-free survival (PFS) per RECIST 1.1 based on blinded independent central review (BICR)

2.2 Secondary Objectives

2.2.1 Key Secondary Objectives

- To assess the effect of tucatinib vs. placebo in combination with capecitabine and trastuzumab on PFS in patients with a history of brain metastases or brain metastases at baseline or equivocal brain lesions at baseline using RECIST 1.1 based on BICR
- To assess the effects of tucatinib vs. placebo in combination with capecitabine and trastuzumab on overall survival (OS)

2.2.2 Other Secondary Objectives

- To assess the effects of tucatinib vs. placebo in combination with capecitabine and trastuzumab on objective response rate (ORR) per RECIST 1.1 based on BICR and investigator
- To assess the effect of tucatinib vs. placebo in combination with capecitabine and trastuzumab on PFS per RECIST 1.1 based on investigator assessment
- To assess the duration of response (DOR) of tucatinib in combination with capecitabine and trastuzumab per RECIST 1.1 based on BICR and investigator
- To assess the clinical benefit rate (CBR) [stable disease (SD) or non-CR/non-PD for ≥ 6 months, or best response of complete response (CR) or partial response (PR)] of tucatinib vs. placebo in combination with capecitabine and trastuzumab per RECIST 1.1 based on BICR and investigator

- To assess health-related quality of life and health economics associated with tucatinib vs. placebo in combination with capecitabine and trastuzumab based on patient health status collected using the EQ-5D-5L instrument and health care resources utilized in patient care
- To assess the safety and tolerability of tucatinib in combination with capecitabine and trastuzumab
- To evaluate the pharmacokinetics of tucatinib and metabolite ONT-993 when administered in combination with capecitabine and trastuzumab

2.2.3 Exploratory Objectives

- To assess the effect of tucatinib vs. placebo in combination with capecitabine and trastuzumab using RANO-BM by BICR in the subgroup of patients with brain metastases at baseline
- To assess the effect of tucatinib vs. placebo in combination with capecitabine and trastuzumab on progression in brain in the subgroup of patients with brain metastases at baseline.
- To identify potential biomarkers of response, including human epidermal growth factor receptor 2 (HER2) mutations and other mutations by DNA sequence analyses of ctDNA isolated from plasma samples

3 STUDY DESIGN

3.1 Description

This is a randomized, international, multi-center, double-blinded study of tucatinib or placebo in combination with capecitabine and trastuzumab in patients with unresectable locally advanced or metastatic HER2+ breast cancer who have had prior treatment with trastuzumab, pertuzumab and T-DM1 (ado-trastuzumab emtansine or trastuzumab emtansine). After signing informed consent and meeting all eligibility criteria, patients will be randomized to receive tucatinib or placebo in combination with capecitabine and trastuzumab.

Treatment will be administered in cycles of 21 days each. Tucatinib or placebo will be given PO BID. Capecitabine will be given at 1000 mg/m² PO BID on Days 1–14 of each 21-day cycle. Trastuzumab will be given as a loading dose of 8 mg/kg IV followed by 6 mg/kg once every 21 days (or as 600 mg of trastuzumab given subcutaneously once every 3 weeks), except in specific circumstances where it may be given weekly to compensate for modifications in treatment schedule.

3.2 Method of Assigning Subjects to Treatment Arms

Subjects will be assigned to the tucatinib or placebo arms (in combination with capecitabine and trastuzumab) in a 2:1 ratio using a dynamic hierarchical randomization scheme. The randomization scheme will control for the stratification factors:

- Presence or history of treated or untreated brain metastases (Yes, No),
- Eastern Cooperative Oncology Group Performance Status (0, 1)
- Region of world (US, Canada, Rest of World)

The presence or history of brain metastases will be determined based upon investigator assessment of screening MRI and clinical history. Patients who have a documented history of prior brain metastases or unequivocal presence of brain lesions on screening MRI will be considered a “Yes” for stratification purposes, and subsequent efficacy assessments. Patients with brain lesions of equivocal significance on screening MRI will also be considered a “Yes” for purposes of stratification and follow-up.

The dynamic hierarchical randomization scheme includes specifications for a biased-coin assignment when the imbalance at a given hierarchical level (overall treatment group balance, then treatment group balance within each of the listed stratification factors) has exceeded a specified threshold.

3.3 Endpoints

3.3.1 Primary Endpoint

The primary endpoint is progression-free survival (PFS) time defined as the time from the date of randomization to the date of documented disease progression (as determined by BICR assessment using RECIST 1.1) or death from any cause, whichever occurs first.

3.3.2 Secondary Endpoints

Key secondary endpoints

- Progression-free survival (PFS_{BM}) time in the subgroup of patients with a history of brain metastases or brain metastases at baseline, or with brain lesions of equivocal significance on screening MRI, defined as the time from the date of randomization to the date of documented disease progression (as determined by BICR assessment using RECIST 1.1) or death from any cause, whichever occurs first.
- Overall survival (OS) time defined as the time from the date of randomization to the date of death from any cause.

Other Secondary Efficacy Endpoints

- Objective response rate (ORR). Objective response is defined as achieving a best overall response of complete (CR) or partial response (PR) as determined by BICR and by investigator using RECIST 1.1.
- Progression-free survival (PFS_{INV}) time defined as the time from the date of randomization to the date of documented disease progression (as determined by the investigator using RECIST 1.1) or death from any cause, whichever occurs first.

- Duration of response (DOR) defined as the time from the first objective response (CR or PR) to documented disease progression (PD) (as determined by BICR and by investigator using RECIST 1.1) or death from any cause, whichever occurs first.
- Clinical benefit rate (CBR). Clinical benefit is defined as achieving stable disease (SD) or non-CR/non-PD for ≥ 6 months or a best overall response of complete (CR) or partial response (PR) as determined by BICR and by investigator using RECIST 1.1.

Safety Endpoints

- Adverse events (AEs)
- Clinical laboratory assessments
- Vital signs and other relevant safety variables
- Frequency of dose holding, dose reductions, and discontinuations of capecitabine
- Frequency of dose holding, dose reductions, and discontinuations of tucatinib
- Frequency of dose holding and discontinuations of trastuzumab

Pharmacokinetic Endpoints

- Plasma concentrations of tucatinib and metabolite ONT-993

Health Economics and Outcomes

- Cumulative incidence of health resource utilization, including length of stay, hospitalizations, and emergency department (ED) visits.
- Health-related quality of life / health status, assessed using the EQ-5D-5L instrument.

3.3.3 Exploratory Endpoints

- Presence of HER2 mutations or other potential biomarkers of response
- In the subgroup of patients with a history of brain metastases or brain metastases at baseline, the following exploratory endpoints will be evaluated, as assessed by BICR using RANO-BM for brain metastases.
 - Objective response rate (ORR_{BC}) in brain
 - Duration of response (DOR_{BC}) in brain
 - Time to progression in brain (excluding body)

3.4 Data Monitoring Committee

An independent data monitoring committee (DMC) will monitor the safety of subjects and provide an ongoing clinical assessment of the study treatment's evolving safety profile as the

trial progresses. The DMC will review blinded and unblinded data that include deaths, discontinuations, dose reductions, adverse events, events of special interest, and serious adverse events. The DMC will meet on a regular basis and make recommendations to the sponsor regarding the conduct of the trial. Further details regarding the DMC's roles, responsibilities, and operating procedures are described in a separate DMC charter.

3.5 Blinding

This is a double-blinded trial. Patients, site investigators and personnel, the sponsor (except for designated Clinical Drug Safety (CDS) personnel), and all other individuals involved in the monitoring, data management, and/or conduct of the trial will be blinded. Designated CDS personnel may request the treatment assignment of an individual subject in the event of a Suspected Unexpected Serious Adverse Reaction (SUSAR) but will not have access to the overall randomization scheme.

Unblinded data including deaths, discontinuations, dose reductions, adverse events (serious and non-serious) will be monitored regularly by an independent DMC. The independent data coordinating center preparing this output for the DMC will be unblinded and have access to the overall randomization scheme.

At the time of the primary analysis for the primary endpoint (PFS), specific sponsor personnel will be unblinded, however sponsor personnel directly involved in the conduct of the study will remain blinded to individual subject treatment assignments (tucatinib/placebo) until the final analysis for the key secondary endpoint of PFS_{BM}.

4 GENERAL STATISTICAL CONSIDERATIONS

4.1 Analysis Sets

Intent-to-Treat

The intent-to-treat (ITT) analysis set will include all randomized subjects.

Specifically, the primary analyses for the primary endpoint of PFS per BICR will be conducted using the first 480 randomized subjects in the ITT analysis set (henceforth referred to as ITT-PFS set). The analyses of the key secondary endpoint OS will be conducted on all the randomized subjects in the ITT analysis set (henceforth referred to as ITT-OS set). The analysis of the key secondary endpoint PFS_{BM} will be conducted using all the randomized subjects in the BM subgroup (as defined in section 4.2) in the ITT analysis set (henceforth referred to as ITT-PFS_{BrainMets} set).

Subjects will be evaluated by their randomized treatment assignment.

Safety

The safety analysis set will include all randomized subjects who received at least one dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab). Subjects will be evaluated by the study treatment actually received.

Pharmacokinetics

The pharmacokinetic analysis set will include all randomized subjects who received at least one dose of tucatinib and who have at least one evaluable PK assessment. Subjects will be evaluated by the treatment actually received.

4.2 Subgroups

The following subgroup variables will be evaluated for primary and key secondary efficacy endpoints when applicable, as supportive analyses. If the total number of subjects in a subgroup is less than 10% of the total population, the subgroup analysis will not be performed. Subgroup analyses will be conducted using conventional stratified log rank statistical methods (i.e., rerandomization methods will not be used), as well as stratified Cox proportional hazards regression model. If the subgroup is a stratification factor, then the stratified models will use all the other stratification factors.

- History of parenchymal brain metastases or brain metastases at baseline (Yes, No): Patients with target and/or non-target parenchymal brain lesions (per RECIST 1.1) at baseline or who have a history of brain metastases, or with brain lesions of equivocal significance on screening MRI based on screening data collected in eCRF will be assigned to the ‘Yes’ subgroup. This group will henceforth be referred to as “BM subgroup” in this document. Patients not meeting the above criteria will be assigned to the ‘No’ subgroup for this variable. Patients with dural lesions only, i.e. no parenchymal brain lesions, will be assigned to the ‘No’ subgroup. Patients with incomplete screening data and not meeting the criteria for BM subgroup will be not evaluable (NE) for this subgroup determination.
- Geographic Region: North America, Rest of World
- ECOG: 0 vs. 1 as recorded in eCRF at baseline
- Age : <65 vs. ≥65 years
- Race: White, African-American, others
- Hormone Receptor Status (Negative, Positive): Patients ‘positive’ for either or both estrogen receptor and progesterone receptor will be assigned to the ‘positive’ subgroup. Patients not meeting the above criteria will be assigned to the ‘negative’ subgroup.

4.3 Handling of Missing Data

4.3.1 Efficacy

Partial dates of start of post study treatment anti-cancer therapy and dates of concomitant medication or procedure, will be imputed as follows:

- Missing day only: For partial dates with only the day of the month missing, the day will be imputed as the 15th day of the month, provided that a preceding or succeeding

date of interest does not occur in the same month. If the preceding date does occur in the same month, then the missing day will be imputed as half the distance between the preceding date and the end of the month. If the succeeding date does occur in the same month, then the missing day will be imputed as half the distance between the beginning of the month and the succeeding date.

- Missing month and day: For partial dates with the month and day missing, the month and day will be imputed as July 1st, provided that a preceding or succeeding date of interest does not occur in the same year. If the preceding date occurs in the same year, then the month and day will be imputed as half the distance from the preceding date to the end of the known year. If the succeeding date does occur in the same year, then the missing day will be imputed as half the distance between the beginning of the year and the succeeding date.

The date of death, tumor assessment dates, dates of last contact, hospitalization, date of initial diagnosis and first date of metastases will be imputed if only day is missing, following the “Missing day only” rule as above.

4.3.2 Safety

For AEs where the date of onset is during or after administration of the first dose of study treatment, missing or partial start dates will be imputed as the earliest possible date that is on or after the date of the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab) and before the AE end date.

- Missing day only: If the start date and date of first dose share the same month and year, the missing start day will be imputed as the day of first dose. If the start date month is after the month of first dose, day will be imputed as the 1st (i.e., 01-*MMM*-*YY*).
- Missing month and day: If the start date and date of first dose share the same year, day and month will be imputed as the day and month of first dose. If the start date year is after the year of first dose, the month and day will be imputed as January 1st (i.e., 01-*JAN*-*YY*).
- Missing month, day, and year: Missing start dates will be imputed as the date of first dose.

For AEs where the end date is a partial date, the end date will be imputed as:

- Missing day only: AE end date will be imputed as the minimum of (death date, data cutoff date, last day of the end date month/year)
- Missing month and day: AE end date will be imputed as the minimum of (death date, data cutoff date, December 31st of the end date year)
- Missing month, day and year: not imputed.

4.3.3 Pharmacokinetics

Missing values for PK measurements will be listed as missing and excluded from the calculation of summary statistics.

4.4 Multicenter Studies

Approximately 600 subjects will be randomized to the study from approximately 200 sites worldwide.

4.5 Determination of Sample Size

The sample size for this study was calculated based on maintaining 90% power for the primary endpoint PFS and 80% power for OS with an alpha level of 0.02.

For PFS, 288 events are required with 90% power to detect a hazard ratio of 0.67 (4.5 months median PFS in the control arm versus 6.75 months in the experimental arm) using a 2-sided log-rank test and alpha of 0.05.

For OS, 361 events are required with 80% power to detect a hazard ratio of 0.70 (15 months median OS in the control arm vs. 21.4 months in the experimental arm) using a 2-sided log-rank test and alpha of 0.02, taking into account of two interim analyses. With 361 OS events, it will provide 88% power using a 2-sided log-rank test with an alpha of 0.05.

For PFS_{BM}, 220 events are required with 80% power to detect a hazard ratio of 0.67 (4.5 months median PFS_{BM} in the control arm versus 6.75 months in the experimental arm) using a 2-sided log-rank test at alpha of 0.05, taking into account of one interim analysis. The power will be 74% at 2-sided alpha of 0.03.

Approximately 600 subjects will be randomized in a 2:1 ratio to either the experimental arm or the control arm. Assuming an accrual period of 48 months and a 5% yearly drop-out rate, it is expected that 361 OS events will be observed approximately 59 months after first subject randomized.

Sample size and power were calculated using EAST[®] version 6.4, by Cytel Inc.

4.6 Timing of Analyses

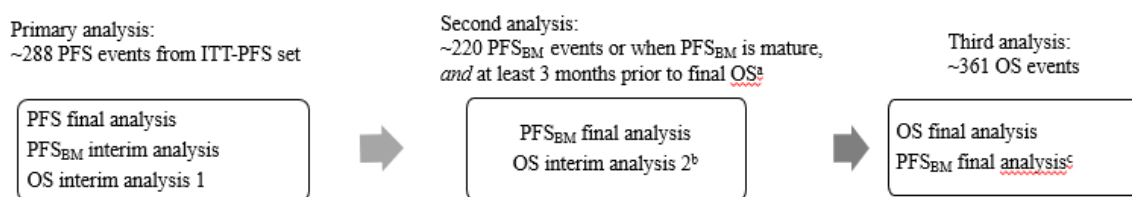
The primary analysis of PFS will occur when approximately 288 PFS events determined by BICR have occurred in the ITT-PFS set and enrollment has been completed for the study. An interim analysis for the key secondary endpoints PFS_{BM} and OS will also be performed at this time if PFS is statistically significant (see Section 5.6 for details).

If PFS_{BM} is statistically significant at the first interim analysis, no further formal testing of PFS_{BM} will be conducted. The second interim analysis for OS will be performed when approximately 75% (271) of total OS events have occurred in the ITT-OS set, and the final OS analysis will be conducted after approximately 361 OS events have occurred in the ITT-OS set.

If PFS_{BM} is not statistically significant at the first interim analysis, a second analysis of the key secondary endpoints will be performed when (a) approximately 220 PFS_{BM} events have occurred in the ITT-PFS_{BrainMets} set or the PFS_{BM} events are sufficiently mature (e.g. approximately less than 6 events are expected with 3 months additional follow up) and (b) at least 3 months before the projected OS final analysis. If both of the above conditions (a and b) are not met, this analysis of PFS_{BM} and OS will not be conducted, and the OS final analysis at 361 OS events will also be the timing of the final PFS_{BM} analysis.

The timing of analyses for the primary and key secondary endpoints are illustrated in Figure 1.

Figure 1: Timing of Primary and Key Secondary Endpoints Analyses



^a If these two conditions are not met, then this analysis will be skipped.

^b If PFS_{BM} is positive at primary analysis, OS interim analysis 2 will be conducted at 75% OS events (~271)

^c only if conditions for second analysis timing not met

4.7 Multiple Comparison/Multiplicity

To maintain strong control of the family-wise type I error rate at 0.05, the PFS will be tested at 0.05 level first in the ITT-PFS set, if it is significant, then the key secondary endpoints will be tested using the group sequential Holm variable (GSHv) procedure (Ye et.al. 2013).

The α split between PFS_{BM} and OS is $\alpha=0.03$ and $\alpha=0.02$, respectively, and each one will be tested at the interim analysis(s) and again at the final analysis, if not rejected at the interim analysis. The information fraction t is the ratio between number of events at interim analysis and number of events at final analysis. For illustration purpose, we assume $t=0.812$ for PFS_{BM} and $t_1=0.626$, $t_2=0.779$ for the two interim analyses for OS.

The boundary at interim analysis is determined according to the Lan-DeMets O'Brien-Fleming (LD(OF)) approximation spending function LD(OF): $\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right)$ for two-sided tests, where $z_{\alpha/4}$ is the upper $\frac{\alpha}{4}$ critical point of the standard normal distribution.

The GSHv procedure operates as follows.

- Begin with a 0.03-level group sequential boundary for PFS_{BM} and a 0.02-level group sequential boundary for OS. The corresponding boundaries for the two endpoints are given in Table 1.

Table 1: Initial LD (OF) boundaries for PFS_{BM} (2 analyses) and OS (3 analyses)

Analysis	PFS _{BM} ($\alpha=0.03$, $t=0.812$)	OS ($\alpha=0.02$, $t_1=0.626$, $t_2=0.779$)
----------	---	--

1	0.0139	0.0023
2	0.0259	0.0069
3		0.0176

If both of the endpoints are found significant at analysis 1 (primary analysis), then no more formal statistical testing for PFS_{BM} and OS will be conducted.

- If only one endpoint is found significant at the analysis 1 (primary analysis) then the α can be recycled to the other endpoint: If PFS_{BM} is significant at interim but OS is not then the α will be recycled from PFS_{BM} to OS and use a 0.05-level LD(OF) boundary for OS. On the other hand, if OS is significant at analysis 1 (primary analysis) but PFS_{BM} is not then the α will be recycled from OS to PFS_{BM} and use a 0.05-level LD(OF) boundary for PFS_{BM}. The corresponding 0.05-level LD(OF) boundaries are given in Table 2. The unrejected hypothesis can be re-tested at the current and future analysis using the modified boundaries.

Table 2: LD (OF) boundaries for PFS_{BM} (2 lo analyses) and OS (3 analyses) at $\alpha=0.05$ level

Analysis	PFS _{BM} ($\alpha=0.05$, $t=0.812$)	OS ($\alpha=0.05$, $t1=0.626$, $t2=0.779$)
1	0.0258	0.0092
2	0.0425	0.0194
3		0.0429

- If neither of the endpoints is found significant at analysis 1 (primary analysis), then both endpoints will be tested again at analysis 2. The initial boundaries for final analysis follow Table 1 (analysis 2). If only one endpoint is found significant by these initial boundaries, then the other one can be tested again using the modified boundary as shown in Table 2 (analysis 2). For example, if PFS_{BM} was found significant at final analysis at $\alpha=0.0259$ level, but OS was not significant at $\alpha=0.0069$ level, then OS can be tested again at the $\alpha=0.0194$ level.
- If PFS_{BM} is significant at analysis 1 or 2, the boundary of OS analysis at analysis 3 is 0.0429; otherwise, the boundary for OS analysis at analysis 3 is 0.0176.
- Note that the boundaries presented in the tables will be adjusted with the actual information fraction.

As detailed in Section 4.6, the second interim analysis for OS may not be conducted, which means both PFS_{BM} and OS will have at most 2 analyses. In that case, LD(OF) boundaries at each analysis will be modified as illustrated in Table 3 and Table 4. Similar to Table 1 and Table 2, the information fraction (t) in Table 3 and Table 4 are for illustration purpose only.

Table 3: Initial LD (OF) boundaries for PFS_{BM} (2 analyses) and OS (2 analyses)

Analysis	PFS _{BM} ($\alpha=0.03$, $t=0.812$)	OS ($\alpha=0.02$, $t=0.626$)
1	0.0139	0.0023
2	0.0259	0.0193

Table 4: LD (OF) boundaries for PFS_{BM} (2 analyses) and OS (2 analyses) at $\alpha=0.05$ level

Analysis	PFS _{BM} ($\alpha=0.05$, $t=0.812$)	OS ($\alpha=0.05$, $t=0.626$)
1	0.0258	0.0092
2	0.0425	0.0471

If both PFS_{BM} and OS are statistically significant, the secondary endpoint of ORR by BICR in the ITT-OS set will be formally tested between two treatment arms at the two sided $\alpha=0.05$ level.

4.8 Data Conventions, Definitions, and Formulas

The following data conventions will be used for the tables, listings, and figures.

- **Study treatment:** tucatinib/placebo, capecitabine or trastuzumab. Subjects who discontinued tucatinib/placebo and only continued with capecitabine and/or trastuzumab are not considered to be receiving study treatment anymore.
- **Baseline:** The last non-missing observation prior to or on the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab), unless otherwise specified. If more than one assessment meet the above criteria and were collected on the same date, the value of the assessment indicating better status will be used as baseline to be conservative, for instance, lower vs. higher lab grade, ECOG 0 vs. 1, ECG normal vs. abnormal. If there are no directional difference (for instance, lab values of the same grade) of observation on the same date, the last record in database (identified by sequence number or visit number) will be marked as baseline. For patients who were randomized but not treated, the observation on screening visit will be marked as baseline.
- **Pre-treatment Period:** Prior to first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab)
- **Study Treatment Period:** Period of time that begins on the date of the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab) through 30 days after the date of the final dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab) as recorded in the End of Treatment eCRF page.
- **Study Day:**
 - **Safety:** Study day will be calculated for safety endpoints relative to the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab). The first dose of study treatment will be Day 1, and the date preceding Day 1 will be Day -1 which is consistent with the Submission Data Standards (Version 3.1) from Clinical Data Interchange Standards Consortium (CDISC).
 - **Efficacy:** Study day will be calculated for efficacy endpoints relative to the date of randomization. The day of randomization will be Day 1.

- **Duration of Exposure:**

- For tucatinib/placebo, duration of exposure (days) = date of last dose – date of first dose + 1
- For capecitabine, duration of exposure (days) = (date of last dose +7) – date of first dose) + 1
- For trastuzumab, duration of exposure (days) = (date of last dose +20) – date of first dose) + 1

Note that for each study drug, the date of last dose is as recorded in the End of Treatment eCRF page.

- **Total Dose of tucatinib/placebo (mg), capecitabine (mg/m²), and trastuzumab (mg/kg):**

$$\text{Total dose (units)} = \sum_{i=1}^n (\text{dose}_i)$$

where i = dose number, $\text{dose}_i = i^{\text{th}}$ dose received (units), n = total number of doses received

- **Intended dose intensity (IDI):** the intended dose of drug per unit of time (day).

For example, tucatinib/placebo: IDI = 300mg BID =600 (mg/day);

capecitabine: IDI = 2000 mg/m²/day * (14 dosing day/21 days in a cycle) =1333.3 mg/m²/day

- **Absolute dose intensity (ADI):** the actual dose per unit of time that the subject received over the duration of exposure for that study drug.

$$\text{ADI} = \text{Total dose} / \text{Duration of exposure (days)}$$

- **Relative dose intensity (RDI):** the percent of the intended dose intensity over the entire treatment period:

$$\text{RDI} = \text{ADI/IDI} * 100\%$$

- **Response assessment date**

For efficacy assessments, the date of response assessment of CR, PR or non-CR/non-PD, SD will be the latest of all radiologic scan dates for the given response assessment. The date of equivocal progression or progression will be the earliest of all radiologic scan dates that showed evidence of PD for the given response assessment.

5 STATISTICAL METHODOLOGY

5.1 Trial Details

5.1.1 Subject Disposition

Patient enrollment and disposition will be summarized by treatment group and total. The table will present the number and percentage of patients who were randomized in each stratum, received study drug, received treatment per randomization assignment, and participated in follow-up visits. The number and percentage of patients who discontinued treatment will be summarized by the reason for treatment discontinuation. The number and percentage of patients who discontinued the study will be summarized by the primary reason for study discontinuation. The summary of disposition will be conducted for ITT-PFS, ITT-OS and ITT-PFS_{BrainMets} sets.

Number of patients who signed informed consent and number of patients in each analysis set will be summarized by treatment group and total.

Number of screen failures and the percentage relative to the total number of subjects screened will be summarized. A listing of subjects who failed screening will also be produced, with reasons for screen failure and available demographic information.

The number of patients enrolled in each country and at each site will be summarized by treatment group and total.

5.1.2 Protocol Deviations

Protocol deviations (as defined in the ONT-380-206 Global Clinical Monitoring Plan) will be identified by site monitors, the medical monitor and by checks of the clinical database. Important deviations will be summarized for each treatment arm by type of deviation. All protocol deviations will be listed.

5.1.3 Baseline Characteristics and Disease History

Baseline characteristics will be summarized for each treatment arm using counts and percentages for categorical variables and summary statistics (mean, median, standard deviation, and range) for continuous variables. Characteristics to be summarized include the following:

- Demographic variables: age, sex, race, and ethnicity
- ECOG performance status
- Disease history:
 - Time (months) from diagnosis of breast cancer to randomization
 - Time (months) from metastatic diagnosis to randomization
 - Unresectable locally advanced breast cancer (yes, no)

- Stage at diagnosis
- History of brain metastases or brain metastases at study entry
- Estrogen/progesterone receptor status
- Non-CNS metastatic disease sites
- Brain metastases treatment status at baseline (treated stable, treated progressive and untreated)
- Time (months) from date of first diagnosis of brain metastases to randomization in subject previously diagnosed with brain metastases
- Prior surgery and/or radiotherapy for brain metastases (yes, no)
- Type of prior radiotherapy for brain metastases (whole brain vs. targeted radiation)
- Prior systemic therapies

The summary of demography and baseline disease characteristics will be conducted for ITT-OS, ITT-PFS and ITT-PFS_{BrainMets} sets.

5.1.4 Concomitant Therapy

Concomitant medications will be listed and coded using the World Health Organization Drug Dictionary (WHODRUG) Version 2009Q3 or higher and summarized for each treatment arm by preferred term and treatment arm using counts and percentages. Multiple occurrences of the same medication within a subject will be summarized only once.

Concomitant systemic corticosteroids, antidiarrheals as well as concomitant procedures will be summarized and listed.

5.1.5 Extent of Exposure

Exposure will be summarized by treatment arm using counts and percentages for categorical variables and summary statistics (mean, median, standard deviation, and range) for continuous variables.

The following information will be summarized separately for capecitabine, trastuzumab, and tucatinib/placebo:

- Total number of treatment cycles per subject
- Duration of exposure
- Total cumulative dose
- Percentage of subjects with interrupted, missed, reduced and discontinued infusions/doses overall and by reason

- Absolute dose intensity (ADI) and relative dose intensity (RDI).

Dose reduction and ADI/RDI will only be summarized for tucatinib/placebo and capecitabine.

For tucatinib/placebo, the type, reason and time to first dose modification will be summarized. The total number of dose modifications and reasons will also be summarized.

The summary of trastuzumab and capecitabine exposure will only include the period when patients are on study treatment, i.e., before tucatinib/placebo is discontinued.

The extent of exposure will be summarized both for the safety set within first 480 randomized subjects and the safety set in all randomized subjects.

5.1.6 Subsequent anticancer treatment

The type and regimen of subsequent anticancer treatment after discontinuation from study treatment will be summarized by treatment arm.

5.2 Analysis of Efficacy

Table 5 summarizes the analysis population at the time of primary analysis for PFS, as well as whether the re-randomization procedure needs to be used to calculate p-value. Subjects will be analyzed based on their randomized treatment arm ('intent-to-treat' analysis).

Table 5: Analysis population and re-randomization procedure for efficacy endpoints

Analysis	Population for primary analyses	Use of re-randomization procedure
Primary endpoint: PFS per BICR ^a	ITT-PFS	Yes
Sensitivity analysis for primary endpoint	ITT-PFS	Yes
Subgroup analysis for primary endpoint	ITT-PFS	No
Key secondary endpoints: PFS _{BM}	ITT-PFS _{BrainMets}	Yes
Sensitivity analysis for PFS _{BM}	ITT-PFS _{BrainMets}	Yes
Subgroup analysis for PFS _{BM}	ITT-PFS _{BrainMets}	No
Key secondary endpoints: OS	ITT-OS	Yes
Sensitivity analysis for OS	ITT-OS	Yes
Subgroup analysis for OS	ITT-OS	No
PFS per INV ^a	ITT-PFS	No
PFS _{BM} per INV	ITT-PFS _{BrainMets}	No
ORR, CBR and DOR ^b	ITT-OS	No
Exploratory efficacy endpoints	ITT-OS	No

^a exploratory analyses will also be conducted using ITT-OS analysis set.
^b exploratory analyses will also be conducted using ITT-PFS analysis set.

For the primary analysis of PFS and OS, the stratification factors per Interactive Response Technology (IRT) system will be used as strata in stratified analysis. For a stratification factor with two strata, if one of the two strata has a sample size too small (e.g., less than 20%), the statistical analysis will not include this randomization stratification factor in the analysis. For a stratification factor with more than two strata, if one of the strata has a sample size too small (e.g., less than 20%), this stratum will be combined with some other strata of this stratification factor. For example, for the region stratification factor, if the stratum Canada has a small sample size of less than 20% per the pooled blinded data, the stratum Canada will be combined with the stratum US in the statistical analysis.

For the primary analysis of PFS_{BM}, the actual stratification factor (brain metastasis as recorded in eCRF) will be used to define the population; the stratification factors per Interactive Response Technology (IRT) system (ECOG and region) will be used as strata in stratified analysis.

Sensitivity analyses using the eCRF values for the actual stratification factors as strata may be performed for the primary endpoint and the two key secondary endpoints if the percentage of mis-stratification exceeds five percent.

5.2.1 Primary Endpoint

5.2.1.1 Primary Analysis

Progression-free survival (PFS) time is defined as the time from the date of randomization to the date of documented disease progression (as determined by BICR assessment using RECIST 1.1) or death from any cause, whichever occurs first. Subjects who are alive and have not progressed at the time of the analysis will be censored at the time of their last tumor

assessment that was a CR, PR, non-CR/non-PD or SD. Details of the censoring scheme for the primary analysis of PFS are described below in Table 6.

Table 6: Censoring Scheme for Primary Analysis of PFS

Scenario	Progression/Censor Date	Outcome
No post-baseline tumor assessments	Date of randomization	Censored
No documented disease progression or death	Date of last tumor assessment of CR, PR, SD, or non-CR/non-PD	Censored
New anti-cancer treatment (systemic, radiation, or surgery) started before PD or death observed	Date of last CR, PR, SD, or non-CR/non-PD on or prior to date of new anti-cancer treatment	Censored
Progressive disease (PD)	Date of PD	Event
Death before first PD assessment	Date of death	Event
Death or progression right after two or more consecutive missed tumor assessments	Date of last tumor assessment of CR, PR, SD, or non-CR/non-PD	Censored

Note: CT, PET/CT scans are performed every 6 weeks starting at Cycle 1 Day1 through Week 24 and every 9 weeks starting at Week 24 until documented PD or death.

Partial or missing dates of death, dates of last contact, and tumor assessment dates will be imputed as described in Section 4.3.

The two treatment arms will be compared for PFS using a stratified, log-rank test controlling for the randomization stratification factors [i.e., history of brain metastases or presence of brain metastases or lesions of equivocal significance on screening MRI (yes, no), ECOG status (0, 1), and region of the world (US, Canada, Rest of world)]. The p-value for this test will be calculated using a re-randomization based procedure (Rosenberger and Lachin, 2002) to reflect the dynamic, hierarchical allocation scheme (Section 3.2) used for the study randomization. The null hypothesis for this comparison is that the assignment of subjects to the two treatment arms had no effect on response for the subjects randomized to treatment. The procedure for performing the comparison and calculating the p-value are described below. Details of the specific significance level to be used for the treatment arm comparison are described in Section 4.7.

Using the randomized treatment assignments for the trial, a stratified, log-rank test chi-square statistic will be computed for the comparison of the two treatment arms for PFS. This test statistic will be referred to as X_0 and will be calculated based on the following sample SAS code:

```

** Pfstime = PFS time;
** Censor = Censor variable (1 = censored);
** Trt = treatment arm;
** BM = presence or history of treated or untreated brain metastases at baseline or
lesions of equivocal significance (Yes, No)
** ECOG (0,1)
** Region (US, Canada, Rest of World)
ODS OUTPUT HomTests=chisq;

```

```

PROC LIFETEST DATA = pfsdata;
  TIME pfstime*censor(1);
  STRATA bm ecog region / GROUP=armcd TEST=logrank;
RUN;

DATA x0(keep=x0);
SET chisq;
  where test = 'Log-Rank';
  X0=ChiSq;
RUN;

```

Utilizing the dynamic randomization algorithm used to create the original randomization scheme for the trial, 10,000 alternative subject randomizations will then be generated. Each subject randomization will then be merged with each subject’s observed PFS time, censoring status, and values for each stratification variable to produce an analysis data set with the basic structure shown in Table 7.

Table 7: Analysis Data Set Structure based on Alternative Randomizations

Randomization ID Number (i)	Subject ID	Baseline Stratification Variables			PFS Time	Censor	Treatment Arm
		Presence/History of Brain Metastases or lesions of equivocal significance	ECOG	Region			
1	1	Y	0	US	T ₁	0	A
1	2	N	1	Canada	T ₂	1	B
-	-	-	-	-	-	-	-
1	600	Y	1	US	T ₆₀₀	1	A
-	-	-	-	-	-	-	-
10,000	1	Y	0	US	T ₁	0	B
10,000	2	N	1	Canada	T ₂	1	A
-	-	-	-	-	-	-	-
10,000	600	Y	1	US	T ₆₀₀	1	B

For each alternative randomization, a stratified, log-rank test chi-square statistic will then be computed (using the same SAS code above used to calculate X₀) for the comparison of the two treatment arms for PFS. The test statistic for the i-th randomization and comparison will be referred to as X_i where i = 1 to 10,000.

The two-sided p-value (based on the rerandomization procedure), will then be calculated as the number of times X_i ≥ X₀ divided by 10,000.

$$p - value = \sum_{i=1}^{10,000} \frac{I(\cdot)}{10,000}, \text{ where } I(\cdot) = \begin{cases} 1 & \text{if } X_i \geq X_0 \\ 0 & \text{if } X_i < X_0 \end{cases}$$

The critical value (X_c) for a 0.05 significance level test using the rerandomization reference distribution will correspond to the 5th percentile of the X_i’s.

For the purpose of describing the treatment effect, the treatment arm hazard ratio and 95% confidence interval will be estimated using a stratified Cox proportional hazards regression model controlling for the study stratification factors. This analysis will be implemented using the following SAS sample code.

```
PROC PHREG DATA = pfsdata;  
  CLASS armcd(ref='B');  
  STRATA stratvar1 stratvar2 stratvar3;  
  MODEL pfstime*censor(1) = armcd / TIES=BRESLOW RL;  
  HAZARDRATIO armcd;  
RUN;
```

Kaplan-Meier curve will be generated by treatment arms. Kaplan-Meier estimates of the median and quartiles (corresponding 95% confidence intervals), as well as probability of PFS at different timepoints (for instance, 3, 6, 9 months) will be computed for each treatment arm.

In addition to the primary efficacy analysis (i.e., re-randomization model analysis), an analysis of PFS time will also be performed using a stratified, log-rank test based on the randomized treatment assignments. The analysis will be implemented using the following sample SAS code.

```
PROC LIFETEST DATA = pfsdata;  
  TIME pfstime*censor(1);  
  STRATA bm ecog region / GROUP=armcd TEST=logrank;  
RUN;
```

5.2.1.2 Sensitivity Analyses

- **Non-proportional hazard:** In the case that the proportional hazard assumption is violated (by plotting the “log-negative-log” of the Kaplan Meier estimator vs time by treatment group, and the Schoenfeld residuals by treatment group), a restricted mean survival time (RMST) analysis up to 18 months will be performed to compare the mean survival time of the two treatment arms. In addition, the Max-Combo test (Losorok and Lin, 1999; Karrison et al., 2016) will be performed to compare the two treatment arms.

If the primary analysis of PFS is significant, the following analyses may be performed for PFS, using the same rerandomization procedure as for the primary analysis.

- **Missing Assessments of Disease Response:** To explore the potential impact of missing assessments of disease response on the primary analysis of PFS, two sensitivity analyses will be performed
 - (1) Ignoring the missing assessments, i.e., subjects who missed two or more consecutive scheduled assessments before death or PD are considered to have had an event on the date of death or progression.

- (2) Imputing the missing assessment, i.e., subjects who missed two or more consecutive scheduled assessments before death or PD are considered to have events at the time of the next scheduled assessment after the last non-missing assessment.
- **Stratification:** In the case of stratification errors >5% between what is recorded in IRT and eCRF, the hazard ratio and its 95% CI will be estimated using a stratified Cox proportional hazards regression model controlling for the eCRF stratification factors.
- **New therapy before PD/death:** For subjects who received new anti-cancer therapy before PD or death, two sensitivity analyses will be conducted:
 - (1) Not to consider any anti-cancer therapies (whether systemic, radiation, or surgery) as a censoring reason,
 - (2) Not to consider radiation therapies as a censoring reason.

5.2.2 Key Secondary Endpoints

5.2.2.1 Progression-Free Survival in BM subgroup

BM subgroup is defined as subjects with target and/or non-target parenchymal brain lesions per RECIST 1.1) at baseline or who have a history of brain metastases, or with brain lesions of equivocal significance on screening MRI based on screening data collected in eCRF. Progression-free survival time (as defined for the primary efficacy endpoint) in the BM subgroup will be tested if the test of the primary endpoint of PFS is statistically significant (Section 4.7). This analysis will be performed using the same statistical methods and set of alternative subject randomizations used to evaluate the primary endpoint overall PFS. This will be accomplished by selecting only subjects in the BM subgroup from the set of 10,000 randomizations and calculating the stratified log-rank statistics (using stratification factors of ECOG and region) and rerandomization procedure p-value as described for primary analysis of PFS in Section 5.2.1.1. Details of the specific significance level to be used for the treatment arm comparison are described in Section 4.7.

The hazard ratio and 95% confidence interval will be estimated using a stratified Cox proportional hazards regression model controlling for the study stratification factors of ECOG and region.

The sensitivity analyses described in Section 5.2.1.2 for the primary endpoint may be performed if appropriate.

As an exploratory analysis, the Kaplan Meier curve and summary for PFS in the non-BM subgroup among all randomized subjects will also be presented.

5.2.2.2 Overall Survival

Overall survival time (OS) is defined as the number of days from the date of randomization until the date of death from any cause. Subjects who did not achieve the event (death) at the time of the analysis or are lost to follow-up will be censored at the date they were last known to be alive (i.e., right censored). Partial or missing dates of death or last contact will be imputed as described in Section 4.3. Details of the censoring scheme for the primary analysis of OS are described below in Table 8.

Table 8: Censoring Scheme for the Primary Analysis of OS

Scenario	Death/Censor Date	Outcome
Not known to have died by data cutoff date	Date last known alive	Censored
Death	Death date	Death

This analysis will be performed using the same statistical methods and set of alternative subject randomizations used to evaluate the primary endpoint PFS. This will be accomplished by merging each subject's observed survival time and censoring status to each subject randomization to produce an analysis data set. The stratified log-rank statistics and rerandomization procedure p-value will then be calculated as described for the primary efficacy analysis of PFS as described in Section 5.2.1.1. Details of the specific significance level to be used for the treatment arm comparison are described in Section 4.7.

The hazard ratio and 95% confidence interval will be estimated using a stratified Cox proportional hazards regression model controlling for the study stratification factors.

The sensitivity analyses for non-proportional hazard and stratification described in Section 5.2.1.2 for the primary endpoint may be performed if appropriate.

5.2.3 Other Secondary Endpoints

Analyses of the following secondary endpoints will not be subject to formal type I error control and will be analyzed using conventional log rank statistical methods (i.e., rerandomization methods will not be used).

5.2.3.1 Objective Response Rate

Objective response is defined as achieving a best overall response of confirmed complete (CR) or confirmed partial response (PR) per RECIST 1.1. Only response assessments before first documented PD or new anti-cancer therapies will be considered. The proportion of subjects with objective response will be calculated by treatment arm. Comparison of the two treatment arms will be performed using a 2-sided Cochran-Mantel-Haenszel (CMH) test controlling for the study stratification factors. ORR determined by BICR will be summarized for subjects who had at least one measurable target lesion at baseline as assessed by BICR among ITT-OS set. If both of the key secondary endpoints (OS and PFS_{BM}) are statistically significant, then the ORR by BICR will be formally tested between two treatment arms (see Section 4.7), the p-value from the stratified CMH test will be reported.

ORR determined by investigator assessment will be summarized for subjects who had at least one measurable target lesion at baseline as assessed by investigator among ITT-OS set. As exploratory analyses, the same analyses for ORR will also be conducted using ITT-PFS set. The nominal p-value from the stratified CMH tests will be reported for ORR determined by investigator in ITT-OS and ITT-PFS sets and ORR determined by BICR in ITT-PFS set.

5.2.3.2 Investigator Assessed Progression-Free Survival

Progression-free survival (PFS_{INV}) time is defined as the time from the date of randomization to the date of documented disease progression (as determined by the investigator using RECIST 1.1) or death from any cause, whichever occurs first. Subjects who are alive and have not progressed at the time of the analysis will be censored at the time of their last tumor assessment that was a CR, PR, non-CR/non-PD, SD or equivocal progression. Details of the censoring scheme for the analysis of PFS_{INV} are described above in Table 6. The primary analysis of PFS_{INV} will be performed based on ITT-PFS set. PFS_{INV} will also be summarized based on ITT-OS set as an exploratory analysis.

In the cases where an equivocal new lesion was later confirmed to be a truly new disease lesion, or a non-target lesion assessed as equivocal progression changed to unequivocal progression in consecutive later assessment, the PD date should be backdated to the visit when the equivocal new lesion or non-target lesion with equivocal progression was first identified. Note: in cases where PD occurs at a date after an equivocal new lesion or equivocal progression of non-target lesion is identified, but the progression is *not* due to a change of the same equivocal new lesion or equivocal non-target lesion progression to an unequivocal lesion or unequivocal non-target lesion progression, but rather from progression of other lesions, the PD date will not be backdated, but will be the date when definitive PD is recorded.

The treatment arm hazard ratio and 95% confidence interval will be estimated using a stratified Cox proportional hazards regression model controlling for the study stratification factors. Comparison of the two treatment arms will be performed using a stratified log-rank test controlling for the study stratification factors. The nominal p-value from the stratified log-rank test will be provided. Kaplan-Meier estimates of the median (corresponding 95% confidence intervals) will also be computed for each treatment arm.

To explore the potential impact of clinical progression on the analysis of PFS, a sensitivity analysis will be performed using the same censoring scheme and methods described for the primary analysis of PFS with the exception that subjects who discontinued any study treatment due to clinical progression will be counted as ‘progressed’ in the analysis.

PFS_{INV} will also be summarized based on ITT-PFS_{BrainMets} set as exploratory analyses.

In addition, the concordance between BICR and investigator assessed PFS event will be summarized.

5.2.3.3 Clinical Benefit Rate

Clinical benefit is defined as achieving stable disease (SD) or non-CR/non-PD for ≥ 6 months (i.e., subject has been followed for at least 6 months and no documented PD or death within 6 months from date of randomization) or a best overall response of confirmed complete (CR) or confirmed partial response (PR) per RECIST 1.1. A subject will have a best response of SD or non-CR/non-PD if there is at least one SD or non-CR/non-PD assessment ≥ 5 weeks after the date of randomization and the subject does not qualify for CR or PR. The duration of SD or non-CR/non-PD will be calculated for subjects who had a best response of SD or non-CR/non-PD, and is defined as the duration from the date of randomization to documented PD or death. Only response assessments before first documented PD or new anti cancer therapies will be considered. The same derivation of PD date and censoring rules as for primary PFS analysis will apply for duration of SD or non-CR/non-PD.

The proportion of subjects with clinical benefit determined by BICR will be calculated by treatment arm. Comparison of the two treatment arms will be performed using a 2-sided CMH test controlling for the study stratification factors. The nominal p-value from the stratified CMH test will be reported. Similar analysis will be performed for CBR determined by investigator assessment. For investigator assessed CBR, the same algorithm for backdating equivocal progression will be applied as in Section 5.2.3.1. CBR will be summarized for the ITT-OS set. As exploratory analyses, the same analyses for CBR will also be conducted using ITT-PFS set.

5.2.3.4 Duration of Response

Duration of response (DOR) is defined as the time from the first objective response (CR or PR that is subsequently confirmed) to documented disease progression (PD) per RECIST 1.1 or death from any cause, whichever occurs first. Only those who achieve a confirmed response among the ITT-OS set will be included in the analysis. Kaplan-Meier estimates of the median (corresponding 95% confidence intervals) will be computed for each treatment arm. The nominal p-value from the stratified log-rank test will be reported. The same derivation of PD date and censoring rules as for primary PFS analysis will apply for DOR. The analysis of DOR will be repeated based on BICR assessment and investigator assessment. For DOR per investigator assessment, the same algorithm for backdating equivocal progression will be applied as in Section 5.2.3.1. As exploratory analyses, the same analyses for DOR will also be conducted using ITT-PFS set.

5.2.4 Exploratory Efficacy Endpoints

Analyses of the following exploratory endpoints will not be subject to formal type I error control and will be analyzed using conventional log rank statistical methods (i.e., rerandomization methods will not be used). Analyses for the exploratory endpoints may not be performed at time of primary analysis for PFS, depending on data availability.

5.2.4.1 Incidence of HER2 and Other Mutations

The incidence of HER2 mutations and other tumor-related mutations will be summarized by treatment arm using counts and percentages. Cox proportional hazards regressions will be performed for the primary efficacy endpoint and the two key secondary endpoints. Each model will include treatment arm and type of mutation as independent variables and the time-to-event variable as the dependent variable. Hazard ratios and 95% confidence intervals will be estimated.

5.2.4.2 RANO-BM assessment for BM subgroup

In the BM subgroup (as defined in Section 4.2) from the ITT analysis set, disease status in brain will be assessed by BICR using RANO-BM for brain metastases. The following exploratory endpoints will be evaluated for the BM subgroup.

Objective Response Rate in Brain

Objective response in brain is defined as achieving a best overall response of confirmed complete (CR) or confirmed partial response (PR) in brain. The proportion of subjects with objective response in brain (ORR_{Brain}) will be calculated by treatment arm. ORR_{Brain} will be summarized for subjects who had at least one measurable target lesion in brain at baseline as assessed by BICR.

Duration of response in Brain

Duration of response in brain (DOR_{Brain}) is defined as the time from the first objective response (CR or PR that is subsequently confirmed) in brain to documented disease progression (PD) in brain or death from any cause, whichever occurs first. Only subjects who achieve a confirmed response in brain will be included in the analysis. Kaplan-Meier estimates of the median (corresponding 95% confidence intervals) will be computed for each treatment arm. The same censoring rules as for primary PFS analysis will apply.

Time to Progression in Brain

Time to progression in brain is defined as the time from the date of randomization to the date of documented disease progression in brain. This endpoint will be evaluated using cumulative incidence methodology for competing risks treating non-brain progression and death as competing events. The treatment arm hazard ratio and 95% confidence interval will be estimated using a Cox proportional hazards regression model based on the following SAS sample code.

```
** tte = Time to event;  
** Status (0=censored, 1=brain progression, 2=death or non-brain progression);  
** Trt = treatment arm;  
PROC PHREG DATA = tdata;  
    CLASS armcd(ref='B') param=glm;  
    MODEL tte*status(0) = armcd / TIES=BRESLOW EVENTCODE=1;
```

HAZARDRATIO armcd;
RUN;

To designate brain progression (Status = 1) in the model, EVENTCODE = 1 is specified in the MODEL statement.

5.3 Analysis of Safety

All analyses of safety will be produced for the safety analysis set.

5.3.1 Adverse Events

An adverse event (AE) is any untoward medical occurrence in a subject administered a pharmaceutical or investigational product, in this case, tucatinib/placebo, capecitabine or trastuzumab. Per the study protocol all adverse events (AE) occurring during the study, whether or not attributable to study treatment, will be recorded in the subject's source documents and eCRF. AE severity will be assessed and graded by the Investigator using Version 4.03 of the National Cancer Institute's Common Toxicity Criteria for Adverse Events (CTCAE). In addition, the Investigator will also assess the relationship (related, not related) of each AE to capecitabine, trastuzumab, and tucatinib/placebo. All AEs will be coded by the Sponsor to standard "preferred terms" (PT) and system organ classifications (SOC) using MedDRA Version 22.0 or higher.

Definitions

Treatment-emergent adverse events (TEAE): Treatment-emergent AEs are defined as events that are new or worsened on or after receiving the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab) and up through 30 days after the last dose of study treatment.

Treatment-related adverse events: Adverse events assessed by the Investigator as 'related' to capecitabine, trastuzumab, or tucatinib/placebo.

Summaries

Adverse events will be summarized by PT or by SOC and PT for each treatment arm using counts and percentages. Multiple occurrences of the same adverse event within a subject will be summarized only once at the most severe grade level for the time frame under consideration. For summaries by severity, only the worst grade for an AE will be counted for a particular subject.

Summaries to be produced include:

- Incidence of TEAE by SOC and preferred term
- Incidence of TEAE by decreasing frequency of preferred term
- Incidence of TEAE by toxicity grade, SOC and preferred term
- Incidence of grade 3 or higher of TEAE by decreasing frequency of preferred term

- Incidence of TEAEs which lead to premature discontinuation of study treatment by decreasing frequency of preferred term
- Incidence of TEAEs which lead to dose interruption or dose reduction of study treatment by decreasing frequency of preferred term
- Incidence of treatment related TEAEs by decreasing frequency of preferred term

In addition to summary tables, the following listings will be produced.

- All AEs
- AEs which resulted in Death
- AEs which lead to discontinuation of study treatment
- Grade 3 or higher AEs

5.3.2 Serious Adverse Events

Serious adverse events (SAEs) will be summarized by preferred term and SOC for each treatment arm using counts and percentages. The following summaries of SAEs will be produced.

- Incidence of treatment emergent SAEs (TESAEs) by decreasing frequency of preferred term
- Incidence of TESAEs by decreasing frequency of SOC and preferred term
- Incidence of treatment related TESAEs by decreasing frequency of preferred term

In addition to summary tables, listings of SAEs will be produced.

5.3.3 Adverse Events of Special Interest

The incidence of treatment emergent adverse events of special interest will be summarized by treatment arm and preferred term or lab values and listings also produced. Adverse events of special interest to be summarized are described in the protocol Section 9.4.7.1 and are the following:

AE of Special Interest	Search Strategy
Potential drug-induced liver injury	<ul style="list-style-type: none"> • Laboratory abnormalities: LFTs meeting the following criteria - AST/ALT >3xULN and Bilirubin >2xULN occurring within 21 days
Cerebral edema not clearly attributable to progression of disease	<ul style="list-style-type: none"> • Brain oedema (PT) • Cytotoxic oedema (PT) • Vasogenic cerebral oedema (PT)
Left ventricular systolic dysfunction leading to a change in study treatment or discontinuation of study treatment	<ul style="list-style-type: none"> • Cardiomyopathy SMQ (Narrow) • Cardiac failure SMQ (Narrow) • Left Ventricular Ejection Fraction measurements by ECHO/MUGA

5.3.4 Clinical Laboratory Results

Serum chemistry and hematology samples will be collected per protocol specified schedules.

All laboratory results will be converted into Système International (SI) units for analysis and graded using the laboratory reference ranges and the criteria from NCI CTCAE (Common Terminology Criteria for Adverse Events, Version 4.03) by the Sponsor. For lab parameters ALT, AST, BILI, ALP, APTT and INR, the reference ranges from individual laboratories will be used if available; for other lab parameters, the Sponsor's standard normal range will be used to perform toxicity grading.

Incidence of Laboratory Toxicities

The incidence of laboratory toxicities by grade will be summarized by treatment arm. For each test, only the worst (i.e., highest) toxicity grade will be counted for subjects with multiple toxicities within a time period (including scheduled and unscheduled assessments). The change of toxicity grade from baseline to worst baseline grade (shift table) will also be summarized. In addition to summary tables, listings of laboratory test results will be provided.

Incidence of Liver Abnormalities

The incidence of liver abnormalities will be summarized by treatment arm. A liver abnormality is defined as AST or ALT elevations that are >3 x ULN with concurrent elevation (same day or within 21 days following AST and/or ALT elevations) of total bilirubin >2 x the ULN.

5.3.5 Ejection Fraction

The minimal post baseline ejection fraction and the maximum decrease from baseline will be summarized for each treatment group. Time to minimal post baseline ejection fraction may also be tabulated.

5.3.6 Vital Signs

Vital signs (weight, body temperature, respiratory rate, heart rate, and systolic and diastolic blood pressure) will be listed. The frequency and percentage of patients with post baseline clinically significant vital signs will be summarized. The clinically significant vital signs are defined as: heart rate >100 bpm; Temperature >38.0 degrees C (100.4 F) and respiratory rate >20 breaths per min. Blood pressure will be summarized both for subjects with systolic blood pressure >120 mmHg or diastolic blood pressure >80 mmHg, as well as for subjects with systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg.

For weight, the maximum decrease from baseline will also be summarized.

5.3.7 Deaths

Death information will be listed by subject.

5.4 Pharmacokinetic Analyses

The analyses described in this section will be produced for the pharmacokinetics analysis set.

Tucatinib and metabolite drug levels will be summarized with descriptive statistics at each PK sampling time point. The primary efficacy endpoint and the adverse events of special interest will be summarized by the quartiles of tucatinib trough concentration. Additional PK and PK/PD analyses may be performed and described in a separate pharmacometric analysis plan.

5.5 Health Economics and Outcomes

5.5.1 Health Care Resource Utilization

Cumulative incidence of health resource utilization, including length of stay, hospitalizations, and ED visits will be summarized by treatment group in safety analysis set.

5.5.2 Health-Related Quality of Life

Health-related quality of life/health status using the EQ-5D-5L instrument will be analyzed for ITT-OS set. Figures may be produced for utility score and compliance with PRO completion (the number of completed surveys at each measure divided by number expected) over time by treatment group for ITT-OS analysis set.

5.6 Interim Analyses

No interim analyses for efficacy are planned for the primary endpoint. One formal interim analysis for superiority is planned for the key secondary endpoint of PFS_{BM} and two formal interim analyses for superiority are planned for the key secondary endpoint of OS if the primary analysis for PFS is statistically significant. The second interim analysis for OS may not be conducted as described in Section 4.6. The interim analyses will be conducted at the timing described in Section 4.6. The stopping boundaries will be determined using Lan-DeMets spending functions for the O'Brien and Fleming boundaries. See Section 4.7 for control of multiplicity.

5.7 Changes in the Planned Analysis

There are no changes from the planned analysis outlined in the protocol for this trial.

6 REFERENCES

Rosenberger, William F., and John M. Lachin. "Chapter 7." Randomization in Clinical Trials Theory and Practice. Hoboken, NJ: John Wiley & Sons, 2002.

Fleming, Thomas R., Mark D. Rothmann, and Hong Laura Lu. "Issues in Using Progression-Free Survival When Evaluating Oncology Products." *Journal of Clinical Oncology* 27.17 (2009): 2874-880.

Karrison, Theodore G. "Versatile tests for comparing survival curves based on weighted log-rank statistics." *Stata Journal* 16.3 (2016): 678-690.

Kosorok, Michael R., and Chin-Yu Lin. "The versatility of function-indexed weighted log-rank statistics." *Journal of the American Statistical Association* 94.445 (1999): 320-332.

Schnipper, Lowell. "Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received." *Journal of Clinical Oncology*, 20 Aug. 2016, ascopubs.org/doi/full/10.1200/JCO.2016.68.2518.

Ye, Y., Li, A., Liu, L. and Yao, B. (2013). A group sequential Holm procedure with multiple primary endpoints. *Statistics in Medicine* 32, 1112-1124