Official Title: **A RANDOMIZED, MULTICENTER, OPEN-LABEL, PHASE III TRIAL COMPARING TRASTUZUMAB PLUS PERTUZUMAB PLUS A TAXANE FOLLOWING ANTHRACYCLINES VERSUS TRASTUZUMAB EMTANSINE PLUS PERTUZUMAB FOLLOWING ANTHRACYCLINES AS ADJUVANT THERAPY IN PATIENTS WITH OPERABLE HER2-POSITIVE PRIMARY BREAST CANCER**

NCT Number: NCT01966471

Document Date: SAP Version 3: 17-January-2020

STATISTICAL ANALYSIS PLAN

TITLE: **A RANDOMIZED, MULTICENTER, OPEN-LABEL, PHASE III TRIAL COMPARING TRASTUZUMAB PLUS PERTUZUMAB PLUS A TAXANE FOLLOWING ANTHRACYCLINES VERSUS TRASTUZUMAB EMTANSINE PLUS PERTUZUMAB FOLLOWING ANTHRACYCLINES AS ADJUVANT THERAPY IN PATIENTS WITH OPERABLE HER2-POSITIVE PRIMARY BREAST CANCER**

PROTOCOL NUMBER: BO28407

STUDY DRUG: Trastuzumab Emtansine

VERSION NUMBER: 3

IND NUMBER: 71072

EUDRACT NUMBER: 2012-004902-82

SPONSOR: F. Hoffmann-La Roche Ltd.

PLAN PREPARED BY: [redacted]

DATE FINAL: Version 1: 14 November 2017
Version 2: 18 May 2018
Version 3: See electronic date stamp below

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Date and Time(UTC)	Reason for Signing	Name
17-Jan-2020 09:16:30	Company Signatory	[redacted]

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

This Statistical Analysis Plan (SAP) Version 3, for Protocol BO28407, has been amended to specify further sensitivity analyses for the primary endpoint. Further details regarding multivariate Cox regression analysis for the primary endpoint and regarding the efficacy subgroups analyses were provided. Exploratory analysis of the occurrence of CNS metastases has been added. Patient-reported outcomes analyses were extended to the full set of treatment-related symptoms scales and function scales. References for the corresponding thresholds used to define clinically meaningful differences were updated.

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

1. BACKGROUND

Breast cancer remains a highly significant cause of morbidity and mortality worldwide. The use of adjuvant trastuzumab based therapy in HER2-positive early-stage breast cancer (EBC) improves patient outcomes as demonstrated in several large, randomized trials.

A total of 1 year of HER2-directed adjuvant therapy with trastuzumab remains the standard of care on the basis of recent results from studies comparing different adjuvant trastuzumab treatment durations (1 year vs. shorter [6 months] or longer [2 years]) (Pivot et al. 2013, Goldhirsch et al. 2013).

At time of current protocol version 3 (30 July 2015) the 3-year disease-free survival (DFS) rate is approximately 85%–90% for the overall population of patients with HER2-positive EBC treated with adjuvant chemotherapy and trastuzumab in Herceptin® adjuvant trials. Recurrence rate at 10 years is reported to be approximately 26% (Romond et al. 2012).

In the recent reported APHINITY study, the estimated 3-year invasive disease-free survival (IDFS) rate was 93% for patients treated with trastuzumab based regimen and 94% for patients treated with adjuvant pertuzumab added to chemotherapy plus 1 year of treatment with trastuzumab. The study population of APHINITY included node-positive patients or a broader node-negative HER2-positive, operable breast cancer patient population (von Minckwitz et al. 2017). The majority of these patients eventually have disease recurrence at distant sites, which is generally considered to be incurable. Thus, there is an unmet need and potential to further improve outcomes, particularly in higher-risk patient subpopulations, by incorporating newer HER2-directed agents, with manageable toxicity, into the treatment paradigm.

Study BO28407 is a randomized, multicenter, open-label, Phase III trial comparing trastuzumab plus pertuzumab plus a taxane following anthracyclines versus trastuzumab emtansine plus pertuzumab following anthracyclines as adjuvant therapy in patients with operable HER2-positive primary breast cancer.

The purpose of this Statistical Analysis Plan (SAP) is to provide the Sponsor's (F. Hoffmann-La Roche, Ltd.'s [Roche]) proposed analysis plan for this study. The SAP overrides the analyses described in the statistical section of the Protocol BO28407 version 3. The analyses as described in the SAP will be used by the Sponsor for the purpose of regulatory submission. Any major deviation from the SAP will be noted in the Clinical Study Report (CSR).

2. STUDY DESIGN

Study BO28407 is a prospective, two-arm, Phase III, randomized, multicenter, multinational, open-label study in patients with newly diagnosed and centrally confirmed HER2-positive primary invasive breast cancer who have had curative-intent surgery of

their primary tumor and are candidates for adjuvant systemic chemotherapy following surgery. According to current protocol version 3, the study was planned to enroll approximately 1850 patients. Actually, 1846 patients were finally enrolled from 342 study sites worldwide.

2.1 PROTOCOL SYNOPSIS

The Protocol Synopsis is in Appendix 1. For additional details, see the Schedule of Assessments in Appendix 2.

2.2 OUTCOME MEASURES

The Protocol Synopsis in Appendix 1 outlines the primary, secondary, exploratory, pharmacokinetic (PK), and safety outcome measures.

2.3 DETERMINATION OF SAMPLE SIZE

The sample size of the study is primarily driven by the analysis of IDFS in both the node-positive subpopulation and the overall protocol-defined population. The statistical assumptions for the revised sample size calculations are summarized in Table 1.

The IDFS analysis is powered at 80% for the node-positive subpopulation. Additionally, the IDFS analysis with the assumptions stated below is powered at 82.5% for the overall protocol defined population. It is assumed that the overall protocol-defined population has approximately 10% more patients and 6% more IDFS events than the node-positive subpopulation. These percentages were observed in the BCIRG 006 (Slamon et al. 2006, 2009, 2011) trial after subsetting for the respective populations in Study BO28407.

To detect a target hazard ratio (HR) of 0.64 in IDFS in the overall protocol-defined population and the node-positive subpopulation, approximately 171 and 160 IDFS events will be required to achieve 82.5% and 80% power, respectively, in the two populations, at a two-sided significance level of 5% using a log rank test. One interim and one final IDFS analysis are planned in both the overall protocol-defined population and the node-positive subpopulation (see also Section 4.10.1). The Lan-DeMets α-spending function with an O'Brien-Fleming boundary will be used such that the overall type I error will be controlled at the 5% level for the IDFS endpoint (DeMets et al. 1994). Approximately 1850 and 1665 patients were planned to be enrolled in the overall protocol-defined population and node-positive subpopulation, respectively, including a dropout/ineligibility hazard of 0.003, which corresponds to a cumulative drop out of 177 patients (9.5%) over the expected study duration of 48 months. The assumed 3-year IDFS rate for the control arms for both the populations were based on the IDFS rate from the BCIRG 006 data for the proposed populations and the assumed target HR from Study BO25126 (APHINITY). The six piecewise Kaplan–Meier estimates of the IDFS function in the control group for the overall protocol defined population and the node-positive subpopulation were 99.6% and 99.5% during the first 6 months, 98.2%

and 98.0% during the second 6 months, 93.7% and 93.1% during Year 2, 89.5% and 89.1% during Year 3, 87.7% and 87.1% during Year 4, and 86.2% and 85.5% during Year 5. With the assumed target HR of 0.64, the estimated 3-year IDFS rate for the experimental arms in each of the two populations is shown in Table 1.

Table 1 Summary of Sample Size Assumptions for IDFS

Population	BCIRG 006 3-Year IDFS for the proposed population AC-TH	Control Group (Study BO25126 [APHINITY]) 3-Year IDFS HR = 0.75 AC-THP	Experimental Group (Study BO28407 [KAITLIN]) 3-Year IDFS HR = 0.64 AC-KP	Power AC-THP vs. AC-KP	No. IDFS Events	Sample Size
Overall protocol–defined population	86.3%	89.5%*	93.2%	82.5%	171	1850
Node-positive subpopulation	85.8%	89.1%*	92.9%	80.0%	160	1665

AC-T=doxorubicin + cyclophosphamide followed by docetaxel; H=trastuzumab; HR=hazard ratio; IDFS=invasive disease–free survival; K=trastuzumab emtansine; P=pertuzumab.

* Assumption at time of sample size planning. Study BO28407 completed enrolment before Study BO25126 results were available.

Three interim OS analyses and one final OS analysis are planned in both the overall protocol-defined population and the node-positive subpopulation (see also Section 4.10.2). The assumed 3-year OS rate for the control arms for both populations were based on the OS rate from the BCIRG 006 data for the proposed population and the assumed target HR from Study BO25126. The five piecewise Kaplan-Meier estimates of the OS function in the control group for the overall protocol-defined population and the node-positive subpopulation were 99.7% and 99.7% during Year 1, 98.6% and 98.5% during Year 2, 96.1% and 95.7% during Year 3, 94.1% and 93.9% during Year 4, and 92.1% and 91.9% during Year 5. The estimated 3-year OS rate for the experimental arms in each of the two populations is shown in Table 2 with an assumed target HR of 0.80 between the control arm (AC-THP) and the experimental arm (AC-KP).

Table 2 Summary of Assumptions for Overall Survival Analyses

Population	BCIRG 006 3-Year OS for the proposed populations AC-TH	Control Group (Study BO25126 [APHINITY]) 3-Year OS AC-THP HR = 0.80	Experimental Group (Study BO28407 [KAITLIN]) 3-Year OS AC-KP HR = 0.80
Overall protocol–defined population	95.2%	96.1%	96.9%
Node-positive subpopulation	94.7%	95.7%	96.6%

AC-T = doxorubicin + cyclophosphamide followed by docetaxel; H = trastuzumab; HR = hazard ratio; IDFS = invasive disease–free survival; K = trastuzumab emtansine; OS = overall survival P = pertuzumab.

For the OS interim analyses and final OS analysis, the Lan-DeMets α-spending function with an O'Brien-Fleming boundary will be used such that the overall type I error will be controlled at the 5% level for the OS endpoint (DeMets et al 1994). With the study sample size and approximately 10 years of follow-up from first patient initiation (FPI), the Study BO28407 has 38% power in the overall protocol-defined population and 35% power in the node-positive subpopulation to detect an HR of 0.8. In the overall protocol-defined population this corresponds to a 0.8% improvement in 3-year OS, from 96.1% in the control arm to 96.9% in the experimental arm and in the node-positive subpopulation to a 0.9% improvement in 3-year OS, from 95.7% in the control arm to 96.6% in the experimental arm, at a two-sided significance level of 5%.

The sample size calculations were performed using EAST v6 software (Cytel Inc.).

2.4 ANALYSIS TIMING

One interim analysis of IDFS is planned after approximately 128 IDFS events are observed in the overall protocol-defined population and approximately 120 IDFS events are observed in the node-positive sub-population. This is projected to occur approximately 43 months from FPI (see Table 3).

The final analysis of IDFS will be performed after approximately 171 events are observed in the overall protocol-defined population and approximately 160 events are observed in node-positive sub-population. This is projected to occur approximately 57 months from FPI.

The actual timing of the interim and final IDFS analysis will be based on regular event tracking on the clinical database and on event projections.

All OS interim and final analyses will be performed by the Sponsor subsequent to the primary IDFS analysis and after the Sponsor is unblinded. If the final IDFS analysis crosses the statistical boundaries, then the second interim OS analysis will be performed at the time of the final IDFS analysis, followed by the third OS analysis that will be

performed at 84 months and the final OS analyses that will performed at approximately 120 months [10 years], from FPI (see Table 3).

For details regarding hierarchical testing and alpha levels, stopping boundaries please refer to Sections 4.4.3 and 4.10.

Table 3 Summary of Planned Analyses in Overall Protocol-Defined Population and Node-Positive Subpopulation

Endpoint	Analysis	No. of Events in the Overall Population	No. of Events in the Node-Positive Subpopulation	Estimated Timing[a]
IDFS	Interim	128	120	43 months
	Final	171	160	57 months
OS	Interim 1 (Interim IDFS)	50	50	43 months
	Interim 2 (Final IDFS)	87	83	57 months
	Interim 3	153	138	84 months
	Final	224	203	120 months

IDFS=invasive disease-free survival; OS=overall survival

[a] Time from the enrollment of first patient to data cutoff.

After the first 600 patients were randomized and followed for 3 months (this had been anticipated to occur at approximately 13 months after FPI), the independent data monitoring committee (iDMC) performed an interim safety analysis (see Section 3.2).

3. STUDY CONDUCT

3.1 RANDOMIZATION ISSUES

Upon verification of inclusion and exclusion criteria, eligible patients are randomized in a 1:1 ratio to one of the two treatment arms using a permuted block randomization scheme via an interactive voice/web response system (IxRS).

Randomization is stratified by the following stratification factors:

- World region (United States/Canada, Western Europe/Australia/New Zealand, Asia, or rest of the world)
- Nodal status (0, 1-3, or ≥4 positive nodes)
- Centrally assessed hormonal receptor status (ER- and/or PR-positive or both ER- and PR-negative)
- Type of planned anthracycline (doxorubicin or epirubicin)

3.2 DATA MONITORING

3.2.1 Independent Data Monitoring Committee

The iDMC monitors accruing patient safety data at least once every 6 months during the study until the last patient has completed study treatment. In addition, safety data related to concurrent radiotherapy and/or hormonal therapy, SAEs, and deaths is monitored by the iDMC at least once every 3 months during the study until the last patient has completed study treatment. At each iDMC review, relevant safety information from ongoing trastuzumab emtansine and/or pertuzumab studies is provided to the iDMC. The iDMC will also assess safety and efficacy as part of the planned interim efficacy and safety analyses.

The iDMC works according to guidelines defined in the iDMC Charter. The iDMC Charter contains details regarding frequency of meetings, guidelines for decision making, and processes for requesting further information. The iDMC members reviewed and signed off the charter before the first iDMC review.

An independent Data Coordination Center (iDCC) performs unblinded analyses to support the periodic iDMC review of safety data and the interim analysis. Additional details are specified in the iDMC Charter.

3.2.2 Clinical Events Committee Data Monitoring Committee

An independent CEC adjudicates pre-specified safety events of interest (cardiac and hepatic dysfunction events) in a blinded fashion. A separate charter outlines committee composition, meeting timelines, and the roles and responsibilities of members. The committee members review all potential cases of congestive heart failure (CHF) and cardiac death as well as all potential cases of hepatic dysfunction and Hy's law. Ad hoc members may be added to the CEC to review/adjudicate other safety events if a new safety signal emerges. Adjudicated cases by the CEC are forwarded to the iDMC on a regular basis as part of ongoing safety reviews of the Study BO28407.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

4.1.1 Intention-to-Treat Population

The intention-to-treat (ITT) population will be the overall protocol-defined population as mentioned in the protocol and will include all patients who were randomized to the study, whether or not they received any study medication. Patients will be analyzed according to the ITT principle, where all randomized patients will be included in the treatment arm they were originally allocated to by the IxRS at randomization.

4.1.2 Node-Positive Subpopulation

The node-positive subpopulation (ITTLN+) will be a part of the ITT and will include patients with ≥ 1 positive lymph node at baseline as reported in the IxRS.

4.1.3 Pharmacokinetic-Evaluable Population

Pharmacokinetic (PK) evaluable patients are defined as patients who have received at least one treatment dose of trastuzumab emtansine and have at least one post-treatment serum sample.

4.1.4 Anti-Drug Antibody Evaluable Population

Anti-drug antibody (ADA) evaluable patients are defined as patients who have received at least one treatment dose of trastuzumab emtansine and have at least one ADA sample.

4.1.5 Safety Population

The safety evaluable population includes all randomized patients who received at least one full or partial dose of any study medication (chemotherapy or HER2 treatment). Patients will be analyzed based on the treatment they actually received (which may be different from the treatment the patient was randomized to). Patients receiving any dose of trastuzumab emtansine will be included in the trastuzumab emtansine arm; all other treated patients will be included in the control arm.

4.2 ANALYSIS OF STUDY CONDUCT

Patient enrollment, duration of follow-up, discontinuation from study treatment and study, and reasons for discontinuation will be summarized by treatment arm to which patients were randomized. In addition, protocol violations will be summarized by treatment arm.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

The evaluation of treatment arm comparability between the two treatment arms will include summaries of demographics, baseline disease characteristics, and patient medical history. Data will be summarized by treatment arm to which patients were randomized. Descriptive statistics (mean, median, standard deviation, and minimum-maximum) will be presented by treatment arm for continuous variables such as age and time since diagnosis. Frequency counts and proportions will be presented by treatment arm for categorical variables such as gender, race, and age category.

The baseline value of any variable will be defined as the last available data point prior to the first administration of study medication.

Numbers and percentage of patients who received adjuvant radiation therapy and adjuvant hormonal therapy will be summarized by treatment arm. Anti-cancer therapies (including radiotherapy, systemic cancer therapy, surgery) administered during survival follow-up will be summarized by treatment arm.

4.4 EFFICACY ANALYSIS

4.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is IDFS. The co-primary efficacy objectives for this study are to compare IDFS (1) in the ITTLN+ and (2) in the ITT populations.

IDFS is defined as the time between randomization and date of first occurrence of any one of the following events:

- Ipsilateral invasive breast tumor recurrence (i.e., an invasive breast cancer involving the same breast parenchyma as the original primary lesion)
- Ipsilateral local-regional invasive breast cancer recurrence (i.e., an invasive breast cancer in the axilla, regional lymph nodes, chest wall and/or skin of the ipsilateral breast)
- Contralateral or ipsilateral second primary invasive breast cancer
- Distant recurrence (i.e., evidence of breast cancer in any anatomic site [other than the three sites mentioned above]) that has either been histologically confirmed or clinically/radiographically diagnosed as recurrent invasive breast cancer
- Death attributable to any cause, including but not limited to breast cancer, non-breast cancer, or unknown cause

All second primary non-breast cancers and in situ carcinomas (including DCIS and LCIS) and non-melanoma skin cancer are excluded as an event in this endpoint.

If a patient experiences an event on the same day as randomization their time from randomization to event will be set to be 1 day. Data from patients who have not had an event at the time of clinical cut-off will be censored at the date they were last known to be alive and event free on or prior to the clinical data cutoff date.

The analysis of primary efficacy endpoint will be (1) using the ITTLN+, and (2) using the ITT populations. A testing hierarchy will be applied to control the overall type I error rate at 5% (see Section 4.4.3).

The log-rank test, stratified by the protocol-defined stratification factors nodal status (0, 1-3, 4+ nodes), centrally assessed hormonal receptor status (ER and/or PR positive, ER and PR negative) and type of planned anthracycline (doxorubicin, epirubicin) (excluding region), will be used to compare IDFS between the two treatment arms. Region will be excluded because of the potential that some of the strata may have very few patients, which would result in a loss of power. Stratified analysis will use the stratification information provided in the IxRS system (nodal status and planned anthracycline treatment) or central testing laboratory (hormone receptor status). The unstratified log-rank test results will also be provided for sensitivity analysis. If, at the time of analysis, it is deemed that the smallest stratum per arm necessary to conduct robust stratified analyses contains <5 events, unstratified analyses will be used as the primary analysis. The Cox proportional hazards model, stratified by the previously noted

stratification factors, excluding region, will be used to estimate the HR between the two treatment arms and the corresponding 95% CI. The Kaplan-Meier approach will be used to estimate 3-year IDFS rates and corresponding 95% CIs for each treatment arm. Further landmark analyses at other timepoints may be conducted as appropriate.

To assess the impact of prognostic factors, multivariate Cox regression analyses controlling for important baseline characteristics including treatment group, region (United States/Canada, Western Europe/Australia/New Zealand, Asia, rest of the world), nodal status (0, 1-3, 4+ nodes), hormonal receptor status (ER and/or PR positive, ER and PR negative), planned anthracycline (doxorubicin, epirubicin), menopausal status at baseline (pre-menopausal, post-menopausal), primary surgery type (breast conserving, non-conserving), adjuvant cancer radiotherapy (yes, no), age (<40, $40-49$, $50-64$, <65, ≥ 65), histological grade (grade 1, grade 2, grade 3), tumor size at time of diagnosis ($0 \leq 2$ cm, $\geq 2-5$ cm, ≥ 5 cm) and race (White, Black, Asian, all other) will be done for the primary efficacy endpoint IDFS to calculate HRs between treatment groups and corresponding 95% confidence intervals. Baseline characteristics with observed imbalances between treatment arms may also be included into the multivariate Cox regression model. Further multivariate Cox regression models including biomarkers of interest (e.g. PIK3CA mutation status, HER2 mRNA expression) may also be fitted, as deemed appropriate.

4.4.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints are defined as follows:

- IDFS including second primary non-breast cancer: defined the same way as IDFS for the primary endpoint but including second primary non-breast invasive cancer as an event (with exception of non-melanoma skin cancers and carcinoma in situ [CIS] of any site). This definition is to acknowledge 1) the difficulty of distinguishing second primaries from breast cancer metastasis and 2) that second cancers may be treatment related.
- DFS: defined as the time between randomization and the date of the first occurrence of an IDFS event including second primary non-breast cancer event or contralateral or ipsilateral ductal carcinoma in situ (DCIS). This definition is historically the primary endpoint for breast cancer adjuvant therapy trials.
- OS: defined as the time from randomization to death due to any cause. OS is part of the testing hierarchy that will be applied to control the overall type I error rate at 5% (see Section 4.4.3).
- DRFI: defined as the time between randomization and the date of first occurrence of distant breast cancer recurrence. Based on previous published trials, distant recurrence is anticipated to account for the majority of first disease recurrence events; hence DRFI is included.

Data from patients who have not had an event will be censored at the date that they are last known to be alive and event free on or prior to the clinical data cut-off date for the

respective analysis. In case of multiple recurrent events in a patient, only the first event will be analyzed.

Secondary endpoints will be analyzed in a similar manner as the primary endpoint to estimate 3-year event rates (and 5-year survival rate for OS) for each treatment arm and the HR between the two treatment arms with 95% CI. Further landmark analyses at other timepoints may be conducted as appropriate.

The log-rank test, stratified by the protocol-defined stratification factors nodal status (0, 1 – 3 nodes, 4+ nodes), centrally assessed hormonal receptor status (ER and/or PR positive, ER and PR negative), anthracycline type (epirubicin containing regimens, doxorubicin containing regimen) (excluding region) will be used to compare OS between the two treatment arms. The unstratified log-rank test results will also be provided for sensitivity analysis.

The analysis of secondary efficacy endpoints will be (1) using the ITTLN+ and (2) using the ITT populations.

Details of IDFS and OS interim analyses are specified in the interim analyses section. (Section 4.10)

4.4.3 Hierarchical Testing Procedure

To control the overall study Type I error rate at 5%, the hierarchical testing procedure will be used on the primary and secondary endpoints IDFS and OS in the order given below:

1. Primary endpoint of IDFS in the ITTLN+
2. Primary endpoint of IDFS in the ITT population
3. Secondary endpoint of OS in the ITTLN+
4. Secondary endpoint of OS in the ITT population

Details of IDFS and OS interim analyses are specified in the interim analyses section (Section 4.10).

4.4.4 Exploratory Efficacy Endpoints

The relationship between molecular markers (e.g., level of HER family receptors and status of PIK3CA mutations see Section 3.4.4 in the Protocol BO28407) and efficacy outcomes will be evaluated as exploratory analyses. Efficacy outcomes considered for this analysis will include IDFS and OS, as appropriate.

Analyses will also be performed to explore the correlation between anti-drug antibodies (ADAs) to trastuzumab emtansine, and clinical outcomes as appropriate.

An assessment of Breast Cancer-Free Interval (BCFI) will also be made. BCFI is defined as the time between randomization and the date of local, regional, or distant breast cancer recurrence, invasive contralateral breast cancer or DCIS (contralateral or ipsilateral). Patients who have not had a recurrence event at the time of data analysis will be censored at the date when they were last known to be alive or at their date of death.

Analyses will be performed to compare the occurrence of CNS metastases as first IDFS event and as only IDFS event between the treatment arms.

4.4.5 Sensitivity Analyses

Alternative definitions of IDFS, including second primary non-breast cancer; and DFS (including second primary non-breast cancer event or contralateral or ipsilateral ductal CIS) are analyzed. Both of these two alternative definitions are secondary endpoints of the study (Section 4.4.2).

To assess the impact of stratification IDFS rates will be compared using an unstratified log-rank test (see also Section 4.4.1).

A Kaplan–Meier plot of the time to censoring for IDFS in the different treatment arms will be performed to investigate differences in follow-up time. In this analysis, patients who had an IDFS event will be censored at the date of their event and patients without an event will be regarded as having had an event at the censoring date.

If the primary endpoint analysis is positive further sensitivity analyses will be performed to assess the potential for bias in the primary endpoint IDFS due to potential imbalance in assessments and follow-up time, potential bias introduced by non-protocol anti-cancer therapy and to assess the robustness of the primary endpoint by understanding the impact that imbalance on assessments may have on the IDFS result:

1. To assess the impact of non-protocol adjuvant anti-cancer therapy prior to an IDFS event, data for patients will be censored for IDFS at the time of the last disease status assessment before the initiation of non-protocol anti-cancer therapy if they begin a non-protocol anti-cancer adjuvant therapy before reporting invasive disease recurrence. Trastuzumab±pertuzumab after discontinuation of trastuzumab emtansine before 18 cycles without experiencing an IDFS event is considered non-protocol anti-cancer therapy in this analysis.
2. The above sensitivity analysis will be repeated whereby the switch to trastuzumab±pertuzumab after discontinuation of trastuzumab emtansine before 18 cycles without experiencing an IDFS event is not considered non-protocol anti-cancer therapy in this analysis. This is to account for the fact that switching from to trastuzumab after discontinuation of trastuzumab emtansine is allowed per protocol.

3. To further assess the impact of non-protocol adjuvant anti-cancer therapy prior to an IDFS event, patients will be considered to have an IDFS event at the time of the initiation of non-protocol anti-cancer therapy if they begin a non-protocol anti-cancer therapy before reporting invasive disease recurrence. As for the previous sensitivity analysis the switch to trastuzumab±pertuzumab after discontinuation of trastuzumab emtansine before 18 cycles without experiencing an IDFS event is not considered non-protocol anti-cancer therapy in this analysis.
4. The number of assessments that could detect recurrence will be summarized by treatment arm and will include unscheduled assessments to check for any suggestion of a qualitative difference.

   To assess the impact of completeness of protocol-specified disease assessments on IDFS, data for patients who withdrew from the protocol-specified assessments (i.e., discontinued study follow-up) will be censored for IDFS at their latest protocol-specified assessment where they were invasive disease recurrence-free (so that any deaths reported after this point through the yearly survival contact are not included as events).
5. To assess the impact of differential follow-up withdrawal due to all causes the primary analysis will be repeated, but considering patients with incomplete follow-up according to the protocol-defined schedule (who have not had an IDFS event) as having an IDFS event one day after they were last known to be IDFS event free. Therefore, if a patient does not have follow-up for disease recurrence within 3 months prior to clinical cut-off if they are in their first two years of follow-up, 6 months in years 3–5 and 12 months in years 6–10, will be deemed to have incomplete follow-up.
6. To assess the impact of differences between the stratification factors reported in the IxRS system against the randomization stratification factors reported in the eCRF, the primary analysis will be repeated, stratified according to the stratification factors reported in eCRF.

Sensitivity analyses will be carried out on the ITT and ITTLN+ populations for the primary IDFS endpoint.

4.4.6 Subgroup Analyses

Subgroup analyses will be performed for IDFS and OS as appropriate, in the ITT and the ITTLN+ to assess the robustness and consistency of treatment effect of trastuzumab emtansine plus pertuzumab.

The following subgroups analyses will be conducted via forest plots including 3 (5) year IDFS (OS) rates and estimates for HR and 95% CI from unstratified Cox proportional hazard models:

- World region (IXRS/eCRF: United States/Canada vs. Western Europe/Australia/New Zealand vs. Asia vs. Rest of the world)
- Nodal status at baseline (IXRS/eCRF: lymph node positive, 4+positive nodes, 1–3 positive nodes, 0 positive nodes)

- Centrally assessed hormonal receptor status at baseline (IXRS: ER and/or PR positive vs. ER- and PR-negative, Central testing laboratory: ER and/or PR positive vs. ER- and PR-negative, ER/PR unknown)
- Type of anthracycline planned at randomization (IXRS/eCRF: doxorubicin containing regimen vs. epirubicin containing regimen)
- Central ER/PgR status (Central testing laboratory: ER positive PgR positive, ER positive PgR negative, ER negative PgR positive, ER negative PgR negative, ER unknown/PgR positive, ER unknown/PgR negative, ER negative/PgR unknown, ER positive /PgR unknown, ER unknown/PgR unknown)
- Menopausal status at randomization (pre-menopausal vs. post-menopausal)
- Gender (female vs. male). The subgroup of male patients will only be analyzed if $>$5% of the ITT
- Race (White vs. Black vs. Asian vs. All other)
- Age at randomization (<40, $40-49$, $50-64$, <65, ≥ 65 years)
- Histological grade (grade 1 vs. grade 2 vs. grade 3)
- Type of surgery for primary tumor (breast conserving surgery vs. non-conserving breast surgery)
- Tumor size at time of diagnosis ($0-<2$ cm vs. $\geq 2-5$ cm vs. ≥ 5 cm)
- Adjuvant cancer radiotherapy (Yes vs. No)
- ECOG performance status at baseline (0 vs. 1)

Treatment effect (as determined by HRs and corresponding 95% confidence intervals) and 3 (5) year IDFS (OS) rates will be estimated separately for the defined subgroups to assess the robustness and consistency of treatment effect. Further landmark analyses in defined subgroup may be conducted as appropriate.

Subgroups analyses related to stratification factors will be conducted based on stratification factors as reported in the IXRS system and repeated based on stratification factors as reported in the eCRF/Central testing laboratory.

Exploratory tests of interaction between treatment effect and subgroup (at a 10% significance level) will be reported using Cox proportional hazards models.

Additionally, the following biomarker subgroups will be explored:

- HER2 mRNA expression level by qRT-PCR ($\geq$median vs. $<$median)
- HER3 mRNA expression level by qRT-PCR ($\geq$median vs. $<$median)
- PIK3CA mutation status (mutated vs. non-mutated)
- HER2 IHC/ISH subgroups
- Heterogeneity categories based on HER2 expression level (focal vs. heterogeneous vs homogeneous)
- HER2 gene copy number (<4 vs. 4-6 vs. >6)

- HER2 gene ratio (<2 vs. 2-4 vs. >4)
- PTEN Cytoplasm H-Score ($\geq$median vs. $<$median)
- PTEN Cytoplasm Staining status (None vs. decreased vs. slightly decreased, equivalent vs. increased) and (None/decreased/slightly decreased vs. equivalent/increased)

Further biomarker subgroups, notably based on immune markers, may be explored depending on the outcome of the primary analysis, as deemed appropriate.

4.5 PATIENT REPORTED OUTCOME ANALYSES

The patient-reported outcomes of health-related quality of life (HRQoL), function, and disease/treatment-related symptoms will be assessed using the EORTC QLQ-C30 questionnaire and a modified version of its breast cancer module (QLQ-BR23). Only patients with a baseline assessment and at least one post-treatment assessment will be included in the analyses.

Summary of completion rates for the EORTC QLQ-C30 and modified BR23 will be provided by treatment arm by each assessment time point per protocol specified assessment schedule (see schedule of assessments, Appendix 1). Reasons for non-completion will be summarized by treatment arm and assessment time point.

For the subscales of the QLQ-C30 and the BR23 questionnaires, if more than 50% of the constituent items are completed, a pro-rated scale score will be computed as is consistent with the EORTC scoring manual (Fayers et al. 1999). For subscales with less than 50% of the items completed, the subscale will be considered missing.

Summary statistics (mean, SD, median, 25th and 75th percentiles, range and 95% CI) of linear transformed scores will be reported for all scales of the EORTC QLQ-C30 questionnaire and of the BR23. The mean change of the linear transformed scores from baseline (and 95% CIs) will be calculated for each time point. Line charts depicting the mean changes from the baseline assessment (and 95% CI) will be provided for each of the scale at each timepoint.

The number and proportion of patients reporting clinically meaningful differences in treatment-related symptoms scales, global health status/HRQoL, and function scales based on the threshold reported by Cocks et al. (2012) and Osoba (1998) will be provided at each time-point. In addition, the proportion of patients reporting “a little” or “quite a bit” for additional neuropathy, joint/muscle pain and skin single items will be reported by treatment arm.

Scale scores of the EORTC QLQ C30 and BR23 scores will be compared between the two treatment arms at the ‘C1D1 of HER2 targeted treatment period’ timepoint to descriptively assess treatment group comparability between the two treatment arms.

The ‘C1D1 of HER2 targeted treatment’ timepoint will be utilized as the reference for time-to-event Kaplan-Meier analysis.

Kaplan-Meier analysis will be used to assess the time from first HER2 targeted treatment to clinically meaningful deterioration in the global health status/HRQoL subscale (question 29 and 30 of the QLQ-C30) in the treatment arms. An event for a given patient is a decrease from baseline in subscale score of QLQ-C30 by ≥10 points (Osoba et al. 1998). Time-to-event analyses to investigate the time to clinically meaningful deterioration in function scales of QLQ-C30 will also be assessed using the published thresholds by Cocks et al. (2012). Stratified log-rank tests will be used to test the differences between treatment arms. Data of patients with no event will be censored at the time of the last completed questionnaire.

Additionally, in order to elucidate if the taxane-sparing arm reduces patient treatment burden over time, a repeated measures mixed-effects model (MMRM) may be used to compare change from baseline in the global health status/HRQoL and functional scales of the EORTC QLQ-C30. In each mixed model, change from baseline will be the response variable; treatment, visit, and treatment by visit interaction terms will be the fixed factors; and the patient will be denoted as a repeated factor. The ‘C1D1 of HER2 targeted treatment’ timepoint will be utilized as baseline for these analyses. If substantial interaction effect is present, pair-wise comparison will be conducted. Only patients with a baseline assessment and at least one post-treatment assessment will be included in the analyses. Missing values will be handled according to the scoring manual.

Health economic data, as assessed by the EQ-5D-5L, will be evaluated for patients with a baseline assessment and at least one post-baseline EQ-5D-5L assessment. The results from the health economic data analysis will be reported separately from the clinical study report.

4.6 PHARMACOKINETIC ANALYSES

Serum samples for measurement of trastuzumab emtansine and total trastuzumab will be obtained at pre- and post-trastuzumab emtansine infusion in Cycle 1, pre-trastuzumab emtansine infusion in Cycle 4 as well as at study treatment termination from patients randomized to the trastuzumab emtansine containing arm (Arm 2).

Individual and mean serum trastuzumab emtansine and total trastuzumab will be tabulated and summarized (e.g., mean, standard deviation, coefficient of variation, median, minimum, and maximum).

Analyses will be performed in the PK evaluable population (see Section 4.1.3).

4.7 ANTI-DRUG ANTIBODY ANALYSES

Serum samples for the measurement of anti-drug antibodies ADA (also previously referred to as anti-therapeutic antibodies or ATA) to trastuzumab emtansine will be obtained at Cycle 1 and Cycle 4 before infusion of trastuzumab emtansine, at study treatment termination and 3 months after last dose of trastuzumab emtansine. Incidence of ADAs (ATAs) will be assessed at baseline and post-baseline and characterized to determine if the positive response is primarily towards trastuzumab or trastuzumab emtansine. The impact of ADA on PK, safety, and efficacy will be explored in the subgroup of patients with positive ADA results (as data allow).

4.8 SAFETY ANALYSES

Safety analyses will be performed on the safety population (see Section 4.1.5).

4.8.1 Exposure to Study Medication

The number of patients who experience any dose modification, including dose delay, dose interruption and dose discontinuation will be summarized for each treatment arm and regimen component. Reasons for treatment discontinuation will also be summarized.

Descriptive statistics will be presented for treatment duration, total cumulative dose, number of cycles received, dose intensity, and weeks of exposure by treatment arm and regimen component.

4.8.2 Adverse Events

Verbatim descriptions of adverse events (AEs) will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0. All AEs, SAEs, NCI CTCAE Grade 3 or above AEs, AEs leading to death, and AEs leading to study treatment discontinuation, modification or delay, occurring on or after the first dose of study treatment (i.e., treatment-emergent AE), AEs related to study treatment and AEs related to radiotherapy, for patients who received radiotherapy, will be summarized by NCI CTCAE grade. For repeated events of varying severity in an individual patient, the highest grade will be used in the summaries.

Selected AEs for trastuzumab emtansine and pertuzumab will be summarized by NCI CTCAE grade for each treatment arm based on project wide pre-specified category definitions. In addition, any AEs occurring within 24 hours of the first dose of each treatment cycle will be summarized to help characterize and distinguish between potential infusion-related reactions, hypersensitivity, and anaphylaxis.

AEs, NCI CTCAE Grade 3 or above AEs, and AEs leading to study treatment discontinuation will be summarized by study treatment phases.

4.8.3 Deaths

The number and percentage of patients who died and reasons for death will be summarized by treatment arm.

4.8.4 Laboratory Data

Clinical laboratory tests will be performed at local laboratories. Laboratory abnormalities will be defined based on local laboratory normal ranges and NCI CTCAE Version 4.0.

Summary tables for worst NCI CTCAE toxicity grade and shifts from baseline to the worst in NCI CTCAE toxicity grade during treatment will be presented. The baseline measurement is defined as the latest assessment prior to receiving any study medication.

4.8.5 Cardiac Safety

Primary Cardiac Endpoint:

The number and percentage of patients who died from cardiac cause or who experienced severe CHF (New York Heart Association [NYHA] Class III or IV) with a decrease in left ventricular ejection fraction (LVEF) of 10 percentage points or more from baseline to an LVEF of $< 50\%$, will be summarized by treatment arm.

Secondary Cardiac Endpoint:

The number and percentage of patients suffering from other cardiac-related events (e.g., any symptomatic congestive heart failure CHF [NYHA II] associated with a $\geq 10\%$ drop in LVEF to $< 50\%$; asymptomatic declines in LVEF requiring dose delay or discontinuation) will also be summarized by treatment arm.

Comparison of incidence of each of the primary and secondary cardiac endpoints by treatment arm will be made by summarizing the proportions in each arm and calculating the difference in proportions with 95% CI using Hauck-Anderson correction.

Change in LVEF over time will be summarized by treatment arm. Baseline LVEF will be the latest assessment prior to receiving any study medication.

4.8.6 Pulmonary Safety Toxicity

The number of patients who died from pulmonary causes, and number of patients suffering from pneumonitis or interstitial lung disease (ILD) will be summarized by treatment arm.

4.8.7 Hepatotoxicity

Hepatotoxicity (i.e., deaths from hepatic causes, patients with severe drug-induced liver injury [Hy's law cases] and patients with nodular regenerative hyperplasia [NRH]) will be summarized by treatment arm.

Analyses of liver function laboratory test (LFT) results by treatment arm will include the following:

- Shift in NCI CTCAE grade from baseline to worst post-baseline level in ALT, AST, total bilirubin (TBILI), and alkaline phosphatase (ALK)
- Summary of number/percentage of patients with AST, ALT, TBILI, and ALK elevation by NCI CTCAE grade and by treatment cycle
- Scatterplots of worst LFT value (AST vs. ALT, ALT vs. TBILI, AST vs. TBILI) relative to upper limit normal (ULN)
- Kaplan-Meier plots of time to first ALT $>3\times$ ULN, first ALT $>5\times$ ULN, and first ALT>8×ULN events;
- Patient LFT profiles (including ALT, AST, TBILI, and ALK) over time points for patients satisfying one of the following two conditions:
  - TBILI elevation $>2\times$ ULN within 21 days after AST/ALT elevation $>3\times$ ULN
  - ALT elevation $>8\times$ ULN

4.8.8 Safety Subgroup Analyses

A summary of safety (number of patients with SAEs, NCI CTCAE Grade 3 or above AEs, selected AEs, number of deaths) and a summary of AEs by MedDRA system organ class preferred term and NCI CTCAE grade will be provided by age at randomisation (<65 vs. $\geq 65-74$ vs. ≥ 75 years and ≥65 years), by race (White, Asian, Black, Other), for patients with ADA positive results and patients who received adjuvant radiotherapy, adjuvant hormonal therapy, and adjuvant radiotherapy or hormonal therapy.

A summary of safety will be provided for the subgroup of patients who received at least one full or partial dose of trastuzumab, trastuzumab emtansine or pertuzumab study medication during the HER2-targeted treatment period. Only HER2-targeted treatment emergent AEs will be considered.

4.9 MISSING DATA

For the analyses of IDFS, IDFS including second primary non-breast cancer, disease free survival (DFS), DRFI, and BCFI, data for patients who do not experience an event will be censored as described in Sections 4.4.1 and 4.4.2.

For patients who died with only a partial death date available, the missing date parts will be imputed with the earliest possible date part such that the date is after the date of randomization.

For the QLQ-C30 and QLQ-BR23 questionnaires, if less than 50% of the constituent items of a subscale are completed, data for the subscale will be considered to be missing. No imputation will be performed.

For PK parameters, analyses will be restricted to patients with available data. No imputation will be performed.

4.10 INTERIM ANALYSES

One interim analysis of IDFS and three interim analyses of OS in both the ITTLN+ and the ITT are planned.

4.10.1 Interim IDFS Analyses

One interim analysis of IDFS will be performed after approximately 75% of the targeted IDFS events are observed in the ITT (i.e., 128 of the 171 target events) and ITTLN+ (i.e., 120 of 160 target events) and is projected to occur approximately 43 months from FPI. The final analysis of the primary endpoint IDFS will be performed after approximately 171 events are observed in the ITT and approximately 160 events are observed in ITTLN+ and is projected to occur approximately 57 months from FPI (see Table 4). The actual timing of the interim IDFS analysis will be based on regular event tracking on the clinical database and on event projections.

A hierarchical testing procedure will be used for the primary endpoint IDFS for the ITTLN+ and the ITT populations as defined in Section 4.4.3. The type I error will be controlled at the 5% level at the interim and final analyses within the ITTLN+ and the ITT populations using a Lan-DeMets α-spending function with an O'Brien-Fleming boundary.

Table 4 Summary of Planned Analyses of IDFS in ITT and ITTLN+

	ITT		ITTLN+		
Analysis of IDFS	**No. of Events**	**Efficacy Stopping Boundary[a,b]**	**No. of Events**	**Efficacy Stopping Boundary[a,b]**	**Estimated Timing[c]**
Interim	128	$p < 0.0193$ or observed HR < 0.6618	120	$p < 0.0193$ or observed HR < 0.6528	43 months
Final	171	$p < 0.0442$ or observed HR < 0.7353	160	$p < 0.0442$ or observed HR < 0.7279	57 months

HR=hazard ratio; IDFS=invasive disease–free survival; ITT=intention-to-treat; ITTLN+=node-positive subpopulation.

[a] p-value will be based on two-sided stratified log-rank test.

[b] Efficacy stopping boundaries follow O'Brien-Fleming design.

[c] Time from the enrollment of first patient to data cutoff.

The IDFS interim analysis will be performed by the independent Data Coordination Center (iDCC) statistician, and the results will be presented to the iDMC by the iDCC statistician. The Sponsor will not be unblinded to the study results.

If the interim analyses in both populations cross the interim efficacy boundaries of the O'Brien-Fleming design, on the basis of the totality of both efficacy and safety data, the

iDMC may recommend releasing the primary endpoint results before the targeted number of 160 and 171 events have occurred in the ITTLN+ and ITT populations, respectively. In this case, the Sponsor will be unblinded to study results, and a full data package including the first OS interim analysis results will be prepared for discussion with the regulatory authorities. The Study BO28407 is planned to continue until 10 years of follow-up from FPI have occurred (unless the sponsor terminates the study earlier), and the IDFS analysis will be updated descriptively. If the IDFS interim analyses in the ITTLN+ and in the ITT fail to cross the statistical efficacy boundaries of the O'Brien-Fleming design, then the study will continue as planned. The Sponsor will conduct the final analyses.

4.10.2 Interim OS Analyses

Three interim OS analyses and one final OS analysis are planned in both the ITT and the ITTLN+.

All OS interim and final analyses will be performed by the Sponsor subsequent to the primary IDFS analysis and after the Sponsor is unblinded. If the interim IDFS analyses in both the ITTLN+ and the ITT populations cross the statistical efficacy boundaries, the first OS interim analysis will be performed in both populations hierarchically at that time (approximately at 43 months from FPI). If the final IDFS analyses cross the statistical boundaries, then the second interim OS analysis will be performed hierarchically at the time of the final IDFS analysis (approximately 57 months from FPI), followed by the third and the final OS analyses performed hierarchically in both populations (planned to occur at 84 months [7 years] and 120 months [10 years], respectively, from FPI) (see Table 5). If, at any OS interim analysis, the O'Brien-Fleming efficacy boundary is crossed, that analysis of OS will be considered as confirmatory and all subsequent analyses of OS will be considered as descriptive. If the final IDFS analyses are not statistically significant, no formal second and third interim and no formal final OS analyses will be done. Secondary efficacy endpoints other than OS will be analyzed at the time of the final analysis of the primary endpoint IDFS.

Table 5 Summary of Planned Analyses of Overall Survival in ITT and ITTLN+

	ITT		ITTLN+		
Analysis of OS	**No. of Events**	**Efficacy Stopping Boundary [a,b]**	**No. of Events**	**Efficacy Stopping Boundary[a,b]**	**Estimated Timing[c]**
Interim 1 (Interim IDFS)	50	$p < 0.0000033$ or observed HR < 0.2647	50	$p < 0.000013$ or observed HR < 0.2908	43 months
Interim 2 (Final IDFS)	87	$p < 0.000643$ or observed HR < 0.4811	83	$p < 0.000908$ or observed HR < 0.4827	57 months
Interim 3	153	$p < 0.01316$ or observed HR < 0.6697	138	$p < 0.01281$ or observed HR < 0.6545	84 months
Final	224	$p < 0.04585$ or observed HR < 0.7658	203	$p < 0.04591$ or observed HR < 0.7556	120 months

HR=hazard ratio; ITT=intention-to-treat; ITTLN+=node-positive subpopulation; IDFS=invasive disease–free survival; OS=overall survival.

[a] p-value will be based on two-sided stratified log-rank test.

[b] Efficacy stopping boundaries follow O'Brien-Fleming design.

[c] Time from the enrollment of first patient to data cutoff.

4.10.3 Interim Safety Analysis

An iDMC monitors accruing patient safety data at least once every 6 months during the study until the last patient has completed study treatment. In addition, safety data related to concurrent radiotherapy and/or hormonal therapy, serious adverse events, and deaths are monitored by the iDMC at least once every 3 months during the study until the last patient has completed study treatment. At each iDMC review, relevant safety information from ongoing trastuzumab emtansine and/or pertuzumab studies is provided to the iDMC.

After the first 600 patients were randomized and followed for 3 months, the iDMC performed an interim safety analysis regarding overall numbers of deaths (all causes, including cardiac deaths) and hepatic events defined as confirmed Hy's law cases. The CEC communicated their findings to the iDMC to aid iDMC review. If an absolute increase of >3% in the percentage of deaths (from any cause) was observed in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, the iDMC was to consider a recommendation of pausing enrollment for further data review, stopping the trial, or modifying the trial.

If the true difference in the percentage of deaths was >3% (e.g., 2% vs. 6%), then there was approximately a 70% chance of observing an absolute difference of >3% at the interim with 600 patients. The Protocol BO28407 presents the probability of observing

an increase of > 3% in the percentage of deaths in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, with different assumptions for the percentage of deaths in the two treatment arms.

If an absolute increase of > 3% in the percentage of Hy's law cases (confirmed by the independent CEC) was observed in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, the iDMC was to consider a recommendation of pausing enrollment for further data review, stopping the trial, or modifying the trial.

If the true difference in the percentage of confirmed Hy's law cases was > 3% (e.g., 0.33% vs. 3.67%), then there was approximately a 54% chance of observing an absolute difference of > 3% at the interim with 600 patients. Protocol BO28407 presents the probability of observing a > 3% increase in the percentage of Hy's law cases in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, with different assumptions for the number of Hy's law cases in the two treatment arms.

5. REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–376.

Cocks K. King MT, Velikova G, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer. 2012 ;48(11):1713-21.

DeMets D, Lan KKG. Interim analysis: the alpha spending function approach. Stat Med 1994;13:1341–52.

Fayers PM, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. The EORTC QLQ-C30 Scoring Manual (2nd Edition). Published by: European Organization for Research and Treatment of Cancer, Brussels 1999.

Goldhirsch A, Gelber RD, Piccart-Gebhart MJ, et al., and the Herceptin Adjuvant (HERA) Trial Study Team. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. Lancet 2013;382:1021–8.

von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. N Engl J Med, 2017.

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

Pivot X, Romieu G, Debled M, et al., and the PHARE trial investigators. 6 months versus 12 months of adjuvant trastuzumab for patients with HER2-positive early breast cancer (PHARE): a randomised phase 3 trial. Lancet Oncol 2013;14:741–8.

Romond E, Suman VJ, Jeong J-H, et al. Trastuzumab plus adjuvant chemotherapy for HER2-positive breast cancer: final planned joint analysis of overall survival (OS) from NSABP B-31 and NCCTG N9831. Cancer Res 2012;72(24 Suppl 3):S5–5.

Slamon DJ, Clark GM, Wong SG, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neo oncogene. Science 1987;235:177–82.

Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 1989;244:707–12.

Slamon D, Eiermann W, Robert N, et al. San Antonio Breast Cancer Conference 2006. Phase III trial comparing AC-T with AC-TH and with TCH in the adjuvant treatment of HER2 positive early breast cancer patients: second interim efficacy analysis. San Antonio Breast Cancer Symposium, 2006. Available from: http://www.bcirg.org/NR/rdonlyres/eqkdodg2dy7t557o7s6uvj7ytpe6gcfg5gmh2ely6hnhh5pjlabz3nd6jddlnao7qoikej3edohsijyiisfvp367uuc/BCIRG006+2nd+Interim+Analysis.pdf.

Slamon DJ, Eiermann W, Robert N, et al. Phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel (AC→T) with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab (AC→TH) with docetaxel, carboplatin and trastuzumab (TCH) in Her2neu positive early breast cancer patients: BCIRG 006 study. Cancer Res 2009;69(24 Suppl 3):62.

Slamon D, Eiermann W, Robert N, et al. Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 2011;365:1273–83.

Appendix 1
Protocol Synopsis

PROTOCOL SYNOPSIS

TITLE:	**A RANDOMIZED, MULTICENTER, OPEN-LABEL, PHASE III TRIAL COMPARING TRASTUZUMAB PLUS PERTUZUMAB PLUS A TAXANE FOLLOWING ANTHRACYCLINES VERSUS TRASTUZUMAB EMTANSINE PLUS PERTUZUMAB FOLLOWING ANTHRACYCLINES AS ADJUVANT THERAPY IN PATIENTS WITH OPERABLE HER2-POSITIVE PRIMARY BREAST CANCER**
PROTOCOL NUMBER:	BO28407
VERSION NUMBER:	3
EUDRACT NUMBER:	2012-004902-82
IND NUMBER:	71072
TEST PRODUCTS:	Trastuzumab Emtansine (RO5304020) and Pertuzumab (RO4368451)
PHASE:	III
INDICATION:	HER2-positive operable primary breast cancer
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives

Efficacy Objectives

The co-primary efficacy objective for this study is as follows:

- To compare invasive disease–free survival (IDFS) (1) in the node-positive subpopulation and (2) in the overall protocol-defined population of patients with human epidermal growth (HER) factor 2–positive breast cancer randomized to receive either a taxane and 1 year of trastuzumab plus pertuzumab following anthracycline-based chemotherapy or 1 year of trastuzumab emtansine plus pertuzumab following anthracycline-based chemotherapy

The secondary efficacy objectives for this study are as follows:

- To compare IDFS plus second non-breast primary cancers, disease-free survival (DFS), and distant recurrence–free interval (DRFI) (1) in the node-positive subpopulation and (2) in the overall protocol-defined population between the two treatment arms
- To compare overall survival (OS) (1) in the node-positive subpopulation and (2) in the overall protocol-defined population between the two treatment arms

Safety Objective

The safety objective for this study is as follows:

- To compare overall safety, cardiac safety, hepatic, and pulmonary safety in the overall protocol-defined population between the two treatment arms

Appendix 1
Protocol Synopsis (cont.)

Patient-Reported Outcome Objectives

Patient-reported outcome (PRO) objectives in the overall protocol-defined population for this study are as follows:

- To compare PROs of treatment-related symptoms, patient functioning, and health-related quality of life (HRQoL) to better understand treatment impact and tolerability, as measured by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire–Core 30 (QLQ-C30) and the modified EORTC Breast Cancer module (Quality of Life Questionnaire–Breast Cancer 23 [QLQ-BR23]), between 'Trastuzumab + Pertuzumab + Taxane following Anthracyclines' and 'Trastuzumab Emtansine + Pertuzumab following Anthracyclines' treatment arms.

Exploratory Objectives

The exploratory biomarker objectives for this study are as follows:

- To evaluate the impact of HER2 mRNA level on treatment benefit using the efficacy endpoints
- To evaluate the impact of the PIK3CA mutation status on prognosis and treatment benefit using the efficacy endpoints
- To assess correlations between candidate biomarkers or biomarker panels and efficacy and/or safety in the overall protocol-defined population
- To identify whether changes in expression levels of biomarker or biomarker panels during treatment correlate with treatment efficacy in the overall protocol-defined population

Efficacy endpoints considered for these objectives will include IDFS and OS, as appropriate.

The exploratory anti-therapeutic antibodies (ATAs) objective in the trastuzumab emtansine–treated patient population for this study is as follows:

- To assess the incidence of ATAs to trastuzumab emtansine and the effect of ATAs on safety and efficacy

The health economic exploratory objective is as follows:

- To assess health status as measured using the EuroQol 5-Dimension Questionnaire (EQ-5D) questionnaire for health economic modeling

Study Design

Description of Study

This is a prospective, two-arm, Phase III, randomized, multicenter, multinational, open-label study in patients with newly diagnosed HER2-positive primary invasive breast cancer who have had curative-intent surgery of their primary tumor and are candidates for adjuvant systemic chemotherapy following surgery. HER2-positive status of the primary tumor will be confirmed by the central pathology laboratory prior to enrollment of the patient in the study. Approximately *1850* patients *are anticipated to* be randomized to one of the two treatment arms listed below in a 1:1 ratio.

- Arm 1: Anthracycline chemotherapy of choice followed by trastuzumab at 6 mg/kg every 3 weeks (q3w; 8-mg/kg loading dose) in combination with pertuzumab 420 mg q3w (840-mg loading dose) and paclitaxel (80 mg/m^2) weekly (qw) or docetaxel q3w (see protocol for details of dose and duration). After the taxane-concurrent phase, HER2-targeted therapy (i.e., trastuzumab at 6 mg/kg q3w in combination with pertuzumab 420 mg q3w) will continue for up to 1 year.

- Arm 2: Anthracycline chemotherapy of choice followed by trastuzumab emtansine 3.6 mg/kg q3w in combination with pertuzumab 420 mg q3w (840-mg loading dose). HER2-targeted therapy (i.e., trastuzumab emtansine at 3.6 mg/kg q3w in combination with pertuzumab 420 mg q3w) will continue for up to 1 year.

Randomization will be stratified by world region (United States/Canada, Western Europe/Australia/New Zealand, Asia, or rest of the world); nodal status (0, 1–3, or ≥4 positive nodes); centrally assessed hormonal receptor status (estrogen receptor [ER] and/or progesterone receptor [PR] positive or both ER and PR negative); and type of anthracycline (doxorubicin or epirubicin).

In Arm 1, HER2-targeted study therapy with trastuzumab plus pertuzumab must start concurrently with the taxane component of chemotherapy following anthracycline therapy. In both arms, after anthracycline treatment, a minimum interval of 3 weeks from the last dose of anthracycline to initiation of HER2-targeted therapy is required. Prior to commencing the HER2-targeted therapy, patients must have a left ventricular ejection fraction (LVEF) ≥50% and must not have experienced any clinical symptoms suggesting heart failure or asymptomatic LVEF declines of 15 percentage points or more from baseline *and below the lower limit of normal*.

Patients will receive up to 1 year of HER2-targeted therapy. Study treatment will be discontinued in the event of invasive disease recurrence, unacceptable toxicity, withdrawal of consent, or study termination by the Sponsor.

Adjuvant radiotherapy is to be given as clinically indicated at the end of chemotherapy (end of taxane in Arm 1; after four cycles of trastuzumab emtansine plus pertuzumab in Arm 2, to be consistent with Arm 1 in terms of timing of initiation) while receiving HER2-targeted therapy. For patients with ER-positive and/or PR-positive tumors, hormonal agents should be administered at the end of chemotherapy (end of taxane in Arm 1; after four cycles of trastuzumab emtansine plus pertuzumab in Arm 2, to be consistent with Arm 1 in timing of initiation).

Number of Patients

Approximately 1850 patients *are anticipated to* be enrolled at approximately 350 sites worldwide.

Target Population

The target population for this study will be patients with newly diagnosed primary invasive breast cancer that is HER2 positive (as determined by the central pathology laboratory) and who will be treated with adjuvant systemic chemotherapy following definitive surgery.

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Age ≥18 years
- Eastern Cooperative Oncology Group Performance Status ≤1
- Non-metastatic histologically confirmed primary invasive breast carcinoma that was operable
- HER2-positive breast cancer prospectively determined on the primary tumor by a central pathology laboratory and defined as follows:

  Immunohistochemistry (IHC) score of 3+ and/or positive by in situ hybridization (ISH), as defined by ISH ratio of ≥2.0 for the number of *HER2* gene copies to the number of chromosome 17 copies. Both IHC and ISH assays will be performed; however, only one positive result is required for eligibility.

  Availability of formalin-fixed paraffin-embedded (FFPE) tissue block or partial block (for minimum dimensions, see laboratory manual) with a representative invasive part of the tumor for central pathology laboratory confirmation of HER2 eligibility, hormonal receptor status, and additional biomarker analysis is required.

Appendix 1
Protocol Synopsis (cont.)

- Known hormone receptor status of the primary tumor determined by a central pathology laboratory

  Hormone receptor-positive status can be determined by either known positive ER or known positive PR status. Hormone receptor–negative status must be determined by both known negative ER and known negative PR.

- Adequately excised: Patients must have undergone either breast-conserving surgery or mastectomy/nipple- or skin-sparing mastectomy.

  For patients who undergo breast-conserving surgery, the margins of the resected specimen must be histologically free of invasive tumor and ductal carcinoma in situ (DCIS) as determined by the local pathologist. If pathologic examination demonstrates tumor at the line of resection, additional operative procedures may be performed to obtain clear margins. If tumor is still present at the resected margin after re-excision(s), the patient must undergo total mastectomy to be eligible. Patients with margins positive for lobular carcinoma in situ (LCIS) are eligible without additional resection.

  For patients who undergo mastectomy/nipple- or skin-sparing mastectomy, margins must be free of gross residual tumor. *It is recommended that p*atients should have a negative microscopic margin in accordance with local pathology protocol. Patients with a microscopic positive deep margin are eligible (see radiation therapy [RT] requirements in the protocol).

- Pathological tumor-node-metastasis staging (Union for International Cancer Control/American Joint Committee on Cancer [UICC/AJCC], 7th edition): Patients must have had sentinel lymph node biopsy and/or axillary lymph node dissection for evaluation of pathologic nodal status. Pathological classification of regional lymph node micrometastases (tumor deposits >0.2 mm and ≤ 2 mm) is considered to be pN1, and isolated tumor cells are considered to be pN0.

  Eligible patients must have one of the following:

  Node-positive disease (pN ≥ 1), any tumor size except T0, and any hormonal receptor status

  Enrollment of patients with 1–3 nodes *was planned to* be limited to no more than 50% of the total number of randomized patients. *However, no formal capping of enrollment will be implemented (see Section 3.3.2.1 for details).*

  There is no prespecified limit for the enrollment of patients with ≥ 4 nodes.

  Node-negative disease (pN0) with pathologic tumor size >2.0 cm by standard local assessment AND negative for ER and PR as determined by a central pathology laboratory

  Enrollment of patients with node-negative disease *was planned to* be limited to no more than 10% of the total number of randomized patients. *However, no formal capping of enrollment will be implemented (see Section 3.3.2.1 for details).*

- Patients with synchronous bilateral invasive disease are eligible only if both lesions are HER2 positive.
- No more than 9 weeks (63 days) may elapse between definitive breast surgery (or the last surgery if additional resection required for breast cancer) and randomization.
- Baseline LVEF $\geq 55\%$ measured by echocardiogram (preferred) or multiple-gated acquisition scans

Documentation of hepatitis B virus (HBV) and hepatitis C virus (HCV) serologies is required. This includes hepatitis B surface antigen and/or total hepatitis B core antibody in addition to HCV antibody testing. The most recent serologic testing must have occurred within 3 months prior to randomization. If such testing has not been done, it must be performed during screening.

Patients who have positive HBV or HCV serologies without known active disease must meet the eligibility criteria for ALT, AST, total bilirubin (TBILI), INR, activated partial thromboplastine time (aPTT), and alkaline phosphatase (ALP) on at least two consecutive occasions, separated by at least 1 week, within the 30-day screening period. The second of these evaluations must be performed within 3 days prior to the first administration of study drug. *Note: positive serology markers that indicate immunity will not be considered as clinically meaningful positive serology to trigger these tests.*

- Female patients of childbearing potential must be willing to use one highly effective form of nonhormonal contraception or two effective forms of nonhormonal contraception. For male patients with partners of childbearing potential, one highly effective form of contraception or two effective forms of contraception must be used (see protocol for descriptions of highly effective and effective contraception). Contraception must continue for the duration of study treatment and for 7 months after the last dose of study treatment.

  The above contraception is not a requirement in the case of any of the following:

  The patient or partner of the patient is surgically sterilized.

  The female patient is $>$45 years of age and is postmenopausal (has not menstruated for at least 12 consecutive months).

  The patient truly abstains from sexual activity and when this is the preferred option to avoid conception and contraception and/or usual lifestyle of the patient.
- Male patients whose partners are pregnant must use condoms or truly refrain from sexual activity for the duration of the pregnancy
- Willing and able to comply with the requirements of the protocol
- Signed informed consent

Exclusion Criteria

Patients who meet any of the following criteria will be excluded from study entry:

- History of any prior (ipsilateral and/or contralateral) invasive breast carcinoma
- History of non-breast malignancies within the 5 years prior to randomization, except for the following:

  Carcinoma in situ (CIS) of the cervix

  CIS of the colon

  Melanoma in situ

  Basal cell and squamous cell carcinomas of the skin
- Any clinical T4 tumor as defined by tumor-node-metastasis classification in UICC/AJCC, 7th edition, including inflammatory breast cancer
- For the currently diagnosed breast cancer, any previous systemic anti-cancer treatment (e.g., neoadjuvant or adjuvant), including but not limited to chemotherapy, anti-HER2 therapy (e.g., trastuzumab, trastuzumab emtansine, pertuzumab, lapatinib, neratinib, or other tyrosine kinase inhibitors), hormonal therapy, OR anti-cancer RT (intraoperative radiotherapy as a boost at the time of primary surgery is acceptable)
- Previous therapy with anthracyclines, taxanes, or HER2-targeted therapy for any malignancy
- History of DCIS and/or LCIS that was treated with any form of systemic chemotherapy, hormonal therapy, or RT to the ipsilateral breast where invasive cancer subsequently developed. Patients who had their DCIS/LCIS treated with only surgery and/or contralateral DCIS treated with radiation are allowed to enter the study.
- Patients with contraindication to RT while adjuvant RT is clinically indicated
- Concurrent anti-cancer treatment in another investigational trial

Appendix 1
Protocol Synopsis (cont.)

- Cardiopulmonary dysfunction as defined by any of the following prior to randomization:

  History of National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0 Grade ≥ 3 symptomatic congestive heart failure (CHF) or New York Heart Association (NYHA) criteria Class $\geq$ II

  Angina pectoris requiring anti-anginal medication, serious cardiac arrhythmia not controlled by adequate medication, severe conduction abnormality, or clinically significant valvular disease

  High-risk uncontrolled arrhythmias (i.e., atrial tachycardia with a heart rate > 100/min at rest, significant ventricular arrhythmia [ventricular tachycardia], or higher-grade atrioventricular [AV]-block [second degree AV-block Type 2 [Mobitz 2] or third degree AV-block])

  Significant symptoms (Grade ≥ 2) relating to left ventricular dysfunction, cardiac arrhythmia, or cardiac ischemia

  Myocardial infarction within 12 months prior to randomization

  Uncontrolled hypertension (systolic blood pressure > 180 mmHg and/or diastolic blood pressure > 100 mmHg)

  Evidence of transmural infarction on ECG

  Requirement for oxygen therapy

- Other concurrent serious diseases that may interfere with planned treatment, including severe pulmonary conditions/illness, uncontrolled infections, uncontrolled diabetes, or known infection with HIV
- Any known active liver disease, including but not limited to disease due to HBV, HCV, autoimmune hepatic disorders, or sclerosing cholangitis. For patients who are known carriers of HBV/HCV, active hepatitis B/C infection must be ruled out on the basis of negative serologic testing and/or determination of HBV DNA/HCV RNA viral load per local guidelines.
- Any of the following abnormal laboratory tests prior to randomization:

  Serum TBILI > 1.0 times the upper limit of normal (ULN). In cases of known Gilbert's syndrome, direct bilirubin should be within the normal range.

  ALT and/or AST $>$ ULN

  ALP $> 1.5 \times$ ULN

  Serum creatinine $> 1.5 \times$ ULN

  Total WBC $< 2500/\mu L$ ($< 2.5 \times 10^9$/L)

  ANC $< 1500/\mu L$ ($< 1.5 \times 10^9$/L)

  Platelets $< 100{,}000/\mu L$ ($< 100 \times 10^9$/L)

  INR or aPTT $> 1.5 \times$ ULN

- Pregnant or lactating women or women of childbearing potential without a negative serum pregnancy test result, within 7 days prior to randomization, regardless of the method of contraception used
- Hypersensitivity to any of the study medications or any of the ingredients or excipients of these medications, including hypersensitivity to benzyl alcohol
- Chronic immunosuppressive therapies, including systemic corticosteroids

Length of Study

The total length of this study will be from randomization of the first patient to completion of the last follow-up assessment of the last patient, which is estimated to occur approximately 10 years after the first patient is randomized.

Appendix 1
Protocol Synopsis (cont.)

End of Study

To enable long-term follow-up for survival and safety information, the study is planned to end approximately 10 years after the first patient is randomized.

The Sponsor has the right to terminate this study*, including long-term follow-up,* at any time (e.g., if emerging safety signals indicate a potential health hazard to patients).

Outcome Measures

Efficacy Outcome Measures

The primary efficacy outcome measures for this study are listed below.

- IDFS, defined as the time from randomization until the date of the first occurrence of one of the following events:

  Ipsilateral invasive breast tumor recurrence (i.e., an invasive breast cancer involving the same breast parenchyma as the original primary lesion)

  Ipsilateral local-regional invasive breast cancer recurrence (i.e., an invasive breast cancer in the axilla, regional lymph nodes, chest wall, and/or skin of the ipsilateral breast)

  Contralateral or ipsilateral second primary invasive breast cancer

  Distant recurrence (i.e., evidence of breast cancer in any anatomic site [other than the three sites mentioned above]) that has either been histologically confirmed or clinically/radiographically diagnosed as recurrent invasive breast cancer

  Death attributable to any cause, including breast cancer, non-breast cancer, or unknown cause

The secondary efficacy outcome measures for this study are as follows:

- IDFS plus second primary non-breast cancer, excluding non-melanoma skin cancers and CIS of any site
- DFS, defined as the time between randomization and the date of the first occurrence of any of the IDFS events described above, second primary non-breast cancer event (excluding non-melanoma skin cancers and CIS of any non-breast site), and contralateral or ipsilateral DCIS
- DRFI, defined as the time between randomization and the first occurrence of distant breast cancer recurrence
- OS, defined as the time from randomization to death due to any cause

Safety Outcome Measures

The safety outcome measures for this study are as follows:

- Incidence, type, and severity of all adverse events based on NCI CTCAE v4.0
- Incidence, type, and severity of serious adverse events
- Incidence, type and severity of $\geq$ Grade 3 adverse events
- Incidence and type of adverse events leading to dose discontinuation, modification, or delay
- Cause of death
- Abnormal laboratory values
- Decrease in LVEF from baseline over time
- Cardiac safety outcome measures

  Primary cardiac endpoints: cardiac events defined as death from cardiac cause or severe CHF (NYHA Class III or IV) with a decrease in LVEF of $\geq$ 10 percentage points from baseline to an LVEF of $<$ 50%

Secondary cardiac endpoints: other cardiac-related events (e.g., any mild symptomatic CHF [NYHA Class II] associated with a ≥10% drop in LVEF to <50%; asymptomatic declines in LVEF requiring dose delay or discontinuation)

- Hepatic safety outcome measures
  - Death from hepatic cause
  - Severe drug-induced liver injury (Hy's law cases)
  - Nodular regenerative hyperplasia
- Pulmonary safety outcome measures
  - Death from pulmonary cause
  - Pneumonitis and interstitial lung disease

Patient-Reported Outcome Measures

The PRO outcome measures for this study are as follows:

- HRQoL, including bothersome side effects of therapy (e.g., peripheral neuropathy, joint/muscle pain, or skin problems), and patient functioning as measured using the EORTC QLQ-C30 and the modified breast cancer module QLQ-BR23
- Time from first HER2-targeted treatment, ± a taxane, to *clinically meaningful deterioration in the* global health status/quality of life (QoL) *and functional (physical, role, and cognitive)* subscale*s* of the QLQ-C30. The event of worsening of global health status/QoL for a given patient is defined as a decrease in *baseline* mean score by 10 points or more at two consecutive timepoints. A 10-point or greater change in mean score is defined as being a "moderate" to "very much" *and* perceived *an* important change from the patient's perspective. *Deterioration in function will be assessed using the published corresponding MIDs by* Cocks et al. 2011.

Exploratory Outcome Measures

The exploratory biomarker outcome measures for this study are the relationship between molecular markers and efficacy and/or safety outcomes. Efficacy outcomes considered for this analysis will include IDFS and OS, as appropriate.

Correlations between biomarker status and efficacy and/or safety will include but not be limited to the following:

- Level of HER2 mRNA expression assessed by quantitative real-time polymerase chain reaction (qRT-PCR) with efficacy outcome
- Status of PIK3CA mutations assessed by PIK3CA allele–specific polymerase chain reaction assay with efficacy outcome
- Level of *HER2* gene amplification assessed by ISH with efficacy outcome
- Level of HER2 protein expression assessed by IHC with efficacy outcome
- Level of HER3 mRNA expression assessed by qRT-PCR with efficacy outcome
- Changes in expression levels of biomarker or biomarker panels over time with efficacy outcome

The ATA outcome measures to be assessed in patients receiving trastuzumab emtansine are the following:

- Incidence of ATAs to trastuzumab emtansine
- Effect of ATAs on safety and efficacy

The EQ-5D will be used to obtain health status information for health economic modeling.

Appendix 1
Protocol Synopsis (cont.)

The EQ-5D is a generic, preference-based health utility measure with questions about mobility, self-care, usual activities, pain/discomfort, and anxiety/depression that are used to build a composite of the patient's health status. A single summary index from the EQ-5D health states will be utilized in this study for economic modeling, and the results will not be reported in the clinical study report.

Investigational Medicinal Products

Study treatment is defined as non-hormonal systemic adjuvant (post-operative) treatment. Trastuzumab emtansine, pertuzumab, and trastuzumab are considered investigational medicinal products (IMPs) in this study. Paclitaxel and docetaxel are also considered IMPs in this study; however, paclitaxel and docetaxel may be considered non-IMPs on the basis of local legislation.

Doxorubicin, epirubicin, cyclophosphamide, and 5-fluorouracil are considered non-IMPs in this study. Depending on local legislation, doxorubicin, epirubicin, cyclophosphamide, and 5-fluorouracil may be considered IMPs. If considered an IMP, then appropriate information on formulation, packaging, handling, and administration will be provided.

Test Product

Trastuzumab emtansine will be given at a dose of 3.6 mg/kg by intravenous (IV) infusion q3w in combination with pertuzumab 420 mg IV q3w (840 mg loading dose) for up to a total duration of 1 year.

Comparator

Pertuzumab will be given at a dose of 420 mg IV q3w (840 mg loading dose) in combination with trastuzumab at 6 mg/kg IV q3w (8 mg/kg loading dose) and paclitaxel (80 mg/m^2) qw or docetaxel q3w (details regarding docetaxel dose and duration are described in the protocol). After the taxane concurrent phase, administration of trastuzumab/pertuzumab will continue for up to a total duration of 1 year

Non-Investigational Medicinal Products

Standard-of-care chemotherapy backbone treatments should include three to four cycles of an anthracycline-based regimen. Either 5-fluorouracil, epirubicin, and cyclophosphamide (FEC) or doxorubicin and cyclophosphamide (AC)/epirubicin and cyclophosphamide (EC) regimens that are protocol approved may be selected at the discretion of the investigator. Paclitaxel and docetaxel may be considered non-IMPs on the basis of local legislation. In Arm 1, three to four cycles of docetaxel or 12 weeks of paclitaxel will be administered.

For non-IMPs, refer to local prescribing information/institutional guidelines for detailed guidelines on administration, premedications, dose delays/reductions for toxicities, contraindications, requirements on contraception duration, and concomitant medications.

Statistical Methods

Primary Analysis

The primary efficacy variable is IDFS, defined as the time between randomization and date of first occurrence of an IDFS event as described in the Efficacy Outcome Measures section. Data from patients who have not had an event at the time of data analysis will be censored on the date on which they are last known to be alive and event free, on or before the clinical data cutoff date of the respective analysis.

The log-rank test, stratified by the protocol-defined stratification factors (excluding region), will be used to compare IDFS between the two treatment arms. Region will be excluded because of the potential that some of the strata may have very few patients, which would result in a loss of power. The unstratified log-rank test results will also be provided for sensitivity analysis. If, at the time of analysis, it is deemed that the smallest stratum per arm necessary to conduct robust stratified analyses contains < 5 events, unstratified analyses will be used as the primary analysis. The Cox proportional hazards model, stratified by the previously noted stratification factors, excluding region, will be used to estimate the hazard ratio (HR) between the two treatment arms

and the corresponding 95% CI. The Kaplan-Meier approach will be used to estimate 3-year IDFS rates and corresponding 95% CIs for each treatment arm.

Determination of Sample Size

The sample size of the study is primarily driven by the analysis of IDFS in both the node-positive subpopulation and the overall protocol-defined population.

To detect a target HR of 0.64 in IDFS in the overall protocol-defined population and the node-positive subpopulation, approximately 171 and 160 IDFS events will be required to achieve 82.5% and 80% power, respectively, in the two populations, at a two-sided significance level of 5% using a log-rank test. Approximately 1850 and 1665 patients will be enrolled in the overall protocol-defined population and node-positive subpopulation, respectively, including a dropout/ineligibility rate of 8% for both arms as estimated from previous trials in this setting.

Hierarchical Testing Procedure

The hierarchical testing procedure will be employed according to the order given below on the primary and secondary endpoints to control the overall study Type I error rate at 5%:

- Primary endpoint of IDFS in the node–positive subpopulation
- Primary endpoint of IDFS in the overall protocol–defined population
- Secondary endpoint of OS in the node–positive subpopulation
- Secondary endpoint of OS in the overall protocol-defined population

Interim Analyses

One interim analysis of IDFS and three interim analyses of OS in both the node-positive subpopulation and the overall protocol-defined population are planned.

A futility analysis may be incorporated to evaluate lack of superiority for the treatment arm of trastuzumab emtansine +pertuzumab following anthracyclines. The timing and details of this futility analysis will be described in the SAP and aligned with the availability of information from other relevant studies (e.g., KRISTINE and/or APHINITY).

Interim IDFS Analyses

An interim efficacy analysis of IDFS will be performed after approximately 75% of the targeted IDFS events are observed in the overall protocol-defined population (i.e. 128 of the 171 target events) and node-positive population (i.e. 120 of 160 target events) and is projected to occur approximately 43 months from first patient initiation (FPI).

A hierarchical testing procedure will be used for the primary endpoint IDFS for the node-positive subpopulation and the overall protocol-defined population as defined in the protocol. The type I error will be controlled at the 5% level at the interim and final analyses within the node-positive subpopulation and the overall protocol-defined population using a Lan-DeMets α-spending function with an O'Brien-Fleming boundary.

The interim analysis will be performed by the independent Data Coordination Center (iDCC) statistician, and the results will be presented to the independent Data Monitoring Committee (iDMC) by the iDCC statistician.

If the interim analyses in both populations cross the interim efficacy boundaries of the O'Brien-Fleming design, on the basis of the totality of both efficacy and safety data, the iDMC may recommend releasing the primary endpoint results before the targeted number of 160 and 171 events have occurred in the node-positive subpopulation and overall protocol-defined population, respectively. In this case, the Sponsor will be unblinded to study results, and a full data package including the first OS interim analysis results will be prepared for discussion with the regulatory authorities. The study will continue until 10 years of follow-up from FPI have occurred, and the IDFS analysis will be updated descriptively. If the IDFS interim analyses fail to cross the statistical efficacy boundaries of the O'Brien-Fleming design, then the study will continue as planned. The Sponsor will conduct the final analyses.

Appendix 1
Protocol Synopsis (cont.)

The protocol summarizes the planned IDFS analyses in the overall protocol-defined population and the node-positive subpopulation; the efficacy stopping boundaries based on the expected number of events; and the estimated timing of these analyses. The boundaries to be used at each interim and final IDFS analysis will depend on the number of IDFS events actually included in the analyses and so may vary from the numbers stated in the protocol.

Interim OS Analyses

Three interim OS analyses and one final OS analysis are planned in both the overall protocol-defined population and the node-positive subpopulation.

For the OS interim analyses and final OS analysis, the Lan-DeMets α-spending function with an O'Brien-Fleming boundary will be used such that the overall type I error will be controlled at the 5% level for the OS endpoint. With the study sample size and approximately 10 years of follow-up from FPI, this study has 38% power in the overall protocol-defined population and 35% power in the node-positive subpopulation to detect an HR of 0.8. This in the overall protocol-defined population corresponds to a 0.8% improvement in 3-year OS, from 96.1% in the control arm to 96.9% in the experimental arm and in the node-positive subpopulation to a 0.9% improvement in 3-year OS, from 95.7% in the control arm to 96.6% in the experimental arm, at a two-sided significance level of 5%.

All OS interim and final analyses will be performed by the Sponsor subsequent to the primary IDFS analysis and after the Sponsor is unblinded. If the interim IDFS analyses in both the node-positive subpopulation and the overall protocol-defined population cross the statistical efficacy boundaries, the first OS interim analysis will be performed in both populations hierarchically at that time (approximately at 43 months from FPI). If the final IDFS analysis crosses the statistical boundaries, then the second interim OS analysis will be performed hierarchically at the time of the final IDFS analysis (approximately 57 months from FPI), followed by the third and the final OS analyses performed hierarchically in both populations (planned to occur at 84 months [7 years] and 120 months [10 years], respectively, from FPI). If, at any OS interim analysis, the O'Brien-Fleming efficacy boundary is crossed, that analysis of OS will be considered as confirmatory and all subsequent analyses of OS will be considered as descriptive.

The protocol summarizes the planned OS analyses in the overall protocol-defined population and the node-positive subpopulation, respectively; the efficacy stopping boundaries based on the expected number of events; and the estimated timing of these analyses. The boundaries to be used at each interim and final IDFS analysis will depend on the number of IDFS events actually included in the analyses and may vary from the numbers stated in the protocol.

Interim Safety Analyses

An iDMC will monitor accruing patient safety data at least once every 6 months during the study until the last patient has completed study treatment. In addition, safety data related to concurrent radiotherapy and/or hormonal therapy, serious adverse events, and deaths will be monitored by the iDMC at least once every 3 months during the study until the last patient has completed study treatment. At each iDMC review, relevant safety information from ongoing trastuzumab emtansine and/or pertuzumab studies will also be provided to the iDMC.

After the first 600 patients have been randomized and followed for 3 months (anticipated to occur at approximately 13 months after FPI), the iDMC will perform an interim safety analysis regarding overall numbers of deaths (all causes, including cardiac deaths) and hepatic events defined as confirmed Hy's law cases. The Clinical Events Committee (CEC) will communicate their findings to the iDMC to aid iDMC review.

If an absolute increase of $>$ 3% in the percentage of death (from any cause) is observed in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, the iDMC will consider a recommendation of pausing enrollment for further data review, stopping the trial, or modifying the trial.

If the true difference in the percentage of death is $>$ 3% (e.g., 2% vs. 6%), then there is approximately a 70% chance of observing an absolute difference of $>$ 3% at the interim with 600 patients. The protocol presents the probability of observing an increase of $>$ 3% in the

percentage of deaths in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, with different assumptions for the percentage of deaths in the two treatment arms.

If an absolute increase of > 3% in the percentage of Hy's law cases (confirmed by the independent CEC) is observed in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, the iDMC will consider a recommendation of pausing enrollment for further data review, stopping the trial, or modifying the trial.

If the true difference in the percentage of confirmed Hy's law cases is > 3% (e.g., 0.33% vs. 3.67%), then there is approximately a 54% chance of observing an absolute difference of > 3% at the interim with 600 patients. The protocol presents the probability of observing a > 3% increase in the percentage of Hy's law cases in the trastuzumab emtansine + pertuzumab arm compared with the trastuzumab + pertuzumab arm, with different assumptions for the number of Hy's law cases in the two treatment arms.

The iDMC will work according to guidelines defined in the iDMC Charter. The iDMC Charter will contain details regarding frequency of meetings, guidelines for decision making, and processes for requesting further information. The iDMC members will review and sign off the charter before the first iDMC review.

Appendix 2
Schedule of Assessments

	Screening[a] (30 days)	Anthracycline Treatment Period	HER2-Targeted Treatment Period within 3 Days of Day 1 of Cycle Number X				End-of-Treatment Visit[b]	Follow-Up[c] (±28 days)		Follow-Up[c] (±42 days)
		Cycles 1[a]–3 or Cycle 4	Cycles 1–4 (concurrent taxane in Arm 1 only)			Cycles 5–18		Year 1–2	Year 3–5	Year 6–10
		Day 1	Day 1	Day 8	Day 15	Day 1				
Informed consent[d]	x									
Mandatory tumor tissue sample for determination of HER2, ER/PR; and exploratory biomarkers[e]	x	x[e] (at disease recurrence)								
Bilateral mammogram[f,g]	x (within 6 months)	x (q12mo)								
Chest X-ray[h]	x (within 2 months)	As clinically indicated								
Demography, medical history	x									
Disease status assessment[g]	x	x (q3mo)						x (q3mo)	x (q6mo)	x (q12mo)
Physical examination[g,i]	x	x (Cycle 1)	x			x	x	x (q3mo)	x (q6mo)	x (q12mo)

Appendix 2
Schedule of Assessments (Cont.)

	Screening [a] (30 days)	Anthracycline Treatment Period	HER2-Targeted Treatment Period within 3 Days of Day 1 of Cycle Number X				End-of-Treatment Visit [b]	Follow-Up [c] (±28 days)		Follow-Up [c] (±42 days)
		Cycles 1[a]–3 or Cycle 4	Cycles 1–4 (concurrent taxane in Arm 1 only)			Cycles 5–18		Year 1–2	Year 3–5	Year 6–10
		Day 1	Day 1	Day 8	Day 15	Day 1				
Vital signs[j]	x	x	x	x (for paclitaxel only)	x (for paclitaxel only)	x				
ECOG Performance Status	x						x			
ECG	x	As clinically indicated					x			
ECHO/MUGA[k]	x	x[k] (end of last cycle)	x[k] (in last week of Cycle 2, and every 4 cycles thereafter)				x[k]	x[k]	x[k]	
PRO HRQoL assessment[l]		x[l]	x[l]			x[l]	x	x[l] (q6mo)		
Hematology[m]	x[n]	x[o]	x[o]	x[o] (for paclitaxel only)[p]	x[o] (for paclitaxel only)[p]	x[o]	x	x[m]		
Serum chemistry[q]	x[n]	x[o]	x[o]			x[o]	x	x[q]		
HBV and HCV serology[r]	x	As clinically indicated								
INR and aPTT	x[n]	As clinically indicated								
Pregnancy test[s]	x[n]		x[o,s] (C1D1 then every 3 cycles thereafter)					x[s]		
Clinical genotyping whole blood sample[t]		x[t]								

Appendix 2
Schedule of Assessments (Cont.)

	Screening[a] (30 days)	Anthracycline Treatment Period	HER2-Targeted Treatment Period within 3 Days of Day 1 of Cycle Number X				End-of-Treatment Visit[b]	Follow-Up[c] (±28 days)		Follow-Up[c] (±42 days)
		Cycles 1[a]–3 or Cycle 4	Cycles 1–4 (concurrent taxane in Arm 1 only)			Cycles 5–18		Year 1–2	Year 3–5	Year 6–10
		Day 1	Day 1	Day 8	Day 15	Day 1				
Mandatory serum and plasma sample for exploratory biomarker analysis[u,v]		x (C1D1 predose)	x (C1D1 predose)				x	x[s]		
		x[u] (at disease recurrence)								
OBRP whole blood sample for genetic analysis (optional)[v]		x[v](to be done at baseline if possible)								
Adverse events[w]	x	x	x	x (for paclitaxel only)	x (for paclitaxel only)	x	x	x[w]		
Concomitant medications[x]	x	x	x	x (for paclitaxel only)	x (for paclitaxel only)	x	x			
Record post-recurrence anti-cancer–related therapies[y]		As required								
Survival post-recurrence[y]		q12mo								

AE = adverse event; ALP = alkaline phosphatase; aPTT = activated partial thromboplastine time; ATA = anti-therapeutic antibody; C1D1 = Cycle 1 Day 1; CT = computed tomography; eCRF = electronic Case Report Form; ECHO = echocardiogram; ECOG = Eastern Cooperative Oncology Group; ER = estrogen receptor; EORTC = European Organisation for Research and Treatment of Cancer; EQ-5D = EuroQol 5-Dimension Questionnaire; FFPE = formalin-fixed paraffin-embedded; FPI = first patient initiation; HBcAb = hepatitis B core antibody; HBsAg = hepatitis B surface antigen; HBV = hepatitis B virus; HCG = human chorionic gonadotropin; HCV = hepatitis C virus; HER2 = human epidermal growth factor-2; HRQoL = health-related quality of life; IDFS = invasive disease–free survival;

Appendix 2
Schedule of Assessments (Cont.)

IVRS/IWRS = interactive voice response system/interactive web response system; LVEF = left ventricular ejection fraction; MRI = magnetic resonance imaging; MUGA = multiple-gated acquisition; OBRP = optional biomarker research program; pac = paclitaxel; PK = pharmacokinetic; PR = progesterone receptor; PRO = patient-reported outcome; QLQ-BR23 = Quality of Life Questionnaire–Breast Cancer 23; QLQ-C30 = Quality of Life Questionnaire–Core 30; qw = weekly; SAE = serious adverse event; SoC = standard of care; TBILI = total bilirubin; ULN = upper limit of normal.

[a] Screening to be performed within 30 days prior to randomization. The randomization visit may be combined with anthracycline Cycle 1 Day 1 visit. Patients should receive their first dose of study treatment on the day of randomization, if possible, but no later than 7 days after randomization. No more than 9 weeks (63 days) may elapse between definitive breast surgery (or the last surgery if additional resection required for breast cancer) and randomization.

[b] End-of-treatment visit or early termination visits will optimally be scheduled within 28 to 42 days following the last dose of study treatment.

[c] The follow-up period begins from the date of the end-of-treatment visit, with a duration of up to 10 years from the date of randomization of the first patient.

[d] Informed consent may be obtained at any time (including prior to the 30-day screening period) but must be obtained prior to the performance of any screening assessments. Results of screening tests or examinations performed as SoC prior to obtaining informed consent and within 30 days prior to randomization may be used rather than repeating required tests unless the tests are required to be within 7 days prior to randomization.

[e] FFPE tumor tissue must be obtained for central laboratory assessment of HER2 status for eligibility. ER/PR determination will also be evaluated by a central laboratory. A tissue block (FFPE material) collected at definitive breast cancer surgery is acceptable. Sections and/or slides are not acceptable. Biomarker analysis will only be performed on tissue from randomized patients. At recurrence, tumor biopsy sample should be taken if the tumor is accessible for biopsy sample collection without (in the investigator's opinion) significant risk to the patient. In case tissue becomes available at time of recurrence, this tissue will be used for exploratory biomarker analysis. If a biopsy is collected as part of routine medical practice at relapse/recurrence, a tissue block or up to five unstained slides should be sent for biomarker analysis in order to gain better understanding of resistance mechanisms.

[f] Bilateral mammogram (or breast MRI if indicated) is to be performed within 6 months prior to randomization. *If mammogram is not available, then it could be performed after surgery.* During treatment and follow-up, mammograms of any remaining breast tissue should be performed at least annually *(±28 days)* since the previous mammogram. *If there is no remaining breast tissue, then mammograms are not mandated.*

[g] Disease status based on all available clinical assessments should be documented from the date of randomization at the following timepoints: every 3 months *(±3 days)* during study treatment and for the first 2 years after study treatment completion/early termination, every 6 months *(±7 days)* from Years 3 to 5 after study treatment completion, and annually *(±*28 days) from Year 6 up to Year 10 after study treatment completion, unless a recurrence that is defined as an IDFS event in Section 3.4.1 has occurred. Patients who have a diagnosis of in situ breast cancer or second non-breast cancer should be maintained on a regular follow-up schedule as described above wherever possible in order to fully capture any subsequent recurrent invasive disease events. Such patients should continue with adjuvant study treatment, if not yet completed and considered by the investigator to be in the patient's best interest, whenever possible. In addition to physical examinations and mammograms, liver function tests, bone scans, chest X-rays/diagnostic CT, liver imaging, or other radiographic modality may be considered when clinically indicated to exclude metastatic disease and within a timeline as per current local standard of practice. Whenever possible, disease recurrence should be confirmed histopathologically. In cases of first disease recurrence (an IDFS event as defined in Section 3.4.1) diagnosed at any time during the study, patients will be out of the study schedule and will be followed annually starting 1 year after first recurrence until Year 10 from FPI for survival, anti-cancer medications, and new relapse (recurrence or disease progression) events.

[h] If a CT scan is already available within 2 months prior to randomization, it may be used in lieu of chest X-ray.

Appendix 2
Schedule of Assessments (Cont.)

[i] A complete physical examination, including height and weight, should be measured at baseline. Throughout the study, limited symptom-directed physical examination focusing on organ systems related to AE or disease may be performed. Weight is to be measured on Day 1 of the specified cycles and compared to baseline. If ±10% or greater variation occurs, then study treatment doses will be recalculated. Dose must be readjusted for ±10% or greater weight change based on the previous weight used for dose recalculation. *For anthracyclines and taxanes, local standards for dose calculations will be followed.*

[j] Vital signs include blood pressure, pulse rate, and body temperature. Vital signs should be obtained and reviewed, but it is not required that they should always be entered into the eCRF. Abnormal vital signs at any time during the course of study treatment should be recorded as AEs or SAEs if clinically significant.

[k] Cardiac monitoring (ECHO/MUGA) will be performed to assess LVEF. LVEF assessment by ECHO is preferred. The same method should be used throughout the study for each patient and, preferably, performed and assessed by the same assessor. At baseline, LVEF must be done within 14 days prior to randomization. At the end of anthracycline therapy, LVEF should be assessed to determine if patient can start HER2-targeted therapy. During HER2-targeted study treatment, ECHO/MUGA should be obtained at Cycle 2, and every 4 cycles thereafter (Cycles 6, 10, 14, and 18). All assessments will be performed within the last week of the treatment cycle to allow evaluation of the results before the next treatment cycle. ECHO/MUGA should be obtained at the end-of-treatment visit if not performed within the previous 6 weeks and at 3, 6, 12, 18, 24, 36, 48, and 60 months of follow-up regardless of the occurrence of invasive disease recurrence. Note: Cardiac signs/symptoms and an additional LVEF assessment must be completed after the last cycle of anthracycline is administered, but prior to the first cycle of targeted therapy. Patients treated with anthracyclines must have an LVEF ≥50% and must have not experienced any clinical symptoms suggesting heart failure or asymptomatic LVEF declines by an absolute point of >15% from baseline *and below the lower limit of normal* prior to commencing the HER2-targeted component of therapy.

[l] The PRO questionnaires (EORTC QLQ-C30, modified QLQ-BR23) and EQ-5D will be completed by the patients at the investigational site. Assessments are to be completed on Cycle 1 Day 1 of the anthracycline treatment period; Day 1 of Cycles 1 through 5, Cycle 9, and Cycle 14 of the HER2-targeted treatment period; at treatment completion/treatment discontinuation visit; and at the 6- and 12-month follow-up visits after the end-of-treatment visit regardless of the occurrence of invasive disease recurrence. All PRO questionnaires are required to be completed by patients prior to administration of study drug and prior to any other study assessment(s) or health care provider interactions to ensure that the validity of the instrument is not compromised and to ensure that data quality meets regulatory requirements. Study personnel should review all questionnaires for completeness before the patient leaves the investigational site, and the hard copy originals of the questionnaires must be maintained as part of the patient's medical record at the site for source data verification.

[m] Hematology tests (complete blood count with differential) including hemoglobin, hematocrit, counts of WBCs, platelets, and differential including absolute neutrophils. Abnormal CBC, including platelet counts, will be followed until resolution, until they are stable as assessed by the investigator, until the patient is lost to follow-up, or until the patient withdraws consent. Only if abnormal platelets have not returned to ≥100,000/mm^3 at the time of end-of-treatment visit, it should be checked at a minimum of every four weeks (monthly) during the follow-up period until ≥100,000/mm^3, until they are clinically stable as assessed by the investigator, until the patient is lost to follow-up, or until the patient withdraws consent, regardless of the occurrence of invasive disease recurrence. Clinicians should consider more frequent evaluation (qw, etc.) for increasing grade of toxicity.

[n] Screening laboratory tests to be performed within 7 days prior to randomization. Screening laboratory assessments may be done on the day of randomization, and their results may be used for randomization visit purposes. Results must be reviewed and documented prior to administration of first dose of study treatment. (Local laboratory [hematology, biochemistry, INR, and aPTT assessments] to be used for these assessments.)

[o] Prior to the first dose of study treatment (i.e., antracycline treatment), screening tests within 7 days prior to randomization will be utilized. Protocol-specified subsequent tests will be performed within 72 hours preceding administration of study treatment; results must be reviewed and documented prior to administration of study treatment. (Local laboratory [hematology, biochemistry, INR, and aPTT assessments] to be used for these assessments.)

p For patients receiving paclitaxel, sample for hematology should be collected qw during treatment. Paclitaxel will be administered every $7 + 3$ days.

q Serum chemistry tests at baseline include sodium, potassium, chloride, glucose, BUN or urea, creatinine, TBILI (and direct bilirubin when TBILI $>$ ULN), total protein, albumin, ALT, AST, and ALP. Patients who have positive HBV or HCV serology without known active disease must meet the eligibility criteria for ALT, AST, TBILI, INR, aPTT, and ALP on at least two consecutive occasions, separated by at least 1 week, within the 30 day screening period. Assessments at each treatment and at the end-of-treatment visit include potassium, TBILI, ALT, AST, and ALP; other assessments may be obtained as clinically indicated. For patients who have positive HBV or HCV serology without known active disease, ALT, AST, TBILI, INR, aPTT, and ALP need to be assessed within 72 hours prior to the first dose on C1D1. Abnormal ALT, AST, and total and/or direct bilirubin will be re-tested at a minimum of every 4 weeks (monthly) until normalization (i.e., until they are clinically stable for at least 3 months, as assessed by the investigator), until the patient is lost to follow-up or until the patient withdraws consent, regardless of the occurrence of invasive disease recurrence. Clinicians should consider more frequent evaluation (qw, etc.) for increasing grade of toxicity. *Note: positive serology markers that indicate immunity will not be considered as clinically meaningful positive serology to trigger these tests.*

r Documentation of HBV and HCV serologies is required: This includes HBsAg and/or total HBcAb in addition to HCV antibody testing. The most recent serologic testing must have occurred no more than 3 months prior to randomization. If such testing has not been done, it must be performed during screening. For patients who are known carriers of HBV and/or HCV, active hepatitis B infection and active hepatitis C infection must be ruled out on the basis of negative serologic testing and/or determination of HBV DNA viral load/HCV RNA load per local guidelines.

s For all women of childbearing potential and for those who do not meet the definition of postmenopausal status (see definition in Section 4.5.1.2) or who have not undergone surgical sterilization, a serum β-HCG must be performed within 7 days prior to randomization. During the treatment period, in all treatment arms, a urine pregnancy test must be performed in women of childbearing potential within 72 hours prior to C1D1 of HER2-targeted therapy, then every 3 cycles thereafter during study treatment, at 3 and 6 months after the end-of-treatment visit, and as clinically indicated regardless of the occurrence of invasive disease recurrence. All positive urine pregnancy tests must be confirmed by a serum β-HCG test.

t A whole blood sample for DNA isolation will be collected at baseline or at any time (after randomization) during the conduct of the clinical study.

u Collection of serum and plasma samples for biomarker analysis is mandatory. After the end-of-treatment visit, these samples will be collected once a year for 5 years, whenever possible, unless a recurrence that is defined as an IDFS event in Section 3.4.1 has occurred. Biomarker samples (or processed samples) will be stored for up to 5 years after final database lock unless the patient consents to long-term storage for 15 years.

v If the patient gives OBRP consent, the serum and plasma samples for biomarker analysis will undergo long-term storage in the study biosample repository for use in future exploratory biomarker analyses. For patients that consented to the OBRP, a whole blood sample for DNA analysis will be collected as well at baseline if possible or any time after randomization.

w AE and SAEs will be recorded from the start of study screening procedures if related to protocol-mandated intervention. All non-serious AEs occurring prior to Day 1 (administration of study treatment) will be reported in the medical history, unless AE reporting is deemed more appropriate. AEs are to be monitored continuously during study treatment. All AEs occurring during the study and until the end-of-treatment/treatment discontinuation visit 28 days after the last dose of study treatment are to be recorded; thereafter, only drug-related SAEs and AEs/SAEs that qualify for long-term reporting should continue to be collected. The investigator should notify the Sponsor of any death, SAE, or other AE of concern occurring at any time after a patient has discontinued study treatment or study participation if the event is believed to be related to prior study drug treatment or study procedures.

x Concomitant medication will be recorded in the interval beginning 7 days prior to the patient being randomized into the study until the end of the treatment period, except for medications that are also collected during follow-up (Section 4.6).

Appendix 2
Schedule of Assessments (Cont.)

[y] In cases of invasive disease recurrence (an IDFS event as defined in Section 3.4.1) diagnosed at any time during the study, patients will be out of the study schedule for disease status assessments and will be followed annually $(\pm 28\ days)$ starting 1 year after first recurrence until 10 years after FPI for survival, anti-cancer medications, and new relapse (recurrence or disease progression) events. If a patient has a recurrence, any anti-cancer medication given after the date of diagnosis must be recorded on the post-treatment anti-cancer medication page.