Statistical Analysis Plan Version 3 I3Y-MC-JPBZ

A Phase 2, Randomized, Multicenter, 3-Arm, Open-Label Study to Compare the Efficacy of Abemaciclib plus Trastuzumab with or without Fulvestrant to Standard-of-Care Chemotherapy of Physician's Choice plus Trastuzumab in Women with HR+, HER2+ Locally Advanced or Metastatic Breast Cancer

NCT02675231

Approval Date: 8-Feb-2019

1. Statistical Analysis Plan: I3Y-MC-JPBZ (monarcHER): A Phase 2, Randomized, Multicenter, 3-Arm, Open-Label Study to Compare the Efficacy of Abemaciclib plus Trastuzumab with or without Fulvestrant to Standard-of-Care Chemotherapy of Physician's Choice plus Trastuzumab in Women with HR+, HER2+ Locally Advanced or Metastatic Breast Cancer

Abemaciclib (LY2835219) Breast cancer

This is a randomized, multicenter, 3-arm, open-label Phase 2 study to compare the efficacy of abemacicl b plus trastuzumab plus fulvestrant and abemaciclib plus trastuzumab to standard-of-care single-agent chemotherapy plus trastuzumab in women with HR+, HER2+ locally advanced or metastatic breast cancer.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I3Y-MC-JPBZ
Phase 2

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 08-Feb-2019 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

Statistical analysis plan (SAP) Version 1 was approved prior to the first visit when a subject received study drug or any other protocol intervention.

In SAP Version 2, SAP Version 1 was amended to update the SAP due to protocol amendment JPBZ(b).

SAP was amended to Version 3 due to protocol amendment JPBZ(c). Clarification was added to the censoring rules for new anticancer treatment.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to compare the efficacy of abemaciclib plus trastuzumab plus fulvestrant and abemaciclib plus trastuzumab to standard-of-care (SOC) single-agent chemotherapy of physician's choice plus trastuzumab with respect to progression-free survival (PFS).

4.2. Secondary Objectives

The secondary objectives of the study are to compare the 3 arms with respect to each of the following:

- overall survival (OS) rate at 1, 2, and 3 years
- objective response rate (ORR)
- duration of response (DoR) (complete response [CR] + partial response [PR])
- disease control rate (DCR) (CR + PR + stable disease [SD])
- clinical benefit rate (CBR) (CR + PR + SD ≥6 months)
- safety and tolerability of abemaciclib in combination with trastuzumab and fulvestrant
- impact on pain, disease symptoms, and overall quality of life using the modified Brief Pain Inventory-Short Form (mBPI-sf), the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30), and the health status scores from the EuroQol 5-Dimension 5 Level (EQ-5D 5L)
- pharmacokinetics (PK) of abemaciclib and its metabolites, fulvestrant, and trastuzumab in the target patient population
- relationship between abemaciclib, trastuzumab, and fulvestrant exposure and response for safety and efficacy endpoints

4.3. Exploratory Objectives

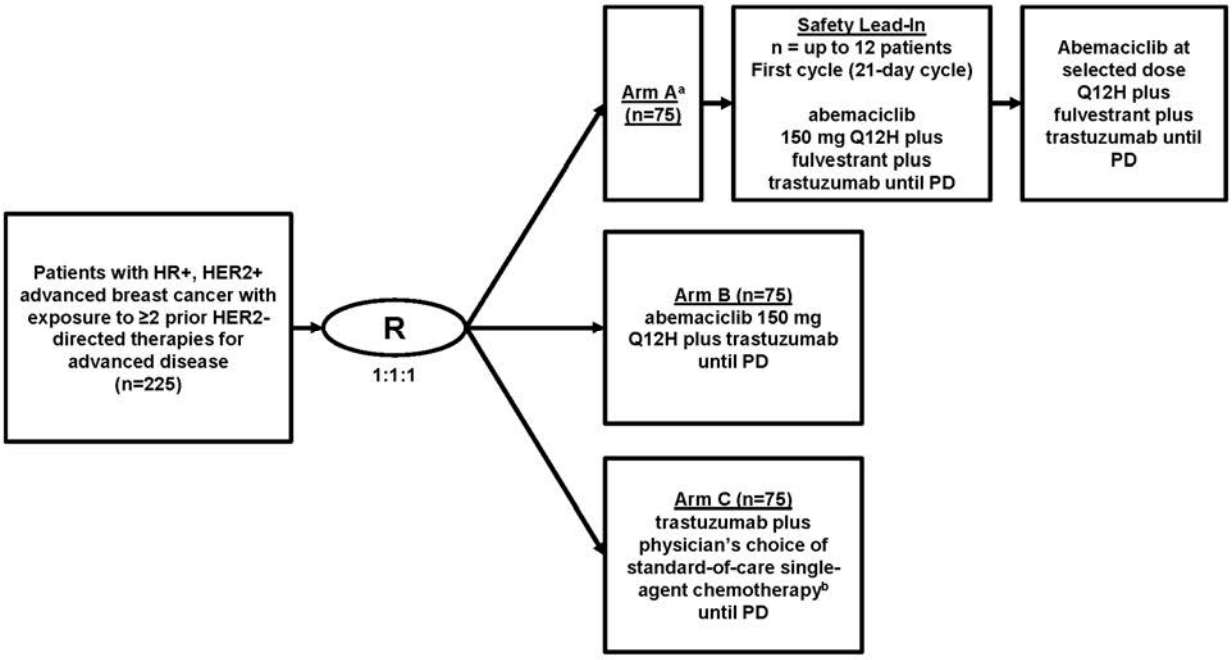
- to explore potential biomarkers related to the mechanism of action of abemaciclib, the cell cycle, and/or the pathogenesis of breast cancer and their association with clinical outcome

5. Study Design

5.1. Summary of Study Design

Study I3Y-MC-JPBZ (JPBZ) is a randomized, multicenter, 3-arm, open-label Phase 2 study to compare the efficacy of abemaciclib plus trastuzumab plus fulvestrant and abemaciclib plus trastuzumab to SOC single-agent chemotherapy of physician's choice plus trastuzumab in women with hormone receptor-positive (HR+), human epidermal growth factor receptor 2-positive (HER2+) locally advanced or metastatic breast cancer.

Figure JPBZ.5.1 illustrates the study design.



Abbreviations: HER2+ = human epidermal growth factor receptor 2-positive; HER2- = human epidermal growth factor receptor 2-negative; HR+ = hormone receptor-positive; n = number; PD = progressive disease; Q12H = every 12 hours; R = randomization.

[a] If the dose of abemaciclib in Arm A is reduced to 100 mg, patients that initiated treatment at the 150-mg dose will be dose-reduced to 100 mg, and patients who started at the 150-mg abemaciclib dose will be replaced so that there are a total of 75 patients in Arm A with a starting dose of 100-mg abemaciclib.

[b] Standard-of-care single-agent chemotherapy should include an approved drug in breast cancer.

Figure JPBZ.5.1. Illustration of study design.

Approximately 225 patients will be randomized 1:1:1 across the 3 arms:

- Arm A: Abemaciclib + trastuzumab + fulvestrant
- Arm B: Abemaciclib + trastuzumab

- Arm C: Standard-of-care single-agent chemotherapy + trastuzumab

Patients will be randomized using the following stratification factors: number of previous systemic regimens (excluding single-agent endocrine therapy) for advanced breast cancer (2 to 3 vs. more than 3) and the status of disease (measurable vs. nonmeasurable).

Terms used to describe the periods during the study are defined below:

- **Baseline:** begins when the informed consent form is signed and ends at the first study treatment dose (or at discontinuation, if no treatment is given). This may be up to 28 days prior to the first study treatment dose.
- **Study Period:** begins at the first study treatment dose and ends at study completion. The study period does not include the continued access period.
    - **Study Treatment Period:** begins at the first study treatment dose and ends when the patient and the investigator agree that the patient will no longer continue study treatment. This date is to be reported on the electronic case report form (eCRF) as the Date of Discontinuation from study treatment.
- **Postdiscontinuation Follow-Up:** begins the day after the patient and the investigator agree that the patient will no longer continue study treatment.
    - ***Short-term follow-up*** begins the day after the patient and the investigator agree that the patient will no longer continue study treatment and lasts approximately 30 days.
    - ***Long-term follow-up*** begins the day after short-term follow-up is completed and continues until the patient's death or overall study completion.
- **Continued Access Period:** begins after study completion and ends at the end of trial. During the continued access period, patients on study treatment who continue to experience clinical benefit may continue to receive study treatment until one of the criteria for discontinuation is met.
    - The continued access period includes continued access period short-term follow-up.
    - Continued access short-term follow-up: begins the day after the patient and the investigator agree that the patient will no longer continue treatment in the continued access period and lasts approximately 30 days.

5.2. Determination of Sample Size

The study will enroll approximately 225 patients in [1:1:1] randomization (75 patients in abemaciclib plus trastuzumab plus fulvestrant; 75 patients in abemaciclib plus trastuzumab; and 75 patients in SOC single-agent chemotherapy of physician's choice plus trastuzumab). The primary analysis will be performed after approximately 165 PFS events have occurred. Assuming a hazard ratio (HR) of 0.667, this sample size yields at least 80% statistical power to detect superiority of the abemaciclib plus trastuzumab plus fulvestrant and abemaciclib plus trastuzumab arms over SOC single-agent chemotherapy of physician's choice plus trastuzumab arm with the use of an experiment-wise 1-sided alpha level of 0.10.

If the true median PFS for the SOC single-agent chemotherapy of physician's choice plus trastuzumab arm is 4 months, then the HR of 0.667 amounts to an approximately 2-month improvement in median PFS for the abemaciclib plus trastuzumab plus fulvestrant and abemaciclib plus trastuzumab arms under an additional assumption of exponential survival distribution. Assuming a censoring rate of approximately 30%, the study will randomize approximately 225 patients.

5.3. Method of Assignment to Treatment

Patients who meet all criteria for enrollment will be randomly assigned to receive abemaciclib plus trastuzumab plus fulvestrant, abemaciclib plus trastuzumab, or SOC single-agent chemotherapy of physician's choice plus trastuzumab at Visit 1. Randomization will be stratified by the number of previous systemic regimens (excluding single-agent endocrine therapy) for advanced breast cancer (2 to 3 vs. more than 3) and status of disease (measurable vs. nonmeasurable).

Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS).

6. A Priori Statistical Methods

6.1. General Considerations

6.1.1. Populations

Statistical analysis of this study will be the responsibility of Lilly.

Disposition summaries will be performed on the **entered** population, defined as all patients who sign the informed consent document.

Efficacy analyses and patient characteristic summaries will be based on the **enrolled** or **intention-to-treat (ITT)** analysis set. This population is defined as all patients randomized to study treatment. Patients will be grouped according to randomized treatment.

Safety analyses and exposure summaries will be based on the **safety** population, defined as all enrolled patients receiving at least 1 dose of any study drug. Patients will be grouped according to treatment received in Cycle 1.

Pharmacodynamic and/or tailoring biomarker analyses will be based on the subset of patients from the above populations from whom a valid assay result (according to laboratory guideline) has been obtained.

6.1.2. Definitions and Conventions

Study drug refers to abemaciclib (LY2835219).

Study treatment refers to abemaciclib plus trastuzumab plus fulvestrant or abemaciclib plus trastuzumab or SOC single-agent chemotherapy of physician's choice plus trastuzumab.

The **date of randomization** is the date the patient was randomly assigned to study treatment using the IWRS.

The **date of first dose** is the date of the first dose of study drug, trastuzumab, fulvestrant, or single-agent chemotherapy.

The **baseline value of a safety assessment** is the last value observed prior to the first dose of study drug, trastuzumab, fulvestrant, or single-agent chemotherapy.

The **baseline value of an efficacy assessment** is the last value observed prior to the date of randomization. If a patient's first assessment occurs after randomization but prior to the first dose, this assessment will be used as the baseline.

The **study day of a safety event or assessment** will be calculated as:

- the difference between the date of the event or assessment and the date of first dose plus 1 for all events or assessments occurring on or after the day of first dose. For example, if an event occurs on 08MAR2016 and the date of first dose was 06MAR2016, the study day of the event is 3.

- the difference between the date of the event or assessment and the date of first dose for all events or assessments occurring before the day of first dose. For example, if an event occurs on 05MAR2016 and the date of first dose was 06MAR2016, the study day of the event is -1.

The **study day of an efficacy event or assessment** will be calculated as:

- the difference between the date of the event or assessment and the date of randomization plus 1 for all events or assessments occurring on or after the date of randomization.
- the difference between the date of the event or assessment and the date of randomization for all events or assessments occurring before the date of randomization.

One **month** is defined as 365/12 days.

Unless otherwise noted, **summaries of continuous variables** will include a mean, median, standard deviation, minimum, and maximum.

Unless otherwise noted, **summaries of categorical variables** will include the frequency and percentage (relative to the population being analyzed) of each category.

The assumptions for each statistical method will be evaluated. If there is violation of assumptions, alternative statistical methods may be used.

Additional exploratory analyses of the data will be conducted as deemed appropriate.

6.2. Handling of Dropouts or Missing Data

With the exception of dates, missing data will not be imputed. The method of imputation for any dates that are imputed is described in the relevant section. Rules for determining date of progression or censor for PFS are defined in Table JPBZ.6.1.

6.3. Multiple Comparisons/Multiplicity

Type I error for PFS will be controlled by testing the treatment arms sequentially. The abemaciclib triplet arm will be compared to the single-agent chemotherapy plus trastuzumab arm first, and the abemaciclib doublet arm will be compared to the single-agent chemotherapy plus trastuzumab arm only if the test for the triplet is significant. OS will be tested only if both tests of PFS are significant.

6.4. Patient Disposition

A detailed description of patient disposition will be provided. It will include a summary of the number and percentage of patients entered into the study, rescreened after screen failure, enrolled in the study, and treated in the study, reasons for discontinuation from study treatment (safety population only), and reasons for discontinuation from study (ITT population only). Reason for discontinuation from both study treatment and the study will be summarized by pre-determined categories.

6.5. Patient Characteristics

6.5.1. Demographics and Performance Status

Patient demographics will be summarized. Patient demographics will include age, race, ethnicity, country, height, weight, and body mass index.

6.5.2. Baseline Disease Characteristics

Disease characteristics will be summarized. Disease characteristics will include the following:

- study entry diagnosis
- disease extent at study entry (metastatic disease versus locally advanced recurrent disease)
- nature of disease (visceral metastases or other)
- measurable disease (yes versus no)
- number of organs involved (1, 2, or 3+)
- metastatic site(s)
- estrogen receptor status
- progesterone receptor status
- baseline Eastern Cooperative Oncology Group (ECOG) performance status (PS)

Nature of disease will be reported directly from the 'Nature of Disease' eCRF. Disease measurability and number of organs involved will be derived from the 'Target Tumor Identification and Results' and 'Non-Target Tumor Identification and Results' eCRFs at baseline. All patients with at least 1 lesion on the target lesion form will be counted as having measurable disease. The number of organs involved will be derived from the location codes of the target and non-target lesions.

6.5.3. Historical Illnesses

Historical illnesses are clinically relevant events in the past that ended before the screening visit. Historical illnesses (using Preferred Term(s) [PTs] from the most current version of the Medical Dictionary for Regulatory Activities [MedDRA®]) will be summarized.

6.5.4. Prior Therapies

Prior radiotherapy, surgery, and systemic therapy will be summarized. Prior radiotherapy and surgery will be categorized by reason for regimen. Prior systemic therapies will be categorized by type of regimen (endocrine therapy, chemotherapy, HER2 directed therapy, or other) and specific therapy within reason for regimen (neoadjuvant, adjuvant, locally advanced/metastatic, or any). Frequency of each specific therapy will be tabulated within each type of therapy and per treatment setting (neoadjuvant, adjuvant, locally advanced/metastatic, or any). Percentage of

patients with prior taxane use and percentage of patients with prior anthracycline use will be summarized. Time to progression on prior-anti HER2 therapies (that is, duration between progression and the start of prior anti-HER2 therapies) will also be summarized for the patients in advanced disease.

6.5.5. Post-Study Treatment Discontinuation Therapies

Therapies received following study treatment discontinuation will be summarized by arm. Therapies will be summarized overall and by category: endocrine therapy or targeted/chemotherapy.

6.6. Treatment Compliance

Treatment compliance information for abemaciclib will be collected through pill counts at each cycle. Compliance will be calculated as the ratio of total dose taken to the total assigned dose (minus any dose adjustments and doses omitted/withheld for medical or logistical reasons). Dosing information for trastuzumab, fulvestrant, and single-agent chemotherapy of physician's choice will be collected at each cycle/visit.

6.7. Concomitant Therapy

All medications will be coded to the generic preferred name according to the current World Health Organization drug dictionary. All concomitant medications will be summarized by number and percentage of patients for the safety population using the base name (without esters or salts).

6.8. Efficacy Analyses

Unless otherwise noted, all efficacy analyses will be performed on the ITT population.

The stratification factors for the analysis of primary and secondary analyses are the number of previous regimens for advanced breast cancer (2 to 3 vs. more than 3) and status of disease (measurable vs. nonmeasurable). The stratification factors will be derived based on CRF data.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.20, unless otherwise stated. All confidence intervals (CIs) will be given at a 2-sided 95% level, unless otherwise stated. For regulatory purposes, tests may be performed at a 1-sided alpha level of 0.025.

6.8.1. Primary Outcome: Progression-Free Survival

6.8.1.1. Definition

The primary endpoint of this study is PFS. PFS time is measured from the date of randomization to the date of investigator-determined objective progression as defined by Response Evaluation Criteria In Solid Tumors Version 1.1 (RECIST v1.1), or death from any cause. Patients who have neither progressed nor died will be censored at the day of their last radiographic tumor assessment (if available) or date of randomization if no post-initiation (that is, post baseline) radiographic assessment is available. The detailed censoring rules are described in Table JPBZ.6.1.

Table JPBZ.6.1. Rules for Determining Date of Progression or Censor for Progression-Free Survival

Situation	Date of Event or Censor	Event / Censor
Tumor progression or death	Earliest date of PD or death	Event
No tumor progression and no death	Date of last adequate radiological assessment or date of randomization (whichever is later)	Censored
Unless		
No baseline radiological tumor assessment available	Date of randomization	Censored
No adequate postbaseline radiological tumor assessment available and death reported after 2 scan intervals following randomization	Date of randomization	Censored
New anticancer treatment started (excluding maintenance endocrine therapy following chemotherapy) and no tumor progression or death within 14 days	Date of adequate radiological assessment prior to start of new therapy +14 days or date of randomization (whichever is later)	Censored
Tumor progression or death documented immediately after 2 or more scan intervals following last adequate radiological tumor assessment or randomization (whichever is later)	Date of last adequate radiological assessment or date of randomization (whichever is later)	Censored

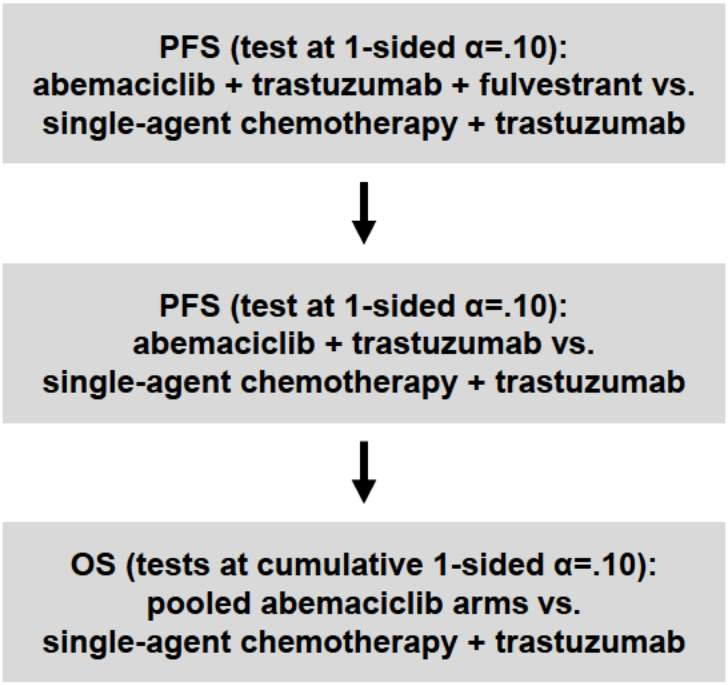
Abbreviation: PD = progressive disease.

Note: Progression-free survival and associated outcome is determined by the earliest of the dates above, if more than 1 situation applies.

6.8.1.2. Primary Analysis

The PFS analysis to test the superiority of abemaciclib plus trastuzumab and abemaciclib plus trastuzumab plus fulvestrant to chemotherapy plus trastuzumab in improving PFS time will be performed on the ITT population and will use the log-rank test stratified by number of previous regimens (excluding single-agent endocrine therapy) for advanced breast cancer and the status of disease (measurable vs. nonmeasurable).

There is 1 planned primary analysis for PFS in this study, which will be performed after approximately 165 events have been observed in the ITT population based on investigator assessment. The primary PFS analysis will compare each arm against the control using a log-rank test stratified by the randomization factors. The primary objective of PFS will be tested at an experiment-wise 1-sided alpha level of 0.10. The abemaciclib treatment arms will be tested sequentially against the trastuzumab plus chemotherapy arm, with the abemaciclib plus trastuzumab plus fulvestrant arm tested first. A comparison of the 2 abemaciclib arms will be considered exploratory. See Figure JPBZ.6.1 for a depiction of alpha spending.



Abbreviations: OS = overall survival; PFS = progression-free survival; vs. = versus.

Figure JPBZ.6.1 Alpha spending for Study I3Y-MC-JPBZ.

6.8.1.3. Other PFS Analyses

6.8.1.3.1. Progression-Free Survival Curves and Hazard Ratio

The Kaplan-Meier method (Kaplan and Meier 1958) will be used to estimate the PFS curves as well as PFS rates at every 3 months for each treatment group. These rates will be compared based on a normal approximation for the difference between the rates.

A Cox proportional hazard model (Cox 1972) stratified by number of previous regimens (excluding single-agent endocrine therapy) for advanced breast cancer and the status of disease (measurable vs. nonmeasurable) with treatment as a factor will be used to estimate the HR and corresponding CI with Wald p-value (Agresti 2002). An additional unstratified Cox regression model will be employed to explore the effects of prognostic variables, such as of the stratification variable and intrinsic/extrinsic factors, on treatment response.

6.8.1.3.2. Restricted Mean Difference

The common method for describing benefit on the time scale is to calculate the difference in median event time between the 2 treatment arms. An alternative method for describing benefit on the time scale is to estimate the average difference between the Kaplan-Meier (KM) curves. This corresponds to calculating the difference in the average time to event for the 2 treatment arms (Irwin 1949; Karrison 1997; Meier et al. 2004). Similar to the HR, this method uses all of the available information across the KM curves, but has the additional advantage of assessing benefit on the time scale.

To estimate an improvement in PFS with abemaciclib, we will follow the method of Irwin (1949) detailed in Karrison (1997) and Meier (2004) for estimating the 'difference in average PFS', which we will refer to more formally as the restricted mean difference in PFS.

The area under each KM curve will be calculated using numerical integration (trapezium rule) per Karrison and implemented in SAS using PROC LIFETEST. The difference between treatment arms and a 95% CI for the difference will be formed.

Since the KM curve may be ill-determined beyond a certain range, or even undefined (if the longest observation is censored), for evaluation and comparison of means, the area under each KM curve will be calculated between time 0 and restriction time T, which is why this is referred to as a "restricted mean". Following the suggestion of Karrison, the restriction time T will be chosen as largest time point t such that the standard error (SE) of the survival estimate at time t in each treatment group is no more than 0.075. For this purpose, we will use the simple, albeit conservative, formula proposed by Peto et al. (1977) for calculating the SE of S(t) as $SE(S(t)) = S(t)\sqrt{(1-S(t))/n(t)}$, where n(t) is the number of patients still at risk at time t.

6.8.2. Gated Secondary Efficacy Endpoint: Overall Survival

Overall survival (OS) is an important secondary endpoint for this study. The OS time is measured from the date of randomization to the date of death from any cause.

OS will be tested only if PFS is significant for both abemaciclib arms. OS will be tested by pooling the 2 abemaciclib arms and comparing them against the SOC systemic therapy arm. Up to a total of 2 interim analyses and a final analysis for OS may be performed in this study. The type I experiment-wise error rate will be controlled at 10% by using the Lan-Demets method with the following O'Brien-Fleming like alpha-spending function:

$$\alpha^*(t_k) = 2\left(1-\Phi\left(\Phi^{-1}(1-\alpha/2)\Big/\sqrt{t_k}\right)\right)$$

Here, t_k is the information fraction at time k, Φ is the standard normal cumulative distribution function, and Φ^{-1} is the standard normal quantile function.

The actual alpha spent will be calculated based on the actual number of events observed at the time of analysis using software that implements the alpha-spending function noted above (for example, ADDPLAN 6.0 or SAS 9.2).

To maintain the experiment-wise type I error rate, OS will be hierarchically tested in the following way: only if the test of PFS is significant will OS also be tested inferentially for significance (Glimm et al. 2010); specifically:

- The first potential time point for OS analysis will be at the time of the primary PFS analysis. If PFS is significant at this stage, the first interim analysis of OS will also be performed. If OS is not significant at this stage, the second interim analysis of OS will be performed after approximately 105 deaths have been observed in the ITT population

(information fraction of 0.667). If OS is not significant at this stage, a final analysis will be performed after approximately 158 deaths have been recorded in the ITT population.

- If the primary analysis for PFS is not significant, OS will not be statistically evaluated.

The OS analysis to test the superiority of abemaciclib plus trastuzumab and abemaciclib plus trastuzumab plus fulvestrant to chemotherapy plus trastuzumab in improving OS time will use the log-rank test stratified by the number of previous regimens (excluding single-agent endocrine therapy) for advanced breast cancer and the status of disease (measurable vs. nonmeasurable).

The following additional analyses will be conducted for OS:

- Kaplan-Meier curves (Kaplan and Meier 1958) will be generated; medians, quartiles, and appropriate point probabilities will be calculated. Interval estimates will be calculated.
- The Cox regression stratified by the randomization factors will be used to estimate the HR between the 2 treatment groups, along with CI.

In addition, OS rate at 1 year in each treatment arm will be calculated by determining OS time for each patient and using KM techniques to assess OS time for each treatment arm. The KM estimate of the OS rate at 1 year will be used to compare treatment arms using a standard normal test of the difference in OS rate at 1 year. The same techniques will be used to calculate and compare OS rates at 2 years and 3 years between arms.

6.8.3. Other Secondary Efficacy Analyses

6.8.3.1. Objective Response Rate, Disease Control Rate, and Clinical Benefit Rate

Objective response rate, DCR, and CBR are summary measures of best overall response (BOR) as defined by RECIST v1.1. BOR is derived from time point responses. All time point responses observed while on study treatment and during the short-term follow-up period (but before the initiation of postdiscontinuation therapy) will be included in the derivation. The one exception includes patients who receive surgery and/or radiotherapy for locally advanced breast cancer. For these patients, only those time point responses occurring prior to surgery/radiotherapy will be included in the derivation. The same rule will be applied to palliative radiation and palliative surgery.

Each patient's BOR will be categorized as CR, PR, SD, progressive disease (PD), or not evaluable (NE). For patients with bone-only nonmeasurable disease (see Section 6.5.2), BOR will be limited to CR, SD, PD, and NE. Patients with SD will be further classified as SD ≥6 months or SD <6 months. Stable disease ≥6 months includes all patients with a best response of SD and a PFS time of ≥6 months. A BOR of CR or PR will not require confirmation.

Objective response rate is the proportion of patients with a BOR of CR or PR. Clinical benefit rate is the proportion of patients with a BOR of CR or PR, or SD ≥6 months. Disease control rate is the proportion of patients with a BOR of CR, PR, or SD. Patients with bone-only

nonmeasurable disease cannot have a best response of PR, thus ORR will be reported for both the ITT population and the subset of patients with measurable disease.

For each of these rates, point estimates and CIs (using the normal approximation to the binomial) will be calculated by treatment arm. Stratified tests comparing these rates between treatment arms will be conducted using a Cochran-Mantel-Haenszel test using randomization factors as strata.

6.8.3.2. Duration of Response

The DoR time is defined only for responders (patients with a BOR of CR or PR). It is measured from the date of first evidence of CR or PR to the date of objective progression or the date of death due to any cause, whichever is earlier. It is calculated as date of progression or death – date of first response evaluation of CR or PR + 1. The DoR will be censored according to the same rules as PFS.

A KM analysis of DoR will be performed to estimate the DoR curve for each arm. Point estimates and CIs for DoR quartiles and DoR rates will be calculated every 3 months for the first 12 months.

6.8.4. Sensitivity Analyses

Sensitivity analyses will be undertaken for calculation of the primary endpoint in order to evaluate the robustness of the analysis. The following sensitivity analyses will be performed for PFS:

Progression-Free Survival Sensitivity Analysis 1 (patients with centrally confirmed HER2+ disease): Progression-free survival will be analyzed including only patients with centrally confirmed HER2+ disease.

Progression-Free Survival Sensitivity Analysis 2 (nonobjective progression as a PFS event): if a patient is discontinued from study treatment due to investigator-determined non-objective progression (for example, symptomatic deterioration), then the patient's PFS time will be calculated using the date of non-objective progression as the progression date.

6.8.4.1. Overall Survival

One sensitivity analysis on OS is planned. Overall survival time for this analysis is defined as the time from the date of study enrollment to the date of death due to study disease. Survival time will be censored on the date the patient was last known to be alive for patients who have no reported event. For patients that have died due to reasons not related to disease, survival time will be censored at the date of death.

6.9. Health Outcomes/Quality-of-Life Analyses

6.9.1. Instruments

Patient-reported outcomes are measured through paper versions of the following:

- mBPI-sf

- EORTC QLQ-C30
- EQ 5D 5L

6.9.2. Pain Intensity and Pain Assessment

Pain as assessed with the mBPI-sf questionnaire (that is, worst, least, average, and current pain over the last 24 hours) will be characterized with descriptive non-inferential statistics. Treatment arm comparisons will be conducted through use of mixed model repeated measures (MMRM) specified with an unstructured covariance matrix and estimated with restricted maximum likelihood (REML) to address the small sample negative bias in the variance component when using the maximum likelihood (ML) estimator (that is, the ML estimator usually has lower mean-squared error [MSE] than the REML estimator).

Corresponding analyses will be conducted for each of the 7 individual pain interference with function items (mBPI-sf items 5a-5g). Additionally, a composite score will be calculated if a majority of items are completed (that is, at least 4 out of 7 pain interference with function items must have a valid recorded response).

6.9.3. Quality of Life

Data from the EORTC QLQ-C30 instrument will be scored as described by Aaronson and colleagues (Aaronson et al. 1993).

Treatment arm comparisons will be conducted through use of MMRM specified with an unstructured covariance matrix and estimated with REML to address the small sample negative bias in the variance component when using the ML estimator (that is, the ML estimator usually has lower MSE than the REML estimator).

6.9.4. Health State Utility

The EQ-5D 5L data will be scored as described by van Hout and colleagues (van Hout et al. 2012). The index score is calculated from a set of item weights to derive a score of 0 to 1, with 1 representing the best health status. Geographic-specific weights will be used as appropriate and when available. The Visual Analog Scale (VAS) is scored from 0 (*worst imaginable health state*) through 100 (*best imaginable health state*) to represent the patient's self-reported health state "today". The EQ-5D 5L responses for each item will be summarized by frequency and corresponding percentages. Descriptive statistics for the index and VAS will be calculated by arm.

For both the index score and VAS, treatment arm comparisons will be conducted through use of MMRM specified with an unstructured covariance matrix and estimated with REML to address the small sample negative bias in the variance component when using the ML estimator (that is, the ML estimator usually has lower MSE than the REML estimator).

6.9.5. Exploratory Analyses of Health Outcomes/Quality-of-Life

Following exploratory analyses will be done for each of the 3 scales, mBPI-sf, EORTC QLQ-C30, and EQ 5D 5L.

For each of these 3 scales, summaries of patient-reported outcomes will also be calculated as part of a strategy to demonstrate postdiscontinuation treatment effect —an issue that continues to be discussed in both clinical and health technology assessment evaluations and which has a demonstrable impact on the results derived from the economic model. To deliver this information, we will conduct the following:

1.a. For each patient with a baseline and at least 1 on-therapy observation, we will calculate the difference between the baseline assessment and the last on-study assessment.

1.b. For each patient with a baseline and an 801 visit, we will also calculate the difference between the baseline assessment and the 801 assessment.

1.c. Difference scores will be calculated between 1.a. and 1.b. mean difference scores (that is, differences between baseline→last on-study vs. baseline→801 assessments) for patients with data from those assessments.

Mean differences will be calculated and treatment difference will be compared using a t-test, and effect sizes will be produced to aid interpretation.

2.a. The duration in days from the last on-study assessment and the 801 assessment will be summarized.

6.9.6. Utilization

Utilization data will be summarized by category by arm. The following categories will be described:

- analgesics (at baseline, on study treatment, and during short-term follow-up)
- anti-diarrheal therapy (at baseline, on study treatment, and during short-term follow-up)
- transfusions (on study treatment and during short-term follow-up)
- surgery (on study treatment and during short-term follow-up)
- hospitalizations (on study treatment and during short-term follow-up)
- postdiscontinuation radiotherapy and systemic therapy

For categorical variables, frequency and the corresponding proportions will be calculated and tests for differences in proportion between groups will be performed using a Fisher's Exact test. Continuous variables will be described by the mean, median, and standard deviation. A t-test will be used to compare mean utilization.

6.10. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic and pharmacodynamic analyses will be performed according to a separate PK analysis plan.

6.11. Safety Analyses

6.11.1. Extent of Exposure

Drug exposure, dose intensity, and drug adjustment (dose omissions, reductions, and delays) for abemaciclib, trastuzumab, fulvestrant, and SOC single-agent systemic therapy of physician's choice will be summarized for all treated patients per treatment arm. Drug exposure will include summaries of cycles received per patient, duration on therapy, and cumulative dose. Dose intensity will be calculated as the actual cumulative amount of drug taken divided by the duration of treatment. Relative dose intensity will be calculated as the actual amount of drug taken divided by the amount of drug prescribed times 100% (expressed as a percentage). The summary of dose adjustments, omissions (oral drugs), and delays (intravenously or intramuscularly administered drugs) will include the reason for adjustment, omission, or delay.

For abemaciclib, extent of exposure will be measured by pill counts. Dose intensity will be expressed in mg/day. The assigned cumulative dose while on study is 2 × 150 mg × number of days on treatment.

For trastuzumab, extent of exposure and assigned cumulative dose while on study will be measured based on data reported on the infused drug form.

For fulvestrant, extent of exposure will be measuring using the fulvestrant administration eCRF. The assigned cumulative dose while on study is 500 mg + [500 mg × ceiling (number of days on treatment/28)].

For SOC single-agent systemic therapy of physician's choice, extent of exposure and assigned cumulative dose while on study will be measured based on data reported on the eCRF appropriate for the route of administration.

For Arm C, SOC physician choice will be summarized by drug. The number of patients treated with maintenance endocrine therapy will be summarized as well.

6.11.2. Adverse Events

Adverse event (AE) terms and severity grades will be assigned by the investigator using Common Terminology Criteria for Adverse Events (CTCAE) Version 4. In addition, AE verbatim text will also be mapped by the sponsor or designee to corresponding terminology within MedDRA. Adverse events will be reported using the following reporting process:

- In CTCAE Version 4, each CTCAE term is a MedDRA lower level term (LLT), except in the case where the CTCAE term is a MedDRA system organ class (SOC) followed by 'Other – specify'.
- The CTCAE Version 4 term reported by the investigator will be mapped to the MedDRA PT and SOC of the corresponding MedDRA LLT, unless the reported CTCAE term is 'Other – specify'.
- If the reported CTCAE term is 'Other – specify', the MedDRA LLT, PT, and SOC mapped from the verbatim AE term will be used.

- All listings and summaries will use the PT resulting from this process.

Preexisting conditions are defined as AEs that begin prior to the first dose of study drug.

A treatment-emergent adverse event (TEAE) is defined as any AE that begins on or after the day of first dose or any preexisting condition that increases in CTCAE grade on or after the day of first dose. Comparisons of preexisting conditions to on-treatment events at the LLT level will be used in the treatment-emergent computation.

A serious adverse event (SAE) is any AE during this study that results in one of the following outcomes:

- death
- initial or prolonged inpatient hospitalization
- a life-threatening experience (that is, immediate risk of dying)
- persistent or significant disability/incapacity
- congenital anomaly/birth defect
- considered significant by the investigator for any other reason

Important medical events that may not result in death, be life-threatening, or require hospitalization may be considered SAEs when, based on appropriate medical judgment, they may jeopardize the patient and may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

The following summaries and listings will be produced:

- Overview of adverse events
- Summary of TEAEs by PT (any grade and Grade ≥ 3)
- Summary of TEAEs by SOC and PT (any grade and Grade ≥ 3)
- Summary of TEAEs by SOC and PT and maximum grade (1-5)
- Summary of SAEs by SOC and PT (any grade and Grade ≥ 3)
- Summary of AEs as reason for study treatment discontinuation by SOC and PT
- Listing of SAEs

The TEAE and SAE summaries will be produced for all TEAEs/SAEs and repeated for TEAEs/SAEs related to study treatment, where relationship of the AE to the study treatment will be assessed by the investigator (yes or no). Fisher's Exact test may be used for comparing AEs between treatment arms.

6.11.3. Deaths, Other Serious Adverse Events

A summary of all deaths, including reasons for deaths, will be provided. All deaths, deaths on therapy, deaths within 30 days of discontinuation of study therapy, deaths on therapy or within

30 days of discontinuation of study therapy, and deaths after 30 days of discontinuation of study therapy will be summarized by reason for death. For deaths due to AE, the preferred term will be provided. In addition to the tabular summary, a by-patient listing of all deaths on study not attributed to study disease by the investigator will be provided.

6.11.4. Clinical Laboratory Evaluation

All relevant hematology and chemistry laboratory values will be graded according to CTCAE Version 4. These calculated grades will be summarized by cycle and maximum postbaseline grade over the entire study for each treatment arm. Treatment-emergent changes will be summarized by the maximum postbaseline grade, and a shift table of baseline grade by maximum postbaseline grade will be produced.

6.11.5. Vital Signs and Other Physical Findings

Temperature, blood pressure, pulse rate, respiration rate, weight, and ECOG PS will be summarized by cycle.

6.11.6. Electrocardiograms

Local electrocardiograms (ECGs) will be summarized by planned cycle and overall. The summary by cycle will classify patients as having normal or abnormal ECG and summarize AEs identified by ECG within each cycle. The overall summary will classify patients as having an abnormal ECG at any point and summarize all AEs identified by ECG.

Left ventricular ejection fraction percentage as recorded on the eCRF (based on echocardiogram or multiple gated acquisition scan) will be summarized by planned cycle for each treatment arm.

6.12. Subgroup Analyses

Subgroup analyses of PFS and OS will be performed for each of following potential prognostic subgroup variables:

- number of previous regimens for advanced breast cancer (2 to 3 vs. more than 3)
- status of disease (measurable vs. nonmeasurable)
- disease setting (metastatic versus locally advanced recurrent)
- nature of disease (visceral or other)
- measurable disease at baseline (yes versus no)
- number of organs involved (1 versus 2 versus 3+)
- age (<65 years versus ≥65 years)
- region (North America, Europe, Asia, and Other)
- race (Caucasian, Asian, and Other)
- progesterone receptor status (positive versus negative)
- baseline ECOG PS (0 versus 1)

- number of prior HER2 directed therapy in advanced disease setting (2 to 3, 4 to 5, and more than 5)

If a level of a factor consists of fewer than 10% of randomized patients, analysis within that level will be omitted.

Analyses will be done within subgroup and, separately, across subgroups with a test of interactions of subgroups with treatment performed. Estimated HRs and CIs for the within-subgroup analyses will be presented as a Forest plot along with p-values for tests of interactions between subgroup variables and treatment.

Other subgroup analyses may be performed as deemed appropriate. If any safety analyses identify important imbalances between arms, subgroup analyses of these endpoints may be performed.

6.13. Protocol Violations

Significant protocol violations that potentially compromise the data integrity and patients' safety will be summarized by treatment group for all randomized patients. These violations will include deviations that can be identified programmatically and those which can only be identified by the clinical research associate during monitoring. Significant protocol deviations are described in another document within the study Trial Master File.

6.14. Interim Analyses and Data Monitoring

6.14.1. Safety Interim Analyses

There are 2 planned interim analyses to evaluate the safety and tolerability of the combination of abemaciclib plus trastuzumab plus fulvestrant in patients enrolled to the safety lead-in for Part A; these analyses will take place when data from the first 6 and 12 qualified patients who complete 1 cycle have been obtained. These interim analyses will be performed by an assessment committee (AC) made up of Lilly members not involved in the day to day study conduct (the medical director, a Global Patient Safety physician, and a statistician) plus the global principal investigators. If $\geq$2 of 6 patients or $\geq$4 of 12 patients in the safety lead-in experience a dose-limiting toxicity (DLT) or DLT-equivalent toxicity as defined in the protocol, the dose of abemaciclib will be reduced to 100 mg for these patients and for all subsequent patients enrolled to Arm A. Even if the AC does not observe the specified number of DLTs or DLT-equivalent toxicities that would lead to dose reduction, Lilly may elect to reduce the dose of abemaciclib in Arm A if warranted by safety concerns. No protocol amendment will be needed to proceed with a lower dose level of abemaciclib in Arm A. If the dose of abemaciclib in Arm A is reduced to 100 mg, patients that initiated treatment at the 150-mg dose will be dose-reduced to 100 mg, and patients who started at the 150-mg abemaciclib dose will be replaced so that there are a total of 75 patients in Arm A with a starting dose of 100-mg abemaciclib. If the 100-mg dose is found to be intolerable, enrollment into Arm A will be discontinued.

In addition to the safety lead-in interim analyses for Arm A, there are 3 planned interim analyses for safety for all treatment arms in Study JPBZ. A safety interim analysis is planned after

approximately 36 patients (a minimum of 12 patients in experimental arms A and B) have been treated for 1 cycle. There will be no prespecified rules for stopping the trial due to safety concerns. The AC members will review safety data at the interim analysis to determine whether there are sufficient safety concerns to justify the termination of study treatment and/or enrollment. Similar interim safety analyses will be performed after approximately 75 patients and after approximately 150 patients have been treated for 1 cycle (a minimum of 25 and 50 patients in experimental arms A and B).

Each safety evaluation will be based, at least, on the following data reports:

- summary of treatment discontinuations and reasons for discontinuation
- summary of SAEs
- summary of TEAEs
- summary of CTCAE-graded central laboratory results
- summary of drug exposure and dose adjustments (including delays and reductions) and reasons for adjustment
- listing of treatment discontinuations due to AE or death
- Lilly Safety System reports for all patients with SAEs (as requested).

6.14.2. Efficacy Interim Analyses

An interim analysis for futility is planned after 75 PFS events have been observed. Futility for the interim analysis will be determined in terms of PFS. As guidance, an AC may recommend stopping the trial or closing an experimental arm for futility if the observed HR>1.3. The stopping guidance should be viewed as only guidance, not an absolute rule. The AC will consider all evidence, including safety and other efficacy parameters, in making this decision.

6.15. Annual Report Analyses

Annual report analyses, including Developmental Safety Update Report and Investigator's Brochure analyses, are described in the LY2835219 Program SAP.

6.16. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset that will be converted to an XML file. Both SAEs and 'Other' AEs are summarized by treatment group and by MedDRA PT.
- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:

  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced

- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the clinical study report, manuscripts, and so forth.

7. References

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KC, Sullivan M, Takeda F. The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst.* 1993;85(5)**:**365-376.

Agresti A. *Categorical Data Analysis*. 2nd ed. New York: Wiley, 2002.

Cox DR. Regression models and life-tables. *J Royal Stat Soc Ser B*. 1972;74(2):187-220.

Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. *Stat Med*. 2010;29(2):219-228.

Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *J Hyg (Lond).* 1949;47(2);188.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.

Karrison TG. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups-- interpretation and power considerations. *Control Clin Trials*. 1977;18(2);151-167.

Meier P, Karrison T, Chappell R, Xie H. The price of Kaplan-Meier. *J Am Stat Assoc*. 2004;99(467):890-896.

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, II. analysis and examples. *Br J Cancer.* 1977;35(1):1-39.

van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708-715.

Leo Document ID = 6a13a30e-1be5-472d-99cb-ca1a50c621f4

Approver: PPD
Approval Date & Time: 07-Feb-2019 15:33:51 GMT
Signature meaning: Approved

Approver: PPD
Approval Date & Time: 08-Feb-2019 16:54:24 GMT
Signature meaning: Approved