1. Statistical Analysis Plan I3Y-MC-JPCF: A Randomized, Open-Label, Phase 3 Study of Abemaciclib combined with Standard Adjuvant Endocrine Therapy versus Standard Adjuvant Endocrine Therapy Alone in Patients with High Risk, Node Positive, Early Stage, Hormone Receptor Positive, Human Epidermal Receptor 2 Negative, Breast Cancer

Abemaciclib (LY2835219)

This is a multicenter, randomized, open-label, Phase 3 study of standard adjuvant endocrine therapy of physician's choice with or without abemaciclib, in patients with high risk, node positive, early stage, HR+, HER2- breast cancer, who completed definitive locoregional therapy (with or without neoadjuvant or adjuvant chemotherapy).

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I3Y-MC-JPCF
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 11-Jul-2017
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 06-Dec-2018
Statistical analysis Plan Version 3 electronically signed and approved by Lilly: 09-Jul-2019
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly: 11-Dec-2019
Statistical Analysis Plan Version 5 electronically signed by Lilly on the date below:

Approval Date: 05-Jun-2020 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 first patient visit. Statistical Analysis Plan Version 2 was approved after the first unblinded safety analysis by the Data Monitoring Committee (DMC), but prior to unblinding of the sponsor and prior to the first interim analysis for efficacy. Statistical Analysis Plan Version 3 was approved after the first unblinded safety analysis by the DMC, but prior to the first futility analysis and prior to unblinding of the sponsor.

Statistical Analysis Plan Version 4 was approved after the first unblinded safety analysis and futility analysis by the DMC, but prior to the unblinding of the sponsor. The overall changes and rational for the changes incorporated in Version 4 are as follows:

- Added details on the analysis of pre-treatment Ki-67 data to keep the highest value when multiple pre-treatment Ki-67 test results are available for a patient.
- Added censoring rule to the derivations of IDFS/DRFS for events occurred prior to the randomization date.
- Added one sensitivity analysis on IDFS/DRFS/OS for control arm patients who receive a CDK4/6 inhibitor prior to their first IDFS events.

Statistical Analysis Plan Version 5 was approved prior to the DMC meeting for the second efficacy interim analysis and prior to the unblinding of the sponsor. The overall changes and rationale for the changes incorporated in Version 5 are as follows:

- Added futility boundary to the second efficacy interim analysis to allow early stop if the chance of achieving statistical significance at final analysis is low. This update was made following the disclosure (29 May 2020) of the futility outcome of another CDK4/6 inhibitor Phase 3 trial in early breast cancer.

4. Study Objectives

4.1. Primary Objective

To evaluate the efficacy of abemaciclib plus adjuvant endocrine therapy versus adjuvant endocrine therapy alone, in terms of invasive disease-free survival (IDFS), for patients with HR+, HER2- early stage breast cancer.

4.2. Secondary Objectives

To evaluate the efficacy, in terms of IDFS, for patients with HR+, HER2- early stage breast cancer with pre-treatment Ki-67 index ≥20% by central lab.

To evaluate the efficacy of abemaciclib plus adjuvant endocrine therapy versus adjuvant endocrine therapy alone in terms of distant relapse-free survival (DRFS) and overall survival (OS).

To assess the safety profile of abemaciclib plus adjuvant endocrine therapy compared to adjuvant endocrine therapy alone.

To evaluate the relationship between abemaciclib dose, concentration, and efficacy and safety outcomes.

To evaluate abemaciclib plus adjuvant endocrine therapy, versus adjuvant endocrine therapy alone, in terms of general oncology and breast cancer self-reported health-related quality of life (FACT-Breast 37-item questionnaire), endocrine therapy-specific symptoms (the FACT-ES 19-item subscale and 2 FACIT-sourced items of cognitive symptoms and 3 FACIT-sourced items for bladder symptoms), and in terms of fatigue experienced during abemaciclib and/or endocrine therapy (FACIT-Fatigue 13-item subscale).

To evaluate health status to inform decision modelling for health economic evaluation using the EQ-5D-5L questionnaire.

4.3. Exploratory Objectives

CCI

[redacted]

5. A Priori Statistical Methods

5.1. Summary of Study Design

MonarchE is a Phase 3 multicenter, randomized, open-label trial in patients with node positive, early stage, resected HR+, HER2- breast cancer who completed definitive locoregional therapy (with or without neoadjuvant or adjuvant chemotherapy) and have high risk of disease recurrence. Treatment with abemaciclib will be given for up to 2 years or until discontinuation criteria are met. Endocrine therapy will be taken as prescribed during the on-study treatment period (Years 1-2). In Year 3 and beyond, standard adjuvant endocrine therapy should continue for a duration of at least 5 years, if deemed medically appropriate.

Standard adjuvant endocrine therapy (SET) is per physician's choice (such as tamoxifen or an aromatase inhibitor, with or without ovarian function suppression per standard practice). Adjuvant treatment with fulvestrant is not allowed at any time during the study. Patients currently receiving standard adjuvant endocrine therapy at time of study entry may not have received more than 12 weeks of standard adjuvant endocrine therapy after completion of their last non-endocrine therapy (surgery, chemotherapy, or radiation) prior to randomization. Randomization must occur within a maximum of 16 months following the definitive breast cancer surgery.

In both treatment arms (Arm A: abemaciclib plus endocrine therapy; Arm B: endocrine therapy alone), Day 1 is the first dose of treatment following randomization, that is abemaciclib and/or endocrine therapy (for Arm A) or endocrine therapy alone (for Arm B), regardless if the patient is receiving endocrine therapy at time of randomization. The first dose of study treatment (abemaciclib and/or endocrine therapy) should be taken no later than 3 days of randomization.

Patients will be enrolled into 2 cohorts: (1) those eligible based on degree of involvement of axillary lymph nodes, tumor size, or grade regardless of Ki-67 status, Cohort 1; and (2) those with at least 1 positive node and eligible exclusively based on a Ki-67 status, Cohort 2 (that is, those patients not eligible based on degree of involvement of axillary lymph nodes, primary tumor size or histologic grade). Note that patients in Cohort 1 may also be eligible by Ki-67, but Ki-67 testing prior to randomization is not required for this cohort. Approximately 4580 patients will be randomized, within cohort, in a 1:1 ratio to either up to 2 years of abemaciclib plus standard adjuvant endocrine therapy or standard adjuvant endocrine therapy alone. The standard adjuvant endocrine therapy will be determined by physician's choice. CCI [redacted]

The intention-to-treat (ITT) population includes all randomized patients in Cohort 1 and Cohort 2. This is the primary analysis population for all efficacy analyses including the primary endpoint (IDFS) and a key gated secondary endpoint (OS). The study was powered based on this population.

5.2. Determination of Sample Size

The study will be powered to approximately 85% assuming an IDFS hazard ratio (HR) of .73 at a cumulative 1-sided alpha of .025. This requires approximately 390 events from across Cohort 1 and Cohort 2 by the time of the primary analysis after accounting for the interim efficacy and futility analyses described in Section 5.9.2.3. The number of patients required to observe approximately 390 events was calculated using Cytel East 6 and the following additional assumptions about pooled population in the two cohorts:

- Patients will enroll at a rate of 2, 8, 32, 60, 102, 140, 164, 188, 198, 206, 218, 238, 256, 260/month for the first 14 months respectively and at 276/months for the remainder of the enrollment period.
- The time from first patient randomized to the observation of approximately 390 events will be approximately 4 years under the alternative hypothesis (HR of .73).
- The probability of a patient dropping out over the first 5 years following randomization is 10%.
- The 5 year IDFS rate for the control arm is 82.5%.

Under these assumptions, approximately 4580 patients will be enrolled.

Patients will be randomized 1:1 within each cohort, using the following stratification factors:

- Prior treatment: neoadjuvant chemotherapy vs adjuvant chemotherapy vs. no chemotherapy
- Menopausal status: premenopausal vs. postmenopausal
- Region: North America/Europe vs. Asia vs. Other

If a patient received both neoadjuvant and adjuvant chemotherapy, the patient will be stratified as neoadjuvant chemotherapy. Male patients will be stratified as postmenopausal at the time of randomization. For a complete definition of menopause, see JPCF Protocol Appendix 5.

5.3. General Considerations

5.3.1. Populations

The following populations will be defined for this study:

Intention-to-Treat population: will include all randomized patients in Cohort 1 and Cohort 2. The ITT analysis of efficacy data will consider allocation of patients to treatment groups as randomized and not by actual treatment received. This population will be used for baseline, efficacy, and health economics analyses.

Safety or Randomized and Treated (RT) population: will include all randomized patients in Cohort 1 and Cohort 2 who received any quantity of study treatment, regardless of their eligibility for the study. The safety evaluation will be performed based on the study regimen a patient actually received, regardless of the arm to which he or she was randomized. The safety population will be used for the primary analysis of dosing/exposure, safety, and resource utilization analyses (See Section 5.12).

Ki-67 High (KI67H) population: will include all randomized patients in Cohort 1 and Cohort 2 with a centrally assessed Ki-67 index ≥20%. Secondary efficacy analyses will be performed on this population.

CCI [redacted]

Pharmacokinetic (PK) population: will include a subset of approximately 20% of patients randomized to Arm A who received at least 1 dose of abemaciclib and have at least 1 post-baseline evaluable PK sample.

5.3.2. Definitions and Conventions

Study drug refers to abemaciclib.

Study treatment refers to abemaciclib + SET or SET alone.

The **date of randomization** is the date the patient was randomly assigned to the abemaciclib + SET arm or SET alone arm using the Interactive Web Response System (IWRS).

The **date of first dose** is the date of the first dose of study drug or SET. For patients receiving SET at the time of randomization, date of first dose is the date of randomization.

The **baseline value of a safety assessment** is the last value observed prior to the first dose. This may occur on the day of first dose.

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 08JUN2017 and the date of first dose was 06JUN2017, 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 05JUN2017 and the date of first dose was 06JUN2017, 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.

5.4. 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.

5.5. 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, enrolled in the study, treated in the study, reasons for discontinuation from study treatment (safety population [Section 5.3.1]), and reasons for discontinuation from study (ITT, KI67H and C1-KI67H populations only). Reasons for discontinuation from both study treatment and the study will be summarized by pre-determined categories. If the reason for discontinuation is adverse event (AE), the associated AE term will be reported.

5.6. Patient Characteristics

5.6.1. Demographics and Performance Status

Patient demographics will be summarized. Patient demographics will include the following:

- Race
- Ethnicity
- Age
- Height
- Weight
- Body mass index (BMI)
- Baseline ECOG PS
- Sex

- Menopausal Status

5.6.2. Baseline Disease Characteristics

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

- Pathological diagnosis
- Primary tumor size by radiology prior to any systemic treatment
- Primary tumor size by pathology following definitive surgery
- Number of involved axillary lymph nodes
- Tumor stage
- Tumor grade
- Most extensive surgery received
- Surgical margin status
- Estrogen receptor status
- Progesterone receptor status
- HER2 Status
- Pre-treatment Ki-67 (central results)

Tumor stage will be derived based on pathologic results.

The pre-treatment Ki-67 results performed by central testing will be summarized. The Ki-67 data from pre-treatment tumor tissues are identified by COLL_BEFORE_SYS_THERAPY = "Yes" in the data from lab requisition form. If a patient has more than one central Ki-67 test results available for untreated tissue, the highest value will be used in the analysis.

5.6.3. 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.

5.6.4. 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 reason for regimen and specific therapy. Frequency of each specific therapy will be tabulated within each reason for therapy.

5.7. Treatment Compliance

Treatment compliance of abemaciclib will be measured by pill counts and summarized. Compliance will be calculated as the ratio of total dose taken to the total assigned dose (minus

any dose adjustments and doses omitted/withheld). The total assigned dose for a patient with no adjustments or omissions is 150 mg per dose × 2 doses per day × number of days on treatment.

Treatment compliance of standard endocrine therapy will be calculated using the Exposure Compliance Endocrine Study Treatment form.

5.8. Concomitant Therapy

All medications will be coded to the generic preferred name according to the current World Health Organization (WHO) drug dictionary. All concomitant medications will be summarized using the preferred name.

5.9. Efficacy Analyses

5.9.1. General Considerations

5.9.1.1. Population

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

5.9.1.2. Stratification Factors

The stratification factors for the analysis of primary and secondary analyses for both cohorts are:

- Prior treatment: neoadjuvant chemotherapy vs. adjuvant chemotherapy vs no chemotherapy
- Menopausal status: premenopausal vs. postmenopausal
- Region: North America/Europe vs. Asia vs. Other

If a patient received both neoadjuvant and adjuvant chemotherapy, the patient will be stratified as neoadjuvant chemotherapy. Male patients will be stratified as postmenopausal at the time of randomization.

The stratification factors are captured in the IWRS and on electronic case report forms (eCRFs). Unless otherwise specified, all stratified analyses will be based on the stratification factors per IWRS. A cross tabulation of the frequency of each level of each stratification factor per IWRS and eCRF will be produced.

5.9.1.3. Hypothesis Tests and Confidence Intervals for Efficacy Data

Unless otherwise noted, all hypothesis tests will be performed at the 1-sided .025 level and all confidence intervals (CIs) will utilize a 95% confidence level.

5.9.2. Primary Endpoint: Invasive Disease-Free Survival

5.9.2.1. Definition

The primary endpoint of this study is IDFS as defined by the STEEP System (Hudis et al. 2007). IDFS time is measured from the date of randomization to the date of first occurrence of:

1. Ipsilateral invasive breast tumor recurrence

2. Regional invasive breast cancer recurrence
3. Distant recurrence
4. Death attributable to any cause
5. Contralateral invasive breast cancer
6. Second primary nonbreast invasive cancer

Patients for whom no event has been observed will be censored at the day of their last post baseline assessment for disease recurrence, as reported on the assessment for disease recurrence CRF, or date of randomization if no post baseline assessment for disease recurrence occurred. The detailed censoring rules are described in Table JPCF.1.

Table JPCF.1. Rules for Determining Date of Event or Censor for IDFS

Situation	Date of Event or Censor	Event / Censor
IDFS event	Date of earliest IDFS event	Event
No IDFS event	Date of last assessment for disease recurrence	Censored
Unless		
IDFS event prior to the randomization date	Date of randomization	Censored
No post baseline disease recurrence assessment	Date of randomization	Censored
IDFS event documented after more than 12 months (+28 days)* following the last disease recurrence assessment or randomization (whichever is later)	Date of last assessment for recurrence prior to the documented IDFS event, or date of randomization (whichever is later)	Censored

*12 months (+28 days) is the longest allowed interval between visits in long-term follow up after Year 5 defined by the schedule of activities
Abbreviation: IDFS = invasive disease-free survival.

5.9.2.2. Description of IDFS events

Invasive disease-free survival events will be summarized as local, regional, contralateral, distant, second primary nonbreast invasive cancer, and death. Location of non-death events will be summarized as reported on the assessment for disease recurrence form. In accordance with the STEEP guidelines, events occurring in the same patient within 2 months of each other will be considered simultaneous, and will be classified as worst event, e.g., simultaneous local and distant events will be classified as distant.

In the analysis, the identification of IDFS events (that are not defined by death) is based on the tumor location of the recurrent disease as recorded by the investigator. If a patient has disease recurrence identified by the investigator (DRCRIND = "Yes") and the corresponding location of tumor recurrence (DRCRLTLC) is local, regional, distant, contralateral, or second primary non-breast neoplasm, that is defined as an invasive disease event (recurrence of noninvasive breast cancer is not counted as an event).

The date of each tumor recurrence per STEEP criteria is defined as the earliest date of the tumor assessment that confirms the recurrent tumor, using the method of radiographic examination or biopsy/FNA. If the date of tumor assessment is missing or no such method is used to identify the tumor recurrence, the date of disease recurrence assessment with DRCRIND = "Yes" will be used. If multiple tumor recurrences occur, IDFS events date will be the date of first tumor recurrence, or the date of death if no tumor recurrence is identified.

5.9.2.3. Hypotheses and Analysis

Letting $S_A(t)$ and $S_S(t)$ denote the IDFS functions of abemaciclib + SET and SET alone respectively, the null hypothesis

H_0: $S_A(t) = S_S(t)$

will be tested against the 1-sided alterative hypothesis

H_1: $S_A(t) > S_S(t)$.

The IDFS analysis to test the superiority of abemaciclib plus standard endocrine therapy to standard endocrine therapy will be performed on the ITT population (all randomized patients in Cohort 1 and Cohort 2) and will use the log-rank test stratified by randomization factors from IWRS. The first futility analysis for IDFS will be conducted when approximately 130 events have been observed in the ITT population. Futility should be declared if the observed IDFS HR is greater than 1.05. There are 2 planned efficacy interim analyses and 1 planned final analysis for IDFS in this study, which will be performed after approximately 195, 293, and 390 events have been observed in the ITT population. The second efficacy interim analysis at approximately 293 IDFS events includes both an efficacy criterion for statistical significance and a futility boundary. The cumulative 1-sided alpha will be controlled at .025, with an alpha split of 0.00000001 for the first futility analysis and 0.02499999 for the planned efficacy analyses. The cumulative 1-sided type I error rate of .02499999 for the 2 planned efficacy interim analyses and 1 planned final analysis will be maintained using the Lan-Demets method (Demets and Lan 1994). Specifically, the alpha spent at each efficacy interim analysis will be based on the exact number of IDFS events observed using the following O'Brien-Fleming type stopping boundary:

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

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. At the first efficacy interim analysis, the nominal one-sided alpha level will be .0015 if exactly 195 IDFS events are observed. At the second efficacy interim analysis, the nominal one-sided alpha level will be .0092 if exactly 293 IDFS events are observed. If the analyses are performed at exactly 195, 293, and 390 events, then the one-sided boundary p-value at the final analysis will be .0220. In addition to the p-value boundary for positive efficacy results, there is a prespecified futility HR boundary of 0.95 in the second interim analysis at approximately 293 IDFS events. Futility should be declared at the second interim analysis if the observed IDFS HR is greater than 0.95.

In the scenario of meeting the futility boundary at the futility-only analysis (approximately 130 IDFS events) or the second efficacy interim analysis (approximately 293 IDFS events), the study may be stopped for futility. However, the sponsor has no intent to stop the study if an efficacy boundary is met at an efficacy interim analysis, and all patients will continue to be followed up for IDFS, DRFS, and OS until study close.

The efficacy and futility boundaries and properties of the design are found in Table JPCF.2.

Table JPCF.2. Efficacy Information

Analysis Point	Approximate Number of IDFS Events	Hazard Ratio for Futility	One-sided Boundary P-value for Efficacy	Cumulative Power Under H_1
Futility	130	1.05	N/A[a]	N/A
Interim 1	195	N/A	.0015[b]	.222
Interim 2	293	0.95	.0092[b]	.634
Final	390	N/A	.0220[b]	.861

Abbreviations: IDFS = invasive disease-free survival; N/A = not applicable.

[a] An arbitrary alpha split of 0.00000001 is applied at the futility analysis.

[b] Dependent on the actual number of events observed at each analysis.

5.9.2.4. Other Analyses

5.9.2.4.1. IDFS Curves and Hazard Ratio (HR)

The Kaplan-Meier (KM) method (Kaplan and Meier 1958) will be used to estimate the IDFS curve for each treatment arm. The difference between IDFS rates for each arm will be reported with 95% confidence interval (CI) estimated by normal approximation, at the end of year 1 & 2, followed by yearly IDFS rates difference until approximately 200 patients in total were at risk.

A stratified Cox proportional hazard model (Cox 1972) with treatment as a factor will be used to estimate the HR between the two treatment arms and the corresponding CI and Wald p-value.

The following sensitivity analyses will be performed on IDFS:

- A log-rank test without stratification by randomization factors will be performed to test the superiority of abemaciclib plus standard endocrine therapy to standard endocrine therapy on the ITT population.
- An unstratified Cox proportional hazard model (Cox 1972) with treatment as a factor will be used to estimate the HR between the 2 treatment arms and the corresponding CI and Wald p-value.
- Censoring for control arm patients receiving CDK4/6 Inhibitor: If a patient on control arm receives a CDK4/6 inhibitor prior to their first IDFS event, IDFS will be censored at the date of the last disease assessment prior to the CDK4/6 inhibitor start date.

CCI

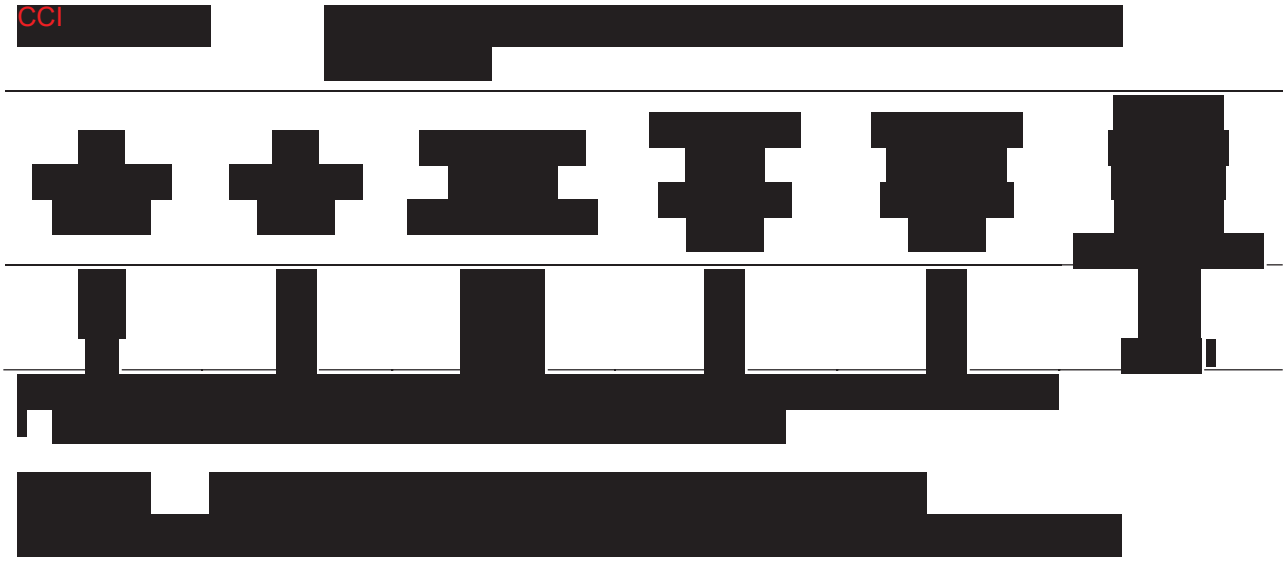
[redacted]

CCI

[redacted]

CCI

CCI
CCI



5.9.3. Second Gated Secondary Endpoint: Overall Survival

5.9.3.1. Background

A sequential gate-keeping strategy will be utilized to control the overall type I error at 0.025 (one-sided) for the secondary endpoint OS in all randomized patients in Cohort 1 and Cohort 2.

CCI

More details concerning gatekeeping and alpha spending across multiple analyses of OS are provided in Section 5.9.3.3.

5.9.3.2. Definition

The OS time is measured from the date of randomization to the date of death from any cause. For each patient who is not known to have died as of the data-inclusion cut-off date for a particular analysis, OS will be censored for that analysis at the date of last contact prior to the data inclusion cut-off date.

5.9.3.3. Hypotheses and Analysis

Letting $S_A(t)$ and $S_S(t)$ denote the OS functions of abemaciclib + SET and SET alone respectively, the null hypothesis

H_0: $S_A(t) = S_S(t)$

will be tested against the 1-sided alterative hypothesis

H_1: $S_A(t) > S_S(t)$.

There are 4 planned interim analyses and 1 final analysis to test the null hypotheses, which will occur at the following time points:

- The first efficacy interim IDFS analysis (195 IDFS events)
- The second efficacy interim IDFS analysis (293 IDFS events)
- The final IDFS analysis (390 IDFS events)

- 2 years after the final IDFS analysis
- Final OS analysis: 390 OS events or 10 years after last patient randomized, whichever occurs earlier.

CCI [redacted]

At each analysis, the null hypothesis above will be tested using a 1-sided stratified log rank test, stratified by the randomization factors.

The cumulative 1-sided type I error rate of .025 will be maintained using the Lan-Demets method (Demets and Lan 1994). Specifically, an α-spending function corresponding to the following O'Brien-Fleming type stopping boundary will be used for this interim efficacy analysis:

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

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 boundary p-value at each analysis will be calculated based on the actual number of events observed at the time of analysis using software that implements this alpha-spending function (for example, ADDPLAN 6.0 or SAS 9.2).

5.9.3.4. Other Analyses

The KM method will be used to estimate the OS curve for each treatment arm. The OS rates for each arm will be compared using a normal approximation for the difference between the rates, at the end of year 1 & 2, followed by yearly OS rates difference until approximately 200 patients in total were at risk.

A stratified Cox proportional hazard model with treatment as a factor will be used to estimate the HR between the 2 treatment arms and the corresponding CI and Wald p-value.

Follow up time for OS will be defined from the date of randomization and will use the inverse of the censoring rules for OS. The median follow-up time will be calculated using the KM method.

The sensitivity analyses described above for IDFS will be repeated for OS.

5.9.4. Distant Relapse-Free Survival

DRFS is measured from the date of randomization to the date of first occurrence of:

1. Distant recurrence
2. Death attributable to any cause

Patients for whom no event has been observed will be censored at the day of their last assessment for disease recurrence or date of randomization if no post baseline disease recurrence assessment occurred. Distance relapse free survival events documented prior to the randomization date will be censored at the date of randomization. Distance relapse free survival events documented after more than 12 months (+28 days) following the last disease recurrence assessment or randomization will be censored at the last assessment for disease recurrence prior to the documented DRFS event, or date of randomization, whichever is later.

In the DRFS analysis, if a patient has disease recurrence per STEEP criteria confirmed by the investigator assessment (DRCRIND = "Yes") and the corresponding location of tumor recurrence (DRCRLTLC) is distant (DRCRLTLC = "Distant"), that is defined as a distant relapse event. The date of the distant recurrence is defined as the earliest date of the tumor assessment that identifies the distant recurrence, using the method of radiographic examination or biopsy/FNA. If the date of tumor assessment is missing or no such method is used to identify the distant tumor recurrence, the date of disease recurrence assessment with DRCRIND = "Yes" will be used. In accordance with the STEEP guidelines, events occurring on the same patient within 2 months of each other will be considered simultaneous, and will be classified as worst event, e.g., simultaneous local and distant events will be classified as distant. If multiple distant relapse events are observed, DRFS events date will be the date of first distant tumor recurrence, or the date of death if no distant tumor recurrence is identified.

The analyses, including the sensitivity analyses, described above for IDFS will be repeated for DRFS.

5.10. Health Outcomes/Quality-of-Life Analyses

Health outcomes and quality of life analyses will be described in a separate Statistical Analysis Plan.

5.11. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

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

5.12. Safety Analyses

All patients who receive at least 1 dose of any study treatment will be evaluated for safety and toxicity.

The Medical Dictionary for Regulatory Activities (MedDRA) Version 19.0 (or higher) will be used to map reported AEs to MedDRA terms. The MedDRA Lower Level Term (LLT) will be used in the treatment emergent computation. Treatment-emergent adverse events (TEAEs) will be summarized by System Organ Class (SOC) and by decreasing frequency of Preferred Term (PT) within the SOC. Preferred terms identified by Medical as clinically identical or synonymous will be grouped together under a single consolidated PT. For example, 'Neutropenia' and 'Neutrophil count decreased' will be reported as 'Neutropenia'. All listings and summaries will use the PT resulting from this process. The National Cancer Institute (NCI)

Common Terminology Criteria for Adverse Events (CTCAE) v 4.0 will serve as the reference document for grading the severity of all AEs and other symptoms.

Safety analyses will be performed in the RT population (Section 5.3.1) and will include summaries of the following:

- TEAEs, including severity and possible relationship to study drug and/or study treatment
- TE-SAEs, including possible relationship to study drug and/or study treatment
- AEs leading to dose adjustments
- discontinuations from study treatment due to AEs or death
- treatment-emergent abnormal changes in laboratory values
- vital signs.

5.12.1. Extent of Exposure

For abemaciclib, extent of exposure will be measured by pill counts and summarized cumulatively. The summary will include total dosage taken and dose intensity. Dose intensity will be calculated as the ratio of total dose taken to the assigned cumulative dose. The assigned cumulative dose while on study is 2 × 150 mg × number of days on treatment.

For co-administered SET, data are reported on Exposure Compliance Endocrine Study Treatment form and will be summarized cumulatively. The summary will include total doses taken and dose intensity. Dose intensity will be calculated as the ratio of total doses taken to the assigned number of doses. The assigned number of doses while on study is 1 dose per day × number of days on treatment.

Dose adjustments and omissions, along with the reason for adjustment or omission, will be summarized for abemaciclib.

5.12.2. Adverse Events

Adverse event verbatim text will also be mapped by the sponsor or designee to corresponding terminology within MedDRA LLT. Severity grades will be assigned by the investigator using CTCAE Version 4. Adverse events will be reported according to PT resulting from this process.

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

A TEAE is defined as any AE that begins between the day of first dose and thirty days after treatment discontinuation (or up to any time if serious and related to study treatment), or any pre-existing condition that increases in CTCAE grade between the day of first dose and 30 days after treatment discontinuation (or up to any time if serious and related to study treatment).

Comparisons of pre-existing conditions to on-treatment events at the LLT level will be used in the treatment-emergent computation.

A 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 TEAE/SAE listings and summaries will be produced:

- Overview of TEAEs
- Summary of TEAEs by PT (all grade and grade ≥3)
- Summary of TEAEs by SOC and PT (all grade and grade ≥3)
- Summary of TEAEs by PT and maximum grade (1-5)
- List of SAEs
- Summary of SAEs by SOC and PT (all grade and grade ≥3).

The 4 summaries will be produced for all TEAEs/SAEs and repeated for TEAEs/SAEs related to study treatment.

5.12.3. Deaths

All deaths on study not attributed to study disease by the investigator will be listed along with the reason for death, if known. For those deaths attributed to an AE, the listing will include the PT of the AE. A summary of deaths including reasons for death will be produced.

5.12.4. 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 post-baseline grade over the entire study.

5.12.5. Vital Signs and Other Physical Findings

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

5.12.6. Electrocardiograms

Local electrocardiograms (ECGs) will be summarized by 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.

5.13. Subgroup Analyses

Subgroup analyses of IDFS will be performed for each of following potential prognostic subgroup variables:

- All baseline stratification factors
- Primary tumor size by pathology following definitive surgery
- Number of involved axillary lymph nodes
- Tumor stage
- Tumor grade
- Progesterone receptor status
- Age
- Race

If a level of a factor consists of fewer than 5% 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.

5.14. Protocol Violations

Major protocol violations that can be derived from the data or that are observed from clinical monitoring and are related to inclusion/exclusion criteria, treatment, or efficacy will be summarized. Major protocol violations will also be listed. These violations will include those defined by:

- Inclusion/Exclusion Criteria
- Treatment
- Efficacy evaluation

5.15. Analyses for the Japanese Regulatory Authority

Analyses conducted specifically for the Pharmaceuticals and Medical Devices Agency (PMDA) will be described in a separate SAP.

5.16. Interim Analyses and Data Monitoring

5.16.1. Safety Interim Analyses

The Data Monitoring Committee (DMC) is responsible for providing external oversight of patient safety in Study JPCF independently of the Lilly study team and Lilly Global Patient Safety (GPS).

Safety interim analyses will be reviewed by the DMC at a frequency described in the DMC charter, but no less than approximately every 6 months. The safety interim analyses will be conducted to evaluate the overall safety profile of abemaciclib when given in combination with standard endocrine therapy.

At each interim analysis, the DMC may recommend the trial continue without modifications, continue with specific modifications, or be stopped for safety concerns. There will be no prespecified rules for stopping the trial due to safety concerns. The DMC members will review safety data at each interim analysis. If a significant safety signal is identified, the DMC may recommend a protocol amendment, termination of enrollment, and/or termination of study treatment. The recommendations of the DMC will be communicated to the Lilly Senior Management Designee (SMD) and, if necessary, an Internal Review Committee (IRC).

A limited number of Lilly representatives external to the study team may have access to treatment assignments as required for evaluation of selected SAEs for determination of regulatory reporting.

5.16.2. Efficacy/Futility Interim Analyses

One futility-only analysis and two efficacy interim analyses are planned. The second efficacy interim analysis includes both an efficacy boundary and a futility boundary.

The efficacy interim analyses will be conducted to provide early efficacy information and could potentially result in early communication with regulatory agencies. The DMC should release the results to the sponsor if the analysis of IDFS is significant as described in Section 5.9.2.3 and any additional criteria specified in the DMC charter are met.

If the study futility boundary is met at the first futility analysis or the second efficacy interim analysis, the DMC should recommend that the study be stopped for futility.

If the analysis of IDFS is statistically significant or futile at an interim analysis, the DMC will be instructed to recommend to the SMD that the results be released to the sponsor. The SMD may convene an IRC to review the DMC's recommendation prior to releasing the results.

The sponsor has no intent to stop the study if an early efficacy boundary is crossed at an efficacy interim analysis, and all patients will continue follow-up for IDFS, DRFS, and OS until study close. Patients randomized to the control group will not be permitted to cross over to the experimental group in case early efficacy is observed during interim review, as this will confound the assessment of OS. If the DMC makes a recommendation counter to this at an interim analysis, for example, the DMC recommends crossing all patients over to the

experimental treatment, FDA will be consulted before any action is taken, as well as other regulatory agencies if deemed appropriate.

The unblinded analysis, including review of the efficacy along with the safety data, will be conducted by the DMC. Study sites will receive information about interim results ONLY if they need to know for the safety of their patients.

6. Unblinding Plan

Study JPCF is a randomized open label study. Randomization will occur using an IWRS system. Assignment to treatment groups will be determined by a computer-generated random sequence. Each patient in this study will be aware of her or his own treatment group. At each investigative site, all staff involved in treating and caring for study patients will have full knowledge of treatment assignments for the patients under their care.

In order to maintain the scientific integrity of this trial, access to study data will be strictly controlled prior to interim and final analyses. For the accumulated aggregate database, treatment assignment will not be included, and other parameters that can disclose treatment will be scrambled. Dummy treatment assignment will be used in the reporting database to maintain the blinding. Analyses using the unblinded treatment assignment will only be performed for DMC review at the prespecified safety and efficacy interim analysis detailed in the protocol and SAP. At each of those analyses for the DMC, only the designated Statistical Analysis Center, who is independent of the sponsor, will perform analyses on unblinded data. For interim PK analyses to occur prior to interim/final analyses, the list of individuals that will have access to unblinded data will be provided with the PK/pharmacodynamics analysis plan, and documentation concerning their access to the data will be retained.

Therefore, the sponsor and all investigative sites will be blinded to treatment assignments for the aggregate database until the database lock for the final analysis of IDFS, or until the DMC recommends the sponsor to be unblinded for efficacy or futility.

7. References

Cox DR. Regression models and life-tables. *J Royal Stat Soc Ser B*. 1972;34(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.

Demets DL, Lan KG. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13(13-14):1341-1352.

Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, Sparano JA, Hunsberger S, Enos RA, Gelber RD, Zujewski JA. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25(15):2127-2132.

Kaplan EL, Meier P. Nonparametric estimation of incomplete observations. *J Amer Stat Assoc*. 1958;53:457-481.

[NCCN] National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology for Breast Cancer Version 2. 2016. Available at: https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. Accessed February 01, 2017.

Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F; ESMO Guidelines Committee. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015;26(suppl 5):v8-v30.

Slud E, Wei LJ. Theory and method: two-sample repeated significance tests based on the modified Wilcoxon statistic. *J Amer Stat Assoc*. 1982;77(380):862-868.

Leo Document ID = b954a778-5d0b-427e-8621-ee99c8e2767d

Approver: PPD

Approval Date & Time: 05-Jun-2020 15:57:01 GMT

Signature meaning: Approved