I3Y-MC-JPBL Statistical Analysis Plan Version 4

A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study of Fulvestrant with or without LY2835219, a CDK4/6 Inhibitor, for Women with Hormone Receptor Positive, HER2 Negative Locally Advanced or Metastatic Breast Cancer

NCT02107703

Approval Date: 08-Jul-2016

1. Statistical Analysis Plan: 13Y-MC-JPBL: A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study of Fulvestrant with or without LY2835219, a CDK4/6 Inhibitor, for Women with Hormone Receptor Positive, HER2 Negative Locally Advanced or Metastatic Breast Cancer

Abemaciclib (LY2835219)

This study is a global, multicenter, double-blind, placeba-controlled, Phase 3 trial for women with hormone receptor positive, HER2 negative locally advanced or metastatic breast cancer randomized to receive fulvestrant with or without LY2835219.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly on 16-Jul-2014
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly on 09-Nov-2015
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date 26 April 2016.

Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below.

Approval Date: 08-Jul-2016 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

This Statistical Analysis Plan (SAP) Version 1 was approved prior to first patient visit.

Statistical Analysis Plan Version 2 was approved on 09 Nov 2015. The overall changes and rationale for the changes incorporated in Version 2 were as follows:

- Additional safety analyses based on the dose change in Protocol Amendment (a).
- Changes to the sample size and primary analysis population corresponding to Protocol Amendment (b).
- Updates to the interim analysis plan and the analysis of overall survival in Protocol Amendment (c).

Statistical Analysis Plan Version 3 was approved on 26 Apr 2016. The overall changes and rationale for the changes incorporated in Version 3 are as follows:

- Updates to the interim analysis plan and the analysis of overall survival in Protocol Amendment (d).

Statistical Analysis Plan Version 4 will be approved after first patient visit but prior to unblinding of the sponsor and prior to the first interim analysis for efficacy. The overall changes and rationale for the changes incorporated in Version 4 are as follows:

- Clarification to the definition of a treatment emergent adverse event.
- Updates to the overall survival analysis plan. Specifically, the pooled (JPBL and JPBM) overall survival analysis was reclassified as an exploratory analysis.

4. Study Objectives

4.1. Primary Objective

The primary objective of Study I3Y-MC-JPBL (JPBL) is to compare abemaciclib plus fulvestrant versus placebo plus fulvestrant with respect to progression-free survival (PFS) for women with hormone receptor (HR)+, human epidermal growth factor receptor (HER)2- locally advanced or metastatic breast cancer.

4.2. Secondary Objectives

The secondary objectives of the study are to compare abemaciclib plus fulvestrant versus placebo plus fulvestrant with respect to each of the following:

- overall survival (OS)
- overall survival (OS) rate at 1, 2, and 3 years
- objective response rate [complete response (CR) + partial response (PR)]
- duration of response (DoR) [CR + PR]
- disease control rate (DCR) [CR + PR + stable disease (SD)]
- clinical benefit rate (CBR) [CR + PR + SD ≥6 months]
- safety and tolerability
- pain and symptom burden using the Brief Pain Inventory (BPI), the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) and the EORTC QLQ-BR23 (breast) questionnaires, and health status scores from the EuroQol 5-Dimension 5 Level (EQ-5D 5L)
- pharmacokinetics (PK) of abemaciclib, its metabolites, and fulvestrant.

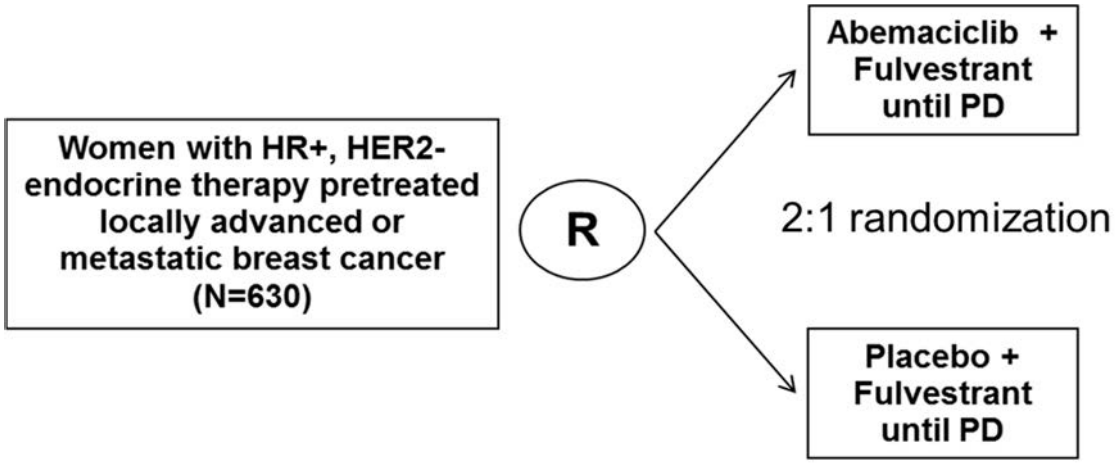
4.3. Exploratory Objectives

- To explore potential biomarkers related to the retinoblastoma (Rb) pathway and/or the pathogenesis of breast cancer
- To explore if change in tumor size is associated with PFS and OS
- To explore time to progressive bone metastases by treatment arm.
- To evaluate time to worsening of Eastern Cooperative Oncology Group performance status (ECOG PS) of ≥2
- time to first skeletal-related event (SRE; defined as either pathological fracture, spinal cord compression, radiation to the bone, or surgery to the bone).

5. Study Design

5.1. Summary of Study Design

Study JPBL is a multicenter, randomized, double-blind, Phase 3 trial for women with HR+, HER2- locally advanced or metastatic breast cancer. Figure JPBL.5.1 illustrates the study design.



Abbreviations: HR+ = hormone receptor positive; HER2- = human epidermal growth factor receptor 2 negative; N = number; PD = progressive disease; R = randomized.

Figure JPBL.5.1. Illustration of study design.

Approximately 630 endocrine therapy pretreated (EP) patients will be randomized 2:1 between the 2 arms:

- Experimental Arm A: abemaciclib 150 mg orally every 12 hours on Days 1 to 28 of a 28-day cycle plus fulvestrant 500 mg intramuscularly on Days 1 and 15 of Cycle 1, then on Day 1 of Cycle 2 and beyond
- Control (Placebo) Arm B: placebo orally every 12 hours on Days 1 to 28 of a 28-day cycle plus fulvestrant 500 mg intramuscularly on Days 1 and 15 of Cycle 1, then on Day 1 of Cycle 2 and beyond.

Enrollment will close when approximately 450 EP patients have been randomized at a starting dose of 150 mg every 12 hours (Q12H). Patients will be randomized 2:1 using the following stratification factors: nature of disease (visceral metastases versus bone only metastases versus other) and sensitivity to endocrine therapy (primary resistance versus secondary resistance). Detailed definitions of the stratification factors can be found in the protocol.

5.2. Determination of Sample Size

This study will enroll 2 strata of patients according to prior endocrine therapy: endocrine therapy pretreated (EP) patients and endocrine therapy naïve (EN) patients.

The primary statistical analyses will be performed on all randomized patients in the EP stratum.

The initial study protocol specified an enrollment of 450 EP patients. Amendment (a) changed the starting dose of blinded study drug from 200 mg Q12H to 150 mg Q12H, and reduced the dose of all patients still on study receiving 200 mg Q12H down to 150mg Q12H. In addition to describing the safety profile in the full EP safety population, safety analyses will be conducted in the subgroup of patients enrolled under Amendment (a) with a starting dose of 150 mg Q12H. In order to do this robustly, enrollment to the study will continue until 450 EP patients are enrolled at a starting dose of 150 mg Q12H. As of the implementation of Amendment (a), approximately 180 EP patients had been enrolled at a starting dose of 200 mg Q12H. Including these 180 patients, the final size of the EP stratum will be approximately 450 + 180 = 630 patients.

A 2-look group-sequential design of the primary endpoint of investigator-assessed PFS will be used to accommodate an event-driven plan for the interim and final PFS analyses (see Section 6.7.2 for details). There is 1 planned interim analysis and 1 final analysis for PFS in this study. The interim analysis is planned to take place after approximately 265 (70% of the 378 planned) investigator-assessed PFS events have occurred. The final PFS analysis will be performed after 378 PFS events have occurred (corresponding to a 40% censoring rate, relative to the anticipated 630 patients enrolled in the EP stratum). The cumulative 1-sided type I error rate of .025 will be maintained using the method described in Section 6.7.2.2. Assuming a hazard ratio (HR) of 0.703, 378 events yields approximately 90% statistical power to detect superiority of the abemaciclib plus fulvestrant arm over the placebo plus fulvestrant arm with the use of a 1-sided log-rank test and a type I error of 0.025. If the true median PFS for the placebo plus fulvestrant arm is 6.5 months, then the HR of 0.703 amounts to an approximately 2.75-month (42%) improvement in median PFS for the abemaciclib plus fulvestrant arm under an additional assumption of exponential survival distribution.

6. A Priori Statistical Methods

6.1. General Considerations

6.1.1. Populations

The **entered** population includes all patients who sign the informed consent document.

The **enrolled** or **intent-to-treat (ITT)** population includes all randomized patients within the EP strata (either primary endocrine resistance or secondary endocrine resistance), per interactive web response system (IWRS).

The safety or **randomized and treated (RT)** population includes all randomized EP patients who received at least one dose of abemaciclib, placebo, or fulvestrant.

Patients randomized within the 'no prior endocrine therapy' stratum comprise the EN population. Exploratory efficacy analyses will be performed on this population. The EN safety population will include all randomized EN patients who receive at least one dose of any study drug.

Unless otherwise noted, all disposition analyses will be performed on the entered population, all patient characteristic and efficacy analyses will be performed on the ITT population, and all safety and exposure analyses will be performed on the RT population.

All analyses will be performed by treatment arm. Unless otherwise noted, all analyses on the ITT population will be performed by assigned treatment arm and all analyses on the RT population will be performed by actual treatment received. All disposition, safety, exposure and patient characteristic analyses will include three groups: abemaciclib treated patients started at the 200 mg dose, abemaciclib treated patients started at the 150 mg dose, and placebo treated patients. Analyses of disposition, safety, exposure and patient characteristics will also be performed on the combined group of abemaciclib treated patients. All efficacy analyses will include two groups, patients randomized to abemaciclib and patients randomized to placebo.

6.1.2. Definitions and Conventions

Study drug refers to abemaciclib or placebo.

Study treatment refers to abemaciclib + fulvestrant or placebo + fulvestrant.

The **date of randomization** is the date the patient was randomly assigned to abemaciclib + fulvestrant arm or placebo + fulvestrant arm using the IWRS.

The **date of first dose** is the date of the first dose of study drug or fulvestrant.

The **baseline value of a safety assessment** is the last value observed prior to the first dose of study drug or fulvestrant.

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 08JUN2014 and the date of first dose was 06JUN2014, 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 05JUN2014 and the date of first dose was 06JUN2014, 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.

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.

6.3. 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 (RT 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. If the reason for discontinuation is adverse event (AE), the associated AE term will be reported.

6.4. Patient Characteristics

6.4.1. 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

6.4.2. Baseline Disease Characteristics

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

- Study entry diagnosis
- Disease stage at study entry
- Endocrine therapy sensitivity (primary resistance, secondary resistance)
- Nature of disease (visceral metastases, bone only metastases, or other)
- Measurable vs non-measurable disease
- Number of organs involved (1, 2, or 3+)
- Estrogen receptor status
- Progesterone receptor status

Nature of disease will be reported directly from the 'Nature of Disease' electronic case (clinical) report form (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 one 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.

6.4.3. 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.

6.4.4. 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 ([neo]adjuvant therapy or therapy for locally advanced or metastatic disease) and specific therapy. Frequency of each specific therapy will be tabulated within each reason for therapy.

Most recent systemic therapy and the duration of that therapy will be summarized within each of the following subgroups:

- Patients whose most recent systemic therapy was an adjuvant therapy
- Patients whose most recent systemic therapy was for locally advanced or metastatic disease.

This summary will include median duration of treatment (date of end of therapy – date of start of therapy + 1), median time to progression (date of progression – date of first dose + 1), and frequency of each specific therapy. If only the month and year of a treatment date or progression date is available, the day will be imputed to the 15th.

6.4.5. 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.

6.5. Treatment Compliance

Treatment compliance of abemaciclib/placebo 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 × 28 days = 8400 mg.

Fulvestrant is administered in the clinic. For analysis of fulvestrant exposure, see Section 6.11.1.

6.6. 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 for the ITT population using the preferred name.

6.7. Efficacy Analyses

6.7.1. General Considerations

6.7.1.1. Population

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

6.7.1.2. Stratification Factors

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

- nature of disease (visceral metastases versus bone only metastases vs other)
- sensitivity to endocrine therapy (primary resistance versus secondary resistance).

The stratification factors are captured in the IWRS and on 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.

6.7.1.3. 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.

6.7.2. Primary Endpoint: Progression Free Survival

6.7.2.1. Definition

The primary efficacy measure is progression-free survival as defined by RECIST Version 1.1 and determined by the investigator. The PFS time is measured from the date of randomization to the date of objective progression or the date of death due to any cause, whichever is earlier.

If a patient is not known to have progressed or died at the time of analysis, PFS time will be censored at the last known progression-free assessment. The detailed censoring rules are described in the table below (Table JPBL.6.1).

Table JPBL.6.1. Rules for Determining Date of Progression or Censor for Progression-Free Survival (PFS)

Rule	Situation	Date of Progression or Censor	Outcome
1	No baseline tumor assessments	Date of Randomization	Censored
2	No post baseline assessments and no death	Date of Randomization	Censored
3	No documented progression and no death (with a post-baseline tumor assessment)	Date of last adequate tumor assessment	Censored
4	Patient lost to follow-up (or withdrew consent from study participation) before documented progression or death	Date of last adequate tumor assessment	Censored
5	Documented progression	Date of documented progression. If a tumor assessment was done on multiple days, use the earliest date for that visit.	Progressed
6	Death without documented progression	Date of death	Progressed
7	Documented progression or death after missing ≥2 consecutive post-baseline tumor assessments*	Date of last adequate tumor assessment before missed assessments or date of randomization, whichever is later	Censored

Note: Progression-free survival and associated outcome is determined by the earliest of the dates above, if more than 1 situation applies. *Two consecutive post-baseline tumor assessments refers to the next two protocol scheduled tumor assessments. Time is measured from the last adequate tumor assessment date. The window of the tumor assessment is also considered. For example if the last adequate scan occurs during cycle 3, the next two protocol mandated assessments are cycle 5 and cycle 7 and the window around each scan is 7 days, thus a patient who goes more than 2*(28*2+7) =126 days has missed two consecutive assessments.

6.7.2.2. Hypotheses and Analysis

Letting $S_A(t)$ and $S_P(t)$ denote the progression free survival functions of abemaciclib + fulvestrant and placebo + fulvestrant respectively, the null hypothesis

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

will be tested against the 1-sided alterative hypothesis

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

There is 1 planned interim analysis and 1 final analysis to test these hypotheses. At each analysis, the hypotheses above will be tested using a 1-sided stratified log rank test, stratified by nature of disease and prior sensitivity to endocrine therapy.

The interim analysis is planned to take place after approximately 265 (70% of the 378 planned) investigator-assessed PFS events have occurred. The cumulative 1-sided type I error rate of .025 will be maintained using the following method. At the interim analysis, the nominal alpha level will be .00001. The remaining alpha will be spent at the final analysis. The resulting boundary p-value for the final analysis is dependent on the exact number of events observed at each analysis and can be calculated using the method of Slud and Wei (1982). If the analyses are performed at exactly 265 and 378 events then the boundary p-value at the final analysis will be .02499996.

The actual p-value required to reject the null hypothesis at the final analysis will be calculated based on the actual number of events observed at the time of each analysis using software that implements the α--spending scheme noted above (for example, ADDPLAN 6.0 or SAS 9.2).

If statistical significance is not observed at the interim analysis, the final PFS analysis will be performed after 378 PFS events have been observed based on investigator assessment. Once statistical significance is declared at the interim analysis or the final analysis, the study will be positive based on the primary endpoint of PFS, and testing of secondary endpoints will proceed as described in Section 6.7.3.

The interim PFS analysis will be performed by the Data Monitoring Committee (DMC). The requirements for unblinding the sponsor at the interim analyses are found in Protocol Section 12.2.15.

6.7.2.3. Other Analyses

6.7.2.3.1. Progression-Free Survival (PFS) Curves and Hazard Ratio (HR)

The Kaplan-Meier (KM) method (Kaplan and Meier 1958) will be used to estimate the PFS curve for each treatment arm. Point estimates and CIs for the first quartile, median, and third quartile for the PFS curve of each arm will be estimated. The PFS rates for each arm will be compared at 3 months intervals up to 15 months using a normal approximation for the difference between the rates.

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.

6.7.2.3.2. Restricted Mean Difference

The common method for describing benefit on the time scale is to calculate the difference in median event time between the two treatment arms. An alternative method for describing benefit on the time scale is to estimate the average difference between the KM curves. This corresponds to calculating the difference in the average time to event for the two 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 LY2835219, 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 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.

6.7.3. Gated Secondary Endpoint: Overall Survival

6.7.3.1. Background

Overall survival(OS) is an important secondary endpoint for this study. A gate-keeping strategy will be utilized to control the overall type I error at 0.025 (one-sided) for the secondary endpoint OS. That is, OS will be tested only if PFS is significant. More details concerning gatekeeping and alpha spending across multiple analyses of OS are provided in Section 6.7.3.3.

6.7.3.2. 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 cutoff date.

6.7.3.3. Hypotheses and Analysis

Letting $S_A(t)$ and $S_P(t)$ denote the overall survival functions of abemaciclib + fulvestrant and placebo + fulvestrant respectively, the null hypothesis

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

will be tested against the 1-sided alterative hypothesis

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

There are 3 planned interim analyses and 1 final analysis to test the null hypotheses which will occur at the following time points:

- The first interim PFS analysis (265 PFS events)
- The final PFS analysis (378 PFS events)
- 331 OS events
- Final OS analysis: 441 OS events

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 PFS interim analysis. If PFS is significant at this stage, the first analysis of OS will also be performed. If OS is not significant at this stage, the second analysis of OS will be performed at the final analysis of PFS (378 PFS events). If OS is not significant at this stage, a third analysis of OS will be performed after 331 deaths. If OS is not significant at this stage, a final analysis of OS will be performed after 441 deaths have been recorded.
- If PFS is not significant at the time of the interim analysis of PFS but is significant at the final analysis for PFS, the second analysis of OS will be performed. In terms of alpha spending, this analysis will be performed as if the first analysis of OS had occurred at the interim PFS analysis (Glimm et al. 2010). If OS is not significant at this stage, the next analysis on OS will be performed when a total of 331 deaths have been recorded. If OS is not significant at this stage, a final analysis will be performed after 441 deaths have been recorded.
- If PFS is not significant after either the interim PFS analysis or the final PFS analysis, OS will not be statistically evaluated.

At each analysis, the null hypothesis above will be tested using a 1-sided stratified log rank test, stratified by nature of disease and prior sensitivity to endocrine therapy.

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

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

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 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).

6.7.3.4. Other Analyses

The KM method will be used to estimate the OS curve for each treatment arm. Point estimates and CIs for the first quartile, median, and third quartile for the OS curve of each arm will be estimated. The OS rates at 1, 2, and 3 years for each arm will be estimated and compared using a normal approximation for the difference between the rates.

A stratified Cox proportional hazard model 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.

A restricted mean difference analysis on OS will be conducted as described for PFS in Section 6.7.2.3.2.

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.

6.7.3.5. Exploratory Pooled Overall Survival Analyses

The ITT populations of JPBL and JPBM will be combined to form a pooled population. Overall survival analyses will be performed on this population. The purpose of these analyses is to evaluate whether adding abemaciclib to the appropriate endocrine therapy improves survival for patients with locally advanced or metastatic disease.

6.7.4. Other Secondary Endpoints

6.7.4.1. Objective Response Rate (ORR), Disease Control Rate (DCR), and Clinical Benefit Rate (CBR)

Objective response rate, disease control rate (DCR), and clinical benefit rate are summary measures of best overall response (BOR) as defined by RECIST Version 1.1. Best overall response 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 post discontinuation 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.

Each patient's BOR will categorized as complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), or not evaluable (NE). For patients with bone-only nonmeasurable disease (see Section 6.4.2), 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.

Overall 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 confidence intervals (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 (CMH) test.

6.7.4.2. 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. For clarity, the start date should be determined by the initial assessment of CR or PR, not the date of confirmation of CR or PR. It is calculated as date of progression or death – date of first response evaluation of CR or PR + 1. Duration of Response will be censored according to the same rules as PFS, with the addition of the following rule: if a patient begins post discontinuation therapy, DoR will be censored on the day of the last response evaluation prior to the initiation of post discontinuation therapy.

A Kaplan-Meier 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 6 months for the first 18 months.

6.7.5. Sensitivity Analyses

6.7.5.1. Progression Free Survival

Sensitivity analyses will be undertaken for calculation of the primary endpoint in order to evaluate the robustness of the analysis. Of specific note, a PFS analysis with the subgroup of the ITT population whose starting dose was 150 mg will be performed as a sensitivity analysis. In addition, the following sensitivity analyses will be performed for PFS:

Progression-Free Survival Sensitivity Analysis 1 (censoring for receiving subsequent systemic anticancer therapy): if a patient is initiated on another anticancer therapy prior to objective progression, including any postdiscontinuation treatment systemic therapy, radiotherapy, or surgical intervention, PFS will be censored at the date of the last complete objective progression-free disease assessment before initiation of the new therapy.

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.

Progression-Free Survival Sensitivity Analysis 3 (forward-dating progressions at unscheduled assessments): if a patient has objective progression at an unscheduled disease assessment, then the PFS time for that patient will be forward-dated to the next scheduled disease assessment.

Progression-Free Survival Sensitivity Analysis 4. PFS will also be analyzed after adjusting for selected prognostic factors. Potential prognostic factors include the stratification factors and other factors as outlined in Section 6.12. The HR for treatment effect will be estimated using a multivariate Cox proportional hazard model constructed by selecting variables among all the potential variables, using stepwise selection method with entry p-value 0.05 and exit p-value 0.01. The treatment factor will be kept out of the model throughout the covariate selection process and only added into the final model.

In addition, a PFS analysis based on independent central review data will be conducted. Details can be found in the Central Review SAP.

6.7.5.2. Overall Survival

One sensitivity analysis on OS is planned. Overall survival time for this analysis 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 disease related, survival time will be censored at the date of death.

6.8. Health Outcomes/Quality-of-Life Analyses

6.8.1. Instruments

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

- modified Brief Pain Inventory, Short Form (mBPI-sf)
- EORTC QLQ-C30
- EORTC QLQ-BR23
- EQ-5D 5L

6.8.2. Pain Intensity and Pain Assessment

Individual pain items on the mBPI-sf (that is, worst, least, average, and current pain) will be described using descriptive statistics by treatment arm and cycle. A mixed effects, repeated measures model will be applied to compare treatment arms by cycle with respect to each item. The model will include baseline score as a covariate and an unstructured covariance matrix will be utilized. The analysis will include all cycles for which at least 25% of patients in each arm have an mBPI-sf assessment.

Corresponding analyses will also be conducted for the mean of 7 pain interference with function items. If a patient does not complete Questions 5a through 5g on the BPI-sf, the mean score for the 7 pain interference items will be calculated based on those answered questions when at least 4 out of 7 questions were completed (that is, ≥50% of the questions were answered).

Time to worsening in pain will be described using the KM method and will be compared between 2 arms by a log-rank test. “Worsening” will be defined as either a “worst pain” increase of ≥2 points postbaseline or an analgesic drug class increase of ≥1 level. Patients who do not

have a worsening event will have time to worsening censored at the time of the last mBPI-sf assessment. Time to worsening rate at 1, 2, and 3 years will be estimated and compared between the 2 arms. The number of events due to each criterion will be described.

6.8.3. Quality of Life

Data from the EORTC QLQ-C30 instrument will be scored as described by Aaronson and colleagues (Aaronson et al. 1993). The EORTC QLQ-BR23 data will be scored as described by the EORTC scoring manual (Fayers et al. 2001).

A mixed effects, repeated measures model will be applied to compare treatment arms by cycle with respect to each instrument. The model will include baseline score as a covariate. For each instrument, the analysis will include all cycles for which at least 25% of patients in each arm have an assessment.

6.8.4. Health State Utility

The EQ-5D 5L data will be scored as described in an article that is under review for publication (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-report for each day. 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 and cycle.

For both the index score and VAS, a mixed effects, repeated measures model will be applied to compare treatment arms by cycle. The model will include baseline score as a covariate. The analysis will include all cycles for which at least 25% of patients in each arm have an assessment.

6.8.5. Utilization

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

- Analgesics (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)
- Post discontinuation 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 chi-squared test. Continuous variables will be described by the mean, median, and standard deviation. A t-test will be used to compare mean utilization.

6.8.6. Time to First Skeletal-Related Event

Time from randomization to documentation of the first occurrence of any skeletal-related event (SRE) will be evaluated. Patients not known to have had an SRE at the time of the analysis will be censored at the date of their last complete documented assessment for SRE. A stratified log-rank test will be used to evaluate the difference of time to first SRE between treatments. Stratification factors will include those used at randomization in addition to whether a patient had experienced at least 1 SRE prior to randomization.

6.8.7. Time to Worsening of ECOG Performance Status (PS)

Time from randomization to documentation of the first occurrence of any PS of ≥2 will be evaluated. Any patient who does have a PS ≥2 documented while on study will have time to worsening censored on the date of the last PS evaluation. A stratified log-rank test, stratified by the randomization factors and baseline PS, will be used to evaluate the difference of time to worsening in ECOG between treatments.

6.9. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

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

6.10. Tailoring Biomarker Analyses

The distributions of biomarkers with continuous measures, such as gene or protein expression, will be described. Summary statistics will include means, medians, corresponding standard errors, quartiles, and ranges. Biomarkers with discrete measures, such as genotype locus, will be summarized in frequency tables. Correlative analyses may be performed to investigate associations between biomarkers and clinical endpoints as deemed appropriate.

Further analysis of biomarkers will be described in a separate biomarker SAP.

6.11. Safety Analyses

6.11.1. Extent of Exposure

For abemaciclib/placebo, extent of exposure will be measured by pill counts and summarized by 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 for each patient during each cycle is 150 mg per dose × 2 doses per day × 28 days = 8400 mg. The assigned cumulative dose while on study is 150 mg per dose × 2 doses per day × number of days on treatment.

For fulvestrant, extent of exposure will be measured using the fulvestrant administration eCRF and summarized by cycle and cumulatively. The summary will include total dosage administered and dose intensity. Dose intensity will be calculated as the ratio of total dose administered to the assigned cumulative dose. The assigned cumulative dose for each patient during each cycle is 1000 mg for cycle 1 and 500 mg for cycle 2 and beyond. The assigned cumulative dose while on study is 500 mg + 500 mg × number of cycles started.

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

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:

- The CTCAE Version 4 term reported by the investigator will be mapped to the MedDRA PT and system organ class (SOC) of the corresponding MedDRA lower level term (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.

Pre-existing 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 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 thirty 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 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.

To assess the relationship of the AE to the study treatment, the following terminologies are defined (in Protocol Section 8.1.2):

- **Related**: a direct cause and effect relationship between the study treatment and the AE is likely.
- **Possibly related**: a cause and effect relationship between the study treatment and the AE has not been demonstrated at this time and is not probable, but is also not impossible.
- **Unrelated**: without question, the AE is definitely not associated with the study treatment.

As per Lilly's standard operating procedures (SOPs), all "related" and "possibly related" AEs and SAEs will be defined as related to study treatment.

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

6.11.3. 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.

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 post-baseline grade over the entire study.

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 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.

6.12. Subgroup Analyses

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

- All baseline stratification factors
- Starting dose (200 mg versus 150 mg)
- Measurable disease at baseline (yes versus no)
- Number of organs involved (1 versus 2 vs. 3+)
- Age (<65 years versus ≥65 years)
- Region (North America, Europe, Asia)
- Race (Caucasian, Asian, and Other)
- PgR status (positive versus negative)
- Baseline ECOG PS (0 versus 1).

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.

6.13. 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 or treatment will be summarized. Major protocol violations will also be listed.These violations will include those defined by:

- Inclusion/Exclusion Criteria
    - Diagnosis
    - Prior treatments received

  - Age
  - Performance Status
- Treatment
  - Dose delays
  - Dose reductions

6.14. Interim Analyses and Data Monitoring

6.14.1. Safety Interim Analyses

The DMC is responsible for providing external oversight of patient safety in Study JPBL independently of the Lilly study team and Lilly GPS.

During the study, safety interim analyses will be performed every 3 months. The first safety interim analysis will be triggered by the 90th patient enrolling, with the data cutoff for this analysis occurring 1 month after the trigger. The safety interim analyses will be conducted to evaluate the overall safety profile of LY2835219 when given in combination with fulvestrant.

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

- summary of patient discontinuations and reasons for discontinuation
- summary of SAEs
- Lilly Safety System reports for all patients with SAEs
- summary of TEAEs
- Details pertaining the conduct of these analyses are provided in the JPBL DMC Charter.

6.14.2. Efficacy Interim Analyses

One efficacy interim analysis of PFS and 3 interim analyses of OS are planned, as described in Sections 6.7.2.2 and 6.7.3.3. The interim PFS analysis (and corresponding OS analysis) will be conducted by the DMC. All other efficacy analyses will be conducted by the sponsor.

Rules for unblinding the sponsor at an interim analysis can be found in the protocol.

At the time of the interim PFS analysis, analyses of PFS and OS provided to the DMC will include:

- the boundary values for significance,
- the p-value for a stratified log rank test comparing the two treatment arms, stratified by the randomization factors,
- an estimate of the HR between the two arms based on a Cox proportional hazards model, stratified by the randomization factors, and
- a KM analysis by treatment arm.

Details pertaining the conduct of these analyses are provided in the JPBL DMC Charter.

6.14.3. Pharmacokinetic/Pharmacodynamic Interim Analyses

A limited number of preidentified individuals independent of the study team may receive access to unblinded data, as specified in the unblinding plan, prior to the interim or final database lock, in order to initiate the final population PK/pharmacodynamic model development processes for interim or final analyses. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

6.15. Analyses for the Japanese Regulatory Authority

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

6.16. Annual Report Analyses

Annual report analyses, including Developmental Safety Update Report (DSUR) and Investigational Brochure (IB) analyses, are described in the LY2835219 Program Safety Analysis Plan.

6.17. 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 which will be converted to an XML file. Both SAEs and 'Other' AEs will be summarized by: treatment group, 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).
- Adverse event reporting is consistent with other document disclosures for example, the clinical study report (CSR), manuscripts, and so forth.

In addition, a participant flow will be created that will describe how many enrolled patients completed the study, and for those who did not, the frequency of each reason for not completing. This analysis will be based on study discontinuation, not treatment discontinuation. A patient

will be identified as having completed the study if the patient dies while on the study or the patient had discontinued study treatment and is in follow up at the time of the final OS analysis. Patients who withdraw consent before the final OS analysis or who are still on treatment at the time of the final OS analysis will be identified as not completing the study.

7. Unblinding Plan

Randomization will occur using an IWRS system. Assignment to treatment groups will be determined by a computer-generated random sequence. Security measures will be taken so that treatment group code that can link patients to study arm will be blinded in the database.

In order to maintain the scientific integrity of this double-blind trial, access to study data will be strictly controlled prior to the interim and final analyses of OS. Dummy treatment assignment will be used in the reporting database until the database lock for the final analysis of overall survival. During this time, analyses using unblinded treatment codes will be performed only at the interim analysis points specified in the protocol/SAP. For those safety and efficacy analyses assigned to the DMC, only the designed Statistical Analysis Center (SAC), who is independent of the sponsor, will perform analyses on unblinded data. For the interim PK analysis to occur prior to the interim/final PFS analyses, the list of individuals that will have access to unblinded data will be provided with the PK/pharmacodynamic analysis plan, and documentation concerning their access to the data will be retained.

Data sets will be created for the purpose of aggregate data review by the sponsor in which treatment assignment is scrambled so that personnel involved in the day-to-day conduct of the trial and development and validation of analysis programs will be blinded to patient treatment assignment.

While every effort will be made to blind both the patient and the investigator to the identity of the treatment, the inadvertent unblinding of a patient may occur. This unblinding will not be sufficient cause (in and of itself) for that patient to be excluded from any safety or efficacy analyses.

8. Changes to Planned Analyses in Protocol

SAP version 1 made the following changes to analyses from the original protocol:

- A best response of CR or PR will not require confirmation as RECIST Version 1.1 does not require confirmation for randomized studies.
- The PFS sensitivity analyses were updated in the following way:
    - the analyses including censoring for discontinuation due to toxicity (denoted analyses 2 and 3 in the protocol) were removed as these analyses are not informative.
    - the analysis concerning unscheduled assessments (referred to as analysis 5 in the protocol) has been updated to forward-date progressions observed at unscheduled assessments to the next planned assessment, rather than back-date to the previous assessment, as this is more appropriate.
    - a sensitivity analysis adjusting for potential prognostic factors has been added.
- Minor updates to health outcomes analyses were made.

9. References

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KCA, 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**:**365-376.

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

European Medicines Agency (EMA) Guidance "Points to Consider on Application with 1. Meta-analyses; 2. One Pivotal Study" (CPMP/EWP/2330/99, 2001). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf Accessed 17 July 2015.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

Food and Drug Administration (FDA) Guidance for Industry Integrated Summary of Effectiveness. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatory information/guidances/ucm079803.pdf. Accessed 17 July 2015.

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 expectational life. *Journal of Hygiene.* 1949;47;188-189.

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

Karrison T. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups-- interpretation and power considerations. *Controlled Clinical Trials.* 1997;(18):151-167.

Meier P, Karrison T, Chappell R, Xie H. The Price of Kaplan-Meier. *J Amer 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.

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.

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 = c092c95e-e2fe-4c49-86ec-bdcc1c77261b

Approver: [redacted]
Approval Date & Time: 08-Jul-2016 14:31:34 GMT
Signature meaning: Approved