IBCSG

International Breast Cancer Study Group Statistical Center

SOLE (Trial 35-07/BIG 1-07)

Statistical Analysis Plan

Version	Author	Date	Status
1	[redacted]	Jan 2017	For primary trial report
2	[redacted]	May 2019	For final, updated report

1 INTRODUCTION

1.1 BACKGROUND

In 2006, the standard duration of adjuvant endocrine therapy for breast cancer (either SERMs or AIs) is five years. Patients who receive extended adjuvant letrozole for five years following approximately five years of tamoxifen obtain further benefit compared with the five years of tamoxifen alone. Similarly, benefit has been demonstrated for switching from tamoxifen to an AI after 2 to 3 years of tamoxifen to complete five years of endocrine therapy, as well as initiating therapy with AI following surgery and administering the AI for five years.

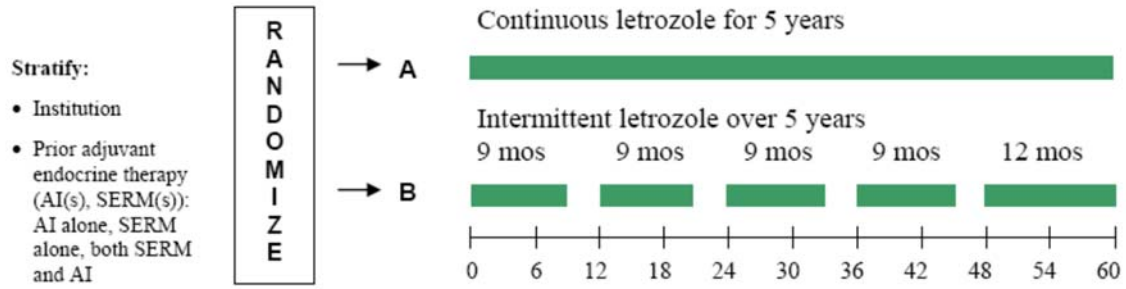
Questions remain about the optimal duration and best schedule of AIs in the extended adjuvant setting. This trial tests the hypothesis that introducing 3-month treatment-free intervals during the course of five years of extended adjuvant letrozole will improve disease-free survival. This hypothesis is based on the theoretical principle that letrozole withdrawal for 3 months will permit some estrogenic stimulation which makes residual resistant disease susceptible to letrozole reintroduction.

1.2 TRIAL DESIGN

SOLE (IBCSG 35-07 / BIG 1-07) OVERVIEW

Title:	**Study Of Letrozole Extension (SOLE):** A phase III trial evaluating the role of continuous letrozole versus intermittent letrozole following 4 to 6 years of prior adjuvant endocrine therapy for postmenopausal women with hormone receptor positive, node-positive early stage breast cancer.
Patient Population:	Postmenopausal women who are disease-free following 4-6 years of prior adjuvant endocrine therapy with selective estrogen receptor modulator(s) (SERM) and/or aromatase inhibitor(s) (AI) for endocrine-responsive, node-positive early breast cancer.
Entry:	Patients must have stopped prior endocrine SERM/AI therapy, and must be randomized within 12 months (1 year) of the last dose of prior endocrine SERM/AI therapy.
Activation Date:	8 November 2007 (First patient randomized on 5 December 2007)
Target Accrual:	4800 patients
Closure Date:	12 July 2012 (Last patient randomized on 8 October 2012)
Final Accrual:	4884 patients
Last Visit:	31 December 2018

Trial Schema:



The SOLE trial is a multinational phase III randomized clinical trial designed to compare continuous letrozole for 5 years with intermittent letrozole over a 5-year period among postmenopausal women who are disease-free following 4 to 6 years of prior adjuvant endocrine therapy with SERM(s) and/or AI(s) for endocrine-responsive node-positive operable breast cancer. The hypothesis is that introducing 3 month treatment-free intervals during the course of 5 years of extended letrozole will improve disease-free survival.

The randomization was stratified according to participating center and prior SERM/AI endocrine therapy (SERM(s) alone, AI(s) alone, both SERM(s) and AI(s)). Patients were randomized 1:1 to one of two extended adjuvant endocrine therapy groups:
A: Continuous letrozole 2.5 mg daily for 5 years
B: Intermittent letrozole 2.5 mg daily for the first 9 months of years 1 through 4, with treatment-free intervals for the last 3 months of years 1 through 4, followed by 12 months of letrozole 2.5 mg daily in year 5
Patients should start trial treatment as soon as possible after randomization. Trial treatment should begin no later than 6 weeks from the date of randomization. All patients will be followed every 6 months for years 1 to 5, and yearly thereafter for assessment of disease status and for survival data collection.

1.3 SAMPLE SIZE CONSIDERATIONS

The sample size was determined to provide 80% power to detect a 20% reduction in the risk of an event defining DFS associated with intermittent letrozole compared with continuous letrozole (hazard ratio = 0.80; 25% increase in 4-year DFS from 90% to 91.917%) using a two-sided 0.05 level test of significance.

To detect the treatment difference 647 DFS events are required, assuming 4800 patients are accrued (1600 patients per year for 3 years), 5% non-assessability at 4 years, and approximately

5 years of additional follow-up. One year of start-up time, as Participating Centers obtain Ethics Committee approval and complete regulatory processes, is anticipated.

1.4 INTERIM MONITORING PLAN (FROM PROTOCOL)

A group sequential design with two interim analyses and one final analysis is used. The target number of events for the final analysis is 647, and interim analyses are planned after 40% and 70% information (259 and 453 events observed respectively). At each interim analysis and at the final analysis, testing will be performed using O'Brien-Fleming boundaries.

1.5 DATA AND SAFETY MONITORING COMMITTEE REVIEWS

The Data and Safety Monitoring Committee (DSMC) reviews the safety information, including adverse events, second (non-breast) malignancies, and deaths without prior cancer event, every six months and has recommended the trial continue as planned. The first interim efficacy analysis was performed for spring 2014 DSMC report when 324 DFS events (50% information) were observed. At that time the O'Brien-Fleming boundary was not crossed and there was not enough evidence to support early stopping of the trial due to efficacy. The DSMC recommended that SOLE treatment administration and follow-up continue as planned. The same results held for the second interim efficacy analysis which was performed for spring 2015 DSMC report when 454 DFS events (70% information) were observed. The final DSMC review was undertaken on 13 November 2017 after all patients had completed the 5-year protocol therapy period, and determined that follow-up should continue as planned.

1.6 PRIMARY RESULTS

After a median follow-up of 5 years (data cut-off, 31 October 2016), DFS events were reported for 665 of 4851 patients in the intention-to-treat population. Patients assigned intermittent letrozole did not have a significantly reduced hazard of a DFS event vs. patients assigned continuous letrozole (HR=1.08; 95% CI, 0.93 to 1.26; P=0.31). The estimated 5-year DFS was 85.8% among patients assigned intermittent letrozole versus 87.5% for patients assigned continuous letrozole. Similar results were observed for secondary endpoints of breast cancer-free interval (431 events, HR=0.98; 95% CI, 0.81 to 1.18; P=0.84), distant recurrence-free interval (338 events, HR=0.88; 95% CI, 0.71 to 1.09; P=0.25), and overall survival (316 deaths, HR=0.85; 95% CI, 0.68 to 1.06; P=0.16). At 60 months median follow-up, 27% of patients continued on protocol treatment. 36.2% and 34.5% of patients assigned to intermittent versus continuous letrozole had grade 3-5 targeted AEs reported, of which grade 3-4 hypertension was most frequently reported.

The results were presented at the 2017 ASCO Annul Meeting in June 2017 [1] and subsequently published in *The Lancet Oncology* [2]. In postmenopausal women with hormone receptor-positive breast cancer, extended use of intermittent letrozole did not improve disease-free survival compared with continuous use of letrozole.

1.7 FINAL, UPDATED ANALYSIS

A decision was made following publication of the primary results to cease follow-up as of 31 December 2018 **(last patient, last FU visit)**, at which point the median follow-up would be extended from 5 to approximately 7 years (range 6 to 11 years). A SOLE Flash newsletter was distributed to investigators, providing information regarding closure of data collection at the end of 2018.

In addition to an update of the primary and secondary endpoints, adverse events and final treatment status (the 5 years of protocol treatment was reached as of October 2017) , it was decided that this final analysis would include also the SOLE-EST substudy and further investigate whether there is heterogeneity of treatment efficacy according to type and duration of prior endocrine therapies, .

2 EFFICACY ANALYSIS PLANS

2.1 OBJECTIVES

The primary objective is to compare continuous letrozole for five years with intermittent letrozole over a five year period for postmenopausal women who are disease-free following 4-6 years of prior adjuvant endocrine therapy with SERM(s) and/or AI(s) for endocrine-responsive, node-positive, operable breast cancer.

2.2 ANALYSIS POPULATIONS

The primary analysis will use an intention-to-treat (ITT) approach. The ITT population will include all randomized patients, regardless of eligibility status; the possible exceptions are patients who immediately withdrew consent prior to treatment initiation and declined all participation, patients determined (e.g., via audit) to be without documented informed consent, and/or patients at a participating center determined not to be compliant with protocol procedures. Any exclusion from the ITT population will be determined prior to the analysis and will be summarized in listing and CONSORT in the trial report. The final, updated analysis will continue to use an ITT approach and the ITT population as originally reported will be used, even if information has changed since the primary report; the only exception would be for a case in inadequately documented informed consent or withdrawal of consent to use any data.

2.3 ENDPOINT DEFINITIONS

2.3.1 Primary Endpoint

• Disease-free survival (DFS) is defined as the duration of time from randomization to the first indication of the following events: invasive recurrence at local (including recurrence restricted to the breast after breast conserving treatment), regional or distant sites; a new invasive cancer in the contralateral breast; any secondary (non-breast) malignancy; or a death without

prior cancer event. Appearance of DCIS or LCIS either in the ipsilateral or in the contralateral breast will not be considered as an event for DFS. In the absence of an event, DFS is censored at the date of last follow-up.

2.3.2 Secondary Endpoints

• Overall survival (OS) is defined as the duration of time from randomization to death from any cause, or is censored at the date last known alive. (Note, for patients who withdrew consent or were lost to follow-up but follow-up for survival was possible through hospital or registry records, OS is censored at the date last known alive rather than date of last follow-up/withdrawn consent).

• Distant disease-free survival (DDFS) was a secondary endpoint in protocol. It will be replaced by a more modern definition -- distant recurrence-free interval (DRFI) defined as the duration of time from randomization to the first indication of invasive breast recurrence at a distant site. In the absence of an event, DRFI is censored at the date of last follow-up or date of death without distant recurrence.

• Breast cancer-free interval (BCFI) is defined as the duration of time from randomization to the first indication of the following events: invasive breast recurrence at local, regional or distant sites; a new invasive cancer in the contralateral breast. In the absence of an event, BCFI is censored at the date of last follow-up or date of death without prior breast cancer event. (Second non-breast malignancies are ignored)

• Site of First Failure (i.e., DFS events): Hierarchy of failures from least to worst, following standard IBCSG definition:

- o Local
- o Contralateral Breast ± above
- o Regional ± above
- o Soft tissues/distant nodes ± above
- o Bone ± above
- o Viscera (including 'other') ± above

as well as,

- o second (non-breast) malignancy,
- o death without prior cancer event (aka death without subsequent cancer event)
- o death with breast cancer recurrence suspected
- o death with no information about recurrence.

For "site of first failure," if there are 2 or more failure sites within 2 months of each other, then the "worst" site is considered as the site of first failure. For example if there is a regional recurrence and a bone recurrence within 2 months of each other, then the bone would be

considered the "worst" site of failure and considered as "site of first failure." If second malignancy is coincident with breast cancer recurrence then site of first failure is reported as the breast cancer recurrence.

Per IBCSG standard, the date of DFS event is the date a proven recurrence was first suspected. In rare cases that the date suspected is prior to randomization, date of randomization is considered as date of event. Since the first report, the 35-RC form SOP was clarified that in cases with multiple proven sites of first recurrence (within 2 months of each other), that the date first suspected should be the earliest of suspected dates from among those proven sites (which may not necessarily correspond to the site with the earlies proven date).

2.4 FOLLOW-UP

Number of DFS events per person-year of follow-up is calculated at each DSMC review time to guide the timing for the database lock. The data cut-off is planned for prior to November 1, 2016 when we expect the number of DFS events will reach the target of 647 for the analysis with database lock in Q1 2017. Data as of the database cut-off will be used for analysis.

Median follow-up is calculated from the Kaplan-Meier estimate of overall survival, with the event/censoring indicator inverted (i.e. alive as event and dead as censored).

The final, updated results will be presented approximately two to three years after the primary report.

2.5 TESTS AND ESTIMATES

The primary endpoint, DFS, will be compared between treatment arms using a two-sided stratified logrank test with an overall experiment-wise alpha-level equal to at most 0.05. The test statistic and p-value will be taken from the stratified Cox PH model score test. Hazard ratios will be estimated from a stratified Cox PH model, with 95% CIs. Kaplan-Meier estimates of the DFS distributions will be calculated for each of the treatment arms, with reporting of the 5yr DFS. We will check the proportional hazards assumption by visually assessing the plot of log(-log(survival)) versus log of survival time for parallelism. This will be done overall, and according to strata.

2.5.1 Stratification

The stratification factor during randomization is prior SERM/AI endocrine therapy (SERM(s) alone, AI(s) alone, both SERM(s) and AI(s)) as reported on randomization (RA) form. Institutions were balanced using dynamic balancing. Logrank test and Cox PH model will be stratified by prior SERM/AI endocrine therapy.

Note about values of strata as provided at randomization vs. actual values: The randomization form (RA form) collects the values entered into the IBCSG randomization system and used for stratification of the randomization assignment. If these values are incorrect then they are amended on the Registration form (A-form). For primary and secondary overall analyses we will use the stratification factor entered on the randomization form. For subgroup analyses, we will use the actual values entered in the A form when available.

2.6 ANALYSIS COMPONENTS

2.6.1 Enrollment, Follow-up Compliance

2.6.1.1 Overview of enrollment

Figures:

- Enrollment over time; x-axis time in 6-montly intervals; y-axis number enrolled

2.6.1.2 Follow-up submission

Tables:

- Institutional follow-up compliance by group/country (rows)

2.6.1.3 CONSORT

In the manner of the CONSORT diagram, the following will be summarized, with changes from the primary report highlighted.

Tables:

- CONSORT diagram content numbers by treatment assignment
    - Number of patients randomized
    - Number of patients included vs excluded from analysis population, with reasons [not expected to change since primary analysis]
    - Number in analysis population who never started protocol treatment [note any changes since primary analysis]
    - Number in analysis population who WC/LFU [update]
    - Number of patients analyzed in analysis population [same as number included; not expected to change since primary analysis]

2.6.1.3 Stratification

Tables

- Distribution of stratification factors, overall and according to treatment assignment
- Listing of any new information about actual (from A-form) values of stratification factors

2.6.1.4 Withdrawn consent/loss to follow-up status

Summary should mention any centers that were lost to follow-up since the primary report.

Tables:

- Withdrawn consent and lost to follow-up status, overall and according to treatment assignment

2.6.2 Patient, Disease and Prior Treatment Characteristics

Characteristics of the analysis population will be summarized overall and by treatment group. Continuous variables are summarized as mean, SD, min/max, and quartiles. Categorical variables are summarized as N(%); for variables with unavailable (missing, unknown, not done) values, the default approach is to include an unknown category that is included in the denominator for percentages (rather than just listing the number of unknowns as a category). The same tables will be re-run, and may differ slightly from the primary report as the data will reflect any updates to the database since the primary database lock.

Tables (overall and by treatment group, unless otherwise specified):

- Patient:
  - Age at randomization (continuous; categorized in 5-year intervals)
  - Race/ethnicity
  - BMI at randomization
  - Menopausal status at primary cancer diagnosis
  - Symptoms (from baseline AE form) [n.b., slight redefinition]
- Treatment:
  - Local therapy (combining surgery [Mx/BCS] and radiotherapy [yes/no]
  - Chemotherapy (whether used)
  - Prior endocrine therapy [stratification factor; actual] and years of prior endocrine therapy [<4.5; 4.5-5.5; >5.5]; more detail will be added, see section 2.6.5 below
  - Duration from end of prior adjuvant therapy to randomization [≤1 vs >1 month]
- Disease:
  - ER/PgR and HER2 status (H form)
  - Nodal involvement (number of positive lymph nodes 0, 1-3, 4-9, 10+, unknown)
  - Tumor size (<1, 1-2, >2-5, >5 cm) and grade (1, 2, 3), primary histology (ductal vs lobular vs other; H form)

2.6.3 Primary Efficacy Analysis

The primary efficacy analysis will proceed as summarized in Section 2.5 above. The data cut-off and database lock dates used for the analyses and the median follow-up duration (and IQR and range; overall and by treatment assignment) will be reported. Numbers of outcome events at primary analysis and at final, updated analysis will be tabulated, for ease of reference.

2.6.3.1 Subgroup Analyses

The protocol pre-specified factors that will be used to characterize the patients enrolled in the study and to provide descriptive statistics of outcomes according to subgroups of the population. These factors include: age at randomization, body mass index, tumor size, tumor grade, number of positive lymph nodes, ER/PgR, HER2 status, type of prior endocrine therapy, duration of prior endocrine therapy by type, interval of time since the cessation of prior endocrine therapy till randomization. These analyses will be considered as secondary and descriptive.

The plans for these variables are summarized below:

- Age (5-year age groups (<55, 55-59, 60-64, 65-69, ≥70))
- BMI (obese (≥30), overweight (25 - <30), normal (<25), unknown)
- Tumor size (≤2 vs >2 cm; unknown)
- Tumor Grade (1,2,3,unknown)
- No. of positive lymph nodes (0, 1-3, 4+, unknown)
- Type of prior endocrine therapy (AI, SERM, AI and SERM)
- Duration of prior endocrine therapy (<4.5, 4.5-5.5, >5.5 years)
- Interval of time since the cessation of prior endocrine therapy till randomization (≤1, >1 mo)

2.6.3.2 Models

Stratified Cox PH regression models will be used to: estimate HRs (95% CI) for treatment effect, unadjusted and adjusted for these covariates in Section 2.6.3.1; and estimate HRs (95% CI) for treatment effect within subgroups by including treatment-by-covariate interaction in the model (but not other covariates) and using contrasts.

2.6.3.3 Tables and Figures

Tables:

- Primary treatment comparison: N events and patients, HR, 95% CI, log-rank test statistic and p-value, 5yr and 7yr DFS, SE and 95% CI
- Treatment effects within subgroups: N events and patients and 7yr rates within each subgroup, treatment HR, 95% CI, p-value for test of treatment-by-variable interaction

Figures:

- KM plot of DFS, by treatment group, for entire analysis population *(y-axis: Percent Alive and Disease-Free; x-axis: Time since Randomization (1-year intervals); x-axis limited to 8 years (median follow-up plus 1 year) and numbers at risk at each yearly interval)*
- Forest plot of DFS, overall and for subgroups

2.6.4 Secondary Efficacy Endpoints

Breast cancer-free interval, distant recurrence-free interval and overall survival will be summarized as described for DFS.

Sites of first treatment failure will be summarized overall and by treatment group as N (%). Second (non-breast) malignancies as site of first failure, and deaths without prior cancer event will be summarized overall and by treatment assignment.

2.6.4.1 Tables and Figures

Tables:

- Primary treatment comparison for each of the 3 secondary endpoints: N events and patients, HR, 95% CI, log-rank test statistic and p-value, 5yr & 7yr DFS, SE and 95% CI
- Sites of first failure, overall and by treatment assignment
- Types of second non-breast malignancies, as site of first failure, overall and by treatment assignment.
- Second non-breast malignancies and Death without prior cancer event (list of patients with type of malignancy or cause of death, respectively)

Figures:

- KM plots of BCFI, by treatment group, for entire analysis population *(y-axis: Percent Free from Breast Cancer; x-axis: Time since Randomization (1-yr intervals); x-axis limited to* 8 *years (median survival plus 1 year) and numbers at risk at each yearly interval)*
- KM plots of DRFI, by treatment group, for entire analysis population *(y-axis: Percent Free from Distant Recurrence; x-axis: Time since Randomization (1-yr intervals); x-axis limited to 8 years (median survival plus 1 year) and numbers at risk at each yearly interval)*
- KM plot of OS, by treatment group, for entire analysis population *(y-axis: Percent Alive; x-axis: Time since Randomization (1-yr intervals); x-axis limited to 8 years (median survival plus 1 year) and numbers at risk at each yearly interval)*

2.6.5 Treatment Efficacy in Relation to Prior Endocrine Therapy Type & Duration (new)

Motivation: Investigate whether duration of prior AI is associated with treatment responsiveness; hypothesize that different duration of prior AI may be differentially sensitive to AI interruption

on basis of heterogeneity of extended AI trial results depending on prior therapy of enrolled population.

Endpoints: BCFI, DRFI, DFS; OS to be analyzed only if other results of interest, to ensure consistency of message.

1) Define prior ET type and duration in more detail:

- number of months of prior AI, number of months of prior SERM;
- first drug taken, most recent drug taken;
- number of weeks of prior AI within the 1 yr before randomization.

Re-tabulate categories of AI duration, in particular looking more closely at those with both prior AI and SERM exposure to better differentiate those with sequential tam>AI (approx. 2>3 yrs) as separate from those that had a short duration of AI and mostly tamoxifen (i.e., more like SERM-only), or had at least 4 yrs of AI (i.e., more like AI-only). [0-<6m; 6m -<2y; 2-<3y; 3-<4y; ≥4y]

2) Assess relation of prior AI use with other covariates, especially age and menopausal status at diagnosis, and disease characteristics.

- Tabulate patient, disease and treatment characteristics reported at randomization, overall and according to AI duration categories above (without regard to treatment assignment).

3) Assess relation of prior AI use (AI duration categories above) with time to protocol therapy cessation (with DFS event as competing risk); with and without regard to treatment assignment.

- There is particular interest in a group of patients who had very short AI exposure (>0 but <6m to 1yr) early in adjuvant treatment and presumably stopped because of tolerability (i.e., do not include pts whose short AI exposure was just prior to randomization); the clinical question would be whether such patients should re-try AI for extended adjuvant or also very likely to cease early.

4) Assess relation prior AI duration with treatment comparison;

- STEPP analysis, on HR and 7-yr event rate, with continuous prior AI duration as x-axis; note this doesn't adjust for covariates.
- Using unstratified Cox model, test prior AI categories (as above) vs treatment assignment interaction, adjusted for selected patient and disease characteristics [age or menopausal status, BMI; nodes, tumor size, tumor grade, HER2 status; prior chemo? prior local-therapy?] (n.b., 3-category prior ET already done in forest plot, unadjusted; may want to repeat this in adjusted Cox model)
- If useful as part of unadjusted analyses, KM plots of subgroups defined by prior AI duration categories (as above) and treatment assignment, with 7-yr estimates.

2.6.6 Adverse Events / Safety

2.6.6.1 Population

The toxicity population is the subset of patients in the ITT analysis population who started protocol treatment. Any patients without at least 1 post-baseline AE form submitted will not be able to contribute; note any such patients in trial report (noting any difference from primary report).

2.6.6.2 Analysis

Targeted AEs, and other grade 3-5 AEs, are collected on CRFs. The grade and causality attribution are recorded according to the NCI Common Terminology Criteria for Adverse Events (CTCAE) version 3.0. The targeted AEs will be summarized by AE type and maximum grade over time, regardless of causality attribution. The maximum grade consolidates the reports of a given type of AE for a patient over time since randomization (i.e., baseline reports are excluded) by taking the maximum across time (i.e., a patient appears only once for a given type of AE). Patients with reports of multiple AEs of different types are reported multiple times under the relevant AE categories. Maximum grade 0 indicates that the AE type has not been reported. 95% exact CIs will be calculated for each targeted AE, according to treatment group.

Other grade 3 or higher AEs are also requested on CRFs, by write-in text. All will be tabulated, similarly according to max grade, but the intention is to focus on those deemed possibly, probably or definitely related to study treatment(s), which will also be tabulated.

Late adverse events (LAEs: cardiac ischemia/infarction, thrombosis/thrombus/embolism, CNS cerebrovascular ischemia, and fractures) were requested on CRFs, but very little was reported, thus we will not automatically summarize these data, as we're concerned that this would underrepresent the prevalence of such events.

2.6.6.3 Tables and Figures

Tables:

- Numbers of patients in the safety population, overall and by stratum and treatment assignment; listing of any changes since primary analysis and listing of any patients who started treatment but do not have AE CRFs submitted
- Targeted AEs by grade, reported according to treatment assignment; N, %
- Targeted AEs, % grade 1-5 AE with 95% CI; % grade 3-5 AE with 95% CI, reported according to treatment assignment
- Targeted AEs by grade, reported according to prior endocrine therapy and treatment assignment; N, frequency
- Other non-targeted grade 3-5 AEs, deemed possibly, probably or definitely related to study treatment(s) according to treatment assignment, N

2.6.7 Treatment

Final Protocol treatment status as of the clinical data cut-off will be summarized. In this report, status will be summarized by treatment assignment. Time from randomization to permanent treatment discontinuation will be summarized using KM curves according to treatment assignment; this will be done without regard to reason to summarize the exposure period, and secondarily with disease events as competing risks as a measure of adherence with 5-yr treatment period. Final Protocol treatment status will be summarized.

Details of scheduled (and unscheduled) interruptions were analyzed and reported in the primary Clinical Trial Report. These analyses below will not be repeated:

Per SOLE Data Manager's Hand Book v3.0 (dated 21Mar13), for Arm B patients, interruptions should be every 9 months and should last for 3 months in years 1-4. Any interruption up to 4 months is considered a "protocol-specified interruption" and the reason for interruption is defined as "protocol-defined interval" on L forms; any interruption of 4 months +1 day (121 days or more) is considered a "non-protocol interruption".

The Idealized Scheduler contains a link to an Excel file on the IBCSG web site. This file calculates the letrozole dispensation and interruptions (Arm B) based off of the Randomization Date.

Treatment variation scenarios:

- Patient on Arm B stops letrozole early (e.g., at Month 6 instead of Month 9). The patient should stop for three months. Patient should re-start letrozole and continue until the pre-defined date of the next interruption. This will get the patient back on schedule.

- Patient on Arm B took medication during the protocol-specified interruption. Patient should continue until Month 9 of the next year and then start the interruption. This will get the patient back on schedule.

- Patient forgot to take some pills and therefore has some left at the date the interruption should start. The patient should not finish the medication. The patient should interrupt on the pre-defined date to stay on schedule.

For Arm A patients, not-per-protocol interruptions are those of over a month in duration.

Reasons for treatment interruptions will be summarized by treatment assignment. Note that there could be more than one reason for an interruption spanning more than one L Form: for example, a patient start interruption per protocol on one L form, but due to adverse event, the interruption continues to the next L form without restarting after 4 months. The first reason would be "protocol-defined interval" and then the second reason would be "Adverse Event".

Based on information on Data Manager's Handbook, to describe adherence with protocol-assigned treatment, the following tables and figures are proposed.

Tables

- Final status of protocol-assigned treatment, overall and according to treatment assignment (same as CTR table 8.1.1)

Figures

- Time from randomization to permanent treatment discontinuation, according to treatment assignment, with yearly estimates: all reasons (CTR Fig 8.2.1), and then with disease event as competing risk (x-axis: months 0-72 since randomization; y-axis: percent on treatment 0-100)

3 SOLE-EST SUBSTUDY

3.1 SUMMARY

Third generation aromatase inhibitors such as letrozole profoundly suppress estrogen levels in postmenopausal women, which is the mechanism of their action against hormone receptor positive breast cancer. The hypothesis under investigation in the SOLE trial is that removing this suppression for a period of time will allow re-sensitization of breast cancer cells to estrogen deprivation resulting in a delay or prevention of the development of resistance to the anti-cancer effect of letrozole.

The estrogen suppression substudy enrolled patients at randomization and aims to describe the changes in serum levels of estrogen levels that occur on letrozole and during the three-month treatment gap without letrozole. The blood sampling time points were over the first year since randomization (see schema). Estrogen suppression and recovery may also be linked with clinical outcomes including toxicity and disease free survival. SOLE-EST will also examine the relationship between estrogen level changes and changes in toxicity, patient-reported musculoskeletal symptoms and QL, and grip strength. As genes important for the metabolism and activity of letrozole may influence estrogen levels, single nucleotide polymorphisms **(**SNPs) related to these will be assessed and analyzed in relation to the variability of estrogen level changes.

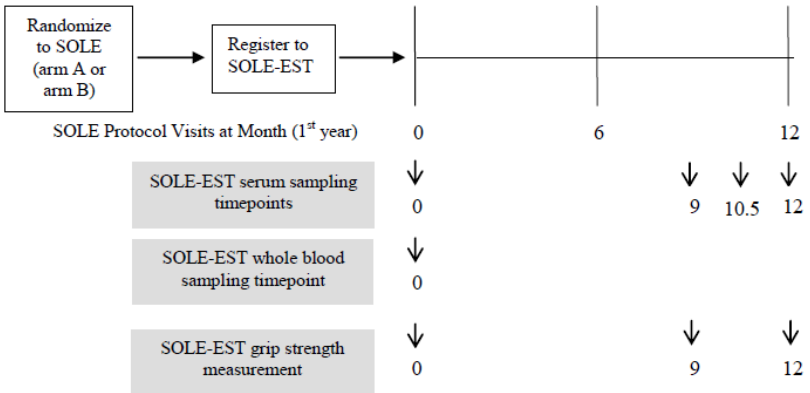
SOLE-EST was activated in IBCSG Participating Centers located in Belgium, Australia, and Milan on 19 November 2010 and the first patient enrolled on 31 January 2011. Enrollment was closed 12 July 2012, with a final accrual of 104 patients (25 from the continuous arm and 79 from the intermittent arm) enrolled at 14 Participating Centers.

All eligible patients randomized to SOLE from these Centers were to be offered participation in SOLE-EST, but inclusion was not mandatory.

SOLE-EST OVERVIEW

Title:	**SOLE-EST: SOLE Estrogen Substudy:** Investigating changes in estrogen levels and grip strength for patients participating in the SOLE Trial.
Patient Population:	Postmenopausal women who are disease-free following 4-6 years of prior adjuvant endocrine therapy with selective estrogen receptor modulator(s) (SERM) and/or aromatase inhibitor(s) (AI) for endocrine-responsive, node-positive early breast cancer, who: have been randomized into the SOLE trial; participate in the Quality of Life substudy; provide written informed consent.
Entry:	Patients should be enrolled to SOLE-EST at the time of entry into the main study and the Quality of Life substudy. The first serum sample must be obtained after randomization but prior to starting letrozole on study according to the guidelines in the SOLE-EST Manual for Blood Sample Logistics and Vigorimeter Use.
Activation Date:	19 November 2010 (First patient enrolled on 31 January 2011)
Target Accrual:	100 patients: 25 from the continuous arm and 75 from the intermittent arm.
Closure Date:	12 July 2012 (Arm A closed 29 Sept 2011)
Final Accrual:	104 patients

SOLE-EST schema and blood sampling timepoints



3.2 BACKGROUND

There are three biologically important circulating estrogens present in postmenopausal women: 17 β-Estradiol (E2) is a product of both estrone and testosterone and the most abundant circulating estrogen in premenopausal women; estrone (E1) is the major product of aromatase enzyme activity in postmenopausal women; estrone sulphate (E1S) is the most stable estrogen fraction measured and provides a robust surrogate-parameter for suppression of estrogen biosynthesis in vivo. Mean plasma levels of E2, E1 and E1S in postmenopausal women are reported to be 20, 80 and 4-500 pmol/L respectively. Suppression of estrogen production by letrozole reduces levels close to or below the level of detection in many assays.

Musculoskeletal side effects are amongst the most common and most troublesome side effects associated with aromatase inhibitors and have been assessed in a number of ways. A simple and validated measure is grip strength using a modified sphygmomanometer. The effect of a three month break on QL was reported for the overall trial population. It was hypothesized that the increase in estrogen levels during the three month gap would reduce side effects, especially musculoskeletal side effects and patient-reported symptoms; and it is hypothesized that this will correlate with the absolute estrogen levels. It is also hypothesized that the increase in estrogen levels will improve QoL overall. The SOLE-EST substudy will test these hypotheses by using the toxicity data collected in the main study and the QoL data collected in the QoL substudy and by measuring the grip strength of the patients.

Letrozole is metabolized by cytochrome P450 (CYP) isoenzymes including CYP2A6 and CYP3A4. Genetic variations in the genes encoding for those enzymes can alter the metabolism and therefore plasma levels of letrozole. The gene encoding CYP19 (the aromatase enzyme that produces estrogen from testosterone) is a highly polymorphic gene that may influence the efficacy of letrozole through an altered (increased or decreased) activity. There is also some evidence that genetic variation may explain the variation of the degree of side effects amongst women on letrozole. Ingle *et al.* found 4 SNPs that were associated with the degree of musculoskeletal symptoms using a genome-wide association case-control study.

3.3 OBJECTIVES AND ENDPOINTS

Primary objective:

- To determine the serum level of estrogens E2, E1 and E1S and Sex Hormone Binding Globulin (SHBG) during letrozole treatment
- To determine the degree of recovery of E2, E1 and E1S during the 3 month gap.

Also noted, describe the within-patient variability of E2, E1 and E1S across time among patients continuously taking letrozole.

Secondary objectives:

- To link the estrogen level changes with the clinical outcomes of toxicity (musculoskeletal toxicity [as assessed by grip strength], arthralgia, hot flushes and insomnia) and QoL.

- To determine the effect of the following factors on estrogen levels: prior adjuvant endocrine therapy (type, duration, duration without ET prior to study entry); age; BMI; type of menopause.
- To explore variability of estrogen level changes and link these with germline SNPs.

Primary endpoints

- Levels of E2, E1 and E1S at 0, 9, 10.5 and 12 months from randomization to the core protocol.
- % change (suppression or recovery) of E2, E1 and E1S from baseline at 9 months, 10.5 months and at 12 months from randomization to the core protocol.

Secondary endpoints

- Toxicity grade changes (for arthralgia, hot flushes and insomnia) between 6 months (on letrozole) and 12 months (off letrozole for 3 months) and correlation with % recovery of estrogen levels.
- Quality of life score change between 6 months (on letrozole) and 12 months (off letrozole for 3 months) and correlation with % recovery of estrogen levels.
- Changes in grip strength score at 9 months and 12 months.

3.4 SUBSTUDY DESIGN AND PROCEDURES

3.4.1 Blood Samples and Assays

A whole blood sample was obtained at baseline. Serum samples were obtained at 4 time points (0, 9, 10.5 and 12 months); they were taken at any time of day and were not fasting. Samples were taken using provided blood collection kits and stored locally at -20°C. Assays were performed centrally.

Estrogen levels must be measured using a highly sensitive assay. A central laboratory (Reproductive Endocrine Research Lab of Frank Stanczyk, Los Angeles, USA) will do the assays, and the sensitivity they quote for their RIA using a 1.0 mL aliquot of serum is as follows:

- for estradiol (E2) is 2 pg/ml (7.34pmol/L)
- for estrone(E1) is 3 pg/ml (11.01pmol/L)
- for estrone sulphate (E1S) is 9 pg/ml (33.03pmol/L)
- for SHBG .

It is considered that a 50% increase in estradiol (E2) levels or a level of >12pmol/L (for those whose levels were below the lower limit of sensitivity for the assay) after cessation of letrozole would be clinically significant.

The protocol stated the SNPs were to be analyzed using the Sequenom MassARRAY® technology in the laboratory of Vesalius Research Center, K.U. Leuven and VIB, Belgium:

- SNPs on alleles CYP2A6*2, CYP2A6*9, CYP2A6*12 and CYP2A6*35
- In the CYP3A4 gene, only the CYP3A4*1B and CYP3A4*7 SNP have a known effect on the enzymatic activity and a frequency of more than 1% in the Caucasian population. Therefore, for the CYP3A4 gene, only those two SNPs will be genotyped.
- The CYP19 SNPs with a frequency of >10% in the Caucasian population and with a known effect on the enzymatic activity will be analyzed, namely rs4646, rs10046, rs727479, rs10459592, rs4775936, rs6493497 and rs7176005.
- SNPs previously associated with musculoskeletal symptoms: rs7158782, rs7159713, rs2369049 and rs6637820.

3.4.2 Measurement of toxicity and quality of life

Adverse event information will be collected at baseline and at 6 and 12 months as per the main study. Primary AEs are arthralgia, hot flushes and insomnia.

Patient-reported symptoms will be assessed by the Breast Cancer Prevention Trial (BCPT) Symptom Scales on Form 35-PRS. In addition, several global QoL indicators will be assessed using the IBCSG Trial 35-07 QL Form.

Grip strength will be measured using the Martin Vigorimeter (a modified sphygmomanometer , provided or reimbursed by IBCSG) at 0, 9, 12 months. To perform the hand grip test, the patient is asked to squeeze the balloon of a modified sphygmomanometer three times with maximal force and the maximal value of three trials of each hand will be used for evaluation. This assessment has high inter-rater reliability and takes a minimal amount of time.

3.4.3 Sample size considerations

Sample size was determined based on a conservative approach to the primary analysis in which patients are classified as to whether or not the change in E2 between 9 and 12 months exceeds a certain threshold indicative of clinically significant recovery of E2. Clinically significant recovery is defined as a 50% increase in estradiol (E2) levels or a level of >12pmol/L (for those whose levels were below the lower limit of sensitivity for the assay). With a sample size of 100 patients, 25 in arm A (continuous letrozole) and 75 in arm B (intermittent letrozole), and assuming conservatively that only 80% of patients have paired samples at 9 and 12 months, then there is over 90% power to detect a difference between 10% vs. 50% of patients having recovery of E2 during the 3-month interval (Fisher's exact test; two-sided α=0.05).

Further, the sample size will allow exploration of the relation of other factors – specified patient characteristics, clinical measures and germline SNPs – with E2 changes measured on the continuum. Again focusing on the change in E2 between 9 and 12 months among patients in arm

B, and considering a range of the frequency of a binary factor in the population from 50-50% to 10-90% , the detectable effect size between two groups ranges from about 0.75 to 1.25 SD (80% power, two-sided α=0.05, two-sample t-test). For a continuous factor a Pearson's correlation coefficient of 0.36 vs. null value of 0 would be detectable.

3.4.4 Analysis plan (from protocol)

Estrogen levels (E2, E1, E1S) will be summarized over time among all patients and by treatment arm using descriptive statistics (mean, SD, median, quartiles) and graphically. It is anticipated based on the literature that log-transformation of estrogen levels may be required, in which case the geometric mean would be reported. Values will also be summarized as percent change from baseline levels.

Using linear mixed modeling of all available data over the 4 timepoints (or of 3 timepoints relative to baseline), each estrogen level will be modeled as a function of time and treatment arm and the interaction of the two factors to investigate the time pattern of estrogen levels and differences in levels between treatment arms. Of primary interest is whether the change from 9 to 10.5 and 12 months differs among patients assigned to intermittent vs. continuous letrozole. It may be the case that estrogen levels while on treatment are below the assay lower limit of detection, in which case the analysis method would be changed to account for this sort of "floor effect" or "left censoring" of data. Relationships of specified patient characteristics (prior endocrine therapy, age at and/or type of menopause, BMI) with the estrogen levels and changes in levels will also be investigated, as well as specific toxicities and quality of life measures and germline SNPs.

3.4.5 Analysis plan for trial reporting

We plan as part of the report of the parent trial to include the analysis of the SOLE-EST primary objective. The secondary objectives will be analyzed separately, with separate SAP.

3.4.5.1 Population and Available Samples

Accrual:
Tabulate by country, site: numbers of patients, first/last enrollment dates

Accounting:
* Patients: Pt participation over the 12 months, incl. protocol treatment status at 9 (arms A & B) and 12 (arm A) months
* Samples: Sample assay results availability, by timepoint; timing relative to planned collection timepoints
* Flow diagram

Characteristics, same tabs as parent trial:
* tabulate SE pts vs all ITT patients
* tabulate SE pts vs all ITT patients at the participating sites
* tabulate SE pts vs all ITT patients at the participating sites during SE enroll period

Additional Characteristics:
In addition to characteristics from the main trial, collected for SOLE-EST are:

- Dominant hand
- Type of menopausal will be derived from additional data collected on SE-A form:
    - menopause information (bilateral oophorectomy, RT ovarian ablation, chemotherapy-induced amenorrhea, natural menopause before BC diagnosis, natural menopause since BC diagnosis)
    - potential for residual ovarian function (whether age<55, ovarian function recovered following chemotherapy, receipt of GnRH/LHRH in past 2 yrs; last menstruation within past 2 yrs, other potential for residual ovarian function specify)

Hormone levels at baseline:
* descriptive statistics in relation to: prior ET; most recent prior ET; duration from end of prior ET to randomization; duration prior AI category, age, type menopause

Hormone levels over time:
** note, may need to look at variability in planned vs actual timing of samples and whether that influences levels;*
* descriptive statistics, boxplots, spaghetti plots according to timepoint and treatment assignment [absolute levels]
* descriptive statistics of percent change and fold-change from baseline, according to timepoint and treatment assignment
* descriptive statistics, graphical summaries of percent change and fold-change from 9m to 10.5 and 12m, according to timepoint and treatment assignment
    * according to other characteristics?