Statistical Analysis Plan (SAP)

GEICAM/2019-01

Phase II, randomized, open-label, international, multicenter study to compare efficacy of standard chemotherapy vs. letrozole plus abemaciclib as neoadjuvant therapy in HR-positive/HER2-negative high/intermediate risk breast cancer patients "CARABELA Study"

Sponsor: GEICAM (Spanish Breast Cancer Research Group Foundation)

Chief Investigators:
Dr. Miguel Martin. H. General Universitario Gregorio Marañón, Madrid, Spain.
Dr. Emilio Alba. H. Universitario Virgen de la Victoria. Málaga, Spain.
Dr. Ángel Guerrero. Instituto Valenciano de Oncología. Valencia, Spain

Pathology central laboratory responsible:
Dr. Federico Rojo. H. Fundación Jiménez Díaz. Madrid, Spain
Sponsor Study Code: **GEICAM/2019-01**
EudraCT Number: 2019-002123-15
Protocol reviewed by FECMA (Spanish Federation of Breast Cancer Patients)

Author: Maribel Casas

Version: 1.0

Date: 30 JUN 2023

Protocol No.: GEICAM/2019-01

Protocol title: Phase II, randomized, open-label, international, multicenter study to compare efficacy of standard chemotherapy vs. letrozole plus abemaciclib as neoadjuvant therapy in HR-positive/HER2-negative high/intermediate risk breast cancer patients “CARABELA Study” (1)

Protocol Version: Version 2.0, 29-Jun-2021

Sponsor: GEICAM

This Statistical Analysis Plan was created according the ICH Good Clinical Practice (2) (3), GEICAM policies and Standard Operating Procedures (SOP); and is consistent with the study protocol (1).

Version Date: 30-JUN-2023

Version No.: 1.0

Previous Version: NA

Prepared by
Statistician:

Maribel Casas

Name	Signature	Date

Reviewed by

Statistician:

Mª José Escudero

Name	Signature	Date

Project Manager:

Victoria Ruiz

Name	Signature	Date

Translational Monitor:

Dra. Rosalía Caballero

Name	Signature	Date

Translational Project Manager

Manager:

Nuria Martín

Name	Signature	Date

Approved by

Medical Monitor:

Dra. Susana Bezares

Name	Signature	Date

Medical Coordinator:

Dr. Miguel Martín

Name	Signature	Date

Medical Coordinator:

Dr. Emilio Alba

Name	Signature	Date

Medical Coordinator:

Dr. Ángel Guerrero

Name	Signature	Date

Pathology central laboratory responsible:

Dr. Federico Rojo

Name	Signature	Date

TABLE OF CONTENTS

ABBREVIATIONS AND DEFINITIONS

A list of abbreviations and acronyms that will be used in the SAP, with their definitions, will be provided.

AC	Doxorubicin + Cyclophosphamide
AE	Adverse Event
AESI	Adverse Event of Special Interest
AI	Aromatase Inhibitor
BC	Breast Cancer
BCS	Breast Conservative Surgery
ctDNA	Circulating tumoral DNA
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CNV	Copy Number Variation
Compliance	Adherence to all the trial-related requirements, good clinical practice (GCP) requirements, and the applicable regulatory requirements.
CR	Complete Response
CRR	Clinical Response Rate
CT	Chemotherapy
CTCAE	Common Terminology Criteria for Adverse Events

DNA	Deoxyribonucleic Acid
EBC	Early Breast Cancer
e-CRF	Electronic Case Report Form
End of Study (Trial)	The end of study (trial) is the date of the last visit or last scheduled procedure shown in the Study Schedule for the last active patient in the study, including follow-up
Enroll	The act of assigning a patient to a treatment. Patients who are enrolled in the trial are those who have been assigned a registration number and treatment.
ET	Endocrine Therapy
ER	Estrogen receptor
ESAG	Exploratory Statistical Analysis Guide
GCP	Good Clinical Practice
GEICAM	Spanish Breast Cancer Research Group
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hormone Receptor or Hazard Ratio depending on the context
ICD	Informed Consent Document
IDMC	Interim Data Monitoring Committee
iEFS	invasive Event Free Survival
IHC	Immunohistochemistry

Investigator	A person responsible for the conduct of the clinical trial at a trial site. If a trial is conducted by a team of individuals at a trial site, the investigator is the responsible leader of the team and may be called the principal investigator
ITT	Intent To Treat
LHRH	Luteinizing hormone-releasing hormone
MedDRA	Medical Dictionary for Regulatory Activities
MRD	Minimal residual disease
MRI	Magnetic Resonance Imaging
NA	Neoadjuvant
NAT	Neoadjuvant therapy
NCI	National Cancer Institute
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
Patient	A subject with a defined disease
PD	Progressive disease
PDL-1	Programmed cell death ligand 1
PEPI	Preoperative Endocrine Prognostic Index
PgR	Progesterone receptor
PP	Per Protocol
PR	Partial Response

PT	Preferred Term
RECIST	Response Evaluation Criteria in Solid Tumors
RCB	Residual Cancer Burden
RD	Residual Disease
RNA	Ribonucleic Acid
RS	Recurrence Score
SAE	Serious Adverse Event
SAR	Serious Adverse Reaction
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
Screen	The act of obtaining informed consent for participation in a clinical trial from patients deemed eligible or potentially eligible to participate in the clinical trial. Patients screened into a trial are those who sign the informed consent document directly or through their legally acceptable representatives.
SD	Standard deviation
SOC	System Organ Class
SOP	Standard Operating Procedure
TEAE	Treatment-Emergent Adverse Event
TLF	Table, Listing and Figure
TNM	Tumor, Node, Metastasis, for the purpose of staging

TMB	Tumor mutational burden
UICC	Union for International Cancer Control's
WHO	World Health Organization

1. INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to give a detailed description of the statistical analysis to be performed to generate the study reports for GEICAM/2019-01 (CARABELA) study (1).

2. OBJECTIVES

2.1 Primary Objective

To assess the Residual Cancer Burden (RCB) 0-I rate in both treatment arms.

2.2 Secondary Objectives

- Changes in Ki67 index value after 2 weeks of treatment in both treatment arms.
- RCB 0+I vs. RCB-II vs. RCB-III in both treatment arms (TNM downstaging).
- Compare RCB values between both treatment arms.
- The rate of PEPI (Preoperative Endocrine Prognostic Index) score 0 at surgery in both treatment arms.
- Clinical response measured by magnetic resonance imaging (MRI) according to RECIST v1.1 in both treatment arms.
- Rate of breast conservative surgery (BCS) in both treatment arms.
- iEFS (invasive Event Free Survival) in both treatment arms.
- Assessment of safety profile by NCI-CTCAE v5.0 classification (4).
- To assess molecular downstaging for high-risk genomic groups defined by a multigene expression panel.

2.3 Exploratory Objectives

- Correlation of Ki67 protein level of change after 2 weeks of treatment with some efficacy variables.
- Genetic changes in sequential tumor samples and tumor evolution during NA therapy (NAT), and its correlation with some efficacy variables and surrogate endpoints for response.
- To explore T-cell functional activation, immune suppression, neoantigens and cytokines production as well as other immune response biomarkers (such as CD4/CD8/FOXP3/PDL-

1), to understand the activity of abemaciclib on the tumor microenvironment and the immune response.

- Circulating tumor DNA (ctDNA) dynamics as a surrogate endpoint for response and prognostic implicationsSequential genomic profiling and quantification of ctDNA samples to explore early response dynamics, minimal residual disease (MRD), tumor tracking, tumor mutational burden (TMB), and clonal diversity along NAT and treatment follow-up.
- To investigate if ctDNA dynamics after 2-3 weeks of treatment predicts PEPI score, molecular downstaging, and MRD in ctDNA samples, and its correlation with some efficacy variables.
- To investigate if ctDNA dynamics after 3-4 weeks post-surgery predicts PEPI score and molecular downstaging, and its correlation with some efficacy variables.
- Genetic changes in pre vs. post NAT tumor samples to identify potential genomic mechanisms of resistance.
- To describe ctDNA mutation landscape in resistant tumors.
- Generation and expansion of *ex vivo* organoid-based models and/or xenografts from post-therapy tumors and if possible, from baseline tumors to perform high-throughput functional screening studies.
- Identification of indirect biomarkers of tumor biology and treatment activity by metabolomic analysis (glutamine pathway).
- To explore other possible biomarkers of clinical activity in tumor and blood samples.
- Rate of conversion from cN1 to ypN0 at surgery.

3. ENDPOINTS AND STUDY VARIABLES

3.1 Primary Endpoint

Evaluation of the number of patients with a Residual Cancer Burden (RCB) 0-I index as a measure of efficacy. RCB (5) is a continuous variable derived from the primary tumor dimensions, cellularity of the tumor bed, and axillary nodal burden. It is estimated from routine pathological sections of the primary breast tumor site and the regional lymph nodes after the completion of NAT. The pathological variables include bidimensional diameters of the primary tumor bed, the proportion of primary tumor area containing invasive carcinoma, the number of positive lymph nodes, and the diameter of the largest nodal metastasis.

3.2 Secondary Endpoints

- The percentage of decrease in the geometric mean of Ki67 index value after 2 weeks of treatment in both treatments' arms.

  Number of patients with cell cycle arrest (Ki67 < 2.7%) after 2 weeks of treatment in both treatment arms.
- RCB is classified in four classes based on the residual disease (RD):
  - RCB-0 is defined as pathological complete response.
  - RCB-I is defined as minimal RD (>0-1.36).
  - RCB-II is defined as moderate RD (>1.36-3.28).
  - RCB-III is defined as extensive RD (>3.28).

  According to W. Fraser Symmans et al (5), two cut-off points were determined sequentially by maximizing the profile log-likelihood of a multivariate Cox model that included the clinical covariates and the dichotomized RCB index. The first cut-off point (RCB-III v RCB-I/II) was selected as the 87th percentile (RCB, 3.28), and the second (RCB-I v RCB-II) corresponds to the 40th percentile (RCB, 1.36). The cut-off points defined subgroups of RCB-0 to RCB-III with increasingly poor prognosis.

  The RCB index cannot be calculated accurately in patients whose disease remains inoperable at the end of NAT (for example, requiring additional treatment before surgical resection can be performed), or in those experiencing progression of the disease and do not undergo surgical resection at the end of NAT. For these patients, residual cancer burden assigned is RCB-III (6) .
- Variation of the value of RCB based on the RD between the two treatment arms.
- PEPI (7) requires pathological stage (tumor size and nodal status), level of Ki67 protein, and Allred ER score measured on the surgical specimen. PEPI score 0 includes pT1 or pT2, pN0, Ki67 ≤ 2.7%, Allred score > 2.
- Clinical Response Rate (CRR) is defined as the proportion of patients with complete or partial radiographic response. Complete Response (CR) and partial response (PR) definitions are assessed by MRI at baseline and prior to breast surgery, with or without regional lymph nodes surgery, and categorized according to percent reduction in breast tumor size.
- Rate of breast conservative surgery (BCS) is defined as the proportion of patients who achieved BCS between both treatment arms.

- Invasive event free survival (iEFS): defined as time from randomization to PD or invasive disease recurrence (local, regional, distant, or contralateral), or death from any cause (8). Invasive disease recurrence is defined as:
    - Ipsilateral invasive breast tumor recurrence (including second primary invasive breast cancer): an invasive breast cancer involving the same breast parenchyma as the original primary lesion.
    - Ipsilateral regional invasive breast cancer recurrence (i.e., an invasive breast cancer in the axilla, other regional lymph nodes, chest wall, and/or skin of the ipsilateral breast).
    - Distant recurrence (i.e., evidence of breast cancer in any anatomic site outside local and/or regional location and that has been either histologically confirmed or clinically diagnosed as recurrent invasive breast cancer).
    - Contralateral invasive breast cancer.
    - Second primary invasive cancer of non-breast origin.
- Safety will be assessed by standard clinical and laboratory tests (hematology, serum chemistry). AEs grade will be defined by the NCI CTCAE (National Cancer Institute Common Terminology Criteria for Adverse Events) version 5.0 (4). AEs terms will be coded according to MedDRA dictionary.
- Gene expression data provided by a multigene expression panel in sequential tumor biopsies.

3.3 Exploratory endpoints

- Variation of the Ki67 levels, determined by immunohistochemistry (IHC), between pre-treatment and after 2 weeks-treatment biopsies, and its correlation with some efficacy variables.
- Mutational and copy number variation (CNV) data analyzed in pre-treatment and surgery tumor samples (large, targeted gene panels including TP53 and MYC mutations and MYC CNVs) and its correlation with some efficacy variables.
- Gene expression data provided by a multigene expression panel in sequential tumor biopsies and its correlation with some efficacy variables.
- Monitoring of ctDNA abundance and specific tumoral mutation and genetic changes observed in ctDNA samples along treatment and follow up.

- Mutations and CNV detected in ctDNA collected after recurrence.
- Values of other proteins, metabolites, RNA or DNA alterations obtained from the tissue or blood samples and its correlation with efficacy variables could be used for assessment of biomarkers related to activity of abemaciclib and breast tumor sensitivity and/or resistance to treatment.

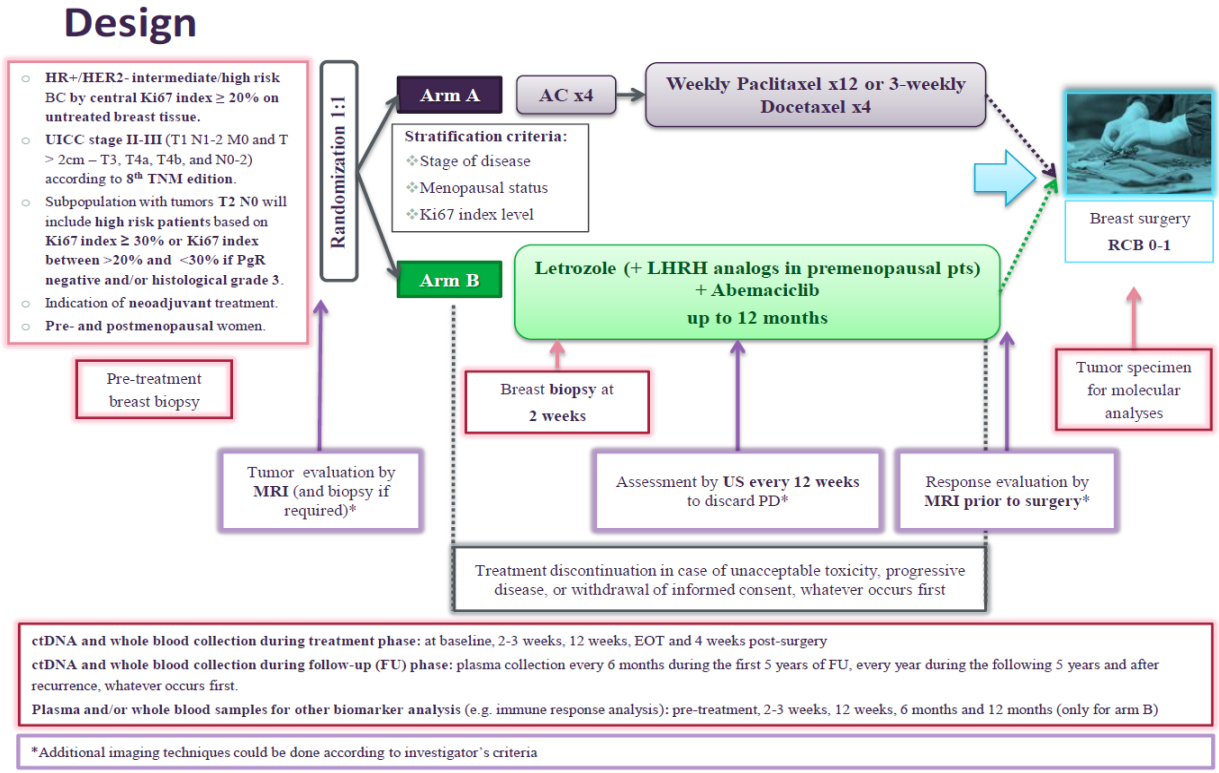
4. STUDY DESIGN

This is an international, multicenter, open-label, randomized phase II study in the NA setting.

Approximately 200 pre- and postmenopausal women with HR-positive/HER2-negative BC of intermediate/high-risk determined by Ki67 index ≥ 20% on untreated breast tissue and centrally assessed, with indication of NAT, were included. Patients with EBC on stages II-III (T1c N1-2 M0 and tumor size [T] > 2cm – T3, T4a, T4b, and lymph node involvement [N] N0-2) according to the 8th edition of the UICC TNM Classification (9). The subgroup with tumors T2N0 included high-risk patients based on Ki67 index ≥ 30% or Ki67 index ≥ 20% and < 30% if progesterone receptor (PgR) negative and/or histological grade 3.

All patients were treated according to the stipulations shown in the Figure 1, unless any of the following occurred: unacceptable toxicity, PD, or withdrawal of informed consent document (ICD), whatever occurred first.

Figure 1. Study design.



For more information review section 3.1 and Attachment 1. Study Schedule in the Study protocol.

4.1 Sample Size

A Bayesian design was used to define the most appropriate sample size:

The major advantage of this Bayesian approach is it allows us to evaluate how similar response rates between both treatment arms are, without using a very large non-inferiority study.

Comparability of RCB0/1 have to be declared if the following Bayesian criterion is achieved in the primary RCB0/1 analysis:

$$\text{Posterior P (true RCB0/1}_{\text{Chemo}} - \text{true RCB0/1}_{\text{Abema+AI}} < 5\%) > 80\%$$

Simulations were conducted to evaluate the probability of declaring comparability based on various scenarios of underlying true RCB0/1 values. Different underlying true RCB0/1 values for abemaciclib plus an AI were considered in the scenarios.

Three scenarios were explored (n=150, n=200, and n=250) and it is considered that the 2nd scenario with a sample size of 200 patients was the best option. Simulation results are shown in Table 1.

Table 1. Simulation results of the three different sample size scenarios.

$RCB0/1_{Chemo}$	$RCB0/1_{Abema+AI}$	N=150	N=200	N=250
0.16	0.16	57%	60%	61%
0.16	0.20	77%	82%	86%
0.16	0.24	90%	94%	96%
0.16	0.28	96%	98%	99%

4.2 Randomization

Approximately 200 pre- and postmenopausal women with HR-positive/HER2-negative intermediate/high-risk BC determined by central Ki67 index ≥ 20% on untreated breast tissue, with indication of NAT, were included.

Eligible patients were randomized in a 1:1 fashion to the **control arm** with standard chemotherapy (CT) based on anthracyclines and taxanes or to the **experimental arm** with letrozole + abemaciclib (+ LHRH analogs in premenopausal women). Patients were stratified according to the disease stage (II vs. III), menopausal status (premenopausal vs. postmenopausal) and Ki67 index (Ki67 < 30% vs. Ki67 ≥ 30%).

5. STUDY POPULATION

The study enrolled a total of approximately 200 pre- and postmenopausal women with HR-positive/HER2-negative intermediate/high-risk BC determined by central Ki67 index ≥ 20% on untreated breast tissue, with indication of NAT.

5.1 Intent To Treat Population (ITT)

The ITT population will include all patients who were enrolled according to the initial treatment assignment. The ITT population will be the primary population for the efficacy analysis. It will be performed a sensitivity analysis using the Per-protocol population.

5.2 Per protocol (PP) population

A subset of the ITT population including patients who received at least one dose of study treatment and completed the study without any major protocol deviations according to the protocol deviation manual.

5.3 Safety population

Safety population will include all patients enrolled in the study who received at least one dose of study treatment, according to the actual treatment received. This population is for the safety analysis.

5.4 Biomarker Population

A subset of the safety population with available samples and clinical data for biomarker analyses and excluding samples according to the following Chief Investigator´s criteria:

- Samples affected by another NAT different to the study treatment.
- No efficacy data available (ICF withdrawal or loss of follow-up).
- Patients with HER2-positive result at diagnosis.

All patients will be examined case by case to confirm exclusion from Biomarker population for each specific analysis.

Additional exclusion criteria may be included in the Exploratory Statistical Analysis Guide (ESAG) for each specific analysis.

6. DATA SCREENING AND ACCEPTANCE

6.1 Missing data

The frequency of missing data will be examined and reported for each variable in the analysis. We will not perform data imputation for missing data.

6.1.1 Missing date

See Appendix 10.1. Date Imputation Rules.

6.2 Statistical Software

All analyses will be performed using the Statistical Analysis Software (SAS) Enterprise Guide 7.1 version or R software.

6.3 Database lock

The data cut-off for the primary analysis is planned when all patients from all arms have available the information of RCB after surgery.

The data cut-off date for the final analysis will occur after all enrolled patients have finalized the study follow-up period, died, or withdraw the informed consent, whatever occurred first.

6.4 Criteria for End of Study

This study will be considered complete following the data cut-off date and data lock for the final analysis. The data cut-off date for the final analysis will occur after all enrolled patients have been followed for at least 10 years or withdrawal of the informed consent, whatever occurred first.

If further data are collected, that are not included as part of the final locked database, the post-lock data will eventually be combined with the locked database and stored in a data library, separated from the locked database.

The end date of study is the date of the last visit of the last patient including follow-up.

Performing exploratory analyses will be independent of the date of the end of the study.

6.5 Duration of the study

It was estimated that the enrollment would be approximately completed in 24 months.

Following the randomization, patients received the treatment for approximately 6 months in **Arm A** and 12 months (±14 days) in **Arm B**, unless PD, unacceptable toxicity, or withdrawal of the informed consent, whatever occurred first. After treatment the patient underwent surgery. After surgery, there was a follow-up of 10 years.

For safety reasons all patients had a visit after finishing treatment with the study medications.

In **Arm A** patients had this visit not earlier than 21 days after finishing treatment with the study medications and before definitive surgery. In **Arm B** the visit had to be done within 7 days from the last abemaciclib and/or letrozole dose.

In some patients, the end of the study treatment (at neoadjuvant setting) may be due to PD or toxicity. In case the beginning of the new therapy cannot be delayed as per the investigator's judgment, the safety visit may be performed in advance and always before starting the new anticancer therapy.

The start date of study is the date of the first site activation.

7. INTERIM ANALYSIS

Not applicable.

7.1 Purpose of Interim Analysis

Not applicable.

7.2 IDMC (Interim Data Monitoring Committee)

Not applicable.

8. STATISTICAL METHODS AND ANALYSES

8.1 Statistical Methods

Continuous data will be summarized using descriptive statistics (number of observations, mean, standard deviation, median, 25th and 75th percentiles [where specified], minimum, and maximum). Frequencies and percentages will be used for summarizing categorical variables.
Confidence intervals, when presented, will generally be constructed at the 95% level. For binomial variables, the normal approximation methods will be employed unless otherwise specified.
Chi-square $\chi 2$ or Fisher Exact test could be used to explore the relationship between qualitative variables. T-test or Mann-Whitney test could be used to compare quantitative variables between treatment arms.
All statistical tests and resulting *P*-values will be reported as 2-sided and will be assessed at α=0.05 significance level unless otherwise stated.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. Confidence intervals about a parameter estimate will be presented using the same number of decimal places as the parameter estimate. Percentages will be presented to 1 decimal place.
A month is operationally defined to be 30.4375 days.

Definitions of Baseline Values

In general, for efficacy endpoints the last observed measurement prior to randomization will be considered the baseline measurement. For safety endpoints, the last observation before the first dose of study drug will be considered the baseline measurement.

Definition of Study Days

For the purposes of efficacy data summary, Day 1 is defined as the date of randomization. For visits (or events) that occur on or after randomization, Day is defined as (date of visit [event] – date of randomization + 1). For visits (or events) that occur prior to randomization, Day is defined as (date of visit [event] – date of randomization). There is no Day 0.
For the purposes of safety data summary or calculations of time since baseline, study Day 1 is defined as the date on which a patient is administered their first dose of study drug. For visits (or events) that occur on or after the first dose of study drug, study day is defined as (date of visit [event] – date of first dose of study drug + 1). For visits (or events) that occur prior to study Day 1, study day is defined as (date of visit [event] – date of first dose of study drug). There is no study Day 0.

8.2 Statistical Analyses

8.2.1 Patient Disposition

Study information including the date when the first patient signed the Informed Consent Form (ICF), date of last patient's last visit/contact, date of last patient's last procedure for collection of data for primary endpoint, Medical Dictionary for Regulatory Activities (MedDRA) version (or/and NCI-CTCAE version), World Health Organization (WHO) Drug version and SAS Version will be generated in a summary table.
A detailed description of patient disposition will be provided. It will include:

- Summary of patients screened and by site.

- Consort flowchart.
- Total number of screened patients.
- Total number of enrolled patients.
- Total number of treated patients.
- Summary of reasons for patients enrolled, but not treated.

A detailed summary of reasons for patient discontinuation from study treatment will be provided.
A summary of all identified important protocol violations will be provided.
All percentages will be based on the number of patients in the ITT population.

8.2.2 Patient Characteristics

Patient characteristics will include a summary of the following:

- Patient demographics.
- Baseline disease characteristics.
- Pre-existing conditions/secondary conditions.
- Stratification factors.

Other patient characteristics will be summarized as deemed appropriate.
Standard descriptive statistics, such as the mean, median, range and proportion, will be used to summarize the patients' sample and to estimate parameters of interest. Ninety-five percent confidence intervals will be provided for estimates of interest, where possible.

8.2.3 Concomitant Therapy

The number and percentage of patients taking concomitant medications will be tabulated by WHO standardized medication name based on the safety population. Concomitant medications are medications taken by the patient within 28 days prior to the start of study drug or medications that started after the first dose and within 30 days of the last dose of study drug.

8.2.4 Treatment Compliance

Summaries and descriptive statistics of duration of treatment in days [(date of last dose – date of first dose + 1)], total number of cycles administered, cumulative dose for each study drug, dose intensity,

relative dose intensity and percent of compliance for all drugs will be summarized for patients in the safety population.

Number of cycles administered = A treated cycle is defined as a cycle in which the patient received any amount of study drug.

Cumulative dose (mg) = Sum of all doses (mg) administered to a patient during the treatment period.

Dose intensity (mg/day) = Cumulative dose of drug divided by the treatment duration.

Relative dose intensity (Doxorubicin) = [Doxorubicin cumulative dose / ([60*BSA*4)]) * 100.
Relative dose intensity (Cyclophosphamide) = [Cyclophosphamide cumulative dose / ([600*BSA*4)]) * 100.
Relative dose intensity (Paclitaxel) = [Paclitaxel cumulative dose / ([80*BSA*12)]) * 100.
Relative dose intensity (Docetaxel) = [Docetaxel cumulative dose / ([100*BSA*4)]) * 100.
Relative dose intensity (Abemaciclib) = [Abemaciclib dose intensity / ([300*365)]) * 100.
Relative dose intensity (Letrozole) = [Letrozole dose intensity / ([2.5*365)]) * 100.

The estimate of percent compliance will be given by:

$$\text{Percent Compliance} = \frac{\text{Actual dose administered}}{\text{Dose expected to be administered}} \times 100$$

No minimal level of compliance will be defined for patient inclusion in efficacy analyses. To be considered compliant patients should have received at least 80% of the planned number of doses. Exploratory analysis of the impact of compliance on selected efficacy endpoints may be performed if deemed necessary.

Action on Study Drug

The reason for dose modification (delay, omission, reduction) of each study drug will be summarized by cycle and overall based on the safety population.

8.2.5 Analyses of Primary Endpoint

The primary endpoint is Residual Cancer Burden (RCB) 0-I rate in both treatment arms.

Residual Cancer Burden (RCB): is a continuous variable derived from the primary tumor dimensions, cellularity of the tumor bed, and axillary nodal burden.

$$RCB = 1.4(f_{inv}d_{prim})^{0.17} + [4(1 - 0.75^{LN})d_{met}]^{0.17}$$

- f_{inv} = the proportion of the primary tumor bed that contains invasive carcinoma.
- d_{prim} = the post-treatment surgical resection specimen could be determined from bidimensional diameters of the primary tumor bed in the resection specimen (d1 and d2); bidimensional measurements of the primary tumor bed (millimeters) were combined as follows: $dprim = \sqrt{d1d2}$.
- LN = the number of axillary lymph nodes containing metastatic carcinoma.
- d_{met} = the diameter of the largest metastasis in an axillary lymph node.

The calculation formula and detailed description can be found at: www.mdanderson.org/breastcancer_RCB. In brief, the variables include cross-sectional dimensions of the residual tumor bed (d1 and d2), estimate of the proportion of that residual tumor bed area that is involved by cancer (% CA), estimate the proportion of the cancer that is in situ component (% CIS), number of positive lymph nodes (LN), and measure of the diameter of the largest nodal metastasis (dmet).

The information about the result of RCB will be provided by the local laboratory. If the patient has no information about the RCB, the patient will be considered as non-responder in the RCB rate analysis.

RCB rate on each treatment arm will be estimated by dividing the number of patients with RCB 0-I by the ITT patients with measurable disease by treatment arm (“response rate”).

$$\text{RCB Rate} = \frac{\text{Number of RCB 0-I}}{\text{ITT population with measurable disease}}$$

Comparability between the two treatment arms will be assessed using the following Bayesian criterion:

Posterior P(true RCB0/1Chemo – true RCB0/1Abema+AI < 5%) > 80%

To obtain a **posterior distribution** $P(\theta|x)$ will use a Bayes Theorem

$$\mathrm{P}(\theta|x) = \frac{\mathrm{P}(X|\theta)\,\mathrm{P}(\theta)}{\mathrm{P}(\mathrm{X})} \propto \mathrm{P}(X|\theta)\mathrm{P}(\theta)$$

Where:

$\mathrm{P}(X|\theta)$ is the **Likelihood**. Quantifies the information provided by the data on the parameter θ.

$\mathrm{P}(\theta)$ is the **prior distribution**. The information of the parameter θ before knowing the data.

The RCB0/1 in each arm follows a binomial distribution k~ Bi (p, n). A non-informative prior, Beta (0.5, 0.5) or Beta (1,1), will be used. Then, the posterior distribution will be obtained by simulations. Posterior medians and Bayesian credible intervals will be reported for the RCB0/1 rate for each arm.

To test Posterior P(true RCB0/1Chemo – true RCB0/1Abema+AI < 5%) > 80%, the differences in the posterior distributions, posterior RCB0/1Chemo – posterior RCB0/1Abema+AI, will be calculated for each simulation. Then, the posterior differences less than 5% found among all the simulations will be summarized to see if there are more than 80%.

In addition, arms will be compared using a Cochran-Mantel-Haenszel (CMH) test using the stratification factors. The Mantel-Haenszel estimate of the odds ratio and the associated 95% CIs will be presented. The accompanying 2-sided 95% confidence intervals (CIs) in each arm and the difference between arms will be computed.

Additionally, a similar analysis will be also performed in the PP population as a sensitivity analysis.

8.2.6 Analysis of Secondary Endpoints

As secondary endpoints, these analyses are considered as supportive of the primary endpoint, and the emphasis of interpretation will be on estimates rather than hypothesis tests.

For patients who have received other neoadjuvant treatment, residual cancer burden score will be 3.8 and other secondary endpoints (PEPI score 0 at breast surgery, clinical response, breast conservative surgery and OncotypeDx post-treatment) will be considered missing.

Changes in Ki67 index value after 2 weeks of treatment in both treatment arms: it will be evaluated the percentage of decrease in the geometric mean of Ki67 index value after 2 weeks of treatment in both treatments' arms. Values after 2 weeks of treatment will be expressed as geometric mean proportion of the baseline and transformed into percentage changes.
Ki67 at baseline (Ki67pre), after 2 weeks (Ki67w2) and the percentage reduction in Ki67 from baseline defined as 100 × (Ki67pre - Ki67w2) / Ki67pre will be summarized by treatment groups using descriptive statistics (such as the number of observations, arithmetic mean, geometric mean, median, standard deviation (SD), minimum, and maximum).

A specific cut-off value will be determined to classify the patients in the low or high categories. The rate of change from high to low-risk category will be calculated and compared between both arms.

An ANOVA analysis will be conducted at a two-sided 5% significance level for a within-treatments and between-treatment comparison. This analysis will be conducted in the ITT population.
Also, to measure the changes in Ki67 after 2 weeks of treatment it will be assessed the number of patients with cell cycle arrest (Ki67 < 2.7%) after 2 weeks of treatment in both treatment arms in the ITT population.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis and in the Biomarker population as an exploratory analysis.

Residual Cancer Burden (RCB): RCB is classified in four classes based on the Residual Disease (RD):

- RCB-0 is defined as pathological complete response.
- RCB-I is defined as minimal RD (>0 – 1.36).
- RCB-II is defined as moderate RD (1.36 – 3.28).
- RCB-III is defined as extensive RD (>3.28).

RCB response for each patient will be summarized; ordered from best to worst.

The differences between RCB 0+I vs. RCB-II vs. RCB-III in both treatment arms will be assessed, and the RCB distribution on each arm will be estimated. For this analysis, the ITT population will be used. Chi-square or Kruskal Wallis test will be used to examine differences between treatment arms.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis.

Residual Cancer Burden (RCB) value: RCB value will be obtained from the residual disease. It will be evaluated the differences of mean between both treatment arms. For this analysis it will be used the ITT population.

The T-test or Wilcoxon test will be used to examine differences between treatment arms.

Additionally, a similar analysis will be also performed in the PP population as a sensitivity analysis.

Rate of PEPI score 0 at breast surgery: Rate of PEPI score of 0 at baseline will be estimated by dividing the number of patients with PEPI score of 0 at surgery by the ITT patients in both arms.

$$\text{Rate of PEPI score 0 at surgery} = \frac{\text{Number of patients with PEPI score 0 at surgery}}{\text{ITT patients}}$$

This rate will be reported, including a 95% confidence interval.

Rate of PEPI score 0 at breast surgery comparison between the two treatment arms will be assessed using a stratified CMH test based on the original stratification factors. The Mantel-Haenszel estimate of the odds ratio and the associated 95% CIs will be presented.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis and in the Biomarker population as an exploratory analysis.

Table 2. PEPI score.

Preoperative Endocrine Prognostic Index (PEPI)*				
Pathology, biomarker status	**Recurrence-Free Survival**		**Breast Cancer-Specific Survival**	
	HR	**Points**	**HR**	**Points**
Pathological tumor size				
T1/2	-	0	-	0
T3/4	2.8	3	4.4	3
Nodal status				
Negative	-	0	-	0
Positive	3.2	3	3.9	3
Ki67 level				
0% - 2.7% (0-1**)	-	0	-	0
> 2.7% - 7.3% (1-2**)	1.3	1	1.4	1
> 7.3% - 19.7% (2-3**)	1.7	1	2.0	2
> 19.7% - 53.1% (3-4**)	2.2	2	2.7	3
> 53.1% (> 4)	2.9	3	3.8	3
ER status, Allred score				
0-2	2.8	3	7.0	3
3-8	-	0	-	0

* The total PEPI score assigned to each patient is the sum of the risk points derived from the pT stage, pN stage, Ki67 level, and ER status of the surgical specimen. The total risk point score for each patient is the sum of all the risk points accumulated from the four factors in the model.
*The natural logarithm interval corresponding to the percent Ki67 values on the original percentage scale.
Abbreviations: T: tumor size. N: nodal status. ER: estrogen receptor.

Points from Breast Cancer-Specific Survival will be used (Table 2). The total PEPI score is the sum of the risk points derived from the pT stage, the pN stage, Ki67 levels and ER status of the surgical regimen. The total score after treatment could be summarized [quantitatively and by 3 risk categories: 0 (low risk), 1-3 (medium risk) and >=4 (high risk)] for each treatment arm.

Clinical Response Rate (CRR): A patient will be considered to have achieved a CR if the patient has a sustained CR or PR assessed by MRI. Otherwise, the patient will be considered as non-responder in the CRR rate analysis. Additionally, patients with inadequate data for tumor assessment (e.g., no baseline assessment or no follow-up assessments) will be considered as non-responders in the CRR rate analysis.

CRR rate on each treatment arm will be estimated by dividing the number of patients with objective response (CR or PR) by the ITT patients with measurable disease by treatment arm ("response rate").

$$\text{Objective Response Rate} = \frac{\text{Number of CRs + PRs}}{\text{ITT population with measurable disease}}$$

The CRR rate will be reported, including a 95% confidence interval.

In addition, the best clinical response for each patient will be summarized by treatment arm. CRR rate comparison between the two treatment arms will be assessed using a stratified CMH test based on the original stratification factors. The Mantel-Haenszel estimate of the odds ratio and the associated 95% CIs will be presented.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis.

Rate of breast conservative surgery (BCS): Defined as the proportion of patients who achieved breast-conserving surgery between both treatment arms.

BCS rate on each treatment arm will be estimated by dividing the number of patients with conservative surgery by the ITT population.

$$\text{BCS Rate} = \frac{\text{Number of patients with conservative surgery}}{\text{ITT population}}$$

The BCS rate will be reported, including a 95% confidence interval.

BCS rate comparison between the two treatment arms will be assessed using a stratified CMH test based on the original stratification factors. The Mantel-Haenszel estimate of the odds ratio and the associated 95% CIs will be presented.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis.

Invasive Event Free Survival (iEFS):

iEFS in months is defined as the time from the date of randomization to the date of first documentation of PD or invasive recurrence or death due to any cause, whichever occurs first.
iEFS (months) = (earliest date of PD or invasive recurrence or death – date of randomization + 1)/30.4375.
iEFS data will be censored on the date of the last tumor assessment on study for patients who do not have PD or invasive recurrence (local, regional, distant, contralateral, or second primary invasive cancer of non-breast origin), and who have not died due to any cause while on study.

iEFS will be analyzed in the ITT population. A stratified and non-stratified log-rank test (two-sided) will be used to compare iEFS time between treatment arms at the final analysis. iEFS for the two arms will be assessed using Kaplan-Meier methods and displayed graphically where appropriate. The median event times and 95% CIs will be estimated. Cox regression models will be used to estimate the treatment hazard ratio and its 95% confidence interval.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis.

Molecular downstaging for high-risk genomic groups defined by a multigene expression panel:

For each patient the Recurrence Score (RS) will be obtained from the OncotypeDx® in pre-treatment and post-treatment. The values of RS are from 0 to 100. It will be calculated the change (post-treatment to pre-treatment) and evaluated the differences between both treatment arms. The RS in pre- and post-treatment will be categorized as low (RS ≤ 25) and high (RS 26 – 100), and the rate of change from high to low-risk category will be calculated and compared between both arms.

OncotypeDx also provides the Estrogen Receptor (ER), Progesterone receptor (PR) and HER2 scores and the cut-off values to classify them as positive or negative. Similarly, to the RS analysis described above, it will be calculated the change (post-treatment to pre-treatment) and evaluated the differences between both treatment arms. The rate of change from positive to negative category will be calculated and compared between both arms.

For these analyses it will be used the patients from ITT population with available score in pre and post treatment.

Descriptive statistics including the 95% CI around mean for pre and post values and for the change (post – pre) in will be calculated. The T-test or Wilcoxon test will be used to examine differences in values between pre and post treatment and between treatments arms. The Chi-square $\chi 2$ or Fisher Exact test will be used to assess if there are differences in RS high (26 -100) and RS low (≤ 25), and in positive and negative between pre- and post-treatment.

Additionally, a similar analysis will also be performed in the PP population as a sensitivity analysis and in the Biomarker population as an exploratory analysis.

8.2.7 Safety Analyses

The toxicity and tolerability of study drugs/medications will be evaluated in the safety population. Safety analyses will include summaries of the incidence of adverse events (AEs) by maximum NCI-CTCAE grade (v5.0; NCI 2017) that occur during the study treatment period or within 30 days of the last dose of study treatment, regardless of causality and according to the relationship to study drug/medication as assessed by the investigator. Additionally, the following safety-related outcomes will be summarized:

- Study treatment discontinuations due to AEs.
- Deaths.
- Serious AEs (SAEs), Serious Adverse Reactions (SARs) and AEs of Special Interest (AESIs).
- Hospitalizations and transfusions.

Use of key concomitant medications, including colony stimulating factors (CSF) of granulocytes (G-CSF).

Analyses for data with discrete dates, for example, deaths, transfusions, and concomitant medications, will be done 30 days after each patient's last dose of study treatment. Adverse events will also be analyzed in this timeframe; that is, if an event starts within 30 days of discontinuation from the study treatment, but 30 days after the last dose of study treatment, it will not be included.
Adverse event summaries to be presented include treatment-emergent AEs (TEAEs) and SAEs by severity grade and causality as well as deaths and discontinuations (of treatment) due to TEAEs. The number and percentage of patients with at least one TEAE will be summarized by system organ class (SOC) and preferred term (PT); further separate summaries will be presented by severity and relationship. AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary (version 24.1 or above), and by severity according to NCI CTCAE criteria v5.0.

Tabular summaries will be provided for the following:

- TEAEs.
- TEAEs by severity.
- Drug-related TEAEs.
- Drug-related TEAEs by severity.
- Grade 3 or higher TEAEs.
- Grade 3 or higher drug-related TEAEs.
- Most reported TEAEs (at least 10% in any arm, sorted by preferred term [PT]).
- SAEs.
- SAEs by severity.
- TEAEs leading to discontinuation.
- TEAEs leading to death.

Patients reporting the same event more than once will have that event counted only once within each body system, and once within each PT.
An internal classification of TEAEs (20230621_AEs_Unification Preferred Term and SOC MedDRA.xlsx) will be used to summarize them.
Hematological and clinical biochemistry toxicities will be assessed from laboratory test parameters.
The safety analysis will be performed in the safety population.

Deaths

All-cause mortality will be tabulated, which includes death of all causes, and deaths related to breast cancer. On-study deaths (from first dose to 30 days after the last dose) will be tabulated including deaths related to breast cancer, deaths due to study treatment, and deaths within 30 days of first dose.

8.2.8 Exploratory Analyses

Exploratory analyses will be outlined in a specific ESAG for each particular analysis.

8.2.9 Other Analyses

8.2.9.1 Subgroup Analysis

Exploratory subgroup analysis may be performed if deemed appropriate.

9. TABLES Y FIGURES

Table . Study Information
Table . Screen Failures
Table . Number of Patients Randomized by Site
Table . Disposition of Patients
Table . Analysis Sets
Table . Demographic Characteristics
Table . Clinical-Pathological Characteristics
Table . Local Laboratory Biomarkers
Table . Central Laboratory Biomarkers
Table . Stratification Factors
Table . Medical History by Preferred Term and SOC
Table . Concomitant Medications by Standardized Medication Name
Table . Concomitant Procedures
Table . Summary of Study Drug Exposure
Table . Summary of Dose Modifications
Table . Breast Surgery
Table . TNM Post-Surgery
Table . Adjuvant Therapies for Breast Cancer
Table . Residual Cancer Burden
Table . Changes in Ki67 index value after 2 weeks of treatment in both treatment arms
Table . PEPI Score

Table . Clinical Response
Figure . Waterfall Plot
Table . Breast Conservative Surgery
Table . Invasive Event-Free Survival (iEFS)
Figure . invasive Event-Free Survival (iEFS)
Table . Recurrence Score from OncotypeDx
Table . Overview of Treatment-Emergent Adverse Events
Table . Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . Treatment-Emergent Adverse Events by System Organ Class, preferred term and worst CTCAE Grade
Table . Most common (at least 10% in any treatment arm) Treatment-Emergent Adverse Events by Preferred Term
Table . Drug-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . Drug-Related Treatment-Emergent Adverse Events by System Organ Class, preferred term and worst CTCAE Grade
Table . Grade 3 or Higher Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . Grade 3 or Higher Drug-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . Serious Treatment-Emergent Adverse Events by System Organ Class, preferred term and worst CTCAE Grade
Table . Treatment-Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term
Table . Adverse Events from Laboratory according NCI CTC v5.0 by System Organ Class, preferred term and worst CTCAE Grade
Table . Most Frequent (>5% in any treatment arm) Non-Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
Table . All-Cause Mortality

10. APPENDIX

10.1 Date Imputation Rules

Incomplete dates in the screening period

1. If only the day-component is missing, the first day of the month will be used if the year and the month are the same as those for the first dose of study drug. Otherwise, the 15th will be used.
2. If only the year is present, and it is the same as the year of the first dose of study drug, the 15th of January will be used unless it is later than the first dose, in which case the date of the first of January will be used.
3. If only the year is present, and it is not the same as the year of the first dose of study drug, the 15th of June will be used.

Incomplete AE onset date

Assumption: For on-study AEs.

If year is missing (or completely missing): set to the date of first dose.

If (year is present and month and day are missing) or (year and day are present and month is missing):

If year = year of first dose: set the date to the first dose date.

If year < year of first dose: set month and day to December 31st.

If year > year of first dose: set month and day to January 1st.

If month and year are present and day is missing:

If year = year of first dose, and:

If month = month of first dose: set day to day of first dose.

If month < month of first dose: set day to last day of month.

If month > month of first dose: set day to 1st day of month.

If year < year of first dose: set day to last day of month.

If year > year of first dose: set day to 1st day of month.

For all other cases: set to date of first dose.

Incomplete Concomitant Medication Start Date

If year is missing (or completely missing): do not impute.

If (year is present and month and day are missing) or (year and day are present and month is missing):

Set month and day to January 1st.

If year and month are present and day is missing:

Set day to 1st day of month.

Incomplete Concomitant Medication End Date

If year is missing (or completely missing): do not impute.

If (year is present and month and day are missing) or (year and day are present and month is missing):

Set month and day to December 31st.

If year and month are present and day is missing:

Set day to last day of the month.

11. BIBLIOGRAFHY /REFERENCES

1. ***CARABELA_Protocol version 2.0_29Jun2021.***
2. *ICH E9, Statistical Principles for Clinical Trials.*
3. *ICH E3, Structure and Content of Clinical Study Reports.*
4. *NCI-CTCAE version 5.0.*
5. *Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy.* Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. J Clin Oncol. 2007; 25(28):4414-22.
6. *Residual Cancer Burden Assessment Manual CARABELA study v1.0 dated 09Jul202.*
7. *Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics.* Ellis MJ, Tao Y, Luo J, A'Hern R, Evans DB, Bhatnagar AS, et al. J Natl Cancer Inst. 2008;100(19):1380-8.
8. *Updated Standarized Definitions for Efficacy Endpoints (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0.* al., Sara M Tolaney et. J. Clin Oncol 2021.
9. *The Union for International Cancer Control's (UICC). UICC TNM classification of malignant tumours 2017 [updated 30th of June 2017. Available from: https://www.uicc.org/news/8th-edition-uicc-tnm-classification-malignant-tumors-published.*

12. SUMMARY OF CHANGES FROM PREVIOUS VERSION

Version No.	Effective Date	Modified section	Description of changes
1	*30-06-2023*	*NA*	*Creation of document*