Document Type: SAP

Official Title: A Phase II Randomized, Double-Blind Study of Neoadjuvant Letrozole Plus GDC-0032 Versus Letrozole plus Placebo in Postmenopausal Women with ER-positive/HER2-negative, Early Stage Breast Cancer

NCT Number: NCT02273973

Document Date: SAP: 22 Jan 2018

ABCSG
Austrian Breast and Colorectal Cancer Study Group
Nussdorfer Platz 8/12-13, A-1190 Vienna, Austria
Phone: +43 (0) 1 408 92 30 / 27, Fax: +43 (0) 1 409 09 90

Statistical Analysis Plan

Drug substances: Letrozole, GDC-0032
Study code: GO28888/BIG-3-13/SOLTI 1205/ABCSG 38

A PHASE II RANDOMIZED, DOUBLE-BLIND STUDY OF NEOADJUVANT LETROZOLE PLUS GDC-0032 VERSUS LETROZOLE PLUS PLACEBO IN POSTMENOPAUSAL WOMEN WITH ER POSITIVE/HER2-NEGATIVE, EARLY STAGE BREAST CANCER

Austrian Breast and Colorectal Cancer Study Group
ABCSG Protocol number 38

Compiled/edited by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	18/JAN/2018
Signature	[redacted]

Reviewed by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	18/JAN/20[redacted]
Signature	[redacted]

Approved by:

VHIO	[redacted]	BIG	[redacted]
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	
Signature		Signature	
ABCSG	[redacted]	Genentech	[redacted]
Date *(dd/mmm/yyy*	[redacted]	Date *(dd/mmm/yyyy)*	
Signature		Signature	

Document status / [redacted] 3.0
Release date: 2017-09-28

ABCSG
Austrian Breast and Colorectal Cancer Study Group
Nussdorfer Platz 8/12-13, A-1190 Vienna, Austria
Phone: +43 (0) 1 408 92 30 / 27, Fax: +43 (0) 1 409 09 90

Statistical Analysis Plan

Drug substances: Letrozole, GDC-0032
Study code: GO28888/BIG-3-13/SOLTI 1205/ABCSG 38

A PHASE II RANDOMIZED, DOUBLE-BLIND STUDY OF NEOADJUVANT LETROZOLE PLUS GDC-0032 VERSUS LETROZOLE PLUS PLACEBO IN POSTMENOPAUSAL WOMEN WITH ER POSITIVE/HER2-NEGATIVE, EARLY STAGE BREAST CANCER

Austrian Breast and Colorectal Cancer Study Group
ABCSG Protocol number 38

Compiled/edited by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Reviewed by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Approved by:

VHIO	[redacted]	BIG	[redacted]
Date *(dd/mmm/yyyy)*	22 JAN 2018	Date *(dd/mmm/yyyy)*	
Signature	[redacted]	Signature	
ABCSG		Genentech	[redacted]
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	
Signature		Signature	

Document status / version: 3.0
Release date: 2017-09-28

ABCSG
Austrian Breast and Colorectal Cancer Study Group
Nussdorfer Platz 8/12-13, A-1190 Vienna, Austria
Phone: +43 (0) 1 408 92 30 / 27, Fax: +43 (0) 1 409 09 90

Statistical Analysis Plan

Drug substances: Letrozole, GDC-0032
Study code: GO28888/BIG-3-13/SOLTI 1205/ABCSG 38

A PHASE II RANDOMIZED, DOUBLE-BLIND STUDY OF NEOADJUVANT LETROZOLE PLUS GDC-0032 VERSUS LETROZOLE PLUS PLACEBO IN POSTMENOPAUSAL WOMEN WITH ER POSITIVE/HER2-NEGATIVE, EARLY STAGE BREAST CANCER

Austrian Breast and Colorectal Cancer Study Group
ABCSG Protocol number 38

Compiled/edited by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Reviewed by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Approved by:

VHIO	[redacted]	BIG	[redacted]
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	[redacted]
Signature		Signature	
ABCSG	[redacted]	Genentech	
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	
Signature		Signature	

Document status / version: 3.0
Release date: 2017-09-28

ABCSG
Austrian Breast and Colorectal Cancer Study Group
Nussdorfer Platz 8/12-13, A-1190 Vienna, Austria
Phone: +43 (0) 1 408 92 30 / 27, Fax: +43 (0) 1 409 09 90

Statistical Analysis Plan

Drug substances: Letrozole, GDC-0032
Study code: GO28888/BIG-3-13/SOLTI 1205/ABCSG 38

A PHASE II RANDOMIZED, DOUBLE-BLIND STUDY OF NEOADJUVANT LETROZOLE PLUS GDC-0032 VERSUS LETROZOLE PLUS PLACEBO IN POSTMENOPAUSAL WOMEN WITH ER POSITIVE/HER2-NEGATIVE, EARLY STAGE BREAST CANCER

Austrian Breast and Colorectal Cancer Study Group
ABCSG Protocol number 38

Compiled/edited by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Reviewed by:

ABCSG	[redacted]
Date *(dd/mmm/yyyy)*	
Signature	

Approved by:

VHIO	[redacted]	BIG	[redacted]
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	
Signature		Signature	
ABCSG	[redacted]	Genentech	[redacted]
Date *(dd/mmm/yyyy)*		Date *(dd/mmm/yyyy)*	1/22/2018
Signature		Signature	[redacted]

Document status / version: 3.0
Release date: 2017-09-28

Table of contents

Glossary of abbreviations

Abbreviation	Definition
ABCSG	Austrian Breast and Colorectal Cancer Study Group
AJCC	American Joint Committee on Cancer
AE	adverse events
AESI	adverse event of special interest
BC	breast cancer
BCS	breast conserving surgery
BTR	BioTelemetry Research
C_{max}	maximum plasma concentration
C_{min}	minimum concentration under steady-state conditions within a dosing interval
CR	complete response
CT	computed tomography
DLCO	Diffusing capacity of the lung for carbon monoxide
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
EOS	End of study
EOT	End of treatment
ER	estrogen receptor
FNA	fine needle aspiration
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
HRQoL	health-related quality of life
ICF	Informed Consent Form
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IDMC	Independent Data Monitoring Committee
IP	Investigational product
IPD	Important protocol deviation
IRF	Independent Review Facility
ITT	Intention-to-treat
IxRS	interactive voice or web-based response system
LPLV	last patient, last visit
LoPO	List of Planned Outputs
MAPK	mitogen-activated protein kinase
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
MT	mutant
NCI CTCAE	National Cancer Institute Common Terminology Criteria

Abbreviation	Definition
	for Adverse Events
ND	no disease
NE	not evaluable
ORR	objective response rate
pCR	pathologic complete response
PD	progressive disease
PEPI	preoperative endocrine prognostic index
PgR	progesterone receptor
PP	Per-protocol
PR	partial response
PT	Preferred Term
PTEN	phosphatase tensin homolog
QD	once daily
QLQ-BR23	Quality of Life Questionnaire Breast Cancer Module
QLQ-C30	Quality of Life Questionnaire Core 30
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAF	Safety
SAP	Statistical Analysis Plan
SD	stable disease
SOC	System Organ Class
TLR	Top Line Results
TSAP	Translational Statistical Analysis Plan
U/S	Ultrasound
WT	Wild type

1. Introduction

This Statistical Analysis Plan (SAP) is based on the Clinical Study Protocol for the ABCSG study 38 (external study code: GO28888/BIG-3-13/SOLTI 1205) in its third amended version from July 27, 2015) and provides a detailed description of all statistical analyses planned to be conducted within this trial at predefined timepoints.

All analyses described herein are performed in accordance to the guidelines of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

2. Study details

2.1. Study objectives

2.1.1. Primary objective

The primary objective of this study is to evaluate the efficacy of letrozole plus GDC-0032 versus letrozole plus placebo in women with ER+/HER2- early stage breast cancer, as measured by the following co-primary objectives:

- Tumor overall objective response rate (ORR) by centrally assessed breast magnetic resonance imaging (MRI) via modified Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 (Appendix 1~~Appendix 1~~) in all enrolled patients and PIK3CA mutant (MT) patients

- pCR rate in breast and axilla (ypT0/Tis, ypN0 as defined in the American Joint Committee on Cancer (AJCC) staging system) by local evaluation in all enrolled patients and PIK3CA MT patients

Information on PIK3CA mutation status is assessed centrally.

2.1.2. Secondary objectives

The secondary efficacy objectives of this study are the following:

- Tumor overall ORR, assessed by centrally assessed breast MRI via modified RECIST 1.1 in PIK3CA wild type (WT) patients

- pCR rate in breast and axilla (total pCR ypT0/Tis ypN0) by local evaluation in PIK3CA WT patients

The following secondary objectives will be performed in all enrolled patients and separated by PIK3CA mutation status:

- Compare letrozole plus GDC-0032 with letrozole plus placebo in terms of locally assessed ORR as measured by modified RECIST 1.1 criteria using the following methods:
    - Breast ultrasound
    - Clinical breast exam (i.e., palpation)
    - Mammography
- Central assessment of changes in Ki67 levels upon treatment with letrozole plus GDC-0032 versus letrozole plus placebo from baseline to Week 3; baseline to surgery; and Week 3 to surgery
- Compare the centrally derived, preoperative endocrine prognostic index (PEPI) score upon treatment with letrozole plus GDC-0032 versus letrozole plus placebo.
- Evaluate the changes in enhancing tumor volume from baseline to surgery as measured by breast MRI via central assessment.
- Evaluate different definitions of pCR including the following: a) ypT0, ypN0, and b) ypT0/is, ypNX (breast pCR).

2.1.3. Safety objectives

The safety objective for this study is to evaluate the safety of letrozole plus GDC-0032 versus letrozole plus placebo.

The safety and tolerability of GDC-0032 will be assessed using the following safety outcome measures:

- Incidence, nature, and severity of adverse events (AEs) graded according to National Cancer Institute Common Terminology Criteria for AEs (NCI CTCAE), v4.0
- Incidence and type of AEs leading to dose discontinuation, modification, or delay
- Serious adverse events (SAEs)
- Protocol-defined AEs of special interest (AESI)
- Clinically significant changes in vital signs and in clinical laboratory results during the AE reporting period

2.1.4. Patient-reported outcome objectives

The patient-reported outcome (PRO) objectives for this study are as follows:

- Evaluate and compare PROs of treatment-related symptoms, patient functioning, and health-related quality of life (HRQoL), including side-effects of therapy (e.g., sore mouth/tongue, difficulty swallowing, diarrhea, skin problems), between treatment arms as measured by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30) and the modified Breast Cancer Module (QLQ-BR23)

2.1.5. Exploratory objectives

The exploratory objectives for this study are as follows:

- ORR, pCR rate, and PEPI scores according to the decrease in Ki67 after 2 weeks of letrozole plus GDC-0032 and letrozole plus placebo.
- To evaluate changes in tumor cellular composition as assessed by diffusion-weighted MRI
- To assess whether biomarkers from tumor tissue or blood, including but not limited to somatic cancer associated mutations, phosphatase tensin homolog (PTEN) expression, pro-survival pathways (such as phosphatidylinositol-3-kinase (PI3K)/AKT, mitogen-activated protein kinase (MAPK) etc.), apoptotic markers, hormone receptor expression levels, and levels of RNA and DNA expression are predictive of response
- To determine whether inhibition of PI3K with GDC-0032 results in changes in downstream markers in tumor tissue and to examine the relationship to anti-tumor activity
- To assess concordance and percentage of PIK3CA mutation status from baseline biopsy and surgical specimen
- To assess emergence of resistance alleles from tumor tissue or blood

- To assess concordance of the different imaging modalities (MRI [volume, enhancement, diffusion metrics], ultrasound, mammography) in measuring tumor response
- To assess the pharmacokinetics and possible drug interaction between letrozole and GDC-0032 upon concomitant administration
- To assess the correlation of GDC-0032 drug levels and GDC-0032 related response (efficacy or AEs [e.g., colitis, rash])
- To assess the influence of pharmacogenetic polymorphisms on GDC-0032 and/or letrozole on pharmacokinetics (e.g. drug metabolizing enzymes and transporters) and response (either efficacy and/or AEs)
- Compare the rates of breast-conserving surgery (BCS) and conversion to BCS in letrozole plus GDC-0032 versus letrozole plus placebo

2.2. Study design

This is a two-arm, randomized, double-blind, multicenter, pre-operative study to evaluate the effect of combining letrozole and GDC-0032 versus letrozole and placebo in postmenopausal women with estrogen receptor positive (ER+) / human epidermal growth factor receptor 2 negative (HER2-) untreated, Stage I-III operable breast cancer whose primary tumors are ≥ 2 cm. Patients with cT4 or cN3 tumors are not eligible. Standard of care assessments/procedures (e.g., bilateral mammogram) performed within 28 days of Day 1 dosing do not need to be repeated for screening purposes. After confirmation of all the eligibility criteria, patients will be randomized to one of the two treatment arms. The study will enroll approximately 330 patients at approximately 110 global sites.

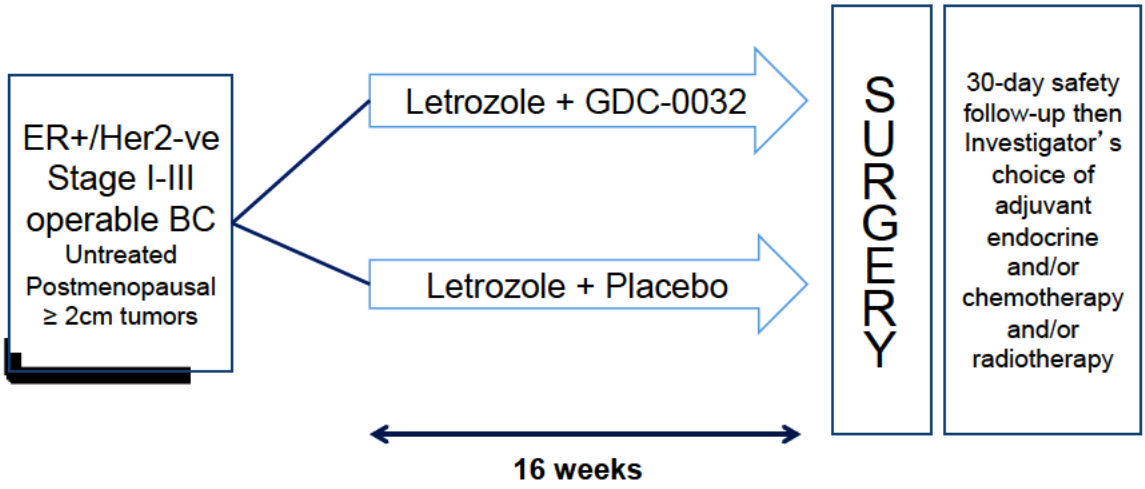
At Weeks 1, 5, 9, 13, and 16 the primary breast tumor and axillary lymph nodes will be assessed by clinical breast examination (palpation and caliper measurement). Suspicion of progression based on clinical exam at any time should be further evaluated. In addition to the safety assessments conducted at the scheduled follow-up visits, patients will be contacted by telephone for a general assessment of adverse events at Weeks 7 and 11.

At Week 9, a breast ultrasound will be performed to ensure that there is no progressive disease (PD) and for the purpose of surgery planning. Suspicion of PD on breast ultrasound should be confirmed by investigator-assessed breast MRI. Patients with primary disease not evaluable (NE) by ultrasound at baseline should be assessed by MRI at Week 9. Suspected progression in nodes should also be confirmed by fine needle aspiration (FNA) if these nodes had not been previously shown to be cytologically positive for cancer. Patients with PD (as defined by modified RECIST 1.1, Appendix 1~~Appendix 1~~), can either proceed directly to surgery or be taken off of the study, according to the investigator's decision. If the patient goes off-study, every reasonable effort should be made to obtain a new biopsy prior to beginning another systemic treatment.

During Week 16, breast MRI will be done for the purpose of primary endpoint analysis. Clinical breast exam, breast ultrasound, and mammography will also be conducted prior to surgery for the purpose of the secondary endpoint analysis.

Blood samples for exploratory endpoint analysis will be collected on Day 1 prior to dosing, at Week 9, and prior to surgery (Week 16 visit) and at the 4-weeks postsurgical follow-up visit.

Figure 1 Study Schema



	Pretreatment	Day 15 (Week 3)	Week 9	Week 16	Surgery (Week 17-18)
Tumor tissue	●	●			●
MRI	●		●	●	
Breast U/S	●		●	●	
Mammogram	●			●	

BC=breast cancer; ER+=estrogen receptor positive; MRI=magnetic resonance imaging; QD=once daily; U/S=ultrasound.

Suspicion of PD on breast ultrasound should be confirmed by investigator-assessed breast MRI. Patients with primary disease not evaluable (NE) by ultrasound at baseline should be assessed by MRI at Week 9.

2.2.1. Surgery

Surgery will take place after at least 16 weeks of treatment, during Weeks 17 – 18. Surgery should be performed within 4 days after the last dose of GDC-0032, if possible, to best observe pharmacodynamic knockdown with GDC-0032 in the surgical specimen. Surgery may be delayed due to toxicity or other safety issues. A maximum of 16 weeks of GDC-0032 can be administered. Letrozole can be continued up to surgery by investigators decision. Information on the type of surgery will be collected and recorded. Surgery specimens will be collected for histological examination to assess for pathologic complete response (pCR) and

for other endpoint analyses. The co-primary efficacy endpoint of the primary objective, i.e. pCR (ypT0/is, ypN0), will be established via a local review following completion of neoadjuvant therapy and surgery.

An Independent Review Facility (IRF) will be used to determine the tumor ORR via MRI. IRF procedures are detailed in the IRF charter.

A post-surgery visit will be performed 4 weeks (+ 1 week) after surgery, and will mark the end of the study. Assessment of AEs and general safety will be collected at this visit.

2.2.2. Study duration – end of study

The end of study (EOS) is defined as the date when the last patient has her post-surgery visit. The total duration of the study is expected to be approximately 24 months for enrollment, plus 5.5 months after last patient in.

2.3. Randomization and stratification criteria

After written informed consent form (ICF) has been obtained and eligibility has been established, the study site will obtain a patient's identification number and treatment assignment using a permuted block randomization algorithm via an interactive voice or web-based response system (IxRS).

Patients will be randomized into one of the two treatment arms in a 1:1 ratio based on the following stratification factors:

- Tumor Size (T1-T2 vs. T3)
- Nodal Status (cytologically positive vs. radiologically or cytologically negative). If on ultrasound examination there is evidence of suspicious axillary lymph nodes at the baseline examination, then FNA or core biopsy is required to confirm nodal status.

2.3.1. Blinding

Investigators and patients will be blinded to treatment assignment of GDC-0032 or placebo.

For emergency situations, the investigator will be able to break the treatment code by contacting the IxRS. The responsibility to break the treatment code in emergency situations resides solely with the investigator. For non-emergency situations, the investigator needs to obtain approval from the Medical Monitor to break the treatment code. Unblinding during the study will result in the withdrawal of a patient from the study. For regulatory reporting purposes, and if required by local health authorities, the Sponsor will break the treatment code for all serious, unexpected or suspected adverse reactions that are considered by the investigator or Sponsor to be related to study drug.

While PK samples must be collected from patients assigned to the comparator arm to maintain the blinding of treatment assignment, PK assay results for these patients are generally not needed for the safe conduct or proper interpretation of this trial. The PK assay group will be unblinded to patients' treatment assignments to identify appropriate PK samples to be analyzed and bioanalytical methodology to employ. However, the PK scientist does not have access to the PK assay results and therefore stays blinded until the PK assay results need to be interpreted and reported. Samples from patients assigned to the comparator arm will be analyzed for letrozole. However, GDC-0032 assay will be analyzed by request (i.e., to evaluate a possible error in dosing).

2.4. Number of subjects - sample size estimation

This study is designed for testing the effect of GDC-0032 on the two co-primary endpoints in all enrolled patients and in the PIK3CA MT patients and plans to enroll 330 patients in total. Assuming the PIK3CA mutation status will not be available (unknown) for approximately 10% of the patients and the prevalence of PIK3CA MT is 40%, there will be approximately 120 patients in the PIK3CA MT cohort.

Given that the PIK3CA mutation status is not a stratification factor for randomization, there might be a possible imbalance between treatment arms within the PIK3CA MT cohort, which may reduce the statistical power in this cohort. To ensure the study provides sufficient statistical power even when the treatment assignment is imbalanced, the sample size was calculated based on a conservative scenario by assuming that the treatment assignment imbalance in PIK3CA MT is 40% vs. 60%. The sample size was calculated based on a chi²-test using continuity correction (Ury and Fleiss 1980).

To control an overall, two-sided, family-wise error rate under 20% for each analysis population, we use a two-sided significance level of 16% and 4% for the co-primary endpoints MRI ORR, and pCR, respectively.

Assuming 10% of the patients are unevaluable for the MRI ORR, approximately 300 enrolled patients and 108 patients in the PIK3CA MT cohort will be evaluable for analyses. This sample size allows us to detect an absolute percentage increase of 24% in MRI ORR rate in the letrozole plus GDC-0032 arm (64%) versus the letrozole plus placebo arm (40%; Smith et al. 2005; Ellis and Ma 2007) in the PIK3CA MT cohort at 80% power and 16% two-sided significance level. The minimal detectable difference for ORR is approximately 15%.

Assuming that all patients are evaluable for pCR (i.e., approximately 330 enrolled patients and 120 in the *PIK3CA* MT cohort), this sample size provides 80% power to detect an absolute percentage increase of 18% in pCR in the letrozole plus GDC-0032 arm (19%) versus the letrozole-only arm (1%, Smith et al. 2005; Ellis and Ma 2007) in the *PIK3CA* MT cohort at the 4% two-sided significance level. The minimal detectable difference for pCR rate is approximately 13%.

The study is considered positive if at least one population shows a statistically significant result regarding the co-primary endpoints.

If the prevalence of the *PIK3CA* mutation is lower than assumed, if there is more substantial treatment assignment imbalance in the *PIK3CA* MT cohort than assumed, or there is an increased number of unevaluable patients for the MRI ORR, the sample size may be increased to obtain the level of power at 80%, and the enrollment may be limited to patients with *PIK3CA* MT.

3. Statistical methods

3.1. Data handling conventions

3.1.1. Data entry errors and potential outliers

Patients may have potential outliers for particular observations. Observations will be checked for correctness by the study team before data freeze and at the time point of data freeze all data should be correct. Remaining potential outliers based on correct values will be included in analysis. If assumptions or data structure do not allow planned analyses, analyses will be updated appropriately. However, sensitivity analyses excluding these measurements may be calculated to evaluate the influence of such extreme values. Values found to be incorrect due to data entry error after database closure will be excluded from all analyses as missing values. Incidence of incorrect values and potential outliers will be listed and summarized by treatment arm.

3.1.2. Missing data

Subjects may have missing specific data points for a variety of causes. In general, data may be missing due to a missed visit, non-evaluability of a specific clinical measurement at its planned clinical visit or a subject's early withdrawal from study. The general procedures outlined below describe what will be done when a data point is missing.

3.1.2.1. Imputations of missing data

Generally, only evaluable measurements are considered for the analyses. No values are imputed for missed or non-evaluable visits. In case of primary and secondary efficacy endpoints possible specific requirements are stated in the according analysis sections.

3.1.2.2. Partial dates

Partially incomplete dates (day or day/month missing) will be imputed for dates related to AEs, to concomitant medication and medical history. Completely missing dates and missing years will not be imputed.

For DOB imputed dates will be used to derive the age of the patients since only month and year are captured due to regulatory reasons in some participating countries. DOB must not

have missing month information and hence, missing months will not be imputed. DOB will be imputed as described in the following table:

	Missing	Impute	Exception
DOB	Day	15	--

Dates belonging to AEs and concomitant medication will be imputed as described in the following table:

	Missing	Impute	Exception
Start date	Day	01	Date of first dose of study drug if imputed date would be prior to date of first dose of study drug
	Day/Month	01JAN	Date of first dose of study drug if imputed date would be prior to date of first dose of study drug
Stop date	Day	Last day of the month	Patient EOS date if the imputed date would be after patient EOS date
	Day/Month	31DEC	

3.1.3. Stratification errors

Stratification errors are defined as different values in the stratification factors compared between IxRS and actual patient values from the electronic case report forms (eCRF). Variables compiled in both systems are:

- Tumor Size (T1-T2 vs. T3)
- Nodal Status (Cytologically Positive vs. Radiologically or Cytologically Negative). If on ultrasound examination there is evidence of suspicious axillary lymph nodes at the baseline examination, then FNA or core biopsy is required to confirm nodal status.

As tumor size has to be measured by MRI for actual patient values in the eCRF, but may be measured by other devices for stratification in IxRS (for which information on device type is not available), there may be several stratification errors for this strata.

Stratified analyses that are intended to evaluate the treatment effect will be based on the randomized stratum (ie, from IxRS) regardless of the subject's actual value throughout the document unless otherwise specified. Sensitivity analyses repeating the main primary and secondary analyses using the actual stratum (i.e. stratum as given in the eCRF) may be performed additionally. If assumptions or data structure do not allow for stratified analyses, appropriate non-stratified analyses (e.g. Fisher's exact test) will be performed. Covariate analyses where covariates are stratification factors should be based on subject's actual value.

3.1.4. Derived endpoints

EOS visit: The date of end of study visit of each patient is defined as the maximum date of all dates during the post-surgery visit for patients who ended the study per protocol and it is the date of early EOS for patients who ended the study due to other reasons. Early EOS occurs

when a patient does not undergo surgery and/or post-surgery visit (independent of the treatment visits).

Early end of treatment (EOT): Early discontinuation from treatment occurs when a patient does not take more than 90% of investigational product (IP; =14.5 weeks=72days of dosing).

Age: Age will be calculated as the time from imputed DOB to date of randomization based on eCRF information.

PEPI score: The PEPI score will be calculated (for relapse-free survival and breast cancer-specific survival) on the basis of pathological tumor size (locally assessed – as ypTStage), node status (locally assessed – as ypNStage), Ki67 level (centrally assessed - VHIO: variable Ki67ResN), and ER status (centrally assessed - VHIO: variable ERResN, Allred score) according to Ellis et al (2008).

Tumor overall ORR assessed by modified RECIST 1.1 criteria:

Lesions should be evaluated and documented at all assessments (even if they have disappeared – laterality and location for the lesion will be set to 'not available'). Documentation should be done at all subsequent visits for the modality where it was detected. For each modality, subject response is always relative to screening (baseline assessment) of that modality. Subject overall response per modified RECIST 1.1 criteria is assessed only at Week 16. Assessments of disease at visits prior to Week 16 visit are to rule out progression.

Target and non-target lesions are identified at screening, while new lesions are only identified after screening. If a new finding is equivocal on a particular modality, and not deemed suspicious enough to warrant clinical follow-up it should not be documented as a new lesion on that modality. Lymph nodes may be identified as non-target lesions at screening or as new lesions at another visit. Lymph nodes are not expected to disappear; a complete response (CR) may be documented if the lymph node becomes completely non-pathologic in appearance (~equivalent to non-lymph node lesion disappearing).

If a target lesion was present at screening the response at Week 16 should be documented as given in Table 1 in Appendix 1~~Appendix 1~~.

If no target lesion was present at screening (but documented as a non-target lesion) the response at Week 16 should be documented as follows:

Non-Target Lesions	New Lesions	Overall Response
CR	No	CR
Non-CR/non-PD	No	SD (non-CR/non-PD)
Not measurable/Not assessed	No	NE
PD (Unequivocal)	Yes or no	PD
Any	Yes	PD

The latter case should not occur on MRI assessments since a measurable lesion is required for study entry, but it may occur on clinical breast examination, mammography, and/or breast U/S assessments. An overall response of PR is not possible per RECIST 1.1 if only non-target lesions are documented at screening.

3.1.5. Statistical considerations

Cochran-Mantel-Haenszel (CMH) tests will be used for several comparisons (e.g. ORR, pCR) between treatment arms, stratified by tumor size and nodal status. No clear rule regarding the requirements for using stratified CMH test based on the combined cell frequencies over all strata could be found in literature (compare Glenn A. Walker 2004: Common Statistical Methods for Clinical Research, 2nd Edition, p 316). Hence, the CMH test will be used only if there is at least 1 responder in each stratum and the number in each cell of the unstratified 2x2 table is greater than 5. Otherwise, unstratified Fisher's Exact test will be used.

3.2. Types and timepoints of analyses

3.2.1. Main analyses

The main analysis used for preparing the Top Line Results (TLR) will be done after all patients underwent their post-surgery visit or dropped out of the study early and after clinical database closure and data management review.

The following data were not captured in the clinical database, but were assessed and/or gathered by external vendors:

- Response by centrally assessed MRI
- PIK3CA mutation status
- Data on KI67, ER status (Allred score) and expression of biomarkers
- PTEN (phosphatase tensin homolog)

- MRI volume
- ECG data

External vendor data available, received and checked by ABCSG prior to clinical database closure will be examined for the main analysis. After clinical database closure unblinding will occur at ABCSG with randomization information received from the IxRS vendor.

ABCSG will receive additional information regarding the concentration of GDC-0032 and letrozole for the pharmacokinetic, pharmakogenetic and biomarker endpoints after unblinding. The final analysis will be performed after all external vendor data are available, received and checked by ABCSG.

3.2.2. Independent Data Monitoring Committee analyses

The Independent Data Monitoring Committee (IDMC) will conduct an interim analysis to review the unblinded safety data after the first 20 patients have either 1) finished the 30-day, follow-up visit after the surgery, or 2) been on study for 20 weeks after the randomization date (for those who do not receive the surgery), whichever occurs first. All available information of all randomized (i.e. enrolled) patients with all available assessments at the respective time point will be included in the interim analyses.

In addition, the IDMC will monitor accumulating patient safety data at a minimum of once every 6 months until the last patient has completed study treatment. Additional details (e.g., IDMC members, communication, affiliations) will be provided in the IDMC charter and the details of the scope of analysis for IDMC will be documented separately. Furthermore, the IDMC or the Medical Monitor may request additional ad hoc meetings of the IDMC at any time during the study to review safety data.

3.3. Analysis populations

3.3.1. Intention-to-Treat population

Primary and secondary analyses, as well as patient-reported and exploratory analyses, will include the Intention-to-Treat (ITT) population; that is, all patients who were randomized regardless of whether they received any study drug (GDC-0032 or placebo). Patients will be analyzed according to the treatment group to which they were assigned at randomization.

3.3.2. Per-protocol population

Sensitivity analyses will be based on the Per-Protocol (PP) population; that is, all patients who were randomized, received study drug (GDC-0032 or placebo), received more than 90% and less than 110% of the per protocol intake of GDC-0032 and who do not have any protocol violations that mandate their exclusion from the PP population. Hence, patients with Important Protocol Deviations (IPDs) that are expected to impact the primary endpoint of the study will be excluded (compare Study GO28888 (LORELEI)_Note-to-File_02-Feb-2016-

Signed.pdf): [ipd-DO01], [ipd-IN04n], [ipd-PE02n], [ipd-EX01], [ipd-EX02], [ipd-EX05], [ipd-EX06], [ipd-EX07], [ipd-EX12], [ipd-EX13], [ipd-EX20] and [ipd-EX26]. Patients will be analyzed according to the treatment actually received. Hence, randomized patients who wrongly received active drug at any time point will be included in the experimental arm in the PP population.

3.3.3. Safety population

Safety analyses will be based on the safety (SAF) population and include all randomized and treated patients (patients who received at least one dose of study treatment). Patients will be analyzed according to the treatment actually received. Hence, randomized patients who wrongly received active drug at any time point will be included in the experimental arm in the SAF population.

3.4. Significance level

A two-sided significance level of 16% and 4% is used for the co-primary endpoints MRI ORR and pCR, respectively, resulting in an overall, two-sided, family-wise type I error rate for the primary endpoints of 20% for each analysis population (all randomized and PIK3CA MT patients). All other tests will be performed at a two-sided significance level of 5%.

3.5. Patient status

3.5.1. Study disposition

Recruitment status and patient discontinuation from study including reasons for discontinuation will be summarized by treatment arm.

3.5.2. Baseline demographic and disease characteristics

Demographic variables, stratification factors and other baseline characteristics will be summarized descriptively by treatment arm. The demographic variables age (metric), gender (female, male), ethnicity (Hispanic/Latino, not Hispanic/Latino) and race (American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian/Other Pacific Islander, White) will be reported. Tumor size (T1-T2, T3) and nodal status (cytologically positive, radiologically or cytologically negative) are the two observed stratification factors by IxRS. Additionally, the tumor specific variables tumor stages (N-Stage; M-Stage) and tumor grading, and hormone receptor statuses (ER and progesterone receptor (PgR)), KI67, ER status (Allred score), PEPI score as well as the Eastern Cooperative Oncology Group (ECOG) status will be analyzed.

Frequencies and percentages, means and standard deviations or medians and interquartile ranges will be given depending on the scale of the variable.Important protocol deviations

IPDs include inclusion and exclusion criteria, as well as selected variables belonging to additional relevant screening assessments, study treatment administration/dispensation, concomitant medication, primary endpoint, study treatment/randomization, study procedures/assessments, withdrawal/termination criteria and safety reporting. The IPDs are defined in a separate document listing all IPDs, their classification and additional details. IPDs will be summarized in total and by treatment arm and will be listed by study site, treatment arm and patient number.

3.5.4. Treatment disposition and exposure to study drug

Patient disposition and early discontinuation from treatment (GDC-0032/Placebo as well as letrozole) including reasons for discontinuation will be summarized by treatment arm.

Study medication intake including dose interruptions and dose reductions will be summarized overall and per treatment arm.

3.6. Primary endpoint evaluation

3.6.1. Primary endpoints

The co-primary efficacy endpoints are (1) tumor ORR (ICON: dataset RVIS, variable RRESPONS – CR, partial response (PR), stable disease (SD), Non-CR/Non-PD, PD, NE, no disease (ND)), centrally assessed by modified RECIST 1.1 criteria by breast MRI after completion of study drug, and (2) the rate of total pCR (tpCR - yes, no, not assessed) in breast and axilla, locally evaluated after completion of study drug and surgery, each in all randomized patients and PIK3CA MT patients. Information on PIK3CA mutation status (HistoGeneX: variable PIK3CAResC) is assessed centrally.

The null- and alternative hypotheses for tumor ORR is as follows:

H0: Tumor ORR is equal in patients receiving letrozole plus GDC-0032 compared to patients receiving letrozole plus placebo

H1: Tumor ORR differs between patients receiving letrozole plus GDC-0032 compared to patients receiving letrozole plus placebo

The pCR endpoint results in the following null- and alternative hypotheses:

H0: Total pCR rate is equal in patients receiving letrozole plus GDC-0032 compared to patients receiving letrozole plus placebo

H1: Total pCR rate differs between patients receiving letrozole plus GDC-0032 compared to patients receiving letrozole plus placebo

3.6.2. Analysis for primary endpoints

The efficacy endpoint tumor ORR will be calculated by treatment arm in all randomized population and in PIK3CA MT population. Within each population, summary tables will be given and the ORR for the two treatment arms will be compared at a two-sided alpha of 16% using a Cochran Mantel-Haenszel test, stratified by tumor size and nodal status. The efficacy endpoint pCR rate will also be calculated and compared at a two-sided alpha of 4% based on the same analytical approach as ORR. The two alpha values account for a family-wise type I error rate of 20% for each population (all randomized patients and PIK3CA MT patients).

Patients with early study termination and hence missing efficacy outcome (ORR or pCR), as well as patients who are NE for MRI by central review (ORR only) will be included and considered as non-responders in the analysis.

3.6.3. Sensitivity analysis for primary endpoints

As sensitivity analysis for the primary efficacy endpoints, according analyses are repeated based on the PP population.

3.7. Secondary endpoint evaluation

3.7.1. Secondary endpoints

As secondary efficacy endpoints (1) tumor ORR (ICON: dataset RVIS, variable RRESPONS – CR, PR, SD, Non-CR/Non-PD, PD, NE, ND), centrally assessed by modified RECIST 1.1 criteria by breast MRI after completion of study drug, and (2) the rate of total pCR (tpCR - yes, no, not assessed) in breast and axilla, locally evaluated after completion of study drug and surgery, both are evaluated in PIK3CA WT patients.

Furthermore, secondary efficacy endpoints contain tumor overall ORR (CR, PR, SD, PD, not measurable/not assessed), locally assessed by modified RECIST 1.1 criteria by clinical breast examination (CaPalpOverResp), mammography (CaMammoOverResp), and breast ultrasound (CaSonoOverResp), after completion of study drug evaluated in all randomized patients and separately by PIK3CA mutation status.

Additionally, (1) change in Ki67 values (VHIO: variable Ki67ResN), centrally assessed, from baseline to Week 3, baseline to surgery, and Week 3 to surgery, (2) PEPI score (for relapse-free survival and breast cancer-specific survival, derived – classified into risk groups according to Ellis et al 2008 as PEPIR and PEPIB, respectively), at surgery, and (3) change in enhancing tumor volume (ICON: dataset VLES, variable TSHOLD1 - metric), centrally evaluated using breast MRI, from baseline to surgery, are considered as secondary efficacy endpoints in all randomized patients and in each of the PIK3CA mutation cohorts.

Last, different definitions of pCR, locally assessed after completion of study drug and surgery, are evaluated in all randomized patients and in each of the PIK3CA mutation cohorts. Included are the following possibilities: a) ypT0, ypN0 (according levels for

ypTStage and ypNStage), and b) ypT0/is, ypNX (breast pCR as nodal stage is not considered; bpCR - yes, no, not assessed).

3.7.2. Analysis for secondary endpoints

The two secondary efficacy endpoint measures in PIK3CA WT population will be summarized by treatment arm and will be analyzed analogous to the primary efficacy endpoints. Information on PIK3CA mutation status (HistoGeneX: variable PIK3CAResC) is assessed centrally. Patients with early study termination and hence missing efficacy outcome, as well as patients who are NE for MRI by central review will be considered as non-responders in the analysis.

The further endpoint measures will be summarized by treatment arm and will be compared between the two treatment arms within each population based on appropriate statistical analyses: locally assessed ORR will be compared using a Cochran Mantel-Haenszel test, stratified by tumor size and nodal status within each method; change in Ki67, PEPI score and change in enhancing tumor volume will be compared by appropriate regression analysis models (depending on the distribution of the variables), e.g. logistic regression models, adjusted for tumor size and nodal status. For each of the different definitions of pCR, comparisons between treatment arms will be done by Cochran Mantel-Haenszel test, stratified by tumor size and nodal status.

3.8. Safety endpoint evaluation

3.8.1. Safety endpoints

Safety endpoints will be assessed through summaries of AEs, changes in laboratory test results and vital signs in all treated patients.

3.8.2. AEs and SAEs

Verbatim descriptions of AEs will be mapped to thesaurus terms (Medical Dictionary for Regulatory Activities (MedDRA) v19. 1). Only treatment-emergent AEs, hence, AEs outlined during protocol treatment period (on or after the first dose of study drug until 30 days after the last dose of study drug or until the end of study visit, whichever occurs later) or after this period, in case of deaths, serious adverse events, or other adverse events of concern deemed related to prior study drug treatment or study procedures, will be tabulated.

All AE data will be given by System Organ Class (SOC), preferred term (PT) and grade and will be summarized overall and by treatment arm. Additionally, serious AEs, deaths, AESIs (protocol defined – based on type and severity of AE or laboratory data), as well as AEs at NCI CTCAE v4.0 Grade 3 to 4 and Grade 5 and AEs leading to discontinuation, modification (i.e. reduction) or delay (i.e. interruption) of treatment will be summarized separately. Furthermore, AE data will be listed by study site, patient number, treatment arm, date of first dose of study drug, onset and resolved study day, seriousness, duration, severity, outcome and the relationship to study drug to understand the cumulative incidence as well as the time

to onset of first event when treated with GDC-0032+letrozole. The time to AE of special interest (time from first dose of study drug to first AE of special interest) will be shown graphically.

3.8.3. Laboratory data and vital signs

For relevant laboratory data (ALT, AST, bilirubin, calcium, glucose, hemoglobin, platelets, serum creatinine, white blood cells and HbA1C – all metric) shifts from baseline to minimum and maximum NCI CTCAE v4.0 grades will be given. Vital sign data (heart rate, blood pressure, and temperature – all metric) will be summarized in total and by treatment arm. Abnormal ECG data (based on BioTelemetry Research (BTR) data) will be listed.

3.9. Patient-reported endpoint evaluation

3.9.1. Patient-reported endpoints

Patient reported outcomes of disease/treatment-related symptoms, patient functioning, and HRQoL will be assessed using the EORTC QLQ-C30 and the modified QLQ-BR23 (Aaronson et al. 1993) for all randomized patients. Hence, patient-reported endpoints are the mean and mean change scores (of the base questionnaire and of the supplementary breast cancer module) of the scales, which are described in the scoring procedures (Fayers et al. 2001).

3.9.2. Analysis for patient-reported endpoints

Completion rates of the measures will be summarized at each time point by treatment arm.

The population for analysis will be defined as all randomized patients who have completed baseline and at least one post-baseline assessment. Summary statistics (mean, mean change, standard deviation, median and 95% CI) of linear transformed scores will be reported for all subscales of the EORTC QLQ-C30 questionnaire, and the modified QLQ-BR23 (including the added items of skin problems, itching, sore mouth/tongue, and trouble swallowing, which will be assessed independently). Both measures and additional scales from the EORTC item bank will be scored according to the EORTC scoring manual guidelines (Fayers et al., 2001). Line charts depicting the mean changes (and standard errors) of items and subscales over time will be provided for each treatment arm from the baseline assessment. Clinically meaningful differences will be assessed relative to established thresholds (Osoba et al., 1998; Cocks et al., 2011).

Additional patient reported outcomes analyses might be conducted.

3.10. Exploratory endpoint evaluation

Details on evaluation (endpoints and analyses) of all translational science objectives will be described in the Translational Statistical Analysis Plan (TSAP). Analyses will include pharmacodynamics and pharmacokinetic, as well as biomarker endpoint evaluation.

3.10.1. Exploratory endpoints

ORR (centrally assessed, ICON: dataset RVIS, variable RRESPONS – CR, PR, SD, Non-CR/Non-PD, PD, NE, ND), pCR rate (local, tpCR - yes, no, not assessed), PEPI score (derived – classified into risk groups according to Ellis et al 2008 as PEPIR and PEPIB, respectively) according to the decrease in Ki67 after 2 weeks of treatment are hypothesized to differ between patients receiving letrozole plus GDC-0032 compared to patients receiving letrozole plus placebo. Each will be evaluated in all randomized patients and, if statistically significant within all randomized patients, also in PIK3CA MT patients.

The concordance of the different imaging modalities in measuring tumor response after completion of study drug (MRI, central, ICON: dataset RVIS, variable RRESPONS – CR, PR, SD, Non-CR/Non-PD, PD, NE, ND; clinical, ultrasound and mammography, local, OverResp for each - CR, PR, SD, PD, not measurable/not assessed) will be observed. Concordance will be evaluated in all randomized patients.

Rates of planned BCS (local, SurgTypePlan - breast conserving, other levels) and conversion to BCS (local, SurgType - breast conserving, other levels) between patients receiving letrozole plus GDC-0032 and patients receiving letrozole plus placebo will be compared. Rates will be evaluated in all randomized patients.

3.10.2. Analysis for exploratory endpoints

Comparisons between treatment arms in ORR (grouped as responder vs. non-responder), pCR rate and PEPI score according to the decrease in Ki67 will be analyzed using appropriate regression analysis models (depending on the distribution of the variables), e.g. logistic regression models , adjusting for tumor size and nodal status in all randomized population and in PIK3CA MT population. Concordance will be examined by frequencies and percentages. Rates will be summarized descriptively.

3.11. Listings

To review the snapshot clinical data for a patient, in case a safety event occurred, additional patient summary listings will be performed. For the following categories listings may be presented: demographics, medical history, history of primary breast cancer, surgery, physical examination, vital signs, radiological examinations, hematology, chemistry, electrolytes, DLCO, GDC-0032/placebo administration, letrozole administration, concomitant medication, treatment completion and surgery at treatment completion.

3.12. Covariates

To adjust for the possible effect of demographic or prognostic factors multivariate analyses will be performed. To avoid multi-collinearity between the covariates, prior correlation analyses will be performed. Depending on the statistical results and the clinical relevance, highly correlated covariates may be excluded from multivariate analyses. Binary dependent variables will be analyzed using logistic regression models. Metric dependent variables will be analyzed using general linear models or a non-parametric equivalent if appropriate. Forest plots will be given where appropriate.

3.13. Software

Analyses will be done by members of the biostatistics group at ABCSG using statistical analysis system (SAS) software (SAS® version 9.3 or higher). If necessary, selected other software will be used.

4. Changes of analysis compared to study protocol

Abbreviation of progesterone receptor changed to PgR as PR is used for partial response.

The following exploratory objective was included to be consistent with the outcome measures given in the protocol: ‘ORR, pCR rate, and PEPI scores according to the decrease in Ki67 after 2 weeks of letrozole plus GDC-0032 and letrozole plus placebo.’.

5. Addendum to the statistical analysis plan

The Addendum to the SAP lists additional analyses for secondary endpoints calculated after finalization of the SAP (V2), including a description of and a rationale for these additions.

As addition to the main results for the two secondary efficacy endpoint measures in PIK3CA WT population, bar charts for rates (%) by treatment arm were given.

Also for further secondary endpoint measures several figures were created:

- Bar charts for rates (%) of locally assessed ORR by alternative methods were given by treatment arm for all three patient populations;
- for change in Ki67 boxplots either by treatment arm or by mutation status grouped by time point, individual patients plots for each treatment arm and either with/without ORR or mutation/wild type by with/without ORR as well as bar charts for proportional

change by treatment arm grouped by time difference (BL-W3, BL-SU, W3-SU) for all three patient populations were shown;

- Bar charts for frequency (%) of PEPI score as well as PEPI score risk groups by treatment arm for relapse-free/breast cancer-specific survival were created;
- Bar charts for change in enhancing tumor volume by treatment arm grouped by mutation status for absolute/percentage change were presented.

Additionally, because of the unexpected results with regard to the changes in Ki67 and the PEPI score, frequencies and percentages of time from EOT to surgery for GDC, Letrozole and overall were calculated.

6. References

Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85:365-76.

Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J. Clin. Oncol. 2011;29:89–9610

Ellis MJ and Ma C. Letrozole in the neoadjuvant setting: the P024 Trial. Breast Cancer Res Treat 2007;105(Suppl 1):33–43.

Ellis MJ, Tao Y, Luo J, A'Hern R, Evans DB, Bhatnagar AS, Chaudri Ross HA, von Kameke A, Miller WR, Smith I, Eiermann W, Dowsett M. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. J Natl Cancer Inst 2008;100:1380–1388

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual. 3rd edn. EORTC, Brussels; 2001.

Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139 44

Smith IE, Dowsett M, Ebbs SR, et al. IMPACT Trialists Group: neoadjuvant treatment of postmenopausal breast cancer with anastrozole, tamoxifen, or both in combination: the immediate preoperative anastrozole, tamoxifen, or combined with tamoxifen (IMPACT) multicenter double-blind randomized trial. J Clin Oncol 2005;23:5108–16.

Ury HK and Fleiss JL. on approximate sample sizes for comparing two independent proportions with the use of yates' correction. Biometrics 1980;36:347–51.

Walker GA. Common Statistical Methods for Clinical Research, 2nd Edition, 2004.

7. Appendices

7.1. Appendix 1

MODIFIED RESPONSE EVALUATION CRITERIA IN SOLID TUMORS: ASSESSMENT OFRESPONSE OF NEOADJUVANT THERAPY IN EARLY BREAST CANCER

Conventional response criteria may not be ideal for the assessment of response in the setting of neoadjuvant therapy in early breast cancer. Therefore, RECIST 1.1 criteria have been modified to specifically address assessment of primary breast lesions along with axillary lymph node disease, using a range of breast imaging modalities. Selected sections from the Response Evaluation Criteria in Solid Tumors (RECIST), Version 1.1[1] are presented below, with modifications and the addition of explanatory text as needed for clarity. For detailed information on the read methodology including how imaging data should be processed prior to reads, please refer to the Study Imaging Charter.

	RECIST v1.1	Modified RECIST Early Breast Cancer Neoadjuvant Therapy
Modalities	CT as primary modality, ultrasound not recommended	No CT; primary assessments by MRI; also assessments by ultrasound, mammography, and clinical exam
Lymph nodes	May be considered target lesions based on size criteria (≥ 15 mm in SAD)	Only axillary lymph nodes assessed; nodes that are considered abnormal on imaging (based on morphological factors including, but not limited to SAD) to be followed as non-target lesions
Possibility of having only non-target disease	Allowed	Not allowed; primary breast lesions must be measurable by MRI

CT = computed tomography; MRI = magnetic resonance imaging; SAD = short axis dimension.

BASELINE DOCUMENTATION OF TARGET AND NON-TARGET LESIONS

To assess objective response or future progression, it is necessary to estimate the overall tumor burden at baseline and to use this as a comparator for subsequent measurements. All baseline evaluations should be performed as close as possible to the treatment start and never more than 4 weeks before the beginning of the treatment.

Method of Measurement

According to RECIST 1.1 guidelines, MRI is the preferred modality to follow breast lesions in a neoadjuvant setting. CT is currently the preferred modality for assessing metastatic disease, but should not be used in this focused setting of neoadjuvant therapy in early breast

[1] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: Revised RECIST guideline (Version 1.1). Eur J Cancer 2009;45:228–47.

cancer. Ultrasound, mammography, and clinical exam are all common and useful modalities for assessing breast lesions, and will also be used to assess response in this protocol, adhering to response criteria as presented in this appendix.

Target Lesions

Target lesions should be selected on the basis of their size (lesions with the longest diameter) and should lend themselves to reproducible repeated measurements. Up to 2 lesions in the breast may be identified as target lesions. A sum of the diameters of all target lesions will be calculated and reported as the baseline sum of diameters. The baseline sum of diameters will be used as a reference to further characterize any objective tumor regression in the measurable dimension of the disease. Lesions that meet the criteria for radiographically defined simple cysts should not be considered malignant lesions (neither target nor non-target) since they are, by definition, simple cysts. Pathologic axillary lymph nodes are not to be designated at target lesions, and lymph node measurements are not to be included in the sum of diameters (see below for more detail).

Bilateral breast imaging studies should be conducted at each study assessment. The same method of measurement and the same technique should be used to characterize each target lesion at baseline and during the study, and all measurements should be recorded in metric notation. Care must be taken in measurement of target lesions with different modalities, since the same lesion may appear to have a different size with each modality. If for some reason the same imaging modality cannot be used at a scheduled assessment time point, then the case should be discussed with the radiologist to determine if substitution of any other approach is possible and, if not, the patient should be considered not evaluable at that time point, for that particular type of imaging assessment.

Non-Target Lesions

Non-target lesions may include any other measurable breast lesions not identified as target lesions, as well as truly non-measurable lesions, such as diffuse skin thickening or other lesions not measurable by reproducible imaging techniques.

Lymph nodes merit special mention since they are normal anatomical structures that may be visible by imaging even if not involved by tumor. Axillary lymph nodes are known to vary widely in size, and signs of abnormality in axillary lymph nodes on imaging include other morphological findings often in addition to changes in nodal size. For these reasons, pathologic axillary lymph nodes on imaging should be identified as non-target lesions at baseline. Change in short-axis dimension may be considered in the assessment of pathology, but measurements are not required, and these lesions should be followed qualitatively, as described below at each response assessment time point.

Signs of lymph node pathology on imaging include the following:

- Increase in short axis dimension
- Thickened cortex, either diffusely or asymmetrically enlarged
- Thinning, or replaced fatty hilum
- Irregular margins or spiculations

- Rim enhancement
- Decreased echogenicity of cortex
- Perinodal edema

EVALUATION OF RESPONSE

Evaluation of Target Lesions

This section provides the definitions of the criteria used to determine objective tumor response for target breast lesions:

- Complete response (CR): disappearance of all target lesions
- Partial response (PR): at least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters
- Progressive disease (PD): at least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (nadir), including baseline

  In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.

  The appearance of one or more new lesions is also considered progression.
- Stable disease (SD): neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum on study

Special Notes on the Assessment of Target Lesions

Target Lesions That Become Too Small to Measure. While on study, all lesions recorded at baseline should have their actual measurements recorded at each subsequent evaluation, even when very small (e.g., 2 mm). However, sometimes lesions that are recorded as target lesions at baseline become so faint on imaging that the radiologist may not feel comfortable assigning an exact measure and may report them as being too small to measure. When this occurs, it is important that a value be recorded on the CRF as follows:

- If it is the opinion of the radiologist that the lesion has likely disappeared, the measurement should be recorded as 0 mm.
- If the lesion is believed to be present and is faintly seen but too small to accurately measure, BML (below measurable limit) should be indicated.

To reiterate, however, if the radiologist is able to provide an actual measure, that should be recorded, and, in that case, BML should not be ticked.

Lesions That Split or Coalesce on Treatment. When non-nodal lesions fragment, the longest diameters of the fragmented portions should be added together to calculate the target lesion sum. Similarly, as lesions coalesce, a plane between them may be maintained that would aid in obtaining maximal diameter measurements of each individual lesion. If the lesions have truly coalesced such that they are no longer separable, the vector of the longest diameter for the coalesced lesion should be recorded.

Evaluation of Non-Target Lesions

This section provides the definitions of the criteria used to determine the tumor response for any non-target lesions identified at baseline. Although some non-target lesions may actually

be measurable, they need not be measured and, instead, should be assessed only qualitatively at the timepoints specified in the protocol.

- CR: disappearance of all non-target lesions
  - All lymph nodes must be non-pathologic in appearance
- Non-CR/Non-PD: persistence of one or more non-target lesion(s)
- PD: unequivocal progression of existing non-target lesions. For pathologic axillary lymph nodes, this may be based on a combination of morphological factors, including a potential increase in short-axis dimension

Special Notes on Assessment of Progression of Non-Target Disease

To achieve unequivocal progression on the basis of the non-target disease, there must be an overall level of substantial worsening in non-target disease in a magnitude that, even in the presence of SD or PR in target disease, the overall tumor burden has increased sufficiently to merit discontinuation of therapy. A modest increase in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status. The designation of overall progression solely on the basis of change in non-target disease in the face of SD or PR of target disease will therefore be extremely rare.

New Lesions

The appearance of new malignant lesions denotes disease progression; therefore, some comments on detection of new lesions are important. There are no specific criteria for the identification of new radiographic lesions; however, the finding of a new lesion should be unequivocal, that is, not attributable to differences in scanning technique, change in imaging modality, or findings thought to represent something other than tumor. This is particularly important when the patient's baseline lesions show partial or complete response. For example, necrosis of a breast lesion may be reported on an MRI scan report as a "new" cystic lesion, which it is not. A lesion identified during the study in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

If a new lesion is equivocal, for example because of its small size, continued therapy and follow-up evaluation will clarify if it represents truly new disease. If repeat scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial scan.

Time point Response (Overall Response)

Table 1 provides a summary of the overall response status calculation at each protocol-specified time point for which a response assessment occurs.

Table 1 Time point Response: Patients with Target Lesions (with or without Non-Target Lesions)

Target Lesions	Non-Target Lesions	New Lesions	Overall Response
CR	CR, or no non-target lesions identified at baseline	No	CR
CR	Non-CR/non-PD	No	PR
CR	NE	No	PR
PR	Any except PD	No	PR
SD	Any except PD	No	SD
NE (Any lesion)	Any except PD	No	NE
PD	Any	Yes or no	PD
Any	PD	Yes or no	PD
Any	Any	Yes	PD

CR = complete response; NE = not evaluable; PD = progressive disease;
PR = partial response; SD = stable disease.

Missing Assessments and Not-Evaluable Designation

When no imaging/measurement is done at all at a particular time point, the patient is not evaluable at that time point. If only a subset of lesion measurements are made at an assessment, usually the case is also considered not evaluable at that time point, unless a convincing argument can be made that the contribution of the individual missing lesion(s) would not change the assigned time point response. Similarly, if one or more non-target lesions are not assessed, the response for non-target lesions should be “not evaluable” except where there is clear progression in non-target lesions that are assessed.

Special Notes on Response Assessment

Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as “symptomatic deterioration.” Every effort should be made to document objective progression even after discontinuation of treatment. Symptomatic deterioration is not a descriptor of an objective response; it is a reason for stopping study therapy. The objective response status of such patients is to be determined by evaluation of target and non-target disease as shown in Table 1.