AMGEN

Statistical Analysis Plan

Product:	Denosumab
Protocol Number:	20050209 / ABCSG 18
Version No.:	6.0
Date:	December 10, 2019

A Randomized, Double-Blind, Placebo-Controlled, Multi-Center Phase 3 Study to Determine the Treatment Effect of Denosumab in Subjects with Non-Metastatic Breast Cancer Receiving Aromatase Inhibitor Therapy

Austrian Breast and Colorectal Cancer Study Group

ABCSG Protocol number 18

ABCSG Lead Statistician

Dr. [redacted] Date

Coordinating Investigator

Univ.Prof. Dr. [redacted] Date

NCT Number: NCT00556374
This NCT number has been applied to the document
for purposes of posting on Clinicaltrials.gov

TABLE OF CONTENTS PAGE

TITLE PAGE 1
LIST OF ABBREVIATIONS / ACRONYMS 5
LIST OF STUDY TERMS 8
1. INTRODUCTION 11
2. STUDY DETAILS 12
2.1 Objectives 12
2.1.1 Primary 12
2.1.2 Secondary 12
2.1.3 Exploratory 12
2.2 Hypotheses 13
2.2.1 Clinical Efficacy Hypothesis: 13
2.2.2 Clinical Safety Hypothesis: 13
2.3 Overall Study Design 13
2.4 Number of Subjects and Sample Size Calculation 15
2.4.1 Primary Endpoint 15
2.4.2 Percent Change in Lumbar Spine BMD at Month 12 16
2.5 End of Treatment 16
2.6 Long-Term Follow-Up 16
2.7 End of Open-label Treatment 17
2.8 End of Study 17
2.9 Study Termination 17
3. DATA SCREENING AND ACCEPTANCE 17
3.1 General Principles 17
3.2 Handling of Missing and Incomplete Data 18
3.2.1 Early Withdrawal from Study and Lost-to-follow-up 18
3.2.2 Missed Visit or Non-evaluability of a Specific Clinical Measurement at its Planned Clinical Visit 18
3.2.3 Dates 18
3.3 Potential Outliers 19
3.4 Distributional Characteristics 20
3.5 Validation and Configuration Management 20
4. ANALYSIS SET 20
4.1 Definition of Analysis Sets 20
4.1.1 Full Analysis Set (FAS) 21
4.1.2 OLP Denosumab (Dmab) Analysis Set 21
4.1.3 BMD Analysis Set 21
4.1.4 Vertebral Fracture Analysis Set 21

4.1.5 Per-Protocol (PP) Population 22
4.1.6 Safety Analysis Set (SAF) 22
4.1.7 OLP Denosumab (Dmab) Safety Analysis Set 23
4.2 Subgroups 23
5. PRIMARY, SECONDARY, SAFETY AND EXPLORATORY ENDPOINTS 24
5.1 Primary Endpoint 24
5.2 Secondary Endpoints 24
5.3 Safety Endpoints 25
5.4 Exploratory Endpoints 25
5.5 Derived Endpoints 26
6. PLANNED METHODS OF ANALYSIS 28
6.1 General Approach / Considerations 28
6.2 Handling of Incorrect Stratification 30
6.3 Analysis Windows 30
6.3.1 Study Month 30
6.3.2 Incidence Endpoints 30
6.3.3 Analysis at Specified Times 31
6.4 Analysis of Key Study Endpoints 32
6.4.1 Efficacy Analysis 32
6.4.2 Safety Analyses 37
6.4.3 Exploratory Analyses 39
6.4.4 Treatment Discontinuation and Lost-to-Follow-Up 41
6.4.5 Subgroup Analysis 42
6.4.6 Covariates 42
6.4.7 SAS Code Fragments 43
7. FURTHER TABULAR SUMMARIES AND LISTINGS 45
8. INTERIM ANALYSES AND EARLY STOPPING GUIDELINES 46
9. DFS ANALYSIS RECOMMENDED BY DMC 47
10. DFS ANALYSIS 2018 48
11. ZA SUBSTUDY 48
11.1.1 Exploratory Objectives 48
11.1.2 ZA Substudy design 48
11.1.3 Number of Subjects and Sample Size Calculation 49
11.1.4 Estimated ZA Substudy Duration 50
11.1.5 End of ZA Substudy 50
11.1.6 Study Termination 50
11.1.7 Exploratory Endpoints 50
11.1.8 Analysis Sets 51
11.1.9 Subgroups 51
11.1.10 Timepoint of Analysis 52
11.1.11 Handling of Missing and Incomplete Data 52

11.1.12 Analysis Windows 53
11.1.13 Analysis Methods 54
12. AMENDMENTS TO THE STATISTICAL ANALYSIS PLAN 55
13. CHANGES FROM THE PROTOCOL PLANNED ANALYSES 56
14. LIST OF PLANNED TABLES AND FIGURES 58
15. REFERENCES 64

1 LIST OF ABBREVIATIONS / ACRONYMS

2

Abbreviation/Acronym	Definition
ABCSG	Austrian Breast and Colorectal Cancer Study Group
AE	Adverse event
AI	Aromatase inhibitor
AIT	Aromatase inhibitor therapy
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATAC	Arimidex, Tamoxifen Alone or in Combination trial
AUC	Area under the serum concentration-time curve
BIG	Breast International Group
BMD	Bone mineral density
BMFS	Bone metastases-free survival
CI	Confidence interval
CDM	Clinical data management
CRF	Case report form
CTCAE	Common terminology criteria for adverse events
CTX	C-terminal telopeptide
DFS	Disease free survival
Dmab	Denosumab
DMC	Data monitoring committee
DMP	Data management plan

Abbreviation/Acronym	Definition
DXA	Dual X-Ray Absorptiometry
ECOG	Eastern Cooperative Oncology Group
EOOLT	End of open-label treatment
EOS	End of study
EOT	End of treatment
ER	Estrogen receptor
FAS	Full Analysis Set
FDA	Food and Drug Administration
FSH	Follicle-stimulating hormone
GCP	Good Clinical Practice
HER-2	Human Epidermal growth factor Receptor 2
HIV	Human immunodeficiency virus
ICF	Informed consent form
ICH	International conference on harmonization
IEC	Independent ethics committee
IP	Investigational product
IRB	Institutional review board
IV	Intravenous
IVRS	Interactive voice response system
LTFU	Long-term follow-up
MedDRA	Medical Dictionary for Regulatory Activities
OL	Open-label
OLP	Open-label phase
[redacted]	[redacted]

Abbreviation/Acronym	Definition
OPG	Osteoprotegerin
ORR	Objective response rate
OS	Overall survival
PADCD	primary analysis data cut-off date
PgR	Progesterone receptor
PIN	Personal identification number
PP	Per protocol
Q6M	Every 6 months
Q12M	Every 12 months
RANKL	Receptor activator for nuclear factor kappa B ligand
RPSFTM	Rank-preserving structural failure time model
SAE	Serious adverse event
SAS	Statistical Analysis System
SERM	Selective Estrogen Receptor Modulator
SC	Subcutaneous
SD	Standard deviation
SOC	Standard of care
SOP	Standard Operating Procedure
TTP	Time to progression
ULN	Upper limit of normal per laboratory reference range
uNTX	Urinary N-telopeptide
ZA	Zoledronic acid

1

1 LIST OF STUDY TERMS

2

Study Term	Definition
Date of randomization	Date subject is randomized; enrollment date
Day 1 in double-blind phase and Open-Label Phase (OLP)	First day investigational product is administered
Day 1 in Zoledronic Acid (ZA) substudy	Defined as 8 months (± 4 weeks) after the last OLP denosumab dose For treatment arm: first day ZA is administered For control arm (SOC treatment): SOC day 1 visit
End of Open-Label Treatment (EOOLT)	Last administration of open-label phase denosumab for each subject
EOOLT Visit	Subjects will complete an EOOLT visit 30 to 45 days after the last dose of open-label denosumab either by clinic visit or telephone contact
End of study (EOS)	The date when the last subject completes the last scheduled long-term follow-up visit
EOS visit	A subject's last formal visit or last formal contact or an unscheduled study visit in case of early withdrawal from study
End of ZA substudy	The date when the last subject participating in the ZA substudy completes the last formal visit or an unscheduled study visit in case of early withdrawal from the ZA substudy
End of ZA substudy visit	A subject's last formal substudy visit or contact or an unscheduled substudy visit in case of early withdrawal from substudy
End of Treatment (EOT) in double-blind phase	Last administration of double-blind investigational product for each subject
EOT visit in double-blind phase	The visit at which the subject receives the last dose of double-blind IP; for all subjects, whose last regular study visit within 6 months (+ 45 days time window) prior to the PADCD was a yearly visit (eg, months 12, 24, etc) and radiological assessments were performed,

	this visit will be considered an EOT visit. Those subjects, for whom the radiological assessments were not performed, will need to attend the next scheduled 6 months visit, when radiological assessments (vertebral x-rays, DXA) will take place and an antibody sample will be taken. For these subjects this visit will be considered as EOT visit. For those whose last regular study visit within 6 months (+ 45 day time window) prior to the PADCD was a 6 months visit (eg, month 6, 18, etc), the next regularly scheduled yearly study visit (eg, Q12M) within 6 months after PADCD will be considered an EOT visit
Fractures	
Clinical fracture	Clinically evident fracture with associated symptoms
Clinical Vertebral Fracture	Clinical vertebral fracture when there is ≥ 1 grade increase from a previous grade of 0 in any vertebra between T4 and L4 at any visit (scheduled or unscheduled) and the subject reported signs and/or symptoms indicative of a fracture. Only fractures with low trauma severity will be included.
Low Trauma Severity	Assessed by the investigator and collected on the Clinical Fracture Summary CRF for each clinical fracture event and includes - Fall from standing height or less - Falls on stairs, steps or curbs - Fall from the height of a stool, chair, first rung on a ladder or equivalent (about 0.5 metres) - Minimal trauma other than a fall - Moderate trauma other than a fall - Unknown (including missing) All asymptomatic vertebral fractures are considered of low trauma severity.
High Trauma Severity or pathologic	Assessed by the investigator and collected on the Clinical Fracture Summary CRF for each clinical fracture event and includes - Fall from higher than the height of a stool, chair, first rung on a ladder or equivalent (> 0.5 metres) - Severe trauma other than a fall - Pathologic fractures
Morphometric fracture	A fracture in the vertebral column that is not clinically evident and that is asymptomatic
New Vertebral	For the study a new vertebral fracture is defined as a fracture in a

Fracture	previously undeformed vertebrae. New vertebral fractures will be classified as either clinical or morphometric depending on whether or not in the investigator's opinion the symptom locality can be linked to the fracture.
Prevalent Vertebral Fracture	Prevalent vertebral fractures will be assessed and recorded at baseline.
Worsening Vertebral Fracture	Worsening is defined as an increase in fracture severity of at least 1 grade on the semiquantitative scale.
New or Worsening Vertebral Fracture	This is the combination of new vertebral fractures and worsening vertebral fractures.
Final analysis	Analysis performed after long-term follow up, and OLP, approximately 66 months after PADCD, after review of the results of the interim analysis
Interim analysis	Analysis for futility for the secondary endpoint DFS to be performed after PADCD by an independent statistician.
DFS analysis	Analysis recommended by the DMC for the secondary endpoint DFS to be performed around 18 months after PADCD.
Long-Term Follow-Up (LTFU) visit	All Q12M visits that are attended after EOT visit
Pre-selected (major) center	Centers were pre-selected (considered major) if they used the Hologic DXA device at site initiation
Primary analysis	Analysis after Primary Analysis Data Cut-off Date (PADCD)
Primary Analysis Data Cut-off Date (PADCD)	The date when approximately 247 subjects have experienced their first clinical fracture and all subjects have either withdrawn or had the opportunity to receive at least 2 doses of IP.
Screening period	Begins with signed informed consent and ends on day investigational product is administered

1

1. INTRODUCTION

The Austrian Breast & Colorectal Cancer Study Group (ABCSG) is a cooperative institution that was set up to conduct controlled clinical trials in breast and colorectal cancer and to facilitate communication and the dissemination of knowledge among scientists and others dedicated to the cancer problem. Since its establishment in 1984, more than 15,000 patients have been enrolled in ABCSG investigations. In certain patient risk groups, ABCSG is currently recruiting up to 40 per cent of all Austrian breast cancer patients to their clinical trials. The ultimate goal of the ABCSG is to enhance the standard of cancer treatment in this country and abroad by developing innovative approaches and testing increasingly more effective therapeutic strategies.

Prospectively randomized clinical investigations have come to be seen as the only instrument to generate valid and reliable clinical data, to gain insights into significant prognostic and predictive factors, and thus to enhance all aspects of evidence-based medicine. As a target of oncological research and practice, the ABCSG has selected two of the most common malignancies affecting women and men in the present-day western world: carcinoma of the breast and bowel.

As a whole, the ABCSG collaborates toward the common goal of controlling, effectively treating, and ultimately curing cancer by means of large, multicenter cancer trials in the (neo-) adjuvant setting. Research results are provided to the medical community through scientific publications and professional meetings.

This Statistical Analysis Plan (SAP) is based on the Amended Clinical Study Protocol for ABCSG study 18 (Protocol Number: 20050209, Superseding Amendment 6, Date: 15 July 2019) and gives a detailed description of all statistical analyses planned to be conducted within this trial at predefined time points.

2. STUDY DETAILS

2.1 Objectives

2.1.1 Primary

To determine whether denosumab compared to placebo will reduce the rate of first (on-study) clinical fracture (ie, clinically evident fracture with associated symptoms) in women with non-metastatic breast cancer receiving non-steroidal aromatase inhibitor therapy (AIT).

2.1.2 Secondary

To assess the effect of denosumab compared to placebo on:

Fracture-related secondary endpoints:

- Bone mineral density (BMD) at lumbar spine, total hip and femoral neck in a subgroup of subjects with evaluable Dual X-Ray Absorptiometry (DXA) scans using the same Hologic device
- Incidence of new vertebral fractures (both clinical and morphometric [ie, a fracture in the vertebral column that is not clinically evident and that is asymptomatic])
- Incidence of new or worsening of pre-existing vertebral fractures (both clinical and morphometric)

Disease outcome-related secondary endpoints:

- Disease-free survival (DFS)
- Bone metastasis-free survival (BMFS)
- Overall survival (OS)

To assess the safety and tolerability of denosumab in this population.

2.1.3 Exploratory

[redacted]

[redacted]

2.2 Hypotheses

2.2.1 Clinical Efficacy Hypothesis:

Denosumab, when administered subcutaneously (SC) at a dose of 60 mg every 6 months (Q6M), will be considered efficacious in breast cancer subjects receiving AIT if the rate of first clinical fractures (clinical vertebral and non-vertebral - consistent with the ATAC trial) in denosumab-treated subjects is lower than that in placebo-treated subjects. It is anticipated that denosumab will reduce the rate by 30% compared with placebo (ie, the true hazard ratio of denosumab compared with placebo is 0.70).

2.2.2 Clinical Safety Hypothesis:

Denosumab, when administered SC at a dose of 60 mg every 6 months will be well tolerated in breast cancer subjects receiving AIT.

2.3 Overall Study Design

This is a multi-center phase 3, randomized, double-blind, placebo-controlled study to determine the treatment effect of denosumab in subjects with breast cancer treated with an approved non-steroidal aromatase inhibitor (AI), e.g. anastrazole. Approximately 3400 subjects will be randomized in a 1:1 ratio to receive either double-blind:

- Denosumab administered at a dose of 60 mg Q6M or
- Placebo SC Q6M

The randomization schedule will use randomly permuted blocks, and will be stratified by

- Type of hospital (pre-selected [major] center or other center),
- Prior AI usage (Yes or No)
- Total lumbar spine BMD score at baseline (T-score < –1.0 or T-score ≥ -1.0).

Subjects will be recruited over an estimated period of 82 months, and remain on investigational product (IP) until the required number of events (where an event is defined as first clinical fracture) is reached and all subjects had the opportunity to receive a minimum of at least 2 doses of IP, whichever occurs later.

The primary analysis data cut-off date (PADCD) is defined as the time at which the required number of events is reached and all subjects have had the opportunity to receive at least 2 doses of IP. The actual timing of the primary analysis will depend on the subject enrollment rate, drop-out rate and the rate at which first clinical fractures are observed. When the PADCD is reached, all subjects will discontinue IP.

Following the study PADCD, subjects will be followed for DFS, BMFS and OS every 12 months (Q12M) by clinic visits or telephone contacts starting from their EOT visit until a maximum of either 18 or 66 months after the PADCD, depending on an interim DFS analysis after PADCD (see Section 8). Based on the Data Monitoring Committee (DMC) recommendation, given that interim futility analysis of DFS did not indicate futility, it was decided to follow subjects until a maximum of approximately 66 months after PADCD.

Additionally, a time-driven analysis for efficacy of the secondary endpoint DFS recommended by the DMC will take place approximately 18 months after PADCD, prior to any unblinding of subjects at the investigator/subject level.

Due to the statistically significant treatment effect in the primary endpoint and fracture-related secondary endpoints between the denosumab arm and the placebo arm, which were demonstrated at the primary analysis (Gnant et al.; Lancet 2015), willing and eligible subjects randomized to placebo during the double-blind phase may participate in an open-label phase (OLP) and receive denosumab 60 mg Q6M for up to 36 months (maximum of 7 doses), as recommended by the DMC. Subjects who do not fulfil the eligibility criteria or do not consent will complete LTFU assessments only.

The final analysis will be performed when all subjects have had the opportunity to attend 5 long-term follow-up (LTFU) visits after approximately 66 months after PADCD or 7 OLP visits after approximately 36 months after first OLP treatment.

A substudy has been added (see section 11) to evaluate the impact of a single intravenous (IV) zoledronic acid (ZA) administration on BMD, fracture incidence, and bone turnover markers. Subjects enrolled in the main study, who completed denosumab during the OLP and are deemed eligible, may choose to participate in this ZA substudy. Subjects that are not included in the ZA substudy will end study as planned. Protocol-defined denosumab administration will complete at end of the open-label (OL) period no matter if subjects participate in the ZA substudy or not.

2.4 Number of Subjects and Sample Size Calculation

Participants enrolled in this clinical investigation shall be referred to as “subjects”.

2.4.1 Primary Endpoint

Approximately 3400 subjects will be enrolled into this trial (1700 per treatment group), which will conclude once approximately 247 subjects have experienced a clinical fracture. This is based on an 82-month accrual period and on a drop-out rate of 3.6% per year.

Approximately 247 subjects must experience a clinical fracture for this study to have 80% power to detect a hazard ratio of 0.70 (denosumab vs. control) with a two-sided significance level of 0.05. A hazard ratio of 0.70 indicates that the fracture rate in the denosumab group is 30% less than in the control group.

The incidence of clinical fracture in the ATAC trial (Jakesz et al, 2005) was 22.6 per 1000 subject years. From this the fracture rate in the control group is estimated at 2.43% per year. In study 20050209, the observed total fracture rate of both arms combined after 61 months is 1.7% per year. Assuming a hazard ratio of 0.70, the fracture rate in the control group is expected to be about 2.0% per year and 1.39% per year in the denosumab group. According to this fracture rate, 247 fractures will be reached after 103 months. All subjects will be followed until they have had the opportunity to receive at least 2 doses of IP, still in a blinded fashion.

All clinical fractures, except those of the skull, face, fingers, and toes, which are typically not associated with osteoporosis, will be included in the analysis of clinical fractures.

2.4.2 Percent Change in Lumbar Spine BMD at Month 12

In addition to the primary endpoint, a comparison of the percent change of lumbar spine BMD between the denosumab treatment and placebo group is planned. In a subset of the ATAC data (Eastell and Adams, 2002) a 1.8% difference was observed between the change from baseline to 12 months in the tamoxifen and anastrozole groups. To have 90% power to detect a 1.8% difference (standard deviation [SD]=3.9%) between denosumab and placebo in the percentage of change of BMD for lumbar spine at 12 months with a two-sided significance level of 0.05, it will be necessary to have BMD data from 102 subjects per treatment arm.

2.5 End of Treatment

The end of treatment (EOT) is defined for each subject as the point in time when the last dose of blinded investigational product is administered.

In order to prevent a lengthy extension to the duration of the study, ABCSG and Amgen will monitor the rate of clinical fracture pooled by treatment group; if it is lower than expected, ABCSG and Amgen may choose to modify the sample size.

2.6 Long-Term Follow-Up

Subjects who do not receive OL denosumab will be followed for DFS, BMFS, and OS Q12M by clinic visits or telephone contacts starting from their EOT visit until approximately 66 months after PADCD.

Subjects who receive OL denosumab will have Q6M on-site visits for administration of denosumab (60 mg SC Q6M) and monitoring of serious adverse events (SAEs) and (serious) adverse events (AEs) of special interest and will be followed Q12M by clinic visits or telephone contacts starting from their EOT visit for DFS, BMFS, and OS until either approximately 66 months after PADCD or completion of treatment, whichever is longer.

In addition, for all subjects, data on clinical fracture recording, concomitant bone affecting medication, anti-cancer related therapy and supplements will be assessed and BMD data will

be collected at any time a DXA scan is performed for BMD analysis as SOCfrom PADCD to end of study (EOS).

To avoid extensive simultaneous visits or contacts each year, the EOT visit is chosen as starting point for the LTFU visits rather than the PADCD.

2.7 End of Open-label Treatment

The end of OL treatment (EOOLT) is defined for each subject as the point in time the last dose of OLP denosumab is administered.

2.8 End of Study

The patient individual EOS visit is defined as the last formal visit or contact for a subject or an unscheduled study visit in case of early withdrawal from study.

EOS reasons include 3 early EOS reasons (death, lost to follow-up, consent withdrawal) and per protocol EOS. At EOS, the EOS electronic Case Report Form (CRF) has to be completed for each subject.

2.9 Study Termination

Study termination will occur when the last subject has completed their last formal visit or last formal contact or an unscheduled study visit in case of early withdrawal from the study. This is expected to occur when the last subject completes their EOS visit or their end of ZA substudy visit (see section 11).

3. DATA SCREENING AND ACCEPTANCE

3.1 General Principles

Data will be continually screened using data acceptance programs in a blinded fashion during the conduct of the study. Before implementing data checks not specified in the Data Management Plan (DMP), the feasibility of adding the checks to the DMP will be discussed. Data issues identified by data acceptance programs will be communicated to ABCSG CDM for review and resolution.

As part of the data acceptance procedure, all planned tables, listings, and graphs will be generated and reviewed to identify any additional data issues. Any critical issues identified must be resolved with CDM before final acceptance of the data.

3.2 Handling of Missing and Incomplete Data

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

3.2.1 Early Withdrawal from Study and Lost-to-follow-up

Reasons for withdrawals are described and censoring patterns are compared between treatment groups in an exploratory manner (see also Section 6.4.4).

3.2.2 Missed Visit or Non-evaluability of a Specific Clinical Measurement at its Planned Clinical Visit

Only evaluable measurements are considered for the analyses. No values are imputed for missed or non-evaluable visits.

3.2.3 Dates

For fractures with a missing x-ray date, the fracture date captured on the eCRF will be used. If the fracture date or a date for an oncological event is completely missing, it will not be imputed. If it is partially missing, imputed dates will be used to derive the time to event for the efficacy analyses.

	Missing	Impute	Exception
Fracture Date	Day	01	Default to Day 1 of the study if an event started the same year and month as Day 1
	Day /Month	01JAN	Default to Day 1 of the study if an event started the same year as Day 1

If a start or stop date for an AE or a concomitant medication use is completely missing, it will not be imputed. If it is partially missing, imputed dates will be used to derive the duration of the AE or the medication use. If the medication is a proscribed therapy per protocol, the imputed start date will be used to identify at which point the data collected for this subject will be excluded from the per-protocol analysis set. Missing years will not be estimated under any conditions.

Imputation Rules for Partial Adverse Event or Concomitant Medication Start and Stop Dates

	Missing	Impute	Exception
Start date	Day	01	Default to Day 1 of the study if an event started the same year and month as Day 1
	Day /Month	01JAN	Default to Day 1 of the study if an event started the same year as Day 1
Stop date	Day	Last day of the month	Default to the EOS visit if the imputed event stop date is after the EOS visit
	Day/Month	31Dec	

3.3 Potential Outliers

Scatter plots will be examined to identify potential outliers in any of the continuous variables. Observations found to be due to data entry errors will be corrected by the study team before data freeze. Potential outliers that are not due to data entry error will be included in the analyses. The validity of any questionable values will be confirmed. No valid measurement will be purposely excluded from descriptive or inferential analyses. However, sensitivity

analyses may be conducted to evaluate the influence of extreme values in the data. These analyses will be documented in the study report.

3.4 Distributional Characteristics

The assumptions underlying the parametric models analyzed for continuous data will be checked. In cases where residuals indicate marked departures from the assumptions, additional sensitivity analyses will be performed using transformations or alternate methods such as nonparametric or robust procedures.

3.5 Validation and Configuration Management

Prior to freezing the database, CDM will perform data testing and ABCSG will run all planned tables, graphs and listings in a blinded fashion, the analysis file creation programs and the output programs. SAS® version 9.1 or higher will be used for all case report tabulations, tables, listings and graph creations. Validated departmental standard macros will be used whenever appropriate for the creation of these items. Output from these standard macros will not be validated further although they will be checked for quality. Output derived from non-standard macros will be fully validated.

4. ANALYSIS SET

4.1 Definition of Analysis Sets

The analyses for the primary, secondary and exploratory endpoints (except BMD and vertebral fracture endpoints) will be conducted on the full analysis set (FAS). Analysis of the BMD endpoints will be based on the BMD analysis set. The vertebral fracture endpoints will be based on the Vertebral fracture analysis set. In addition, exploratory analyses for [redacted] will be performed on the FAS. Analysis of the per-protocol (PP) population will be considered supportive. OLP summaries of efficacy will be based on the OLP Denosumab (Dmab) Analysis Set.

Safety analyses will be conducted on the safety analysis sets, while OLP concomitant medication information will be based on the FAS.

4.1.1 Full Analysis Set (FAS)

The FAS is defined as all subjects who are randomized. Every patient will be analyzed according to the randomized (and not necessarily the actual) treatment.

4.1.2 OLP Denosumab (Dmab) Analysis Set

The OLP Dmab analysis set will consist of all subjects who joined the OLP phase. Hence, all patients received placebo during the double-blind phase (randomized treatment) and fulfill all OLP eligibility criteria.

4.1.3 BMD Analysis Set

This BMD analysis set includes subjects defined in Section 4.1.1 with evaluable DXA scan values for the endpoint of interest (lumbar spine, total hip or femoral neck) at baseline and the post baseline timepoint under consideration (12, 24 or 36 months). DXA scans must be performed using the same Hologic device and be taken on the same side of the body (as the baseline measurement). Subjects in this subset will be analyzed according to their original treatment assignment, regardless of treatment received.

4.1.4 Vertebral Fracture Analysis Set

The Vertebral fracture analysis set includes subjects defined in Section 4.1.1 who have a baseline and $\geq$ 1 post baseline evaluation of vertebral fracture at or prior to the time point under consideration. This analysis set will additionally include subjects who have vertebrae (T4 - L4) with missing Genant semi-quantitative scores at baseline and whose first post baseline spinal radiograph shows no fracture on the same vertebrae because it can be inferred that the baseline scores would have also shown no fracture had they been available. Note that this subset could potentially be different from endpoint to endpoint due to missing data. Subjects in this subset will be analyzed according to their original treatment assignment, regardless of treatment received.

4.1.5 Per-Protocol (PP) Population

This PP population subset includes subjects defined in Section 4.1.1, who are compliant with the protocol, as characterized by the following criteria:

- Subjects who received $\geq$ 1 dose of investigational product, where possible
- The absence of violating any inclusion / exclusion criteria for subject eligibility

Subjects will be analyzed according to the randomized treatment. For subjects who deviate from the randomized treatment, all data collected on or after the first occurrence of incorrect treatment will be excluded from the analysis in the subset.

For subjects who received proscribed medications or therapy on study, all data collected on or after the first occurrence (for those indicated as any use) or after the 30th day of use (for those indicated as > 30 days cumulative) will be excluded from any analyses in the subset.

Proscribed therapies include commercially available denosumab (any use),IV bisphosphonates (any use), oral bisphosphonates (>30 days cumulative), fluoride (for osteoporosis; any use), strontium ranelate (any use), systemic estrogen (> 30 days cumulative), selective estrogen receptor modulators (SERM; e.g. raloxifene; >30 days cumulative), tibolone (> 30 days cumulative), calcitonin (> 30 days cumulative), anabolic steroids (any use), parathyroid hormone (or a derivative; any use), calcitriol (any use), tamoxifen (any use) and any other medication that is known or suspected to have activity on bone metabolism (except calcium and vitamin D; any use).

The time to first on-study clinical fracture will be re-defined as the time from Study Day 1 to the date of event or censoring. Subjects who (i) deviated from randomized treatment or (ii) took prohibited medication, and have not yet had the event of interest, will be censored at PADCD, or the earliest violation date of either (i) or (ii).

4.1.6 Safety Analysis Set (SAF)

The SAF will consist of all subjects who are randomized and receive at least 1 dose of investigational drug. These subjects will be analyzed according to their actual treatment received, where subjects who received at least 1 dose of denosumab will be analyzed in the denosumab treatment group regardless of the randomized treatment.

4.1.7 OLP Denosumab (Dmab) Safety Analysis Set

The OLP Dmab safety analysis set will consist of all subjects who receive at least 1 dose of OLP denosumab.

4.2 Subgroups

The primary endpoint will be analyzed in subgroups by randomization strata and other factors assessed at the screening phase:

- Prior AI usage (Yes or No)*
- Total lumbar spine BMD score at baseline (T-score < –1.0 or T-score ≥ -1.0)*

  *Based on the randomized stratum (ie, from an Interactive Voice/Web Response System [IVRS]), regardless of the subject's actual value.
- Age (<50, 50-59, 60-69, 70-79 or ≥80 years)
- T-stage (T0+Tis+T1 or T2+T3+T4)
- pN-stage (positive (pN1 + pN≥2) or negative (pN0))
- Grading (G1, G2+Gx or G3)
- Primary Tumor (ductal invasive carcinoma, lobular invasive carcinoma or other)
- Receptor status (ER+/PgR+ or other)
- Chemotherapy (yes - adjuvant, yes - neoadjuvant or no)

Subgroups other than stratification variables will be re-examined for appropriateness and may be re-categorized or eliminated to ensure at least 10% of subjects within each subgroup and an adequate number of events for analysis before unblinding.

5. PRIMARY, SECONDARY, SAFETY AND EXPLORATORY ENDPOINTS

5.1 Primary Endpoint

The time to first on-study clinical fracture will be defined as the number of days from randomization to the date of the x-ray confirming the clinical fracture. Subjects who die or withdraw without experiencing a clinical fracture will be censored at the date of last contact before PADCD (including date of scheduled and unscheduled, clinic and telephone visits, of early study termination and of deaths) or EOS visit whichever is earlier. If a subject does not have a post-randomization visit, the subject will be censored at randomization. Only clinical fractures which occur prior to or on PADCD will be included in the analysis for the primary endpoint.

A fracture is defined as clinical if CTCAE grade is greater than 1. Fractures with a missing CTCAE grade will not be considered clinical fractures. All clinical fractures, except those of the skull, face, fingers, and toes, which are typically not associated with osteoporosis, will be included in the analysis of clinical fractures. Clinical fractures of any trauma severity are subject to medical review and all exclusions of fractures as events will be documented prior to unblinding.

For the exploratory analysis of [redacted]

5.2 Secondary Endpoints

Fracture-related secondary endpoints:

- The percent change in total lumbar spine, total hip and femoral neck BMD from baseline to 36 months in subjects with evaluable DXA scans using the same Hologic device
- Subject incidence of new vertebral fractures (morphometric fractures identified from on study x-rays and clinical vertebral fractures confirmed by x-rays) at Month 36.

- Subject incidence of a new or worsening of pre-existing vertebral fractures (morphometric vertebral fractures identified from on study x-rays and clinical vertebral fractures confirmed by x-rays) at Month 36.

Disease outcome-related secondary endpoints:

- DFS determined by the time from randomization to the first observation of disease recurrence or death from any cause
- BMFS determined by the time from randomization to the first observation of bone metastasis or death from any cause
- OS determined by the time from randomization to death from any cause

5.3 Safety Endpoints

- Subject incidence of treatment-emergent AEs
- Clinically significant changes in laboratory values
- Subject incidence of anti-denosumab antibody (binding and neutralizing) formation

5.4 Exploratory Endpoints

[redacted]

[redacted]

5.5 Derived Endpoints

Baseline Value

Baseline value is the latest recorded measurement on or prior to the day of the first dose of IP. If there is no such baseline value available, BMD measurements recorded within 45 days after day 1 are also acceptable as a baseline measurement for BMD. In case of several candidates, the closest value to day 1 will be selected. If day 1 is missing, the measurement closest to randomization date within a window of 45 days is used as BMD baseline measurement.

If a subject doesn't receive IP, baseline is the latest recorded measurement on or prior to the enrollment date.

Change from Baseline Value

The arithmetic difference between a value of interest and a baseline value: Change from baseline value = (value of interest - baseline value)

Percent Change from Baseline Value

The ratio of the arithmetic difference between a value of interest and the baseline value to the baseline value multiplied by 100: Percent change from baseline value = [(value of interest - baseline value) / baseline value] * 100.

Subject Incidence Rate

The subject incidence rate for a given event is defined as the number of subjects with one or more reported occurrence of the event divided by the number of subjects who have the opportunity to report the event.

Time to Event for Efficacy

Time interval (days) from the randomization date to the date of occurrence of the event or censorship during the given period:

Time interval = (date of occurrence of the event or censorship – randomization date) +1

BMFS Time

Time interval (days) from the randomization date to the date of first occurrence of bone metastasis or death from any cause, whichever comes first. Subjects last known to be alive, who have not experienced bone metastasis, are censored at their last assessment (i.e., bone scan) date, or at the end of LTFU whichever comes first. Subjects who had their first occurrence of bone metastasis before randomization will be censored at their randomization date.

DFS Time

Time interval (days) from the randomization date to the date of first evidence of local or distant metastases, contra-lateral breast cancer, secondary carcinoma, or death from any cause (whichever comes first). Subjects last known to be alive, who have not experienced recurrence of disease, are censored at their last contact date (including date of scheduled and unscheduled, clinic and telephone visits, of early study termination and of deaths), or at the end of LTFU/PADCD (for the futility analysis)/interim data cut-off date (for the additional DFS analyses), whichever comes first. Subjects who had first disease recurrence before randomization will be censored at their randomization date.

OS Time

Time interval (days) from the randomization date to the date of death from any cause. Subjects last known to be alive are censored at their last contact date, or at the end of LTFU, whichever comes first.

Treatment Emergent Adverse Events

All AEs occurring on or after the first dose of IP through 30 days after the last dose of IP (EOT) will be considered treatment emergent AEs. The following steps will be applied when AE onset date is partial or missing:

- All cases where onset dates are partial, the AE will be considered treatment emergent unless it is impossible for this to be the case.
- If onset date is missing, and outcome date is on or after first dose, the AE will be considered treatment emergent.
- If onset date is missing, and outcome date is partial and does not clearly indicate that AE falls strictly before first dose, the AE will be considered treatment emergent.

Corrected Calcium

Calcium values need to be corrected when the serum albumin value is < 40 g/L. The corrective formula is indicated below; the raw calcium value is same as the corrected calcium value when the serum albumin is >=40 g/L.

Corrected Calcium (mmol/L) = Total Calcium (mmol/L) + 0.025 [40 (g/L) – Albumin (g/L)]

6. PLANNED METHODS OF ANALYSIS

The analyses will be carried out using statistical analysis system (SAS) software (SAS® version 9.1 or higher), by members of the biostatistics group at ABCSG. If necessary, selected other software will be used, e.g. StatXact (Version 8 or higher) for exact statistical tests not offered by SAS.

6.1 General Approach / Considerations

The primary analysis will take place after PADCD when all subjects attended their last study visit (considered as EOT visit). The primary analysis includes the primary endpoint, the first three secondary endpoints concerning the changes in BMD and vertebral fractures, exploratory endpoints and the safety endpoints.

Final analysis will be performed at the end of the main study, approximately 66 months after PADCD, when both the LTFU and the OLP are finalized for all patients. Hence, after the last LTFU-patients had their final visit in the LTFU and the last OLP-patients received their last treatment in the OLP. The final analysis will include an exploratory [redacted] analysis as well as the main analyses for efficacy of the secondary endpoints BMFS and OS. Additionally,

[redacted]

[redacted] will be evaluated in a final exploratory analysis at this time point. BMD data, as well as data on [redacted], collected during LTFU, will be summarized descriptively. Data collected during the ZA substudy will not be used for the final analysis. Data from patients included to the ZA substudy will be censored at the EOT Visit (OLP) or the EOS Visit (LTFU and OLP), whichever occurs later.

The primary and secondary null hypotheses will be tested using the hierarchical analysis strategy and the Hochberg procedure (Westfall et al, 1999) to control the overall significance level of 0.05.

The primary null hypothesis will be tested first at a significance level of 0.05. If the primary null hypothesis is rejected, the secondary null hypotheses will be tested in a stepwise fashion over 6 steps, in the order A to F as detailed below, at a significance level of 0.05. In case any one of the hypotheses is not rejected at a previous step, all subsequent endpoints will be analyzed in a descriptive manner only. Some steps involve single null hypothesis and some involve multiple null hypotheses (ie, BMD endpoint A). If there are multiple null hypotheses, the Hochberg procedure will be used to control for multiplicity and the testing will proceed to the next step only if all null hypotheses are rejected. The hierarchical analysis strategy will involve the secondary null hypotheses for the following endpoints in the order specified:

Fracture-related secondary endpoints:

A. Percent change in total lumbar spine, total hip and femoral neck BMD from baseline to Month 36 in subjects with evaluable DXA using the same Hologic device (using the Hochberg procedure controlling for multiplicity)

B. Subject incidence of new vertebral fractures (morphometric vertebral fractures identified from study x-rays and clinical vertebral fractures confirmed by x-ray) at Month 36

C. Subject incidence of a new or worsening of pre-existing vertebral fractures (morphometric vertebral fractures identified from on study x-rays and clinical vertebral fractures confirmed by x-rays) at Month 36

Disease outcome-related secondary endpoints:

D. DFS is defined as any evidence of disease recurrence or death from any cause

E. BMFS determined by the time to first occurrence of bone metastasis (either symptomatic or asymptomatic) or death from any cause

F. OS

Statistical analysis procedures and techniques will be reviewed in the light of current practice and new techniques, prior to breaking treatment blind, and if appropriate alternative newer technical procedures will be adopted.

6.2 Handling of Incorrect Stratification

For analyses where stratification needs to be adjusted for, the following general principles will be followed:

- Stratified analyses that are intended to evaluate the treatment effect will be based on the randomized stratum (ie, from IVRS), regardless of the subject's actual value. If the error rate is more than 5%, then a sensitivity analysis using the actual stratum will also be performed for the primary and secondary endpoints.
- Covariate analyses where covariates are stratification factors should be based on subject's actual value.
- For subgroup analyses, additionally
    - If the subgroup variable is a stratification factor, sensitivity analyses according to actual value may be needed.

6.3 Analysis Windows

6.3.1 Study Month

Study Month will be defined as:

CEIL(date of event – date of Study Day 1*))/30.44)

* If Study Day 1 is missing, the randomization date will be used instead

6.3.2 Incidence Endpoints

Incidence between periods of time will use the formula in Section 6.3.1 to derive month from Study Day 1 to event between Months 1 and 12, between Months 1 and 24 and between Months 1 and 36.

6.3.3 Analysis at Specified Times

Per-protocol, all tests and procedures will aim to be performed within the specified windows according to the definitions given in the tables below. To allow for variations in scheduling the following visit windows will be used to assign evaluations to a most appropriate nominal visit for analysis and summarization.

One set of windows will be applied to endpoints that are measured semi-annually such as laboratory assessments (chemistry and hematology). In order to get as many data points as possible for summarization, there will be no gaps between visit windows. If more than one assessment is available within a window, the assessment closest to the target day at each nominal visit (Month 6, Month 12, …) will be used. If Study Day 1 is missing, the randomization date will be used instead. BMD measurements recorded within 45 days after Day 1 are acceptable as baselines (see Section 6.1).

Nominal Visit Window	Target Day	Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1
Month 6	183	Study Day 2 to 274
Month 12	365	Study Day 275 to 456
Month 18	547	Study Day 457 to 634
Month 24	729	Study Day 635 to 821
Month 30	911	Study Day 822 to 1003
Month 36	1093	Study Day 1004 to 1186
Month 42	1275	Study Day 1187 to 1368
Month 48	1457	Study Day 1369 to 1550
Month 54	1637	Study Day 1551 to 1735
Month 60	1821	Study Day 1736 to 1917

A different set of windows will be applied to endpoints that are measured annually such as DXA and antibody assay. In order to get as many data points as possible for summarization, there will be no gaps between visit windows. If more than one assessment is available within a window, the assessment closest to the target day at each nominal visit will be used.

Nominal Visit Window	Target Day	Definition (Study Day)
Baseline*	1	Last evaluation prior to or on Study Day 1
Month 12	365	Study Day 2 to 547
Month 24	729	Study Day 548 to 912
Month 36	1093	Study Day 913 to 1277

Month 48	1457	Study Day 1278 to 1642
Month 60	1821	Study Day 1643 to 2007

* If there is no such baseline value available, DXA (BMD) measurements recorded within 45 days after day 1 are also acceptable as a baseline measurement for BMD. In case of several candidates, the closest value to day 1 will be selected. BMD values assigned as the baseline value cannot also be assigned to month 12.

6.4 Analysis of Key Study Endpoints

6.4.1 Efficacy Analysis

The efficacy analyses for the primary and the secondary endpoints will be based on the FAS and on PP population approach except for the BMD endpoints which are restricted to patients in the BMD analysis set and vertebral fracture secondary endpoints which will be based on the Vertebral fracture analysis set. The PP analyses will be considered supportive. Exploratory analyses will be conducted on the FAS unless otherwise specified.

OLP efficacy analyses will be based on the FAS, while descriptive summaries of OLP specific data will be based on the OLP Dmab Analysis Set.

6.4.1.1 Primary Endpoint

The time to first on-study clinical fracture as defined in Section 5.1 will be analyzed using a Cox model (Cox, 1972) including treatment groups as the independent variable and stratified by the randomization stratification factors. A supportive analysis will be performed using a log-rank test stratified by the randomization stratification factors.

[redacted] (see Section 6.4.6). Summary statistics from the Cox model will include the hazard ratio (95% CI) of denosumab compared with placebo. Furthermore clinical fracture rates (95% CI) and differences between clinical fracture rates (95% CI) at 36 months will be estimated using Kaplan-Meier methodology. The time to first clinical fracture will be estimated and graphically displayed for each treatment group by Kaplan-Meier curves.

The proportionality assumption of the Cox model will be investigated with a time-dependent exploratory variable, which is defined as treatment multiplied by the logarithm (base e) of the

time-to-event. If there is evidence of a departure from the adjusted model assumptions the reason will be explored and reported.

Further sensitivity analyses based on different subgroups will be performed using the main analysis method as described in the first paragraph of Section 6.4.1.1.

- The first sensitivity analysis will consider the subjects who withdraw from the trial due to a >10% loss of BMD in the total hip or lumbar spine over any 1-year period as having had a clinical fracture at the time of withdrawal. This will assess if study monitoring has resulted in informative censoring of subjects within the trial.
- A second sensitivity analysis will be conducted on the PP population to assess the effect of non-protocol compliant subjects.
- The main analysis will consider all available subject data regardless of the subject's time since last treatment with study drug. A third sensitivity analysis will consider the influence of subject experience subsequent to discontinuation from study treatment. This will be assessed by including time varying covariates to the Cox model: A binary time dependent covariate is used to assess the effect of treatment discontinuation. This time dependent covariate has a value of zero during treatment application and it changes to one if the patient discontinues study treatment for any reason (except in case of study termination) prior to PADCD.

If any of the sensitivity analyses cause conflicting results to the main analysis then similar sensitivity analyses will be conducted for the secondary endpoints related to BMD and fractures.

The final exploratory analysis of clinical fractures will use the main analysis method as described in the first paragraph of Section 6.4.1.1.

Additionally, two sensitivity analyses will be performed to account for treatment cross-over associated with the OLP:

First, a rank-preserving structural failure time model (RPSFTM) (Robins and Tsiatis, 1991) to correct the treatment effect estimate for bias introduced by cross-over from placebo to OLP Dmab will be conducted. The RPSFTM provides a randomization-based estimate of treatment effect assuming a multiplicative effect of treatment on time to event. The approach also allows reconstruction of the placebo time to event curve as if all patients initially

randomized to placebo never switched to OLP Dmab. The RPSFTM adjusted HR (95% CI) and the acceleration factor (95% CI) will be presented. Additionally a Kaplan-Meier curve may be used to display the RPSFTM estimate.

One of the key assumptions of the RPSFTM is the "common treatment effect", ie, the treatment effect for the patient is the same regardless of when the patient started taking Dmab (double-blind phase or OLP). To check this key assumption, an analysis will be conducted for two groups of patients, looking at patients who received OLP Dmab on the one hand and who received Dmab during the double-blind phase on the other hand. In addition, major baseline characteristics will be compared between the two groups of patients.

Second, the main analysis method will be repeated, but patients receiving OLP Dmab will be censored at the date of first receiving OLP treatment.

6.4.1.2 Secondary Endpoints

The BMD secondary endpoints will be based on the BMD analysis set. The vertebral fracture secondary endpoints will be based on the Vertebral fracture analysis set. The disease outcome-related secondary endpoints will be based on the FAS.

Sensitivity analyses will be conducted on the PP population to assess the effect of non-protocol compliant subjects. If there are conflicts between the results of the main analysis and the sensitivity analyses of the primary endpoint, additional sensitivity analyses for the secondary endpoints similar to the primary endpoint will be performed.

Percent Change in Lumbar Spine, Total Hip and Femoral Neck Bone Mineral Density from Baseline to Month 36

The percent change in lumbar spine (L1-L4, Vertebrae that are clearly fractured will be excluded), total hip and femoral neck BMD from baseline to 36 months will be calculated for the BMD analysis set. Missing data will not be imputed. There will be no distinction between left and right sides of the total hip and femoral neck BMD. The main analysis of these endpoints will employ an analysis of covariance (ANCOVA) including treatment group as the independent variable and adjusted for baseline value and the randomization stratification factors.

Summary statistics will include the observed and estimated percent changes, 95% confidence interval (CI), differences with 95% CI between the percent changes in the two groups at Month 36. Model fit will be assessed by residual analysis and other diagnostic statistics. In case of heteroscedasticity, appropriate transformations will be performed to assess sensitivity to this violation of the model assumption.

Additional investigatory analyses of covariates will be conducted.

Incidence of new vertebral fractures at Month 36

The presence or absence of a new vertebral fracture (morphometric vertebral fractures identified from on study x-rays and clinical vertebral fractures confirmed by x-rays) will be noted at 36 months (over a 36 month evaluation period) and classified according to Genant semi-quantitative grades.

There are 2 situations where Genant semi-quantitative grades can be missing: either a spine radiograph is not taken or individual vertebrae on a radiograph are not evaluable. Because a vertebral fracture can only get worse or at best remain at the same severity over time, the Genant semi-quantitative grade for a vertebra can only increase or remain the same. This rationale will be used for handling missing data as follows and these imputations steps will be done before any analysis:

- If the Genant semi-quantitative grade on a given vertebra is 0 (ie, no fracture) at a later visit, it is reasonable to impute a grade of 0 for the same vertebra at previous visit(s) including baseline.
- If the Genant semi-quantitative grade on a given vertebra is > 0 at a later visit, any missing value for the same vertebra from previous visits will remain missing because it cannot be certain when the vertebral fracture occurred., i.e., the value > 0 is applicable at the time of first occurrence

Note that baseline grades cannot be carried forward to impute missing post baseline values.

The independent assessment of vertebral fractures classified according to the Genant semi-quantitative grading conducted by the Central Review Committee (CRC) will be used for the main analysis. If two CRC assessments are available, the second reviewer's assessment will be used.

The main analysis of this endpoint will use logistic regression models including treatment groups as the independent variable and stratified by the randomization stratification factors. Additional investigatory analyses of covariates may be conducted.

Summary statistics will include the crude incidences at Month 36 and EOS, 95% CI, differences between incidences in the two groups and 95% CI. Hosmer-Lemeshow tests will be calculated to assess the goodness-of-fit of the models. Deviance and Pearson residuals and the diagonal elements of the hat matrix will be checked to detect outliers and influential observations. Deviations from the model assumptions will result in further sensitivity analyses.

Incidence of new or worsening vertebral fractures at Month 36

The analysis described above will be repeated considering new or worsening vertebral fractures.

Disease-free survival

Main and [redacted] analyses (DFS as defined in section 5.5) employ the same methods as the primary endpoint. Additional investigatory analyses of covariates (see section 6.4.6) may be conducted. [redacted]

An additional sensitivity analysis will be performed censoring subjects who received any bisphosphonates or commercial available denosumab prior to EOT (with EOT reason 'requirements for alternative therapies'). Subjects who took an alternative therapy, and have not yet had a DFS event, will be censored at the time of EOT.

Bone metastasis-free survival

BMFS as defined in section 5.5 will employ the same methods as the primary endpoint. Sensitivity analyses accounting for treatment cross-over associated with the OLP will be performed (as described for the primary endpoint). Additional investigatory analyses of covariates (see section 6.4.6) may be conducted.

Overall survival

OS as defined in section 5.5 will employ the same methods as the primary endpoint. Sensitivity analyses accounting for treatment cross-over associated with the OLP will be performed (as described for the primary endpoint). Additional investigatory analyses of covariates (see section 6.4.6) will be conducted.

6.4.2 Safety Analyses

Safety analyses, prior to OLP, will be based on the SAF. Safety analyses in this study will evaluate the safety profile of denosumab as compared with placebo. No formal statistical testing will be conducted for the safety analyses.

OLP AE data will be based on the OLP Dmab safety analysis set, OLP concomitant medication information will be based on the FAS. Both will further evaluate the safety profile of denosumab.

6.4.2.1 Adverse Events

All AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA version 17.0 or higher). AEs will be classified as treatment emergent following rules outlined in Section 5.5. Only treatment emergent AEs will be tabulated. AEs reported but not classified as treatment emergent will be listed.

The subject incidence rates of treatment-emergent AEs reported during the treatment period will be tabulated by system organ class and preferred term.

Additional summary tables will be provided separately for

- all AEs

- Common Terminology Criteria for AEs (CTCAE) grade 3, 4, or 5 AEs,
- AEs leading to investigational product discontinuation
- AEs leading to study discontinuation
- all investigational product related AEs
- Investigational product related CTCAE grade 3, 4, or 5 AEs
- Investigational product related AEs leading to investigational product discontinuation
- Investigational product related AEs leading to study discontinuation
- SAEs
- Investigational product related SAEs
- Fatal AEs

[redacted]

Narratives of deaths and SAEs will also be provided.

An adjusted AE rate of treatment-emergent AEs reported during the treatment period will be presented as the number of AEs reported divided by subject years in the treatment period. The individual subject years in the treatment period are the number of days between a subject's first dose date to the EOT date, inclusive, divided by 365.25 (ie, [EOT date – first dose date + 1]/365.25). The subject years for a treatment group in the treatment period are the sum of individual subject years in the treatment period for all subjects in the treatment group.

The final safety analysis, as well as safety analyses in an exploratory manner, will include

[redacted]

6.4.2.2 Clinical Laboratory Assessments

Laboratory parameters will be summarized over time using descriptive statistics for recorded values and change from baseline. Furthermore shifts tables will display the incidence of shift

of toxicity grade (CTCAE version 3.0) in recorded values from baseline to “worst” on-study value. Graphs for other parameters of interest may also be presented.

6.4.2.3 Anti-denosumab Antibodies

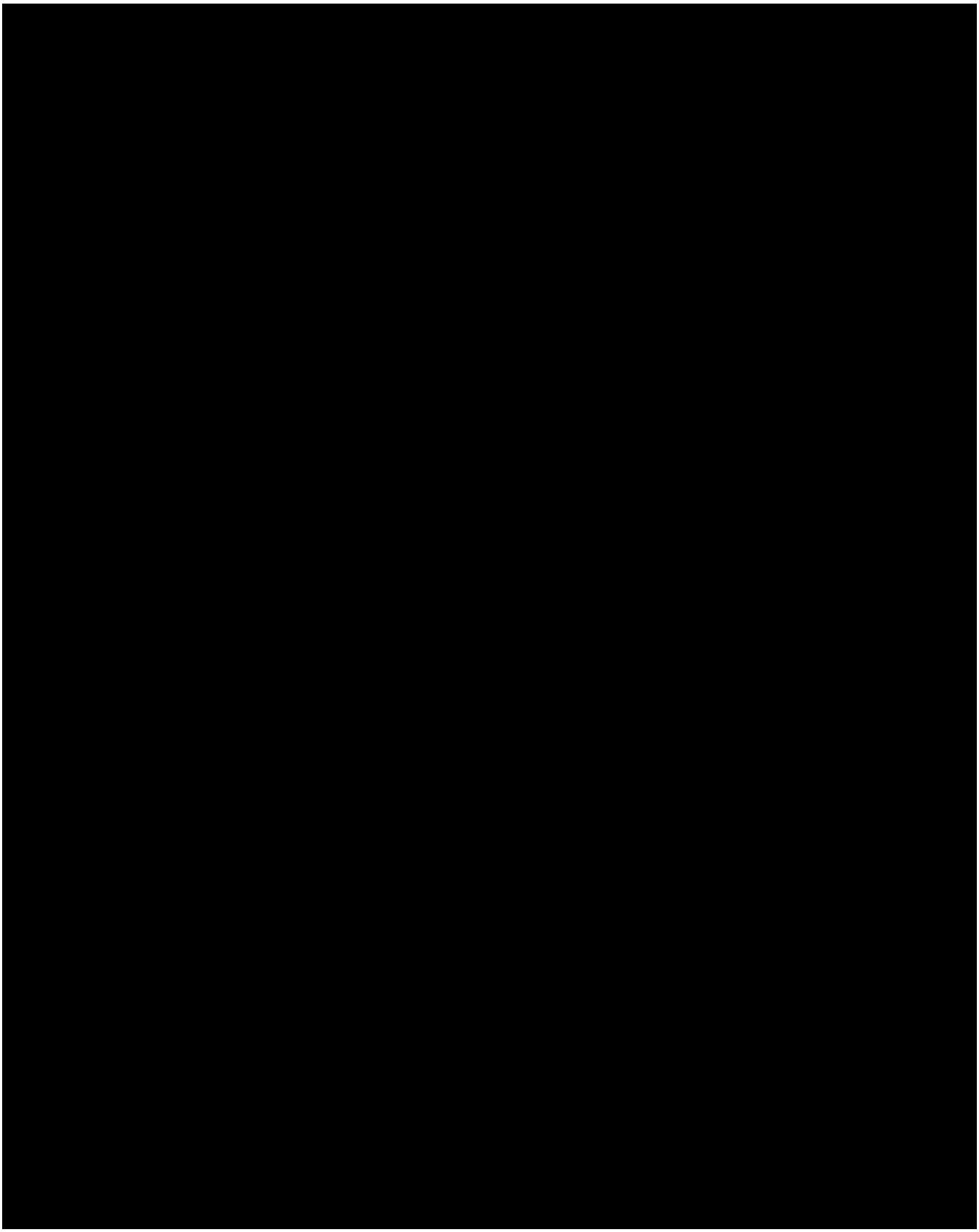
The incidence and percentages of subjects who develop binding and neutralizing anti-denosumab antibodies at any time will be tabulated by treatment group.

6.4.2.4 Concomitant Medication

For the final safety analysis, descriptive summary tables for concomitant bone affecting medication (including bone targeted therapy, glucocorticoids, antiepileptic drugs, antidepressants, insulin), as well as anti-cancer related therapy, will be provided.

6.4.3 Exploratory Analyses

[redacted]



[redacted]

Translational endpoints

[redacted]

6.4.4 Treatment Discontinuation and Lost-to-Follow-Up

The basic assumption that censored (lost-to-follow-up) patients and treatment discontinuations will distribute equally to both treatment groups will be checked in an exploratory manner. For this purpose the numbers of a) censored patients and b) treatment

discontinuations (except Completed Investigational Product) will be compared between both treatment groups. This will be described by Kaplan-Meier curves, where a) censoring and b) treatment discontinuations are handled as interesting events and other observations are censored at their last follow-up time or death time.

6.4.5 Subgroup Analysis

The primary endpoint will be analyzed using the Cox Model described in Section 6.4.1.1 within each subgroup indicated in Section 4.2 (if there are at least 10% of subjects within each sub-category). The primary endpoint will be summarized using descriptive statistics within each sub-category of the subgroup variables.

The treatment by subgroup interaction will be tested as follows: the Cox Model specified in Section 6.4.1.1 will have the subgroup variable and the interaction term of treatment-by-subgroup added as covariates. If the p-value of interaction term is ≥ 0.05, the treatment-by-subgroup interaction is not significant. Otherwise, if the interaction is significant, the Gail and Simon test (Gail & Simon, 1985) will be used to test for qualitative interaction at the significance level of 0.05. The interaction will be described graphically.

6.4.6 Covariates

The relationship of the following covariates to the primary and secondary efficacy endpoints will be explored:

- DXA device type (Hologic or other) [for analyses of BMD endpoints only]*
- Prior AI usage (Yes or No)*
- Total lumbar spine BMD score at baseline (T-score less than –1.0 or T-score ≥ -1.0)*

* Based on subject's actual value.

- Age (<50, 50-59, 60-69, 70-79 or ≥80 years)
- T-stage (T0+Tis+T1 or T2+T3+T4)*
- pN-stage (positive (pN1 + pN≥2) or negative (pN0)
- Grading (G1, G2+Gx or G3)

- Primary Tumor (ductal invasive carcinoma, lobular invasive carcinoma or other)
- Receptor status (ER+/PgR+ or other)
- Chemotherapy (yes - adjuvant, yes - neoadjuvant or no)
- Neoadjuvant trastzumab therapy (yes or no)
- Biphosphonate medication (yes or no)
- Human Epidermal growth factor Receptor 2 (HER-2)-neu (positive, negative or missing)
- Other covariates reported in literature as important or from other ongoing Amgen studies will be considered in the analysis if considered appropriate at the time of analysis.

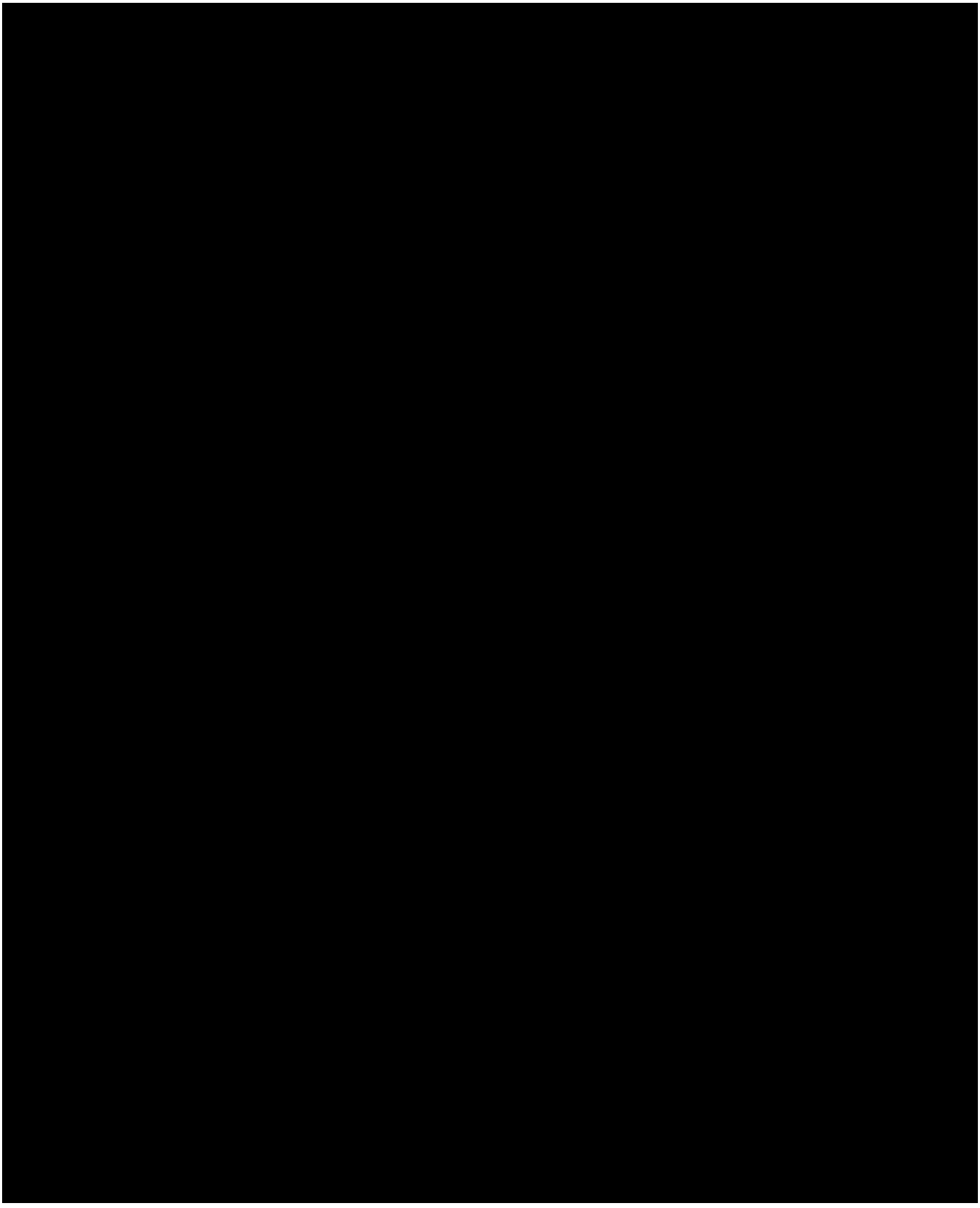
*T-stage will be fitted along with the chemotherapy covariate to account for neo-adjuvant therapy.

The following covariate analyses will be conducted for the primary and secondary endpoints:

- For the primary endpoint, the Cox model defined in Section 6.4.1.1 adding each above indicated covariate individually as a fixed (or continuous) effect will be used to obtain the 2-sided CIs for the covariate-adjusted estimates of the treatment and covariate effect with their respective p-values. The analysis set will be the same as used in the main analysis of the primary endpoint.
- For the primary endpoint, the Cox model defined in Section 6.4.1.1 adding all potential covariates listed above as fixed (or continuous) effects will be used to obtain the 2-sided CI for the covariate-adjusted estimate of the treatment effect and its p-value. The analysis set will be the same as used in the main analysis of the primary endpoint.

6.4.7 SAS Code Fragments

The following statements are code fragments used for the main statistical analyses. They are based on SAS® version 9.1. Italic words will be replaced by correct names of the datasets and variables (according to the database). In time-to-event data events are coded with 1 and non-events (censoring) with 0.



[redacted]

7. FURTHER TABULAR SUMMARIES AND LISTINGS

Description of tabular summaries:

The following demographic information will be summarized by randomized treatment group:

- Age: this summary will include number of observations, mean, SD, minimum, maximum and number of subjects with missing values in case of an approximately normal

distribution. Otherwise, mean and SD are replaced by the median and minimum and maximum.

- Race: grouped into White/Caucasian and others

The following baseline, breast cancer history and BMD information will be summarized by randomized treatment group:

- Stratification criteria for randomization
- T-stage, pN-stage, Grading, primary tumor, estrogen receptor (ER) and progesterone receptor (PgR) expression: the number and percentage of subjects will be summarized.

The following withdrawal information will be summarized by randomized treatment group:

- Subject disposition status (randomized, treated, withdrawn, completed); the number and percentage of subjects will be summarized.
- Reason for withdrawal (according to "End of Treatment" CRF): the number and percentage of subjects will be summarized.
- mean time under observation per patient from randomization to lost-to-follow up date or end-of-study visit

The following exposure / follow-up information will be summarized by randomized treatment group:

- Number of doses

Any data collected relating to important protocol violations will be listed and summarized.

Accesses to individual subject treatment assignments (unblinding) and their reasons are summarized in a descriptive manner.

8. INTERIM ANALYSES AND EARLY STOPPING GUIDELINES

An external DMC will be formed. Details regarding this DMC will be provided in a separate DMC charter. Selected Austrian Breast and Colorectal Cancer Group (ABCSG) and Amgen staff may serve as liaisons to the external DMC, but will not be voting members, and will not be unblinded to the results. The DMC will have access to subject's individual treatment assignments. To minimize the potential introduction of bias, these individuals will not have

any direct contact with the study site personnel or subjects. Members of this external DMC will include, at a minimum, physicians external to Amgen and appropriate statistical representation external to Amgen. The DMC will meet approximately annually. Records of all meetings will be archived. The DMC will communicate major safety concerns and recommendations regarding study modification or termination to Amgen senior management.

This DMC will review unblinded safety data. Safety analyses provided to the DMC will be descriptive in nature.

There will be no interim analysis based on the primary endpoint. An interim analysis for futility for the secondary endpoint DFS will be based on the FAS and performed after PADCD, at the time of primary analysis, by an independent statistician. Only data captured prior to or on PADCD will be included in the futility analysis. This interim analysis will use an informal futility bound. Study termination may be recommended by the DMC if the observed hazard ratio exceeds the hazard ratio for futility (0.85).

Following review of the interim DFS analysis results, the DMC determined that the hazard ratio did not indicate futility and that the study should continue further for approximately 66 months. In addition, the DMC recommended that a time-driven analysis for efficacy of the secondary endpoint DFS should be performed at approximately 18 months after PADCD, before any investigator/subject level unblinding occurred for entry into the OLP.

In the event the study is terminated, it will be terminated when all subjects have had the opportunity to attend one LTFU visit after PADCD (approximately 18 months after PADCD). Otherwise the study will continue further 60 months.

9. DFS ANALYSIS RECOMMENDED BY DMC

Based on the recommendation of the DMC from 28 April 2015 (DMC meeting on 7 April 2015) a time-driven analysis for efficacy of the secondary endpoint DFS will be based on the FAS and performed around 18 months after PADCD. The DFS analysis data cut-off date will be 15 September 2015. Only data captured prior to or on this data cut-off date will be included in the analysis. The analysis must occur before any unblinding on patient/investigator level took place.

As the interim analysis for DFS did not indicate futility, the study continues until approximately 66 months after PADCD. The DFS analysis around 18 months after PADCD will be considered the main DFS analysis for the hierarchical testing strategy (see section 6.1), with [redacted]. Analysis for efficacy of the secondary endpoints BMFS and OS will not be affected and performed after 66 months after PADCD.

10. DFS ANALYSIS 2018

Based on recommendation from the Study Steering Committee in early 2017 due to treatment crossover an additional, exploratory analysis of the secondary endpoint DFS will be performed early 2018. The data cut-off date will be 30 September 2017. Only data captured prior to or on this data cut-off date will be included in the analysis. The analysis will be based on the FAS and will include the main DFS analysis, as well as all above mentioned sensitivity analyses.

11. ZA SUBSTUDY

11.1.1 Exploratory Objectives

[redacted]

11.1.2 ZA Substudy design

[redacted]

Subjects enrolled in the main study who received denosumab during the OLP may choose to participate in this substudy.

Willing and eligible subjects who completed OL denosumab are randomized to 1 single 5 mg IV dose of ZA or to SOC in a 1:1 ratio by an IVRS. The randomization schedule will use

randomly permuted blocks and will be stratified by AI use at the timepoint of OLP denosumab completion (Yes or No).

Follow up will continue for a total of 18 months after day 1. Day 1 for both arms should be 8 months (± 4 weeks) from the last dose of OL denosumab. For the ZA arm, ZA will be administered on this visit.

Subjects who do not fulfill the eligibility criteria for or do not consent to the ZA substudy will complete OLP assessments only and will end the main study as planned.

11.1.3 Number of Subjects and Sample Size Calculation

For the ZA substudy, it is anticipated that approximately 200 subjects could be randomized.

Data is limited on BMD for subjects who discontinue denosumab 60 mg Q6M in the oncology setting. Therefore, data from postmenopausal osteoporosis was used in the sample size.

A phase 2 randomized blinded clinical trial in postmenopausal women with low bone mass assessed the effect of denosumab 60 mg Q6M on BMD and bone turnover markers (CTX and Osteocalcin) after long-term continued, discontinued, and restarting of therapy (Miller et al, 2008). One of the treatment cohorts received denosumab 210 mg Q6M for 24 months then placebo for the next 24 months. Based on this study, a 5% (SD = 4.3%), 5.2% (SD = 2.6%), and 3.9% (SD = 3.8%) decrease in the lumbar spine, total hip, and femoral neck BMD was seen during the 2 years after denosumab discontinuation (data on file). A paper reporting 22 case studies of postmenopausal women who received 5 injections (approximately 2.5 years) of denosumab 60 mg Q6M and were then given a single dose of ZA 6 months after the fifth injection (Lehmann and Aeberli, 2017). A 3.8% (SD = 2.8%), 1.7% (SD = 3.3%), and 0.6% (SD = 5%) decrease in the lumbar spine, total hip, and femoral neck BMD was seen during the 2.5 years after denosumab discontinuation.

A table showing the level of precision (ie, half width of 95% CI) for each treatment arm and each BMD type for different sample sizes, calculated based on the SD estimated from the 2 studies mentioned above, is presented below.

Table 1. Sample Sizes

Number of subjects in	Precision (half width of 95% CI)	
	SOC	IV ZA

each arm	Lumbar Spine	Total Hip	Femoral Neck	Lumbar Spine	Total Hip	Femoral Neck
75	± 1.0%	± 0.6%	± 0.9%	± 0.6%	± 0.8%	± 1.1%
100	± 0.9%	± 0.5%	± 0.7%	± 0.5%	± 0.7%	± 1.0%
125	± 0.8%	± 0.5%	± 0.7%	± 0.5%	± 0.6%	± 0.9%

IV = intravenous

11.1.4 Estimated ZA Substudy Duration

Subjects will be in the ZA substudy for a maximum duration of approximately 27 months (considering that the patient may enter the substudy 9 months after OLP denosumab completion and has a follow up of 18 months).

11.1.5 End of ZA Substudy

The patient individual end of ZA substudy visit is defined as the last formal substudy visit or contact for a subject or an unscheduled substudy visit in case of early withdrawal from the substudy. End of ZA substudy reasons include 3 early EOS reasons (death, lost to follow-up, and consent withdrawal) and per protocol EOS of ZA substudy.

The end of ZA substudy will therefor occur when the last subject participating in the ZA substudy has completed their last formal substudy visit or an unscheduled substudy visit in case of early withdrawal from the ZA substudy.

11.1.6 Study Termination

Study termination will occur when the last subject has completed their last formal visit or last formal contact or an unscheduled study visit in case of early withdrawal from the study. This is expected to occur when the last subject completes their EOS visit or their end of ZA substudy visit.

11.1.7 Exploratory Endpoints

[redacted]

[redacted]

11.1.8 Analysis Sets

11.1.8.1 ZA Analysis Set

[redacted]

11.1.8.2 ZA Safety Analysis Set

[redacted]

11.1.9 Subgroups

[redacted]

11.1.10 Timepoint of Analysis

[redacted]

11.1.11 Handling of Missing and Incomplete Data

For radiologic assessments x-rays should be performed for all patients. Nevertheless, for fractures with a missing x-ray date, the fracture date captured on the fracture recording eCRF will be used. If the fracture date is completely missing, it will not be imputed. If it is partially missing, imputed dates will be used to derive the incidence at each time point.

	Missing	**Impute**	**Exception**
Fracture Date	Day	01	Default to Day 1 of ZA substudy if an event started the same year and month as Day 1
	Day /Month	01JAN	Default to Day 1 of ZA substudy if an event started the same year as Day 1

If a start or stop date for an AE is completely missing, it will not be imputed. If it is partially missing, imputed dates will be used to derive the duration of the AE. Missing years will not be estimated under any conditions.

	Missing	**Impute**	**Exception**
Start date	Day	01	Default to Day 1 of ZA substudy if an event started the same year and month as Day 1 of ZA substudy
	Day /Month	01JAN	Default to Day 1 of ZA substudy if an event started the same year as Day 1 of ZA substudy
Stop	Day	Last day of the	Default to the end of ZA substudy visit

date		month	if the imputed event stop date is after the end of ZA substudy visit
	Day/Month	31Dec	

11.1.12 Analysis Windows

11.1.12.1 Study Month

Study Month will be defined as:

CEIL(date of event – date of ZA substudy Day 1*))/30.44)

* If ZA Substudy Day 1 is missing, the randomization date will be used instead

11.1.12.2 Incidence Endpoints

Incidence between periods of time will use the formula in Section 11.1.12.1 to derive month from ZA substudy Day 1 to event between Months 1 and 6, between Months 1 and 12 and between Months 1 and 18.

11.1.12.3 Analysis at Specified Times

All tests and procedures will aim to be performed within the specified windows according to the definitions given in the tables below. To allow for variations in scheduling the following visit windows will be used to assign evaluations to a most appropriate nominal visit for analysis and summarization.

In order to get as many data points as possible for summarization, there will be no gaps between visit windows. If more than one assessment is available within a window, the assessment closest to the target day at each nominal visit (Month 6, Month 12, …) will be used. If ZA substudy Day 1 is missing, the randomization date will be used instead. BMD, CTX and Osteocalcin measurements recorded within 45 days after Day 1 are acceptable as baselines. In case of several candidates, the closest value to Day 1 will be selected.

Nominal Visit Window	Target Day	Definition (ZA substudy Day)
Baseline	1	Last evaluation prior to or on ZA substudy Day 1
Month 6	183	ZA substudy Day 2 to 274
Month 12	365	ZA substudy Day 275 to 456
Month 18	547	ZA substudy Day 457 to 634

11.1.13 Analysis Methods

[redacted]

12. AMENDMENTS TO THE STATISTICAL ANALYSIS PLAN

Amendments to the Statistical Analysis Plan must include a description of and a rationale for changes to this original plan.

Version 1 of the SAP was issued on 27 April 2010 describing in detail the analysis per the original clinical study protocol for ABCSG study 18 (Protocol Number: 20050209, Date: 13 March 2006).

Version 2 of the SAP was issued on 16 December 2014 updating the planned analysis per the amended clinical study protocol for ABCSG study 18 (Protocol Number: 20050209, Superseding 3 Amendment 3, Date: 25 January 2013).

Version 3 of the SAP was issued on 10 March 2015. The following specific updates were made:

- The vertebral fracture endpoints will be based on the Vertebral fracture analysis set (section 4.1.4) rather than the FAS. Only approximately 60% of the FAS will have available data for analysis due to the high number of non-evaluable or missing independent assessments of vertebral fractures by the Central Review Committee (CRC). An additional exploratory analysis will be conducted on the FAS using time to event analysis methodology.
- Treatment emergent AEs (section 5.5) will include AEs occurring on or after the first dose of IP reported through 30 days after the last dose of IP (EOT). AEs reported but not classified as treatment emergent will be listed.
- The proscribed therapies detailed in the per protocol population definition (section 4.1.5) were updated in-line with the clinical study protocol (Protocol Number: 20050209, Superseding 3 Amendment 3, Date: 25 January 2013),

Version 4 of the SAP was issued on 22 October 2015 updating a DFS analysis per recommendation of the DMC and decision of the Study Steering Committee and the Study Sponsor.

Following review of the main DFS analysis results, version 4.1 of the SAP was issued on 24 November 2015 adding a sensitivity analysis for the DFS analysis to account for treatment cross-over.

Version 5 of the SAP was issued on 15 January 2018 updating for OLP, including additional measurements and analyses, focusing on sensitivity analyses accounting for cross-over. Furthermore, an additional exploratory [redacted] analysis in 2018 was introduced as recommended by the Study Steering Committee.

Due to inclusion of the ZA substudy Version 6 of the SAP was issued on 10 December 2019. Details on the substudy, as well as clarifications regarding the final analysis of the main study and the analysis of the ZA substudy were included. Additionally, abbreviations were updated and made consistent throughout the document.

13. CHANGES FROM THE PROTOCOL PLANNED ANALYSES

This SAP includes the following changes to the statistical analyses planned to be conducted within the amended clinical study protocol for ABCSG study 18 (Protocol Number: 20050209, Superseding 3 Amendment 3, Date: 25 January 2013):

The fracture-related secondary objective and associated endpoint has been updated from:

- Objective: Bone mineral density (BMD) at total lumbar spine, total hip and femoral neck in a subgroup of subjects at pre-selected sites
- Endpoint: The percent change in total lumbar spine, total hip and femoral neck bone mineral density (BMD) from baseline to 36 months (at pre-selected sites)

Revised with:

- Objective: Bone mineral density (BMD) at total lumbar spine, total hip and femoral neck in a subgroup of subjects with evaluable DXA scans using the same Hologic device
- Endpoint: The percent change in total lumbar spine, total hip and femoral neck bone mineral density (BMD) from baseline to 36 months in a subgroup of subjects with evaluable DXA scans using the same Hologic device

The exploratory objective has been updated from:

- Objective: [redacted]

Revised with:

- Objective: [redacted]

Background:

Sites were pre-selected at initiation if they used a Hologic DXA device. The original BMD related endpoints were to be subset on this stratum in order to limit variability across and within subjects. During the blinded review of the data it has become apparent that during the course of the study sites have changed their DXA devices and subjects may switch between DXA machines during the duration of the study. Therefore the stratification factor does not reflect the actual DXA machine type used. Consequently, the analyses of this endpoint will be based on the machine type reported on the subject's eCRF. Subjects will be included in the main BMD analyses if the same Hologic DXA device was used (see Section 4.1.3, BMD analysis set).

An additional section regarding an analysis of DFS at 18 months after PADCD was included in Version 4 of the SAP (15 September 2015) based on a recommendation of the DMC and decision of the Study Steering Committee and the Study Sponsor.

Version 5 of this SAP includes the following changes to the statistical analyses planned to be conducted within the amended clinical study protocol for ABCSG study 18 (Protocol Number: 20050209, Superseding Amendment 4, Date: 22 February 2016):

Based on recommendation from the Study Steering Committee, due to concerns of a potential dilution effect caused by treatment crossover, an additional, exploratory [redacted] analysis will be conducted early 2018, including the main DFS analysis, as well as all above mentioned sensitivity analyses.

14. LIST OF PLANNED TABLES AND FIGURES

List of Tables:

1. Primary Endpoint for FAS - description:
    a) Number of analyzed patients and number of first clinical fractures by treatment groups in total and stratified by a) type of hospital, b) prior AI and c) total lumbar spine BMD score at baseline (Tscore <-1.0 or ≥-1.0)
2. Primary Endpoint for FAS - modelling:
    a) Result of Cox regression analysis with treatment groups as covariate stratified by randomization stratification factors, description with hazard ratio and 95 % confidence interval and estimated fracture rates at 36 months and 95 % confidence intervals for each treatment group
    b) Crude incidences at the end of study for each treatment group
    c) Results of Cox regression analyses including treatment and other covariates: age, prior AI usage, total lumbar spine BMD score at baseline (T-score <-1.0 or ≥-1.0) and potential other covariates; description with hazard ratios and 95 % confidence intervals. Univariate and multivariate analyses
3. Primary Endpoint sensitivity analyses - modelling:
    a) Result of Cox regression analysis with treatment groups as covariate stratified by randomization stratification factors using FAS, description with hazard ratio and 95 % confidence interval; subjects who withdraw from the trial due to a > 10 % loss of BMD in the lumbar spine over any 1-year period are treated as having had a clinical fracture at the time of withdrawal
    b) Result of Cox regression analysis with treatment groups as covariate stratified by randomization stratification factors using per protocol population, description with hazard ratio and 95 % confidence interval
    c) Result of Cox regression analysis with treatment groups as covariate and time from treatment discontinuation and first clinical fracture or date last seen as binary time dependent covariate stratified by randomization stratification factors using FAS, description with hazard ratios and 95 % confidence intervals
    d) Result of RPSFTM using Cox regression analysis with treatment groups as covariate stratified by randomization stratification factors using FAS, description with hazard ratios (95% CI) and acceleration factor (95% CI) (final analysis)
4. Secondary Endpoint for FAS - description:
    a) Observed percent change and corresponding 95 % confidence interval in lumbar spine, total hip and femoral neck BMD from baseline to 36 months for each treatment group (BMD analysis set)
    b) Number of patients, new vertebral fractures and number of new or worsening vertebral fractures at 36 months for each treatment group (Vertebral fracture analysis set)

c) Crude incidences of new vertebral fractures at 36 months with 95 % confidence intervals for each treatment group and their difference with corresponding 95 % confidence interval (Vertebral fracture analysis set)

d) Crude incidences of new or worsening vertebral fractures at 36 months with 95 % confidence intervals for each treatment group and their difference with corresponding 95 % confidence interval (Vertebral fracture analysis set)

5. Secondary Endpoint for FAS (unless otherwise specified) - modelling:

a) Results of logistic regressions modelling the incidence of new vertebral fractures (odds-ratios and 95 % confidence intervals) with treatment group as covariate stratified by randomization stratification factors at 36 months (Vertebral fracture analysis set)

b) Results of logistic regressions modelling the incidence of new vertebral fractures with treatment group as covariate and other covariates found significant (odds-ratios and 95 % confidence intervals) at 36 months (Vertebral fracture analysis set)

c) Results of logistic regressions modelling the incidence of new or worsening vertebral fractures (odds-ratios and 95 % confidence intervals) with treatment group as covariate stratified by randomization stratification factors at 36 months (Vertebral fracture analysis set)

d) Results of logistic regressions modelling the incidence of new or worsening vertebral fractures with treatment group as covariate and other covariates found significant (odds-ratios and 95 % confidence intervals) at 36 months, respectively (Vertebral fracture analysis set)

e) Result of Cox regression analysis after around 18 months after PADCD for DFS; the model includes treatment as covariate and is stratified by the randomization stratification factors, description of the model by with hazard ratios and 95 % confidence intervals (IDMC recommended analysis)

f) Result of Cox regression analysis for DFS analysis 2018; the model includes treatment as covariate and is stratified by the randomization stratification factors, description of the model by with hazard ratios and 95 % confidence intervals (Steering Committee recommended analysis)

g) Result of RPSFTM using Cox regression analysis for DFS analysis 2018; the model includes treatment groups as covariate stratified by randomization stratification factors using FAS, description with hazard ratios (95% CI) and acceleration factor (95% CI) (Steering Committee recommended analysis)

h) Result of Cox regression analysis after long-term-follow-up for a) DFS, b) BMFS and c) OS; the model includes treatment as covariate and is stratified by the randomization stratification factors, description of the model by with hazard ratios and 95 % confidence intervals (final analysis)

i) Result of RPSFTM using Cox regression analysis for a) DFS, b) BMFS and c) OS; the model includes treatment groups as covariate stratified by randomization stratification factors using FAS, description with hazard ratios (95% CI) and acceleration factor (95% CI) (final analysis)

6. Adverse Events for SAF:

a) Subject incidence rates of treatment emergent adverse events reported during treatment period are tabulated by system organ class and preferred term and split by treatment and age categories and by treatment and race (White/Caucasian and others)

b) Investigational product related adverse events split by treatment and age categories and by treatment and race

c) Serious adverse events split by treatment and age categories and by treatment and race

d) Common Terminology Criteria for Adverse Events (CTCAE) grade 3, 4, or 5 adverse events split by treatment and age categories and by treatment and race

e) Adverse events leading to investigational product discontinuation split by treatment and age categories and by treatment and race

f) Adverse events leading to study withdrawal split by treatment and age categories and by treatment and race

g) Investigational product related serious adverse events split by treatment and age categories and by treatment and race

h) Fatal adverse events split by treatment and age categories and by treatment and race

i) Narratives for death and serious adverse events

7. Clinical Laboratory Assessments for SAF:

a) Summary of recorded values by minimum, maximum, median, mean and standard deviation for each nominal visit window (semi-annually)

b) Summary of change from baseline by minimum, maximum, median, mean and standard deviation for each nominal visit window (semi-annually)

c) Tables of shifts of toxicity grades (CTCAE: <3, 3, 4 , 5) for baseline and “worst” on-study values. For a selected subset both minimum and maximum are reported

8. Anti-denosumab Antibodies: incidence and percentages of subjects who develop binding and neutralizing anti denosumab antibodies split by treatment group and visit (SAF)

9. Exploratory Analyses for FAS (unless otherwise specified) – description:

11. Subgroup Analysis for FAS - modelling:

    a) Result of Cox regression analysis with treatment groups as covariate separately for each sub-group combination of a) major academic site, b) Prior AI usage and c) Total lumbar spine BMD score at baseline (Tscore <-1.0 or ≥-1.0); description of model results with hazard ratios and 95 % confidence intervals

    b) Results of three Cox regression analyses with treatment and a subgroup variable (a) major academic site, b) Prior AI usage and c) total lumbar spine BMD score at baseline (Tscore <-1.0 or ≥-1.0)) as covariates; interaction between treatment and corresponding subgroup variable is also included; description of model results with hazard ratios and 95 % confidence intervals

12. Further tabular summaries and listings for SAF unless otherwise specified

    a) Demographic data of patients (age and race), stratification criteria (type of hospital, prior AI usage and total lumbar spine BMD score at baseline (Tscore <-1.0 or ≥-1.0)) and baseline and breast cancer history (T-stage, pN-stage, Grading,

primary tumor, ER and PgR expression) split by randomized treatment groups (FAS)

b) Withdrawal information (subject disposition status, reason for withdrawal and mean time under observation per patient from randomization to lost-to-follow up date or end-of-study visit) summarized by randomized treatment groups

c) Exposure / follow-up information (number of doses) summarized by randomized treatment group

d) Any data collected relating to important protocol violations

e) Accesses to individual subject treatment assignments (unblinding)

f) Listing of Manufacturing Lot Numbers

g) Listing of Unique Manufacturing Lot Numbers

h) Listing of Laboratory Reference Ranges

i) Listing adverse events considered not treatment emergent

List of Tables for ZA substudy:

1. Exploratory Endpoints for ZA analysis set - description:

[redacted]

2. Exploratory Endpoints for ZA analysis set- modelling:

[redacted]

[redacted]

List of Figures:

1. Kaplan-Meier estimates for each treatment group for primary outcome
2. Kaplan-Meier estimates for each treatment group where censoring is the interesting outcome
3. Kaplan-Meier estimates for each treatment group where treatment discontinuation is the interesting outcome
4. Scatter plots to identify outliers in any of the continuous variables – including sensitivity analyses
5. Description of interactions

15. REFERENCES

Cox D.R. (1972) Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B 34, 187-220.

Eastell R. and Adams J.E. (2002) Effect of anastrozole on bone density and bone turnover: Results of the “arimidex” (anastrozole), tamoxifen, alone or in combination (ATAC) trial performed by the ATAC Trialists’ Group. Ann Oncol. 13, 32 (suppl5).

Gail M. and Simon R. (1985) Testing for qualitative interactions between treatment effects and patients subsets. Biometrics 41 (2): 361-372

Gnant M, Pfeiler G, Dubsky PC, et al. Adjuvant denosumab in breast cancer(ABCSG-18): a multicentre, randomised, double-blind, placebo-controlled trial. Lancet. 2015;386(9992):433-43. Epub 2015 May 31.

Jakesz R, Jonat W, Gnant M, et al. Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial. Lancet. 2005;366-455

Kaplan E.L. and Meier,P. (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.

Kenward, M.G. and Roger, J.H. (1997), Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, Biometrics, 53, 983 - 997.

Lehmann T, Aeberli D. Possible protective effect of switching from denosumab to zoledronic acid on vertebral fractures. Osteoporos Int. 2017;28(10):3067-3068.

Miller PD, Bolognese MS, Lewiecki EM, et al. Effect of denosumab on bone density and turnover in postmenopausal women with low bone mass after long-term continued, discontinued, and restarting of therapy: a randomized blinding phase 2 clinical trial. Bone. 2008;43(2):222-229.

Robins, J.M. and Tsiatis, A.A. (1991) Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods. 1991;20:2609-2631.

Westfall, PH, Tobias, R.D., Rom, D. Wolfinger, R.D., Hochberg, Y. (1999) Multiple comparisons and multiple tests using the SAS system, SAS Institute Inc.