SAP CanStem111P

A Phase III Study of BBI-608 plus nab-Paclitaxel with Gemcitabine in Adult Patients with Metastatic Pancreatic Adenocarcinoma

Statistical Analysis Plan (SAP)

Version: 5.0

Author: [redacted]

Date: 16-October-2020

Revision History

Version	Date	Author(s)	Summary of Changes/Comments
Version 1.0	August 6, 2017	[redacted]	Statistical Analysis Plan
Version 2.0	August 10, 2018	[redacted]	• Efficacy claim (no stopping for efficacy) in the first interim analysis has been removed. • Additional covariates have been added. • Laboratory variables have been edited. • Additional subgroup analysis have been specified. • Summary of key efficacy analysis has been added.
Version 3.0	April 12, 2019	[redacted]	• Biomarker endpoints have been moved to exploratory • Baseline value for efficacy has been edited • PFS, DCR, and ORR have been marked as key secondary endpoints and quality of life as other secondary endpoint • Multiplicity adjustment to control for Type I error rate has been incorporated • Quality of Life detailed analysis has been added • More details to the analyses of the primary and secondary endpoints has been added

			• Summarization of new anticancer therapy, physical examination, and ECOG have been added • ECG analysis has been edited • Adverse Events of Clinical Relevance (AECR) have been added • Elevated liver function test results have been edited • Dose intensity section has been edited • By cycle and cycle 1 summaries for treatment compliance have been removed • Sensitivity analysis for PFS has been added • Censoring hierarchy for PFS has been edited
Version 4.0	October 24, 2019	[redacted]	• Changing BBI-608 to napabucasin • Adding language after futility boundary was met in first interim analysis • Per Protocol Analysis Set has been deleted • Coutry subgroup analysis has been deleted • Methods for handling missing data has been edited

			• Baseline assessment window has been edited from 14 days to 28 days • Censoring rules for time to event analysis of New Anti-cancer therapy has been added • Dose Intensity section has been edited • Appendix 1.4 has been added
Version 4.1	April 24, 2020	[redacted]	• Compliance section has been edited
Version 4.2	June 25, 2020	[redacted]	• Primary and Secondary Analyses has been edited
Version 4.3	August 18, 2020	[redacted]	• BBI has been changed to SDPO • Primary and Secondary Analyses has been edited • Added pSTAT3 unknown to potential biomarker analysis • AECR has been changed to Clustered TEAE and associated SMQs and terms have been changed
Version 5.0	October 16, 2020	[redacted]	• Dose Intensity section has been edited • PD/Biomarker Analysis Set has been deleted • Dose Changes section has been edited

STATISTICAL ANALYSIS PLAN APPROVAL

[redacted]

TABLE OF CONTENTS

ABBREVIATIONS & DEFINITIONS

ADR	Adverse Drug Reaction
AE	Adverse Event
ALT	Alanine Aminotransferase
ANC	Absolute Neutrophil Count
AST	Aspartate Aminotransferase
BMI	Body Mass Index
BOR	Best Overall Response
BP	Blood pressure
CI	Confidence Interval
CMH	Cochran-Mantel_Haenszel
CR	Complete Response
CRF	Case Report Form
CRO	Contract Research Organization
DCR	Disease Control Rate
DSMB	Data Safety and Monitoring Board
EDC	Electronic Data Capture
EORTC-QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire
IA	Interim Analysis
ITT	Intent to Treat
M&N	Miettinen and Nurminen
mOS	Median Overall Survival
mPFS	Median Progression-Free Survival
MTD	Maximum Tolerated Dose
NCI CTCAE	National Cancer Institute Common Toxicity Criteria for Adverse Events
NOAEL	No Observed Adverse Effect Level
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PDAC	Pancreatic ductal adenocarcinoma
PFS	Progression-Free Survival
PH	Proportional Hazards
PK	Pharmacokinetics
PR	Partial Response
pSTAT3	Phospho-STAT3
QoL	Quality of Life
RAS	Response Analysis Set
RECIST	Response Evaluation Criteria In Solid Tumors
RP2D	Recommended Phase 2 Dose

SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
ULN	Upper Limit of Normal

1. AMENDMENTS FROM PREVIOUS VERSION(S)

This is the version 5.0 of the statistical analysis plan, based on the CanStem111P protocol amendment dated July 31, 2019. The major changes in this amendment include changing Boston Biomedical Inc to Sumitomo Dainippon Pharma Oncology, Inc, changing BBI-608 to napabucasin, adding language after futility boundary was met during first interim analysis, deleting the Per Protocol Analysis Set and the PD/Biomarker Analysis Set, editing the methods for handling missing data section, deleting country subgroup analysis, editing the baseline assessment window from 14 days to 28 days, adding censoring rules for time to event analysis of New Anti-cancer therapy, editing the dose intensity section, editing the Dose Changes section, editing the Adverse Events of Clinical Relevance section to Clustered TEAEs and the SMQs and sponsor derived terms associated with the analysis, adding pSTAT3 unknown to the biomarker analysis, editing the Primary and Secondary analyses from stratified for subgroup analyses to unstratified, removing PFS sensitivity analyses using alternative censoring hierarchy, editing DCR and ORR unstratified analyses using normal approximation and the z test, and adding Appendix 1.4: Response and Evaluation Endpoints.

As noted in Section 13 in the protocol amendment, the detailed methodology for summary and statistical analyses of the data collected in this study are documented in the statistical analysis plan (SAP), which is maintained by the sponsor. There may be modifications made to the plans outlined in the protocol but these will not include any major modifications of the primary endpoint definition and/or its analysis. Any major modifications will be reflected in a protocol amendment.

2. INTRODUCTION

2.1. Study Design

This is a randomized, open-label, multi-center, phase III study of napabucasin plus weekly nab-paclitaxel with gemcitabine (Arm 1) vs. weekly nab-paclitaxel with gemcitabine (Arm 2) for adult patients with metastatic pancreatic ductal adenoacarcinoma (PDAC).

1132 patients will be randomized in a 1:1 ratio, stratified according to geographical region (North America/Western Europe/Australia vs. Japan/Korea vs. Rest of the World), Eastern Cooperative Oncology Group (ECOG) performance status (0 vs. 1), and presence of liver metastases (yes vs. no).

Metastatic PDAC	RANDOMIZE	↗	Arm 1	napabucasin orally, twice daily *plus* Nab-Paclitaxel + Gemcitabine IV, weekly		Disease progression based on RECIST criteria[1, 2] *or* Unacceptable toxicity occurs		OS
					→		→	

		↘	Arm 2	Nab-Paclitaxel + Gemcitabine IV, weekly				

[1]If no other standard therapies are available at the time of disease progression, and the patient has not experienced any adverse events requiring permanent discontinuation, napabucasin may be continued in monotherapy.
[2]From the time of this amendment, and since the outcome of the interim analysis was communicated to investigators, patients may continue protocol therapy if it is believed to be in their best interest by the investigator and patient, and with the patient's informed consent. Patients will receive napabucasin, nab-paclitaxel and/or gemcitabine at the same dose and schedule that they were receiving prior to the amendment. Patients on Arm 1 may continue napabucasin with the Sponsor's approval. Patients on Arm 1 may discontinue napabucasin but choose to continue with nab-paclitaxel and gemcitabine.

Until the time of this amendment, the study proceeded in 28-day (4-week) cycles. Napabucasin was administered orally, at 240 mg bid (480 mg total daily dose), with doses separated by approximately 8-12 hours. Napabucasin administration began 2-5 days prior to the first nab-paclitaxel with gemcitabine infusion. Nab-paclitaxel 125 mg/m^2 immediately followed by gemcitabine 1000 mg/m^2 was administered on Days 1, 8 and 15 of every 28-day cycle via intravenous infusion. Tumor assessments will be performed every 8 weeks after randomization until objective disease progression or treatment discontinuation due to toxicity.

The study was planned for completion on February 28th, 2020.

2.2. Study Objectives

2.2.1. Primary Objectives

To compare overall survival (OS) of patients with metastatic (Stage IV) PDAC treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine.

2.2.2. Key Secondary Objectives

- To compare PFS in patients treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine.
- To compare DCR in patients treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine.
- To compare ORR in patients treated with napabucasin plus weekly nab-

paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine.

2.2.3. **Other Secondary Objectives**

- To evaluate the safety profile of napabucasin administered daily plus weekly nab-paclitaxel with gemcitabine, with safety assessed according to the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE) version 4.0.
- To compare the Quality of Life (QoL), as measured using the EORTC-QLQ-C30, in patients with treatment-naive metastatic PDAC treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine.

2.2.4. **Exploratory Objectives**

- To compare OS in patients treated with napabucasin plus weekly nab-paclitaxel with gemcitabine (Arm 1) versus weekly nab-paclitaxel with gemcitabine (Arm 2) in biomarker positive PDAC patients. ¥
- To compare PFS in patients treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine in biomarker positive PDAC patients.
- To compare ORR and DCR in patients treated with napabucasin plus weekly nab-paclitaxel with gemcitabine versus weekly nab-paclitaxel with gemcitabine in biomarker positive PDAC patients.

¥This biomarker-positive sub-population is defined as those patients with phospho-STAT3 positivity on immunohistochemical (IHC) staining of Formalin Fixed Paraffin Embedded (FFPE) tumor tissue.

3. ENDPOINTS AND COVARIATES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

Overall survival (OS) of patients with metastatic (Stage IV) PDAC treated with napabucasin plus weekly nab-paclitaxel with gemcitabine and overall survival of patients with weekly nab-paclitaxel with gemcitabine in the general study population.

3.2. Key Secondary Endpoints

- PFS, DCR and ORR in the general study population
- Objective tumor response, as assessed using the Response Evaluation Criteria in Solid Tumor (RECIST) version 1.1

3.3. Other Secondary Endpoints

Mean European Organization for Research and Treatment of Cancer Quality of Life questionnaire (EORTC QLQ-C30 QoL) change scores from baseline at time 2 (~8 weeks post randomization) and time 4 (~16 weeks post randomization) for the physical function and global health status/quality of life subscale scores.

3.4. Exploratory Endpoints

- OS, PFS, ORR and DCR in the predefined biomarker-positive population

3.5. Safety Endpoints

3.5.1. Adverse Events and Laboratory Abnormalities

- Adverse Events (AEs) as characterized by type, frequency, severity (as graded by National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE] version 4.0) timing, seriousness, and relationship to study therapy

- Laboratory abnormalities as characterized by type, change from baseline, and severity (as graded by NCI CTCAE version 4.0)

3.6. Covariates

Stratification factors used for randomization including geographical region (North America/Western Europe/Australia vs. Japan/Korea vs. Rest of the world), ECOG performance status (ECOG 0 vs. ECOG 1), and presence of liver metastases (yes vs. no) will be used as covariates in primary analyses for OS, PFS, ORR and DCR.

Besides the stratification factors, the following factors at patient entry may also be considered as covariates in sensitivity analyses:

- Primary tumor location (head of pancreas *versus* body/tail of pancreas)
- Level of CA19-9 (normal *versus* >ULN and <59 x ULN *versus* $\geq$59 x ULN)
- Age (< 65 *versus* $\geq$ 65)
- Sex (male *versus* female)
- Number of metastatic sites at baseline (1 *versus* $\geq$ 2)
- Smoking status (Yes *versus* No)
- Race (white, black, Asian, other)
- Prior adjuvant chemotherapy (Yes versus No)
- Prior radiotherapy (Yes versus No)

Subgroup analyses and the corresponding forest plot analysis by these covariates may also be conducted.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The null and alternative hypothesis for the primary endpoint (OS) is:

H_0: OS in napabucasin + nab-paclitaxel with gemcitabine ≤ OS in nab-paclitaxel with gemcitabine

H_a: OS in napabucasin + nab-paclitaxel with gemcitabine > OS in nab-paclitaxel with gemcitabine

4.2. Statistical Power and Sample Size Calculation

The primary study endpoint of the study is OS in the general study population.

The study is designed to have a power of 90% and a one-sided alpha of 2.5% to detect a 20% reduction in the continuous risk of death (HR 0.80, which corresponds to an increase of median survival from 8.5 to 10.63 months) in the Intent to Treat (ITT) general study population. It is estimated that 864 events will be required to detect a 20% reduction in the risk of death which would be observed by randomizing 1132 patients over 24 months with patient follow up for an additional 12 months, for total study duration of 36 months. It is anticipated that up to 5% dropout rate will occur. When the required number of events for the primary endpoint has been reached, all randomized patients still alive will continue study follow up through to their deaths.

5. INTERIM ANALYSES, FINAL ANALYSIS, AND UNBLINDING

5.1. INTERIM ANALYSES AND FINAL ANALYSIS

Prior to this amendment, there was one interim analysis after 50% of the required number of events (432 events) had been observed.

The interim analysis was for futility only, with the futility boundary set at HR ≥ 1. Safety data and primary efficacy analysis based on approximately 432 events were presented to the DSMB on June 24th, 2019. The DSMB reviewed the results on July 1st, 2019. Based on the recommendation of the DSMB, the Sponsor informed investigators on July 2nd, 2019 that the study would be discontinued due to futility. For patients enrolled to Arm 1, the Sponsor advised that patients stop treatment with napabucasin but given that no safety concerns were identified by the DSMB, in cases in which a patient appears to be deriving benefit from napabucasin in the opinion of the investigator and approved by the Sponsor and with the patient's informed consent, continuation of protocol therapy was permitted. Patients enrolled to Arm 2 were given the option to continue receiving standard of care therapy within the study.

Patients continued to receive protocol therapy until a discontinuation criterion was met or until February 28th, 2020, whichever occurred first.

As the Interim Analysis data concluded the study, the IA report will be included as an appendix to the CSR. The final analysis including data after July 2nd, 2019 will be exploratory in nature. This is due to a potentially imbalanced dropout rate of the two treatment arms as the study was deemed futile. Therefore, caution is necessary when interpreting the results.

At study completion, analysis will be conducted for primary endpoint of OS as outlined in Section 3. Additionally, all secondary and exploratory analyses will be performed as outlined.

5.2. UNBLINDING AND DATA SAFETY AND MONITORING COMMITTEE

Although this is an open-label study, a blinding plan will be implemented to minimize the risks of potential Sponsor bias for decision making at the interim analyses and the final analysis.

Specifically, there will be an unblinded reporting team to report complete interim analysis results and regular safety updates to an independent Data Safety and Monitoring Board (DSMB). DSMB will review the IA results and make recommendations to the Sponsor based on the decision rules described in Section 5.1. DSMB will also review safety data regularly during conduct of the study. The role and responsibilities of the DSMB will be detailed in a separate Charter.

Aggregated data by treatment will not be reviewed or analyzed by the study team until the final database release or early termination of the study. SDPO will restrict access to the database to only selected team members (or their designees) who have a need to review individual unblinded patient information on an ongoing basis for safety and pharmacovigilance monitoring.

Adequate procedures will also be implemented so that team members (or their designees) who have unblinded information for patients in the study do not disclose and disseminate any unblinded information or unblinded data to any blinded personnel until after the official database release. The roles and responsibilities of the blinded and unblinded staff will be detailed in a separate plan.

6. ANALYSIS SETS

6.1. Intent-to-Treat Analysis Set

The Intent-to-Treat (ITT) population is defined as all patients who are randomized. Patients will be analyzed according to the treatment arm they were randomized to, regardless of any potential errors in dosing. The ITT population will be used for primary and secondary efficacy endpoint analyses.

6.2. Safety Analysis Set

The safety analysis set is defined as all patients who receive at least 1 dose of any study drug. Patients will be analyzed according to the treatment they actually received. That is, those patients who are randomized to the active arm but receive the regimen in the control

arm will be included in the control arm; those patients who are randomized to the control arm but receive the regimen in the active arm will be included in the active arm for safety analyses.

Safety analysis set will be used for all safety related analyses including treatment emergent adverse events (TEAEs), concomitant medication, laboratory tests, and vital signs.

6.3. Response Analysis Set (RAS)

Response analysis set (RAS) is defined as all ITT patients with measurable disease per RECIST 1.1 at randomization. RAS will be used for the primary analysis of response data (ORR and DCR). Response data may also be summarized in the ITT population.

6.4. PK Analysis Set

The PK analysis set will include all patients who received at least one dose of study drug (napabucasin) and have at least one quantifiable concentration.

6.5. Treatment Misallocations

If a patient was:

- Randomized but not treated, then they will be excluded from analyses of safety and PK as they have not received at least one dose of study medication. These patients will be included in ITT analyses, however.

- Treated but not randomized, then they will be excluded from the efficacy analyses since randomized treatment is missing, but will be reported under the treatment they actually received for all safety and PK analyses.

- Randomized but took incorrect treatment, then they will be reported in the ITT analysis under the treatment they were randomized to, and under the treatment they actually received for the PK and all safety analyses.

7. DATA HANDLING

7.1. Methods for Handling Missing Dates

All analyses and descriptive summaries will be based on the observed data. Except for the data otherwise specified in the language below, missing data will not be imputed or "carried forward". For the patient data listings, no imputation of incomplete dates will be applied. The listings will present the incomplete dates without any change.

Missing or Partial Death Dates

- If the entire date is missing, the death date will be imputed as the day after the date last known alive.

- If the day or both day and month is missing, the death date will be imputed to the maximum of the full (non-imputed) day after the date last known alive and the following:
    - If day is missing, day will be 1st of the month
    - If both day and month are missing, death month and day will be January 1st.

Date of Last Dose of All Study Drugs

No imputation will be done for first dose date. Date of last dose of study drug, if unknown or partially unknown, will be imputed as follows:

- If the last date of study drug is completely missing and the End of Treatment (EOT) eCRF page has not been completed and no death date has been entered, the patient should be considered ongoing and the cutoff date for the analysis should be used as the last dosing date.
- If the last date of study drug is completely or partially missing and there is EITHER a completed EOT eCRF page OR a death date available (within the data cutoff date), then impute this date as the last dose date:

  = 31DECYYYY, if only Year is available and Year < Year of min (EOT date, death date)

  = Last day of the month, if both Year and Month are available and Year = Year of min (EOT date, death date) and Month < Month of min (EOT date, death date)

  = min (EOT date, death date), for all other cases.

- If patients were never dosed, the date of last dose will be marked as Not Applicable and there will be a date of discontinuation on the EOT eCRF page and/or a death date on the Death eCRF

Missing Dates in Adverse Events/Concomitant Therapies/Medications

Dates missing the day or both the day and month of the year will adhere to the following conventions:

- The missing day of onset of an AE or start date of a therapy will be set to:
    - the day of the first study treatment, if the onset yyyy-mm is the same as yyyy-mm of the first study treatment
    - otherwise, the first day of the month that the event occurred
- The missing day of resolution of an AE or end date of a therapy will be set to:
    - the last day of the month of the occurrence. If the subject died in the same month, then set the imputed date as the death date
- If the onset date of an AE or start date of a therapy is missing both the day and month, the onset date will be set to:

- the date of the first treatment, if the onset year is the same as the year of the first study treatment
- otherwise, January 1 of the year of onset

- If the resolution date of an AE or end date of a therapy is missing both the day and month, the date will be set to:
  - December 31 of the year of occurrence. If the subject died in the same year, then set the imputed date as the death date

- If date is completely missing, then no imputation will be done and the event will be considered as treatment emergent (for AEs) or concomitant (for medications) unless the end date rules out the possibility.

Missing Dates in Prior Therapies

Dates missing the day or both the day and month of the year will adhere to the following conventions:

- The missing day of start or end date of prior therapy will be set to:
  - the 15th of the month or date of informed consent, whichever is earlier

- If the start or end date of a prior therapy is missing both the day and month, the date will be set to:

  - July 1 of the year or date of informed consent, whichever is earlier

- If date is completely missing, then no imputation will be done

Missing Efficacy Endpoints

For primary and secondary efficacy analyses no values will be imputed for missing data. For time to event endpoints, non-event observations will be censored. For ORR/DCR, patients with no post-baseline tumor evaluations or missing baseline tumor evaluation will be counted as non-responders.

Missing Age

Age will be computed from birth date to the date of Informed Consent, as (Date of Informed Consent – Date of Birth +1)/365.25 rounded down to the nearest integer.

- If birth date is missing day, day will be set to the 15th to minimize bias
- If missing both month and day, birth date will be set to July 1 of the birth year

Missing QoL Data

Missing items in a scale in EORTC QLQ-30 will be handled by the following methods: Values will be imputed for missing items by “assuming that the missing items have

values equal to the average of those items which are present" for any scale in which at least half the items are completed. A scale in which less than half of the items are completed will be treated as missing. No imputation will be conducted at a specific visit if data is completely missing.

7.2. Definition of Baseline Values

- Efficacy: The last non-missing measurement on or prior to the date of randomization will serve as the baseline measurement, so long as this assessment was taken within 28 days of randomization. Any imaging assessments performed outside the time window will be considered protocol deviations and will be evaluated on a case by case basis.
- Safety: The last non-missing assessment prior to the date of the first study drug administration (either napabucasin or nab-paclitaxel or gemcitabine, whichever was administered first) will be used as the baseline assessment. In the case this value is missing, the last non-missing assessment prior to or on the randomization date will be used as baseline assessment.
- Other baseline characteristics: The last non-missing measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such a value is missing, the last non-missing assessment completed prior to the date of the first dose of study drug administration (either napabucasin or nab-paclitaxel or gemcitabine, whichever was administered first) will be used as the baseline assessment.

7.3. Study Day

Study day for safety analysis is calculated as:

- Assessment date – first dose date + 1; if the assessment was performed on or after the first dose date.
- Assessment date – first dose date; if the assessment was performed prior to the first dose date.

Study day for efficacy analysis is calculated as:

- Assessment date – randomization date + 1; if the assessment was performed on or after the randomization date.
- Assessment date – randomization date; if the assessment was performed prior to the randomization date.

7.4. Visit Windows

All data will be categorized based on the scheduled visit at which it was collected. These visit designators are predefined values that appear as part of the visit tab in the eCRF.

7.5. Dropouts

Time to event parameters will be censored if patients drop out (withdraw from consent or lost to follow-up) before documentation of the events (progressive disease / death). Rules for censoring for PFS are detailed in Appendix 1.2 Censoring for Time to Event Data.

7.6. Pharmacokinetics

The handling of BLQ values for nonlinear mixed-effects modeling (AKA: Population-PK, or Pop-PK) will be described in a separate analysis plan. For the purpose of data tabulation, summary statistics, and by-time point graphics, BLQ values will be treated as zero.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

Whilst every effort has been made to pre-specify all analyses in this statistical analysis plan, if any additional exploratory analyses are found to be necessary, the analyses and the reasons for them will be detailed in the clinical study report (CSR).

8.1.1. Analysis for Time to Event Data

Time-to-event curves between the two treatment groups will be compared with a one-sided stratified log-rank test at significance level of 0.025 (one-sided) adjusting for the stratification variables at randomization. Stratified and unstratified Cox proportional hazards model will be used to estimate the hazard ratio as well as 2-sided 95% confidence intervals (CI), and may be used to explore the potential influences of the baseline factors (eg, age, gender and ethnic origin, etc) on the time-to-event data. The time-to-event data will also be summarized using the Kaplan-Meier method and displayed graphically when appropriate (2). Median event times and 2-sided 95% CIs for each time-to-event endpoint will be provided. The 6 month, 1 year, and 18 month survival probability will be estimated using the Kaplan-Meier method and a 2-sided 95% CI will be calculated using the normal approximation to the log transformed cumulative hazard rate.

8.1.2. Analysis for Binary Data

Descriptive statistics (frequency and percentage) for binary endpoints will be presented by treatment group. CMH test will be used for comparison of the two treatment groups, adjusting for the randomization stratification variables. Two-sided Z test will be used for comparison of the two treatment groups without stratification. In addition, the treatment difference and 95% CI based on the Miettinen and Nurminen method (5) adjusting for the stratification factors will be displayed. For unstratified analyses, the treatment difference and 95% CI will be based on the normal approximation method, and may be used to explore the potential influences of the baseline factors on the binary endpoints.

8.1.3. Analysis for Continuous Data

Descriptive statistics, such as the mean, standard deviation, median, minimum, and maximum values, will be provided for continuous endpoints. Linear or non-linear models may be employed to analyze the continuous data.

8.1.4. Analysis for Quality of Life Data

Summary statistics for Quality of Life will be provided by treatment for Functional, Symptom and Global health status raw scores as well as for linearly converted scores on the scale of 0-100. The mean EORTC QLQ-C30 QoL changes scores from baseline at time 2 (~8 weeks post randomization) and time 4 (~16 weeks post randomization) for the physical function and global health status/qualify of life subscale scores will also be presented. Wilcoxon tests will be used to compare the difference at each of these two time points between the two treatment arms for each of these two subscales. In addition, a chi-square test will be used to compare the distribution of patients who have improved, are stable, or have worsened from baseline between the two arms. A Fisher's exact test will be used to compare between the two arms the proportion of patients with deterioration in physical function and global QoL at time 2 and time 4.

8.2. Statistical Analyses

8.2.1. Standard Analysis

Study Conduct and Patient Disposition

Patient disposition includes the number and percentage of patients for the following categories: patients in each of the study populations, patients discontinued from the treatment, primary reason to discontinue from the treatment, patients discontinued from the study, and primary reason to discontinue from the study. All percentages will be based on the number of patients in the ITT population.

Demographic and Baseline Characteristics

Baseline characteristics such as demographics, baseline disease characteristics, prior therapies, and ECOG performance score will be tabulated and listed.

Prior and Concomitant Medications

Prior and concomitant medications will be coded to ATC (Anatomical Therapeutic Chemical) classification and Drug Class using WHO Drug Dictionary (WHO-DD) March 2016, and will be listed.

Medications that start and stop prior to the date of first treatment administration will be classified as 'prior' medications. If a medication starts on or after the date of first treatment administration up to the last dose date of study medication (inclusive), then the medication will be classified as 'concomitant'.

Summaries of prior and concomitant medications will be provided by level 3 ATC classification and preferred term using frequencies and percentages for the ITT population. A separate concomitant medications table using preferred base and level 4 ATC classification will also be provided using frequencies and percentages for the safety analysis set.

Prior Treatment

Prior treatment will be summarized by overall and listed for each patient in the ITT population.

Prior cancer surgery will be coded by System Organ Class (SOC) and Preferred Term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA® v.19.0) and summarized by MedDRA PT within MedDRA SOC. The number and percentage of patients in each type of prior cancer surgery will also be reported by overall.

The total dose of prior cancer radiotherapy will be summarized. The number and percentage of patients with radiation to different anatomical locations will be reported, along with indication.

Prior cancer therapies will be summarized by the number of unique agents per patients, the number and percentage of patients reporting each agent, time from start of first anticancer therapy to first dose date of napabucasin, and time from end of last anticancer therapy to first dose date of napabucasin. Prior cancer therapies will be coded to Agent Received using WHO Drug Dictionary (WHO-DD) March 2016.

Concomitant Palliative and Supportive Care Therapies

Concomitant palliative and supportive care therapies will be summarized by overall and listed for each patient in the ITT population.

Concomitant palliative and supportive care therapies will be coded using WHO Drug Dictionary (WHO-DD) March 2016.

New Anti-cancer Treatment

New anti-cancer therapy will be derived from Post protocol cancer therapy.

New anti-cancer therapy will be summarized by type of therapy and median time to post protocol systemic therapy from randomization. Median time will be calculated using the Kaplan Meier Method and CIs estimated using the Brookmeyer Crowley method.
Patients who have not received a new anti-cancer therapy at the time of interim or final analysis will be censored at the date of data cutoff.

Medical History

Medical history will extracted from both medical history and AE CRFs.

Adverse Events listed on the AE CRF, with the “Is this an ongoing medical history condition present at date of informed consent?” marked as “Yes” will be considered

medical history, as these events were ongoing at the time of informed consent and prior to randomization.

Medical history will be coded by SOC and PT using MedDRA v19.0. Medical history will be summarized by SOC and PT using the number and percentage of patients for the ITT population.

Dose Changes

The details of napabucasin dose change are outlined below.

- Dose Held Due to Non-AE - 2 or more consecutive doses held (AM-PM or PM-AM) due to patient forgot, patient decision or other reason.
- Dose Held Due to AE - 2 or more consecutive doses held (AM-PM or PM-AM) due to adverse event.
    - Restarted at the same dose - Restarted at a dose the same as the previous non-zero dose at the time of dose held.
    - Restarted as reduced dose (dose-decrease) - Restarted at a dose decreased from the previous non-zero dose at the time of dose held.
        - Dose increase to partial dose following reduction - An increase to <240 mg for 2 or more consecutive doses (AM-PM or PM-AM) following restart at reduced dose.
        - Dose increase to full dose following reduction - An increase to 240 mg for 2 or more consecutive doses (AM-PM or PM-AM) following restart at reduced dose.
- Permanently Discontinued Due to AE.

Extent of Exposure Exposure may be summarized (per run-in and/or first 28 days and/or overall) as dose received (cumulative dose or actual dose intensity) or as dose received relative to intended dose (relative dose intensity [RDI]).

The information that will be summarized depends on how the study drug is dosed (e.g., infusion cyclical, oral daily, oral cyclical).

In what follows, “time unit” can be e.g., weeks or days.

Treatment duration in days

For napabucasin:

- Overall intended treatment duration = earliest date of {latest date of {start date of last cycle +27 days, last dose date of Napabucasin}, death date, end of study date} – earliest date of {first dose date of Napabucasin and/or nab-paclitaxel and/or gemcitabine}+1.
- Run-in treatment duration = actual number of days in run-in period.

- First 28 days intended treatment duration = earliest date of {start date of 1st cycle + 27 days, death date, end of study date} – start date of 1st cycle +1.

For backbone treatment:

- Overall intended treatment duration = (earliest date of {latest date of {start date of last cycle +27 days, last dose date of Napabucasin}, death date, end of study date} – earliest date of {first dose date of nab-paclitaxel or gemcitabine (whichever backbone duration you are calculating)} +1)/ 7.
- First 28 days intended treatment duration = (earliest date of {start date of 1st cycle + 27 days, start date of nab-paclitaxel or gemcitabine (whichever backbone duration you are calculating) within Day 28 ±3 window -1 } – start date of 1st cycle +1)/ 7.

Note that the actual end of treatment date in this analysis is not the collected end of treatment date in the eCRFs, as the actual end of treatment date in this analysis is being derived and may be different from what was recorded by sites.

Cumulative dose in a period or overall is the sum of the actual doses received in a period or overall, respectively.

Actual Dose Intensity [DI]

- Overall actual DI = overall accumulative dose in mg or mg/m^2 / overall intended treatment duration in days or weeks.
- Run-in actual DI = actual total dose in mg in run-in / actual treatment duration in days in run-in.
- First 28 days actual DI = actual total dose in mg or mg/m^2 in first 28 days / intended treatment duration in days or weeks in first 28 days.

Relative Dose Intensity (RDI): The basic intent is to evaluate dose per *time unit* factoring in dose reductions, interruptions, or delays.

- Relative dose intensity (RDI) by period and overall
    - Intended DI (*dose unit*/*time unit*) = [intended cumulative dose per period] / [intended number of *time units* in a period]
    - Overall RDI (%) = 100 × [overall actual DI] / [intended DI]

Note:

- *The intended dose level is fixed at the start of treatment rather than the start of a cycle*

  - *One exception is for all backbone therapies. All backbone therapies incorporate BSA to calculate each dose. Therefore, the BSA to be used for calculating each dose of the backbone therapies will be using the Mosteller method (see below) based on the last available weight and height prior to each infusion.*
  - *BSA [m^2] = ((Height [cm] × Weight [kg])/3600) × 0.5*
  - *Calculated doses will be rounded to the nearest integer*
  - *Cumulative dose, actual dose intensity and relative dose intensity will remain missing if they cannot be derived due to missing BSA. Height from the most recent earlier visit can be used if height is missing.*
- *The intended cumulative dose of a combination drug per cycle is constant for all cycles*
  - *One exception is for all backbone therapies after the updated BSA or weight calculation prior to each infusion. The updated dose level based on changing BSA or weight should be considered when calculating intended cumulative dose of these drugs per cycle*
- *The intended cumulative dose of napabucasin per period is calculated by (intended dose level per day) x (intended period duration in days)*
- *Dose unit of napabucasin is mg, dose unit of a combination drug is mg/m^2*

Treatment compliance

Napabucasin compliance overall

Overall: Treatment compliance (%) for napabucasin =
(Cumulative actual total dose /total planned or intended dose) x 100

Daily treatment compliance will be reported for each patient for the following dose-levels and intervals:

- The % of days the patient received a total dose of at least 480 mg napabucasin out of actual treatment duration in days.
- The % of days the patient received a non-zero dose of napabucasin out of actual treatment duration in days.

Daily treatment compliance will be grouped according to the following categories: < 60%,≥ 60% - < 80%, ≥ 80% - < 90%, ≥ 90%, and will be summarized for the treatment group of napabucasin plus Nab-Paclitaxel + Gemcitabine IV.

Combination drugs compliance overall

Overall: Treatment compliance (%) for combination drugs =
(Number of treatments administered /number of treatments that should have been administered) x 100

8.2.2. **Analysis for Primary Endpoint**

8.2.2.1. Overall Survival

Overall Survival in the general study population, the primary endpoint of this study, is defined as the time from randomization to death from any cause. Patients who are alive at the time of interim or final analysis or who have dropped out will be censored at their last date known to be alive on or before the date of data cutoff. Patients will be analyzed in the arm to which they are randomized regardless of the treatment they received (intent-to-treat analysis). OS in months is calculated as (date of death/Last known to be alive date – date of randomization + 1)/30.4375.

The survival experience of patients in both treatment groups will be summarized by the Kaplan-Meier method and compared primarily by a stratified log-rank test adjusting for stratification variables at randomization, including geographical region (North America/Western Europe/Australia vs. Japan/Korea vs. Rest of the World), ECOG performance status (ECOG 0 *versus* ECOG 1), and presence of liver metastases (yes *versus* no). The hazards ratio and 95% CI will be obtained using the Stratified Cox PH Model. If any randomized stratification factor proves to have an inadequate sample size, it will be dropped from the primary analysis or pooling will be applied. Details will be determined in a blinded fashion prior to database lock. The survival probability at 6 months, 1 year, and 18 months will be summarized using the Kaplan-Meier method.

Sensitivity analysis using Cox proportional hazards model based on actual stratified factors will be performed. An unstratified, 1-sided log-rank test will also be performed for sensitivity using an unstratified Cox PH model to obtain HR and the 95% CI. A multi-variate Cox PH model stratified by the stratification factors and including other baseline factors (as described in the subgroup analyses below) will also be conducted. Additional analyses for each stratification level will be performed using the Kaplan-Meier method for each treatment for each stratification level. A stratified, 2-sided log-rank test will be performed along with a stratified Cox PH model to obtain HR and 95% CI. Additional analyses may be conducted using the unstratified Cox proportional model incorporating each baseline factor if deemed adequate and necessary.

Subgroup analysis and the corresponding forest plot will be conducted by the stratification factors at baseline as mentioned above as well as the following factors:

- Primary tumor location (head *versus* body/tail of pancreas)
- Level of CA19-9 (normal *versus* >ULN and <59 x ULN *versus* $\geq$59 x ULN)
- Age (< 65 *versus* $\geq$ 65)
- Sex (male *versus* female)
- Number of metastatic sites at baseline (1 *versus* $\geq$ 2)
- Race (white, black, Asian, other)
- Smoking (Yes *versus* No)
- Prior adjuvant chemotherapy (Yes versus No)

- Prior radiotherapy (Yes versus No)
- Geographical region (North America *versus* Western Euope *versus* Australia *versus* Asia *versus* China *versus* Japan *versus* Korea *versus* Rest of Asia (excluding China, Korea, and Japan) *versus* Other (excluding North America, Western Euope, Australia, and Asia))

8.2.3. Analyses for Secondary Endpoints

8.2.3.1. Key Secondary Efficacy Endpoints Analysis

The following key secondary outcomes will be also be assessed.

PFS in the General Study Population
PFS in the general study population is defined as the time from randomization to the first objective documentation of disease progression or death due to any cause. If a patient has not progressed or died at the time of interim or final analysis (up to the date of data cutoff), PFS will be censored on the date of the last tumor assessment (up to the date of data cutoff). PFS in months is calculated as (First event date/Censored Date – date of randomization + 1)/30.4375.

The difference in PFS between treatment arms will be analyzed using the log rank test stratified by stratification factors. The 1-sided p value from the stratified log-rank test will be reported along with the median event time (and other quartiles) and corresponding 2-sided 95% CI for each treatment arm. Estimates of the PFS curves obtained from the Kaplan-Meier method will be presented along with a graphical presentation of PFS curves. The Cox PH model stratified for randomized stratification factors will be fitted and the estimated hazard ratio and 2-sided 95% CI will be provided.

Additional analyses may be conducted using the unstratified Cox proportional model incorporating each baseline factor if deemed adequate and necessary.

DCR in the General Study Population
DCR is defined as the proportion of patients with a documented complete response, partial response, and stable disease (CR + PR + SD) based on RECIST 1.1. The primary estimate of DCR will be based on the Response Analysis Set (RAS).

DCR will be summarized for each treatment arm along with the corresponding exact 2-sided 95% CI using the Clopper Pearson method. Differences of DCR between the two arms will be compared using a 1-sided Cochran-Mantel-Haenszel test stratified for stratification factors. The treatment difference of DCR and its 95% CI based on the Miettinen and Nurminen (M&N) method adjusting for stratification factors will be provided along with the DCR and 95% CI for each stratification factor level for each treatment arm.

Additional analyses may be conducted using the unstratified one-sided Z test incorporating each baseline factor to compare the differences of DCR between the two arms. The treatment difference of DCR and its 95% CI may also be obtained using the normal approximation method.

ORR in the General Study Population
ORR is defined as the proportion of patients with a documented complete response and partial response (CR + PR) based on RECIST 1.1. The primary estimate for ORR will be based on the RAS.

ORR will be analyzed similarly as DCR.

8.2.3.2. Other Secondary Endpoint Analysis

QoL Analysis
The QoL of patients will be assessed using EORTC QLQ-30. The EORTC QLQ-30 is a self-administered cancer specific questionnaire with multi-dimensional scales. It consists of both multi-item scales and single item measures, including five functional domains, a global quality of life domain, three symptom domains, and six single items. Scoring of the EORTC QLQ-30 data will be completed following the procedures recommended by the EORTC Study Group on Quality of Life. For each domain or single item measure a linear transformation will be applied to standardize the raw score to range between 0 and 100. Questionnaire compliance rates will be ascertained for each group at each measurement time point. Mean baseline scores for each subscale and summary scores will be calculated.

Compliance at an assessment time point is defined as the number of patients who were assessed divided by the expected number of patients at that time point. The expected number of patients:

- at baseline is equal to the number of patients randomized
- at any post-baseline visit is equal to the number of patients who are alive and have not progressed

A subject who answers at least one item at a time point is considered to have been assessed.

The endpoints in QoL analysis are the mean EORTC QLQ-C30 QoL change scores from baseline at time 2 and time 4 for the physical function and global health status/quality of life subscale scores. Wilcoxon tests may be used to compare the difference at each of these two time points between two treatment arms for each of these two subscales.

Summary statistics will be provided by treatment for Functional, Symptom and Global health status raw scores for QoL-C30 as well as for linearly converted scores on the scale of 0-100.

Note the raw score is: RawScore=RS=$(I_1+I_2+\ldots+I_n)/n$, where I_1, I_2,…and I_n are the nth component in the categories of functional, symptom or global health status scales.

The linearly converted score (LCS) for Functional scales is defined as:

Linearly Converted Score =LCS = (1- (RS-1)/Range)) ×100

For Symptom scales or Global health status:

Linearly Converted Score =LCS = (RS-1)/Range ×100

The following gives the LCS for all the scales:

Functional scale's scores:

- Physical functioning: (1 - ((Q1+Q2+Q3+Q4+Q5)/5 -1)/3) * 100
- Role functioning: (1 - ((Q6+Q7)/2-1)/3) * 100
- Emotional functioning: (1 - ((Q21+Q22+Q23+Q24)/4-1)/3) * 100
- Cognitive functioning: (1 - ((Q20+Q25)/2-1)/3) * 100
- Social functioning: (1 - ((Q26+Q27)/2-1)/3) * 100

Global health status score:

- Global QOL: ((Q29+Q30)/2-1)/6 * 100

Symptom scale's scores:

- Fatigue: ((Q10+Q12+Q18)/3-1)/3 * 100
- Nausea and vomiting: ((Q14+Q15)/2-1)/3 * 100
- Pain: (Q9+Q19)/2-1)/3 * 100
- Dyspnea: ((Q8-1)/3 * 100
- Insomnia: (Q11-1)/3 * 100
- Appetite loss: (Q13-1)/3 * 100
- Constipation: (Q16-1)/3 * 100
- Diarrhea: (Q17-1)/3 * 100
- Financial difficulties: (Q28-1)/3 * 100

Missing data handling for missing items in a scale are specified in Section 7.1.

The QoL assessment is performed prior to randomization and during protocol treatment after randomization. Since exact time of assessment may vary from patient to patient, it is necessary to provide a window for each QoL time point. What follows is a description of how to assign a questionnaire to a discrete time point:

Relative Days (REL_Day) from Randomization =

- Assessment date – randomization date+1; if the assessment was performed on or after the randomization date.
- Assessment date – randomization date; if the assessment was performed prior to the randomization date.

Table 1 Analysis Windows for QoL Endpoints

Time Point	Target Date (Rel_Day)	Windows

Baseline	1	-14≤rel_day≤ 1
4 Weeks post randomization	29	1<rel_day≤ 43
8 Weeks post randomization	57	43<rel_day≤ 71
12 Weeks post randomization	85	71<rel_day≤ 99
16 Weeks post randomization	113	99<rel_day≤ 141
24 Weeks post pandomization	169	141<rel_day≤ 197

Note that if more than 1 questionnaire is available for the baseline window, then the latest non-missing measurement, per question, will be considered as the baseline. If more than 1 questionnaire is available at a time point other than baseline, then the average (per question) of the non-missing measurements will be used.

If there are multiple records in the same day to 1 item, the more severe answered will be counted.

QOL Response Analysis for EORTC QLQ-C30

For EORTC-QLQ-C30, QoL response for functional scales and global health status is calculated as follows: A change score of 10 points from baseline is defined as clinically relevant. Patients are considered to have clinical improvement if reporting a score 10-points or better than baseline at any time of QoL assessment. Conversely, patients are considered worsened if reporting a score minus 10-points or worse than baseline at any time of QoL assessment without any clinical improvement. Patients whose scores are between 10-point changes from baseline at every QoL assessment will be considered as stable. In contrast to functional scales, for the determination of patient's QoL response, classification of patients into improved and worsened categories is reversed for symptom scales. A Chi-square test may then be performed to compare the distributions of these three categories between the two arms.

The proportion of patients in either arm with at least a minimum of 10 unit(s) deterioration in change scores at both 8 and 16 weeks will be compared by means of Fisher's exact test.

8.2.3.3. PK Analysis

Plasma napabucasin concentrations will be tabulated by Day within Cycle, nominal time and dose in the case of dose reduction if appropriate. The listing will include actual sampling time relative to the napabucasin dose administered on the day of the PK sample collection. Summary statistics by Day within Cycle, and irrespective of Day and Cycle, will be presented including arithmetic mean, arithmetic standard deviation, arithmetic-based percent coefficient of variation, geometric mean, geometric standard deviation, maximum, median, minimum, and sample size.

Exploratory analyses may be performed by Pop-PK analysis of napabucasin, possibly in conjunction with data from other napabucasin studies. The exposure-response relationship between clinical and safety endpoints and napabucasin/nab-paclitaxel with gemcitabine exposure may also be examined.

The procedures for the population PK modeling, including model evaluation, will be described in a seperate Pop-PK analysis plan. The corresponding results will be reported in a separate document, not in the clinical study report (CSR).

8.2.3.4. Biomarker Analysis - Correlative Studies

Pre-clinical and clinical studies conducted by SDPO have identified several biomarkers in tumor tissues whose levels either increase or decrease upon exposure to napabucasin. Tumor archival tissue and biopsies will be collected as described in Protocol (1) Section 12.3. The correlations between the biomarker results, and pharmacokinetic parameters and measures of anti-tumor/anti-cancer efficacy signals will be explored if data allow and it is deemed appropriate.

Specifically, the relationship between phospho-STAT3 (pSTAT3) and the clinical outcomes with napabucasin therapy may be explored. The following efficacy endpoints may be analyzed by pSTAT3 status (positive and negative and unknown):

- Primary Endpoint: Overall Survival (OS),
- Secondary Endpoint: Progression Free Survival (PFS), and
- Objective Response Rate (ORR) and Disease Control Rate (DCR)

Safety, demographics, baseline disease characteristics, and prior therapy may be summarized by treatment by pSTAT3 status (positive or negative or unknown) in all the patients with biomarker data.

8.2.4. Analyses for Other Safety Endpoints

All treatment emergent adverse events, lab abnormalities, vital signs, ECOG, physical exam, and ECG data analyses will be summarized based on the safety analysis set.

8.2.4.1. Adverse Events

Overall Summary of AEs

An AE will be regarded as **treatment-emergent**, if

- it occurs for the first time on or after the first dose date of either napabucasin or nab-Paclitaxel or gemcitabine up to 30 days after the last dose of study treatment (or up to any time if serious and considered related to study treatment); or
- it occurs prior to first dose date of either napabucasin or nab-Paclitaxel or gemcitabine and worsens in severity on therapy or up to 30 days after the last dose of study treatment (or up to any time if serious and considered related to study treatment)

Adverse events will be coded by SOC and PT using the MedDRA®v19.0. The severity of AEs will be graded by the investigator using NCI CTCAE Version 4.0. The verbatim term will be included in the AE listings. A Fisher's exact test may be used to compare adverse events between the two arms.

An overview of treatment-emergent adverse events (TEAEs) will be provided. The number and percentage of patients with following will be summarized for:

- Patients with at least one TEAE
- Patients with TEAE of CTCAE grade 3 or higher
- Patients with serious TEAE
- Patients with serious TEAE related to napabucasin
- Patients with serious TEAE related to nab-paclitaxel
- Patients with serious TEAE related to gemcitabine
- Patients with napabucasin related TEAE
- Patients with napabucasin related TEAE of CTCAE grade 3 or higher
- Patients with Nab-Paclitaxel related TEAE
- Patients with Nab-Paclitaxel related TEAE of CTCAE grade 3 or higher
- Patients with gemcitabine related TEAE
- Patients with gemcitabine related TEAE of CTCAE grade 3 or higher
- Patients with TEAE related to study drug (napabucasin or nab-paclitaxel or gemcitabine)
- Patients with study drug (napabucasin or nab-paclitaxel or gemcitabine) related TEAE of CTCAE grade 3 or higher
- Patients with TEAEs leading to death
- Patients with TEAEs leading to napabucasin dose modification
- Patients with TEAEs leading to napabucasin dose delay
- Patients with TEAEs leading to napabucasin discontinuation
- Patients with TEAEs leading to dose reduction of napabucasin
- Patients with TEAEs leading to nab-paclitaxel dose modification
- Patients with TEAEs leading to nab-paclitaxel dose delay
- Patients with TEAEs leading to nab-paclitaxel discontinuation
- Patients with TEAEs leading to dose reduction of nab-paclitaxel
- Patients with TEAEs leading to gemcitabine dose modification
- Patients with TEAEs leading to gemcitabine dose delay
- Patients with TEAEs leading to gemcitabine discontinuation
- Patients with TEAEs leading to dose reduction of gemcitabine

Summary of TEAEs by System Organ Class and Preferred Term

The number and percentage of patients who experience any TEAE and TEAEs of grade 3 or higher (Grade 3, 4, 5) will be summarized by SOC and PT. A summary of TEAEs by PT will be presented in the descending order of frequency counts for all grades and grade 3 or higher. A summary of AEs by PT and maximum CTCAE grade will be presented.The

most commonly reported AEs using different cutoffs (e.g., 2%, 5% or 10% or more of patients in either arm) may also be summarized by PT as needed for various reporting purposes. Adverse events associated with permanent discontinuation/dose reduction/dose interruption of either napabucasin and/or nab-paclitaxel and/or gemcitabine will also be summarized by SOC and PT and maximum CTCAE grade.

Treatment Related TEAE

AEs reported with a relationship to a treatment considered by the investigator to be 'possible', 'probable' or 'definite' will be considered "Related" to study treatment or paclitaxel or gemcitabine, respectively. AEs reported with a relationship to a treatment considered by the investigator to be "unlikely" or "unrelated" will be considered as "Not Related" to study treatment or nab-paclitaxel or gemcitabine, respectively. Missing relationship will be considered as "Related". Similar summaries of all causality AEs will be provided. TEAEs related to study drug, leading to dose modification, dose reduction, and dose discontinuation will be summarized by SOC and PT.

Serious TEAE and Death

Treatment emergent SAEs and treatment related TESAEs will be summarized by MedDRA SOC and PT and Maximum CTCAE grade.

Patients who experienced a TESAE during AE reporting period will be listed for all safety patients. The number and percentage of patients who experience any treatment emergent SAE will be summarized by SOC, PT and maximum CTCAE grade. Similar summary for treatment related TESAEs will be provided as well.

Deaths that occur on or after the first dose of study treatment and within 30 days of last dose of any study treatment will be summarized. The number and percentage of patients who died during the study treatment and within 30 days after the last dose will be presented. TEAEs leading to death will be summarized by MedDRA SOC and PT.

A listing of death data will also be provided and it will include all deaths that occurred during the reporting period for deaths which starts from the signing of informed consent to the end of the follow up period. The listing will include date of death, and the number of days relative to the administration of first and last dose.

Clustered Treatment Emergent Adverse Events

Selected treatment emergent adverse events were pre-specified for additional focus due to the potential clinical significance of the event and/or the potential association with the investigational product. These events include those in the standard MedDRA (narrow or broad, as noted) query terms of: non-infectious diarrhea (narrow SMQ), gastrointestinal nonspecific inflammation (narrow SMQ), gastrointestinal haemorrhage (narrow SMQ), gastrointestinal perforation (narrow SMQ), gastrointestinal obstruction (narrow SMQ), acute renal failure (narrow SMQ), haematopoietic leukopenia (broad SMQ),

hematopoietic erythropenia (broad SMQ), and hematopoietic thrombocytopenia (broad SMQ). Other events include the following sponsor derived terms: Abdominal pain, fatigue, hypotension, and torsade de Pointes. Tables listing the incidence, maximum severity, relation to study drug, dose modification of study drug, and death due to these events in the safety analysis set will be generated.

8.2.4.2. Laboratory Data

For the purposes of summarization in both the tables and listings, all laboratory values will be converted to standardized units. If a lab value is reported using a non-numeric qualifier (eg, less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier.

Laboratory test results will be summarized according to the scheduled sample collection time point. Change from baseline will also be presented. Unscheduled laboratory test results will be listed and included in laboratory shift tables. The parameters to be analyzed are as follows:

- Hematology: hemoglobin, hematocrit, ANC, neutrophils, lymphocytes, monocytes, eosinophils, basophils, platelets, and white blood cell (WBC) count
- Biochemistry: blood urea nitrogen, creatinine, total bilirubin, LDH, albumin, alkaline phosphatase, AST, ALT, sodium, potassium, magnesium, phosphoros.

Shift tables will be constructed for laboratory parameters to tabulate changes in NCI CTCAE for toxicity (version 4.0) from baseline to post baseline worst CTC grade. Parameters to be tabulated will include:

- Hematology: ANC, hemoglobin, platelets, WBC
- Biochemistry: ALT, AST, alkaline phosphatase, creatinine, total bilirubin, magnesium, potassium, sodium, and phosphoros.

Listing will also be presented for urinalysis values.

- Urinalysis: specific gravity, protein, occult blood, glucose

Maximum change versus baseline figures and E-DISH scatter plots for key lab parameters will be produced, including but not limited to liver function tests (ALT, AST, ALP and total bilirubin).

A listing of all lab parameters including hematology, biochemistry, urinalysis will be provided, including the test result, units, normal range (H and L) and change from baseline, and CTCAE grades if grading applies. Patients who developed toxicities of grade >= 3 will also be listed.

8.2.4.3. Electrocardiograms

12-lead ECG with categorical results (Normal, Abnormal [Not clinically significant], Abnormal [clinically significant]) will be summarized by treatment and visit. The shift from baseline to worst post baseline will be produced. A patient listing will also be provided.

8.2.4.4. Vital Signs

Vital sign data (heart rate, systolic and diastolic blood pressure, and temperature etc.) change from baseline will be summarized by treatment group and visit. Maximum change from baseline will also be summarized by treatment group.

Summaries of markedly abnormal vital signs parameters, including blood pressure (BP), pulse, and BMI will be presented by treatment group. Values for vital signs for all patients will be presented in a listing, and patients with markedly abnormal values will be flagged.

Markedly abnormal ranges for vital signs parameters are given in the table below.

Vital Sign Parameter	Markedly Abnormal (Low)	Markedly Abnormal (High)
Systolic BP	Absolute value ≤ 90 mmHg, or a decrease from baseline ≥ 20 mmHg	Absolute value ≥ 150 mmHg, or an increase from baseline ≥ 20 mmHg
Diastolic BP	Absolute value ≤ 50 mmHg, or a decrease from baseline ≥ 15 mmHg	Absolute value ≥ 90 mmHg, or an increase from baseline ≥ 15 mmHg
Pulse	Absolute value ≤ 50 bpm, or a decrease from baseline ≥ 15 bpm	Absolute value ≥ 120 bpm, or an increase from baseline ≥ 15 bpm
BMI	Absolute value ≤ 18 kg/m^2	Absolute value ≥ 40 kg/m^2

8.2.4.5. Physical Exam

Physical examination abnormalities will be summarized for each visit by body system and by treatment group. Patients with clinically significant abnormal findings will be flagged in the data listing.

8.2.4.6. ECOG

ECOG performance status will be summarized in a shift table from baseline to worst post baseline for Safety analysis set.

9. REFERENCES

1. Protocol Canstem 111P A Phase III Study of napabucasin plus nab-Paclitaxel with Gemcitabine in Adult Patients with Metastatic Pancreatic Adenocarcinoma, December 4th, 2018
2. Blyth C.R., Still H.A. [1983]. Binomial Confidence Intervals *Journal of the American Statistical Association 78*, 381.
3. Brookmeyer R, Crowley JJ [1982]. A confidence interval for the median survival time. *Biometrics*. 38:29-41.
4. Maurer, W., Bretz, F. (2013). Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research, 5, 311-320.
5. Miettinen O., Nurminen M. (1985) Comparative Analysis of Two Rates. Statistics in Medicine, 4, 213-226.
6. Williams D.A. [1972]. The Comparison of Several Dose Levels With a Zero Dose Control *Biometrics 28*, 519-531.

10. APPENDICES

10.1. Appendix 1.1 DATA DERIVATION DETAILS

Table 2 Response Criteria

Evaluation of target lesions	
Complete Response (CR):	Disappearance of all target lesions Any pathological lymph nodes must have reduction in short axis of <10mm
Partial Response (PR):	At least a 30% decrease in the sum of the LD of target lesions, taking as reference the baseline sum LD
Progressive Disease (PD):	At least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum LD recorded since the treatment started. In addition to the increase of 20%, the sum must also demonstrate an absolute increase of at least 5mm. The appearance of one or more new lesions is also considered progression
Stable Disease (SD):	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started

Evaluation of non-target lesions	
Complete Response (CR):	Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10mm short axis)
Non-CR/Non-PD	Persistence of one or more non-target lesion(s) and/or maintenance of tumor marker level above normal limits
Progressive Disease (PD):	Appearance of one or more new lesions and/or unequivocal progression of existing non-target lesions

The best overall response (BOR) is the best response recorded from the randomization date until disease progression/recurrence or start of new anti-cancer therapy and is determined as indicated in Table 3. BOR is derived from the sequence of overall response. Assessment done after PD or after "new anti-cancer" treatment but prior to PD will not be considered evaluation of best overall response. Tumor assessments will be performed every 8 weeks. BOR derivation (based on unconfirmed response) is derived from the sequence of overall response determined by the following order:

- CR: One objective status of CR documented before progression or start of new anti-cancer therapy.
- PR: One objective status of PR documented before progression or start of new anti-cancer therapy, but not qualifying as CR.
- SD: At least 1 objective status of SD or better documented within at least 1 nominal scan interval (8 weeks – 5 days window = 51 days) after start date and before progression and the start of new anti-cancer therapy, but not qualifying as CR or PR.
- PD: Progression documented within 2 nominal scan intervals (or 16 weeks + 5 days window = 117) after start date and not qualifying as unconfirmed CR, unconfirmed PR, or SD.

- NE: All other cases. Note that reasons for NE should be summarized and the following reasons could be used:
    - Early death (Note: death prior to 8 weeks after start date)
    - No post-baseline assessments
    - All post-baseline assessments have overall response NE
    - New anti-cancer therapy started before first post-baseline assessment
    - SD too early (<8 weeks after randomization date)
    - PD too late (>16 weeks after randomization date)

Special and rare cases where BOR is NE due to both early SD and late PD will be classified as 'SD too early'.

Table 3 Overall Response

Target lesions	Non-Target lesions	Evaluation of New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	not all evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not All Evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

Disease Control Rate (DCR) is the proportion of patients with a documented CR, PR, or SD based on RECIST 1.1 as determined by the investigator assessment collected in the CRFs, relative to the RAS.

Objective response rate (ORR) is defined as the proportion of patients with a documented complete response (CR) or partial response (PR) based on RECIST 1.1 as determined by the investigator assessment collected in the CRFs, relative to the RAS.

The patients will be considered as non-responder until proven otherwise. Thus, the following patients are considered non-responders:

- Do not have CR or PR while on study; or
- Do not have a baseline or post-baseline tumor evaluation; or
- Do not have an adequate baseline tumor evaluation; or
- Receive anti-tumor treatment other than the study medication prior to reaching a CR or PR; or
- Die, progress, or drop out for any reason prior to reaching a CR or PR.

Progression free survival (PFS) is defined as the time from the date of randomization to the date of PD, or death due to any cause, whichever comes first. If neither event has been observed, then the patient will be considered as censored. PFS (in months) will be calculated as (first event date – date of randomization + 1) / 30.4375. Sensitivity analyses for PFS may be conducted if deemed necessary.

Overall survival (OS) is defined as the time from the date of randomization to the date of death from any cause. Patients who are still alive at the time of the final observation, or who have become lost to follow-up will be censored at their last date of contact. OS in month is calculated as (date of death – date of randomization +1) / 30.4375. Sensitivity analyses for OS may be conducted if deemed necessary.

10.2. Appendix 1.2 Censoring for Time to Event Data

Table 4 summarizes the censoring rules for the PFS primary analysis. Table 5 shows the general reasons for PFS censoring for primary analysis and where the primary analysis censoring hierarchy in Table 4 comes from.

Table 4 Event or Censor Time and Censoring Hierarchy for Primary Analysis on PFS

Censoring Hierarchy	Situation	Date of Event or Censor	Primary Analysis
1	**No baseline** radiological tumor assessment available	Date of randomization	Censored

2	**New anticancer treatment** started before tumor progression or death	Date of previous adequate radiological assessment immediately prior to start of new therapy or Date of randomization, whichever comes later	Censored
3	**Tumor progression** (per RECIST 1.1) documented or death after 2 scan intervals following previous adequate radiological tumor assessment, no new anticancer treatment started	Date of previous adequate radiological assessment or Date of randomization, whichever comes later	Censored
4, 5	**No tumor progression** (per RECIST 1.1), no death reported and subject lost to follow-up or withdrawal of consent, no new anticancer treatment started	Date of last adequate radiological Assessment or Date of randomization, whichever comes later	Censored
6	**No tumor progression** (per RECIST 1.1) and no death reported within 2 scan intervals following last adequate radiological tumor assessment or randomization (if no post baseline tumor assessment available), no new anticancer treatment started	Date of last adequate radiological tumor assessment or Date of randomization, whichever comes later	Censored
7	**No tumor progression** (per RECIST 1.1) and no death reported and none of the conditions in the prior hierarchy are met	Date of last adequate radiological tumor assessment or Date of randomization, whichever comes later	Censored

	No tumor progression (per RECIST 1.1) but death reported within 2 scan intervals following last adequate radiological tumor assessment or randomization (if no post baseline tumor assessment available)	Date of death	Event
	Tumor progression (per RECIST 1.1) documented within 2 scan intervals following previous adequate radiological tumor assessment	Earliest of the target, non-target and new tumor assessment dates	Event

Notes: (1) Symptomatic deteriorations (i.e. symptomatic progressions, which are not radiographically confirmed) will not be considered as progressions.

(2) If target, non-target and new lesion assessments have different dates within a visit, then the earliest of those dates will be considered as the date of the tumor assessment if the assessment for that visit is progressive disease (PD); otherwise the latest date will be used.

(3) Adequate radiographical tumor assessment refers to an assessment with overall response of CR, PR, SD or PD.

(4) Comparing date of last tumor assessment to date of randomization is necessary if the last tumor assessment is baseline assessment

Table 5 PFS Censoring Reasons

Hierarchy	Condition	Censoring Reason for Primary PFS Analysis	Censoring Reason for Sensitivity PFS Analysis
1	No adequate baseline assessment	No adequate baseline assessment	No adequate baseline assessment
2	Start of new anti-cancer therapy before event.	Start of new anti-cancer therapy	Event
3	Event more than *16 weeks* from last adequate post-baseline tumor assessment/start date	Event after missing assessments [a]	Event
4	No event and subject withdraws consent	Withdrawal of consent	Withdrawal of consent

5	No event and lost to follow-up in any disposition page	Lost to follow-up	Lost to follow-up
6	No event and no adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment
7	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event	Ongoing without an event

[a] more than 16 weeks after last adequate tumor assessment

For OS, patients last known to be alive are censored at date of last contact.

Date of Last Contact

The date of last contact will be derived for patients not known to have died at the analysis data cutoff date using the latest complete date (non-imputed) among the following:

- All patient assessment dates (e.g., blood draws [laboratory, PK], vital signs, performance status, ECG, tumor assessments, concomitant radiation, surgery)
- Start and end dates of follow-up anti-cancer therapies
- AE start and end dates
- Last date of contact where "Subject Remains in Follow-up" collected on the "Survival Follow-up" eCRF (do not use date of survival follow-up assessment unless status is alive)
- Study drug start and end dates
- Randomization date
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up or death).

Note:

1. This list is not all inclusive and should be agreed upon by the study team according to the data collected in the CRF
2. Only dates associated with patient visits or actual examinations of the patient should be used. Dates associated with a technical operation unrelated to patient status (e.g., the date a blood sample was processed) should not be used.
3. Assessment dates after the cutoff date will not be applied to derive the last contact date.

10.3. Appendix 1.3 Summary of Key Efficacy Analysis

Endpoint (Section)	Analysis Set	Statistical Method	Missing Data	Interpretation
OS (8.2.2)	ITT	Stratified log-rank test (1-sided, all S factors); KM method (median, 95% CI). HR and 95% CI from Stratified Cox PH Model	Censor patients who didn't die or who lack any data beyond randomization	Primary Analysis for primary endpoint
6 months/1 year/18 months survival	ITT	KM method (95% CI)		
OS (8.2.2)	ITT	Stratified log-rank test (1-sided, all S factors on actual stratification factor); KM Method (median, 95% CI). HR and 95% CI from Stratified Cox PH Model		Sensitivity analysis of primary endpoint
OS (8.2.2)	ITT	Un-stratified log rank test (1-sided). HR and 95% CI from un-stratified Cox PH model		Sensitivity analysis of primary endpoint
OS (8.2.2)	ITT	Multi-variate Cox PH Model stratified by S factors and including other baseline covariates (HR and 95% CI)		Sensitivity Analysis of primary endpoint

OS (8.2.2)	ITT	Cox PH Model containing treatment group and one given factor (HR and 95% CI)		Sensitivity Analysis of primary endpoint
OS (8.2.2)	Subgroup	For each baseline factor, KM Method (median, 95% CI). Unstratified log-rank test (2-sided) HR and 95% CI from un-stratified Cox PH Model		Sensitivity Analysis of primary endpoint
PFS(8.2.3)	ITT	Stratified log-rank test (1-sided, all S factor); KM Method (median, 95% CI). HR and 95% CI from Stratified Cox PH Model	Censoring patients following Table 4 criteria in Appendix 1.2	Primary analysis for secondary endpoint.
6 months/1 year/18 months progression free survival	ITT	KM method (95% CI)		
PFS (8.2.3)	ITT	Un-stratified log rank test (1-sided). HR and 95% CI from un-stratified Cox PH model		Sensitivity analysis of secondary endpoint

PFS (8.2.3)	Subgroup	For each baseline factor, KM Method (median, 95% CI). Unstratified log-rank test (2-sided) HR and 95% CI from un-stratified Cox PH Model		Sensitivity Analysis of secondary endpoint
DCR (8.2.3)	RAS	Stratified CMH (1-sided, Stratified by S factors) TrtDiff and 95% CI from M&N method adjusting S factors. Exact CI based on Clopper-Pearson method.	Patients without on-study tumor assessment or who die, progress or drop out for any reason, or receive anti-tumor treatment prior to reaching a CR or PR as non-responders.	Primary analysis for secondary endpoint.
DCR (8.2.3)	RAS	Z test (1-sided). TrtDiff and 95% CI from normal approximation method. Exact CI based on Clopper-Pearson method		Sensitivity analysis of secondary endpoint
DCR (8.2.3)	Subgroup	Z test (2-sided) TrtDiff and 95% CI from normal approximation method. Exact CI based on Clopper-Pearson method.		Sensitivity Analysis of secondary endpoint

ORR(8.2.3)	RAS	Stratified CMH (1-sided, Stratified by S factors) TrtDiff and 95% CI from M&N method adjusting S factors. Exact CI based on Clopper-Pearson method.	Patients without on-study tumor assessment or who die, progress or drop out for any reason, or receive anti-tumor treatment prior to reaching a CR or PR as non-responders.	Primary analysis for secondary endpoint.
ORR(8.2.3)	RAS	Z test (1-sided). TrtDiff and 95% CI from normal approximation method. Exact CI based on Clopper-Pearson method		Sensitivity analysis of secondary endpoint
ORR (8.2.3)	Subgroup	Z test (2-sided) TrtDiff and 95% CI from normal approximation method. Exact CI based on Clopper-Pearson method.		Sensitivity Analysis of secondary endpoint

ITT: Intent-to-Treat Analysis Set; RAS: Response Analysis Set

10.4. Appendix 1.4 Response and Evaluation Endpoints

Response and progression will be evaluated in this study using the revised international criteria (1.1) proposed by the RECIST (Response Evaluation Criteria in Solid Tumors) committee.

1.1.1 Measurable Disease

Measurable *tumor lesions* are defined as those that can be accurately measured in at least one dimension (longest diameter to be recorded) as ≥ 20 mm with chest x-ray and as ≥ 10 mm with CT scan, or clinical examination. Bone lesions are considered measurable only if assessed by CT scan and have an identifiable soft tissue component that meets these requirements (soft tissue component ≥ 10 mm by CT scan). *Malignant lymph nodes* must be ≥ 15mm in the short axis to be considered measurable; only the short axis will be measured and followed. All tumor measurements must be recorded in millimeters (or decimal fractions of centimeters). Previously irradiated lesions are not considered measurable unless progression has been documented in the lesion.

1.1.2 Non-Measurable Disease

All other lesions (or sites of disease), including small lesions are considered non-measurable disease. Bone lesions without a measurable soft tissue component, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis cutis/pulmonis, inflammatory breast disease, lymphangitic involvement of lung or skin and abdominal masses followed by clinical examination are all non-measurable. Lesions in previously irradiated areas are non-measurable, unless progression has been demonstrated.

1.1.3 Target Lesions

When more than one measurable tumor lesion is present at baseline all lesions up to *a maximum of 5 lesions total* (and a maximum of *2 lesions per organ*) representative of all involved organs should be identified as target lesions and will be recorded and measured at baseline. Target lesions should be selected on the basis of their size (lesions with the longest diameter), be representative of all involved organs, but in addition should be those that lend themselves to *reproducible repeated measurements*. Note that pathological nodes must meet the criterion of a short axis of ≥ 15 mm by CT scan and only the *short* axis of these nodes will contribute to the baseline sum. All other pathological nodes (those with short axis ≥ 10 mm but < 15 mm) should be considered non-target lesions. Nodes that have a short axis < 10 mm are considered non-pathological and should not be recorded or followed (see 10.2.4). At baseline, the sum of the target lesions (longest diameter of tumor lesions plus short axis of lymph nodes: overall maximum of 5) is to be recorded.

After baseline, a value should be provided on the CRF for all identified target lesions for each assessment, even if very small. If extremely small and faint lesions cannot be accurately measured but are deemed to be present, a default value of 5 mm may be used. If lesions are too small to measure and indeed are believed to be absent, a default value of 0 mm may be used.

1.1.4 Non-Target Lesions

All non-measurable lesions (or sites of disease) plus any measurable lesions over and above those listed as target lesions are considered *non-target lesions*. Measurements are not required but these lesions should be noted at baseline and should be followed as "present" or "absent".

1.1.5 Response

All patients will have their BEST RESPONSE from the start of study treatment until the end of treatment classified as outlined below:

Complete Response (CR): disappearance of *target* and *non-target*. Pathological lymph nodes must have short axis measures < 10 mm (Note: continue to record the measurement even if < 10 mm and considered CR). Residual lesions (other than nodes < 10 mm) thought to be non-malignant should be further investigated (by cytology specialized imaging or other techniques as appropriate for individual cases *(Eisenhauer, 2009)*) before CR can be accepted.

Partial Response (PR): at least a 30% decrease in the sum of measures (longest diameter for tumor lesions and short axis measure for nodes) of target lesions, taking as reference the baseline sum of diameters. Non-target lesions must be non-PD.

Stable Disease (SD): neither sufficient shrinkage to qualify for PR, nor sufficient increase to qualify for PD taking as reference the smallest sum of diameters on study.

Progressive Disease (PD): at least a 20% increase in the sum of diameters of measured lesions taking as references the smallest sum of diameters recorded on study (including baseline) AND an absolute increase of $\geq$ 5mm. Appearance of new lesions will also constitute progressive disease (including lesions in previously unassessed areas). For patients with on-study tumor assessments with questionable findings of new lesions, such as cases where it is possible that the lesions were present at baseline imaging albeit less visible, the patient may continue treatment (if patient is clinically doing well) for an additional 4-8 weeks, until the next scan which will either confirm presence of a new lesion finding or exclude it. In exceptional circumstances, unequivocal progression of non-target disease may be accepted as evidence of disease progression, where the overall tumor burden has increased sufficiently to merit discontinuation of treatment or where the tumor burden appears to have increased by at least 73% in volume. Modest increases in the size of one or more non-target lesions are NOT considered unequivocal progression. If the evidence

of PD is equivocal (target or non-target), treatment may continue until the next assessment, but if confirmed, the earlier date must be used.

Table 2: Integration of Target, non-Target and New Lesions into Response Assessment

Target Lesions	Non-Target Lesions	New Lesions	Overall Response	Best Response for this Category also Requires
Target Lesions ± Non Target Lesions				
CR	CR	No	CR	Tumor nodes < 10mm
CR	Non-CR/Non-PD	No	PR	
CR	Not all evaluated	No	PR	
PR	Non-PD/ not all evaluated	No	PR	
SD	Non-PD/ not all evaluated	No	SD	Documentation at least once ≥ 6 wks. from baseline
Not all evaluated	Non-PD	No	NE	
PD	Any	Any	PD	
Any	PD	Any	PD	
Any	Any	Yes	PD	
Non-Target Lesions ONLY				
N/A	CR	No	CR	Tumor nodes < 10mm
N/A	Non-CR/non-PD	No	Non-CR/non-PD	
N/A	Not all evaluated	No	NE	
N/A	Unequivocal PD	Any	PD	
N/A	Any	Yes	PD	
Note: Patients with a global deterioration of health status requiring discontinuation of treatment without radiological progression having been observed at that time should be reported as "symptomatic deterioration". This is NOT objective PD. Every effort should be made to document the objective progression even after discontinuation of treatment.				

1.2 Response Duration

Response duration will be measured from the time measurement criteria for CR/PR (whichever is first recorded) are first met until the first date that recurrent or progressive disease is objectively documented, taking as reference the smallest measurements recorded on study (including baseline).

1.3 Stable Disease Duration

Stable disease duration will be measured from the time of randomization until the criteria for progression are met, taking as reference the smallest sum on study (including baseline).

1.4 Methods of Measurement

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Assessments should be identified on a calendar schedule and should not be affected by delays in therapy. While on study, all lesions recorded at baseline should have their actual measurements recorded at each subsequent evaluation, even when very small (e.g. 2 mm). If it is the opinion of the radiologist that the lesion has likely disappeared, the measurement should be recorded as 0 mm. If the lesion is believed to be present and is faintly seen but too small to measure, a default value of 5 mm should be assigned. For lesions which fragment/split add together the longest diameters of the fragmented portions; for lesions which coalesce, measure the maximal longest diameter for the "merged lesion".

Additionally, for optimal tumor assessment scanning options are listed below in the decreasing order of preference:

Order of Preference	Scanning Option
1	Chest-Abdomen-Pelvis CT with oral and I.V. contrast
2	Chest CT without I.V. contrast PLUS MRI Abdomen-Pelvis with oral and I.V. contrast[1]
3	Chest-Abdomen-Pelvis CT with oral contrast[2]

[1] If Iodine contrast media is medically contraindicated.

[2] If Iodine contrast media is medically contraindicated and MRI cannot be performed.

1.4.1 Clinical Lesions

Clinical lesions will only be considered measurable when they are superficial and $\geq$ 10 mm as assessed using callipers (e.g. skin nodules). For the case of skin lesions, documentation by colour photography including a ruler to estimate the size of the lesion is recommended. If feasible, imaging is preferred.

1.4.2 Chest X-ray

Chest CT is preferred over chest X-ray, particularly when progression is an important endpoint, since CT is more sensitive than X-ray, particularly in identifying new lesions. However, lesions $\geq$ 20 mm on chest X-ray may be considered measurable if they are clearly defined and surrounded by aerated lung.

1.4.3 CT/MRI

CT is the best currently available and reproducible method to measure lesions selected for response assessment. This guideline has defined measurability of lesions on CT scan based on the assumption that CT slice thickness is 5 mm or less. When CT scans have slice thickness greater than 5 mm, the minimum size for a measurable

lesion should be twice the slice thickness. MRI is also acceptable in certain situations (e.g. for body scans). Other specialized imaging or other techniques may also be appropriate for individual case *(Eisenhauer, 2009)*. For example, while PET scans are not considered adequate to measure lesions, PET-CT scans may be used providing that the measures are obtained from the CT scan and the CT scan is of identical diagnostic quality to a diagnostic CT (with IV and oral contrast).

1.4.4 Ultrasound

Ultrasound is not useful in assessment of lesion size and should not be used as a method of measurement. If new lesions are identified by ultrasound in the course of the study, confirmation by CT is advised.

1.4.5 Endoscopy/Laparoscopy

The utilization of these techniques for objective tumor evaluation is not advised. However, they can be useful to confirm complete pathological response when biopsies are obtained or to determine relapse in trials where recurrence following complete response or surgical resection is an endpoint.

1.4.6 Histology

These techniques can be used to differentiate between PR and CR in rare cases (for example, residual lesions in tumor types such as germ cell tumors, where known residual benign tumors can remain).

The cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment when the measurable tumor has met criteria for response or stable disease is mandatory to differentiate between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease.