A Phase 3, Multinational, Randomized, Open-Label, Three Parallel-Arm Study of PF-06801591, an Anti-PD-1 Antibody, in Combination With Bacillus Calmette-Guerin (BCG Induction With or Without BCG Maintenance) Versus BCG (Induction and Maintenance) in Participants With High-Risk, BCG Naïve Non-Muscle Invasive Bladder Cancer or PF-06801591 as Single Agent in Participants With BCG-Unresponsive NMIBC

STATISTICAL ANALYSIS PLAN – B8011006

Compounds:	PF-06801591 BCG
Compound Name:	Sasanlimab BCG
Version:	8
Date:	10-Dec-2024

This document contains confidential information belonging to Pfizer. Except as otherwise agreed to in writing, by accepting or reviewing this document, you agree to hold this information in confidence and not copy or disclose it to others (except where required by applicable law) or use it for unauthorized purposes. In the event of any actual or suspected breach of this obligation, Pfizer must be promptly notified.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

TABLE OF CONTENTS

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

LIST OF TABLES

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

LIST OF FIGURES

APPENDICES

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for Study B8011006 is based on the protocol amendment 5 dated 17 June 2024.

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
1	11-Dec-2019	Not applicable
2	10-Aug-2020	Title page and headers - Sasanlimab has been added since it was officially confirmed as the generic name for PF-06801591 Section 2.2 “Study Design”, 3.4.2 “Baseline Characteristics”, Section 5.1.1 “Hypotheses and sample size determination”, Section 6.4 “Subset Analyses”, Section 6.5.1.1 “Demographic characteristics” - updated Western European Union to Western Europe Section 2.2 “Study Design” - Figure 1 added in “or persistence of CIS” Section 3.2.4 “Pharmacokinetic endpoints” - updated definition of pharmacokinetic endpoints to include trough and deleted reporting of during multiple dosing Section 3.2.6 ‘Biomarker endpoints” - in Table 4 replaced inflammatory cells with immune cells. Section 3.5.1 “Adverse Events” - removed BCG AEs Grade >=3 from AESI, it is captured under the AE sections Section 4 “Analysis Sets” - in Table 5 updated PK analysis set definition to indicate that both the PK concentration and PK parameter populations are the same Section 5.3.1.1 “Pharmacokinetic concentrations” - removed log-linear plots Section 5.3.1.2 “Pharmacokinetic parameters” was removed Section 6.1.1.1 “Primary Analysis” and Section 6.2.2.1 “Overall survival (key secondary)” - updated to indicate that ties will be handled using Exact option in SAS proc phreg Section 6.1.1.1 “Primary Analysis” and Section 6.2.2.1 “Overall survival (key secondary)” - removed summaries of Kaplan Meier estimates for time of follow-up for EFS or OS Section 6.1.1.1 “Primary Analysis”, Section 6.2.2.5 “Duration of Complete Response (participants with CIS as randomization only)”, Section 6.2.2.6 “Time to recurrence of Low-Grade disease”, and Section 6.2.2.8 “Time to Cystectomy” - updated time points for analyses Section 6.2.2.2 “Sensitivity analyses for event-free survival” - updated references to Tables 11 and 12. Section 6.2.2.2 “Sensitivity analyses for event-free survival” - added in COVID-19 sensitivity analyses Section 6.2.2.2 “Sensitivity analyses for event-free survival” - removed 2 sets of RMST analyses and updated text referring to three sets of analyses Section 6.2.2.3 “Sensitivity analyses for OS” - added in COVID-19 OS analysis Section 6.2.2.4 “Complete Response (participants with CIS at randomization only)” - added in COVID-19 analysis for C Section 6.2.2.5 “Duration of Complete Response (participants with CIS as randomization only)” - included alternative duration of CR analysis using ITT population

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 6.2.2.7 “Disease Specific Survival” - included DSS calculation and updated censoring language Section 6.2.2.7 “Disease Specific Survival” - for Table 18 included a footnote to clarify bladder cancer deaths Section 6.2.3 “Pharmacokinetic endpoints” - updated pharmacokinetic endpoint summaries referring to collection time and additional analyses Section 6.2.6 “Endpoints for immunogenicity data of PF-06801591” - Table 20 was updated and ADA and Nab analyses further described. Section 6.2.6 “Endpoints for immunogenicity data of PF-06801591” - ADA tier analyses was removed Section 6.3.1.4 “Descriptive Summary” - it was clarified that EQ-5D-5L is index and VAS Section 6.4 “Subset Analyses” - removed DoCR; included additional clarification of the handling of low numbers of participants and events. Section 6.5.1.1 “Demographic characteristics” - removed vital signs, physical exams, and weight references. Section 6.5.1.3 “Disease characteristics” - location of CIS at baseline was removed Section 6.5.1.4 “Prior anti-cancer therapies” - removed listing of anti-cancer surgeries and radiation therapy. Section 6.5.2.1 “Participant disposition” - added in “by reason” for discontinuations from treatment. Added in listings related to COVID-19 discontinuations or study disruptions. Section 6.5.2.2 “Protocol deviations” - added listing of protocol deviations related to COVID-19 Section 6.5.3 “Study treatment compliance and exposure” - clarified BCG dosing. Section 6.5.3.1 “Exposure to study drug” - removed period related duration of exposure for BCG. Section 6.5.3.2 “Dose reduction” - updated text for dose reductions of sasanlimab to note “partial dose” Section 6.5.3.4 “Dose delays” - clarified calculation for BCG dose delays. Section 6.5.5 “Subsequent anti-cancer therapies” - listings of anti-cancer therapies were removed Section 6.6.1.1 “All adverse events” - removed references to summaries for “any study drugs” and added in COVD-19 summary for AEs. Section 6.6.2 “Deaths” - added in COVID-19 death summary and listing flags. Section 6.6.4 “Other significant adverse events” - removed BCG summaries for BCG-related AEs leading to death, BCG-related AEs leading to discontinuation of all study drugs, and Serious BCG-related AEs Section 6.6.6 “Vital Signs” - removed Vital Signs section, Section 6.6.6 in SAP Version 2 is now “Electrocardiogram” Section 6.6.6 “Electrocardiogram” - reduced descriptive ECG summaries Section 6.6.8 “ECOG” - removed entire ECOG section Additional updates were made for consistency or clarity

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
3	09-Mar-2021	Section 2 “Introduction” - added text regarding BICR. Section 2.1 “Study Objectives, Endpoints, and Estimands” - updated error from “Nabs” to “Nabs” Section 3.4.5 “Baseline Characteristics” - removed “physical measurements”. Section 4 “Analysis Sets” - for Table 5 clarifications were added to the biomarker analysis set and immunogenicity analysis sets. Section 5.2.6 “Definition of start of new anti-cancer therapy” - clarified the surgery date to be based on positive biopsies other than low grade Ta or missing biopsies for a lesion. Section 5.2.8 “Standard derivations and reporting conventions” - removed “The integer part of the calculated age will be used for reporting purposes”. Section 5.2.10 “Adequate baseline disease assessment” - the definition was updated to remove cytology and cystoscopy requirements for patients with CIS at baseline. The cytology requirement of within 28 days from randomization was removed for participants without CIS at baseline and to allow missing and not done cytology to be considered adequate if cystoscopy is adequate. Section 5.2.11 “Adequate post-baseline disease assessment” - added in “(participants without CIS at randomization)” for low grade disease and clarified that post baseline disease assessment is based on the IODA eCRF page. Section 5.2.12 “Disease Assessments” - clarified the date to be used for each type of assessment and which assessment apply. Clarified that recurrence of low-grade disease and no evidence of disease is for participants without CIS at randomization only. Clarified recurrence of high-grade disease for participants with CIS at randomization. Section 6.1.1.1 “Primary analysis” - Table 11 added in “adequate” for “No adequate baseline”. Section 6.1.1.1 “Primary analysis” - Table 12 added in “adequate” for “No adequate baseline” and revised text to “EOS present OR all EOT disposition pages of all drugs in the treatment arm say participant will not continue into any”. Added the footnote “[b] EOS here refers to completion of follow-up disposition page.’’ Section 6.2.2 “Efficacy endpoints” - removed text regarding central pathology and independent review of imaging scans. Added in “using BICR results”. Section 6.2.2.1 “Overall survival (key secondary)” - in Table 13 updated lost to follow-up timing to 25 weeks from 24 weeks to match the long-term EFS assessment window. Section 6.2.2.2 “Sensitivity analysis for event free survival” - for the COVID-19 analysis added in excluding COVID-19 related deaths. Removed text regarding central pathology and independent review of imaging scans. Section 6.2.2.3 “Sensitivity Analysis for OS” - for the COVID-19 analysis added in the text “non-COVID-19 related deaths”. Section 6.2.2.4 “Complete Response (participants with CIS at randomization only)” - removed reference to inadequate baseline. Removed text regarding central pathology and independent review of imaging scans. Section 6.2.2.5 “Duration of Complete Response (participants with CIS at randomization only)” - updated the definition for the alternative DoCR. Removed text and referred to Table 14. Removed “persistence of CIS from the

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		summary of events. Removed text regarding central pathology and independent review of imaging scans. Section 6.2.2.6 “Time to Recurrence of Low-Grade Disease” - added in text “according to the “Oncology Biopsy” CRF page” for how low-grade Ta disease will be determined. In Table 17, added in “adequate” for “No adequate baseline” and revised text to “EOS present OR all EOT disposition pages of all drugs in the treatment arm say participant will not continue into any”. Added the footnote “[b] EOS here refers to completion of follow-up disposition page.’’ Section 6.2.2.7 “Disease specific survival” - in Table 18 updated lost to follow-up timing to 25 weeks from 24 weeks. Section 6.2.2.8 “Time to cystectomy” - updated censoring to be “death date or last date known to be alive”. In Table 19, updated lost to follow-up timing to 25 weeks from 24 weeks. Section 6.2.3 “Pharmacokinetic endpoints” - added the following test to the first bullet “will be excluded from”. Section 6.2.6 “Endpoints for immunogenicity data of PF-06801591” - added in “Incidence may also be presented graphically and, if ≥ 10% in a treatment arm, immunogenicity titer data will also be summarized and box plots of PK concentrations by ADA and Nab status will be presented.” Updates were made to Table 20 for Nab negative subject row and removed reference to 2.00 and 0.602 titer in the first four rows. Included a range for EOT samples from 21 to 35 days of last PF-06801591 dose. Section 6.5.1.1 “Demographics and baseline characteristics” - added in the words “and baseline” for clarification. Removed “<65” in the second bullet under “Age categories”. Section 6.5.1.3 “Disease characteristics” - included the T stage that is identified by biopsy according to worst stage: Ta, CIS, T1, T2, T3, T4 Section 6.5.1.4 “Prior anti-cancer therapies” - specified the CRF page name for “Prior Cancer Therapy”. Section 6.5.2.1 “Participant disposition” - removed summary of number of participants entering a period but not that are treated. Added in number of participants entering each BCG period and number of participants ongoing in follow-up. Section 6.5.3 “Study treatment compliance and exposure” - updated text regarding partial doses. Section 6.5.3.4 “Dose delays” - removed the delay of 1-3 days in Cycles 1-3 and 1-7 days in Cycles 4-25 and categorized these as no delays to be consistent with protocol dosing window. Section 6.5.4 “Concomitant medications and non-drug treatments” - the references to prior medications were removed. Section 6.6.6 “Electrocardiogram” - removed text regarding ECG summaries and listings. Also removed referenced to QtcLogP and added that additional corrections may be performed. Section 7.2.1 “Interim analysis for EFS” - deleted the word “vital”. Section 8 “REFERENCES”- added reference #18. Appendix 2 “List of Abbreviations” - removed QtcLogP reference Additional clarifications for consistency as needed were added.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
4	18-Nov-2021	Section 1 "VERSION HISTORY" - updated amendment number and date of amendment. Section 2 "INTRODUCTION" - added in primary analysis for Cohorts B1 and B2. Section 2.1 "Study Objectives, Endpoints, and Estimands" - added in and updated cohort descriptions; added in objectives and endpoints for Cohorts B1 and B2; removed "eg, IL-6, IFN-γ and/or tissue FoxP3" from endpoints in Table 2. Section 2.1.1 "Primary Estimands" - added in estimands for Cohorts B1 and B2; updated table references. Section 2.2 "Study Design" - updated study design description and added in sample size and study schema figure (Figure 2) for Cohorts B1 and B2 Section 3.1 "Primary Endpoint" - added in endpoints for Cohorts B1 and B2. Section 3.2.2 "Efficacy endpoints" - added in EFS as assessed by BICR for Cohort A; added "and by BICR" for CR and DoCR for Cohort A; removed "as assessed by investigator" from DoCR for Cohorts B1 and B2; added in secondary endpoints for Cohorts B1 and B2. Section 3.2.4 "Pharmacokinetic endpoints" - added in PK endpoint for Cohorts B1 and B2. Section 3.2.5 "Immunogenicity endpoints" - added in immunogenicity endpoints for Cohorts B1 and B2. Section 3.3 "Tertiary/Exploratory Endpoints" - removed "(eg. IL-6, IFN-γ and/or tissue FoxP3)". Section 3.4.1 "Study drug, study treatment and baseline definitions" - added in description of cohorts for B1 and B2 and further clarifications of cohorts; defined 'start date' for Cohort A and Cohorts B1 and B2; clarified sampling time for immunogenicity; clarified definition of baseline for immunogenicity. Section 3.5.1 "Adverse events" - added "(also refer to Appendix 1 for ir AEs)". Table 6 "Analysis Sets" - added in "in Cohort A" to FAS and PP definitions for Cohort A; added in FAS for Cohorts B1 and B2; removed "analysis sets will be defined separately by biomarker". Table 7 "Statistical Analyses by Analysis Set" - clarified the per protocol analyses for Cohort A only; added in PRO endpoint. Section 5.1.1 "Hypotheses and sample size determination" - added in Cohort B sample size and Table 8 for CR rate and 95% exact confidence interval for Cohort B1. Section 5.1.2 "Decision rules (Cohort A only)" - updated text regarding testing of OS after IA; updated boundaries for Arm B vs Arm C. Section 5.2 "General Methods" -updated Cohort A descriptions and added in description of Cohorts B1 and B2 Section 5.2.7 "Definition of on-treatment period" - added in reference to irAE appendix. Section 5.2.10 "Adequate baseline disease assessment" - added in adequate baseline definition for Cohorts B1 and B2 Section 5.2.11 "Adequate post-baseline disease assessment" - removed "IODA crf" and reference to Section 6.2.2.4; added reference to Section 5.2.12

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 5.2.12 “Disease Assessments” - clarified persistence of CIS for Cohort B1; changed “CIS at randomization” to “CIS only at baseline” for Cohort A for PD and “CIS at randomization” to “CIS at baseline” for No Evidence of Disease. Section 6 “ANALYSES AND SUMMARIES” - clarified that summaries will be separated by treatment arm in Cohort A and separately for Cohorts B1 and B2. Section 6.1.2 “Complete Response as assessed by BICR (Cohort B1)” - analysis of complete response added for Cohort B1. Section 6.1.3 “Event-free survival as assessed by investigator (Cohort B2)” - EFS analysis added for Cohort B2. Section 6.2.2 “Efficacy endpoints” - analyses updated for Cohort A and Cohorts B1 and B2. Section 6.2.2.1 “Overall survival” - added in OS definition for Cohorts B1 and B2. Table 20 “DoCR Censoring Reasons and Hierarchy” - removed footnote [a]. Section 6.2.2.5 “Duration of Complete Response” - updated duration of complete response for Cohort A and Cohort B1. Section 6.2.2.6 “Complete response at 12 months (Cohort B1 only)” - added in complete response at 12 months analysis for Cohort B1 Section 6.2.2.7 “Event-free survival (Cohort B1)” - added in EFS analysis for Cohort B1 Table 24 “DSS Censoring Reasons and Hierarchy” - updated footnote. Section 6.2.3 “Pharmacokinetic endpoints” - updated PK analyses regarding EOT. Section 6.2.5 “Biomarker endpoints” - updated text to be consistent with PD-L1 endpoint. Section 6.2.6 “Endpoints for immunogenicity data of PF-06801591” - clarified summarization of data and timing of injection for pre-dose; updated Table 26 definitions. Section 6.3.1.3 “Instrument completion rates” - removed text referring to analysis. Section 6.3.1.4 “Descriptive Summary” - removed SE from descriptive summary; changed SE to 95% CI for line chart; added in cumulative patient disposition figure. Section 6.3.1.5 “Health Status” - removed 2-sided 95% CI. Section 6.3.2 “Exploratory Biomarker endpoints” - removed “(eg. IL-6, IFN-γ and/or tissue FoxP3)”. Section 6.4 “Subset Analyses” - clarified by investigator and BICR for EFS and CR; added in subset analyses for Cohort B1. Section 6.5.1.3 “Disease characteristics” - added in reason participant did not undergo cystectomy; clarified baseline data for Cohort B1 is based on the BICR data. Section 6.5.1.4 “Prior anti-cancer therapies” - added in for Cohorts B1 and B2: prior BCG strain, number of prior BCG doses, prior BCG setting, and prior BCG failure. Section 6.5.3.4 “Dose Delays” - added in details for Cohort B. Section 6.6.1 “Adverse events” - removed BCG related AEs.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 6.6.1.1 “All adverse events” - removed BCG related summaries. Section 6.6.2 “Deaths” - updated death summaries to match CRF. Section 6.6.4 “Other significant adverse events” - removed BCG related summaries; added in that AEs of special interest are irAEs. Appendix 1 “Immune-Related Adverse Events” - added in Step 4. Clarifications were made throughout the SAP to identify Cohort A versus Cohorts B1 and B2 descriptions and summaries. Updates were made throughout the SAP to add ‘start date’ when referring to date of randomization for Cohort A or date of first dose for Cohorts B1 and B2. References to tables were updated throughout. Updates to footnotes in tables were made where necessary. Removed abbreviations that no longer apply from Appendix 2. Additional clarifications for consistency as needed were added.
5	15-Feb-2022	Section 2.1 “Study Objectives, Endpoints, and Estimands” - in Table 3 added in “C_{max} (Cohort B2 only)” to the PK endpoint. Section 2.2 “Study Design” - updated Figure 2 for Cohorts B1 and B2 cycles and dosing. Section 3.2.4 “Pharmacokinetic endpoints” - added in C_{max} to the text and Table 4. Section 4 “ANALYSIS SETS (POPULATION FOR ANALYSES)” - in Table 6 added in details for Immunogenicity analysis set that participants must receive at least one dose of sasanlimab. Section 5.2.12 “Disease Assessments” - recurrence of disease definition was clarified. Section 6.2.3 “Pharmacokinetic endpoints” - summary and plot of C_{max} were added and Cycles were added for C_{trough}. Section 6.2.6 “Endpoint for immunogenicity data of PF-06801591” - added in Cycles and Days for blood samples. Section 6.3.1.5 “Heath Status” - added in Cohorts B1 and B2. Section 6.4 “Subset Analyses” - IMURON-VAC was changed to BCG-1. Section 6.4 “Subset analysis” - “at induction” was added to “BCG strain” to clarify which strain is used for analysis. Section 6.5.2.1 “Participant disposition” - removed COVID-19 listing. Section 6.5.2.2 “Protocol deviations” - moved listing for COVID-19 from Section 6.2.2.2. Section 6.5.3 “Study treatment compliance and exposure” - for Cohort B2 the dose was changed to SC Q6W at 600mg. Section 6.5.3.2 “Dose reductions” - for Cohort B2 the dose was changed to SC Q6W at 600mg; text was also updated to reflect that a 300mg dose in Cohort B2 is a partial dose. Section 6.5.3.3 “BCG retention interruption” - text was updated to clarify the timing cut offs to be summarized. Section 6.5.3.4 “Dose delays” - updates were made to account for Cohort B2 where the last cycle is Cycle 17 and cycle length is 42 days and text in the example was modified to accurately reflect how delays would be determined.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 6.6.1.1 “All adverse events” - text was modified for the COVID-19 related TEAS that the data will come from the Adverse Event CRF; the term “related” was removed when referring to COVID-19. Section 6.1.1.1 “Primary analysis” - clarifications were made in Table 12 Section 6.2.2.7 “Event-free survival (Cohort B1) - clarifications were made in Table 21. Appendix 2 “Listing of Abbreviations” - added C_{max} and removed T_{max}. Additional clarifications for consistency as needed were added.
6	6-Jan-2023	Section 2 “Introduction” - text was added in to clarify the decision to close enrollment to Cohort B. Section 2.1 “Study Objectives, Endpoints, and Estimands” - added in text before Table 3 that the Cohort B objectives are no longer required. Section 2.1.1 “Primary Estimands” - removed Cohort B estimands. Section 2.2 “Study Design” - text was added in to clarify decision to close enrollment to Cohort B. Section 2.2 “Figure 2” - Cohort B schema was updated. Section 3.2.2 “Secondary Endpoints” - removed OS, CR at 12 months, DoCR, and Time to Cystectomy for Cohort B as data for these endpoints will not be presented. Section 3.2.3 “Patient reported outcomes (Cohort A)” - removed summaries for Cohort B. Section 3.2.4 “Pharmacokinetic Endpoint” - removed Cohort B endpoints. Section 3.3 “Tertiary/Exploratory Endpoints” - removed Cohort B from tertiary/exploratory endpoints. Section 5.1.1 “Hypotheses and Sample Size Determination” - text was added in to clarify decision to close enrollment to Cohort B. Section 6.1.1 “Primary Analysis” - added text that a Kaplan-Meier plot for EFS will be generated. Section 6.1.2.1 “Primary Analysis” - removed summaries of CR and added text that only a listing will be provided for CR for Cohort B1. Section 6.1.3.1 “Primary Analysis” - removed text regarding summaries for EFS and added text that only a listing will be provided for EFS for Cohort B2. Section 6.2.2 “Efficacy Endpoints” - removed references to analyses for Cohort B and added text indicating that only listing will be provided for secondary efficacy endpoints for Cohort B. Section 6.2.2.1 “Overall Survival (Cohort A only)” - added in text that a Kaplan Meier plot will be generated and updated the section to be for Cohort A only. Section 6.2.2.4 “Complete Response (participants with CIS at randomization)(Cohort A only)” - modified the analysis of CR to be a stratified analysis of risk difference. Section 6.2.2.5 “Duration of Complete Response (Cohort A only)” - summary of DoCR was removed for Cohort B1. Old Section 6.2.2.6 “Complete Response at 12 months” - this section was deleted. Section 6.2.2.6 “Event-free survival (Cohort B1)” - specified that only a listing will be provided for EFS. Section 6.2.2.7 “Time to Recurrence of Low-Grade Disease” - added in text regarding analysis of endpoint using Cox PH and KM plot. Section 6.2.2.8 “Disease Specific Survival” - added in KM plot and a listing of DSS.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 6.2.2.9 “Time to Cystectomy (Cohort A only)” - added in KM plot and removed Cohort B. Section 6.2.3 “Pharmacokinetic Endpoints” - specified that Cohort B concentration data will be provided in listings only. Section 6.2.6 “Endpoints for immunogenicity data of PF-06801591” - removed text pertaining to Cohort B summaries; specified that Cohort B data will be provided in listings only; updated the reporting of ADA and Nab titers and included summaries for combined across arms A and B. Section 6.3.1 “PRO Endpoints” - removed summary of PRO data for Cohort B. Section 6.3.2 “Exploratory biomarker endpoints” - removed summaries for Cohort B. Section 6.4 “Subset Analyses” - removed Cohort B subgroup analyses and further clarified Cohort A analyses regarding CIS at baseline (IRT vs CRF vs BICR). Section 6.5.1 “Baseline analyses” - specified that Cohort B baseline data will be provided in listings only. Section 6.5.2 “Study conduct and participant disposition” - specified that Cohort B data will be provided in listings only. Section 6.5.3.1 “study treatment compliance and exposure” - specified that Cohort B exposure data will be provided in a listing only. Section 6.5.3.2 “Dose reductions” - specified summaries are only for Cohort A. Section 6.5.3.3 “BCG retention interruption” - fixed the interruption categories. Section 6.5.3.4 “Dose Delays” - specified dose delay summaries are for Cohort A only. Section 6.5.4 “Concomitant medications and non-drug treatment” - removed Cohort B. Section 6.5.5 “Subsequent anti-cancer therapies” - specified that Cohort B data will only be provided in listings. Section 6.6.1.1 “All adverse events” - added in “ISRs related to PF-06801591” to the summaries; specified selected summaries are for Cohort A only; specified that a listing of all AEs will be provided. Section 6.6.1.2 “Adverse events leading to discontinuation of study drug” - specified selected summaries are for Cohort A only. Section 6.6.1.3 “Adverse events leading to interruption of study treatment” - removed Cohort B summaries. Section 6.6.4.1 “Immune related adverse events” - added in “worst Grade” and “any Grade” to selected summaries; specified certain summaries for Cohort A only; removed “irAEs, by Cluster and PT”. Section 6.6.4.2 “Injection site reactions related to PF-06801591”- this section was added. Section 6.6.5 “Laboratory Data” - specified that lab data for Cohort B will be provided in listings only; updated text regarding shift table. Updated “gender” to “sex” throughout the SAP. Additional updates were made for consistency with the protocol and throughout the SAP.
7	16-Aug-2024	Section 2 “Introduction” - added summary statement of protocol amendment 5. Section 2.1 “Study objectives, Endpoints, and Estimands”, Section 3.1 “Primary Endpoint”, and Section 3.2.2 “Efficacy Endpoints” - the primary

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		objective of 'To demonstrate that PF-06801591 + BCG (induction) is superior to BCG (induction and maintenance) in prolonging EFS in participants with high-risk NMIBC' was changed to a key secondary objective, and the endpoint and estimand associated with this objective were changed to key secondary endpoint and key secondary estimand accordingly. Section 5.1. "Decision rules (Cohort A only)" - removed IA2 for EFS and updated stopping boundaries for OS. Section 5.2.10. "Adequate baseline disease assessment" - for EFS by the investigator, changed the allowed baseline cystoscopy assessment window from 28 days to 35 days prior to and including the date of randomization; added definition of adequate baseline disease assessment for EFS by BICR. Section 6.1. "Event-free analysis as assessed by investigator (Cohort A only)" and Section 6.2 "Efficacy endpoints" - Changed EFS for Arm B vs Arm C from primary endpoint to key secondary endpoint. Section 6.2 "Efficacy endpoints" - added sensitivity analysis for EFS by the investigator for possibly informative censoring. Section 6.5.4. "Concomitant medications and non-drug treatment" - added the listing of concomitant medications. Section 7.3. "Subgroup Analysis" - clarified that T1 disease status at baseline for subgroup analysis of EFS by BICR is based on BICR assessment. Section 8. "Interim Analysis" - removed IA2 for EFS; clarified that although this study is an open-label randomized study, the treatment is unblinded on participant level, the aggregate/cumulative data summaries by treatment arm will be unavailable to the study team and the external investigators until the database snapshot for the primary analysis; updated boundary for the final efficacy analysis for OS. Additional updates were made for consistency with the protocol and throughout the SAP
8	10-Dec-2024	Added Section 2.1 "Modifications to the Analysis Plan Described in the Protocol". Section 3.2.2. "Efficacy Endpoints" - removed summary of CR and DoCR per BICR, changed the definition of time to recurrence of low-grade disease to exclude death as event; removed EFS as assessed by BICR for Cohort B1. Section 5.2.10. "Adequate baseline disease assessment" - modified the definition of adequate baseline disease assessment for EFS per BICR to follow the same definition of adequate baseline disease for EFS per the investigator. Section 6.1.1. "Event-free survival as assessed by the investigator" - added summary of the type of PD event for EFS per the investigator. Section 6.2.2.2. "Overall survival" – modified the definition of censoring reason of "lost to follow-up". Section 6.2.2.3. "Sensitivity analyses for event-free survival" - added summary of BICR oncology evaluation of investigator-assessed EFS events.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
		Section 6.2.2.5. "Complete response (participants with CIS at randomization)" - removed the summary of CR per BICR; added summary of BICR biopsy evaluation of investigator-assessed CR. Section 6.2.2.7. "Event-free survival (Cohort B1)" – removed EFS by BICR. Section 6.2.2.8 "Time to recurrence of low-grade disease" modified the definition to exclude death as event. Section 6.2.2.9. "Disease specific survival" – modified the definition of censoring reason of "lost to follow-up". Section 6.4. "Subset Analyses" - removed subgroup analysis of EFS by BICR, and subgroup analysis of CR by BICR. Appendix 1: changed "Immune-mediated endocrinopathies" to "Immune-mediated endocrinopathies: Thyroid disorders". Additional editorial updates were made for consistency with the protocol and throughout the SAP.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study B8011006. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

A separate SAP will cover the periodic safety review by the External Data Monitoring Committee (E-DMC).

Statistical analyses will be performed using cleaned electronic case report form (eCRF) data as well as non-CRF data (i.e., PRO, pharmacokinetic (PK) concentration, anti-drug antibody (ADA), neutralizing antibody (Nab), biomarker data, and Blinded Independent Central Review (BICR) of pathology, imaging, and oncology review data).

As of 31 August 2022, enrollment in Cohorts B1 and B2 was closed by the sponsor for business strategy reasons. This decision was not made due to any safety concerns, new emerging data, or regulatory interaction. Due to the decision to close enrollment in Cohorts B1 and B2, the Cohort B1 and B2 study objectives are no longer required.

As of 17 June 2024, the Protocol Amendment 5 revised the Cohort A study analysis plan for the primary and key secondary objectives. The rationale for this change is based on the observation from pooled treatment groups that EFS events are accumulating at a much slower rate than originally projected, together with an increasing rate of participants dropping out from EFS observation. The IA2 will be removed, and the FA for EFS based on a calendar-based data cut-off date will be conducted. Randomization to Cohort A was completed on 16 November 2021 and the EFS FA data cut-off date will be set to allow for approximately 3 years of follow-up after the last participant was randomized. The treatment period was completed on 16 November 2023. Based on the blinded observed EFS events accrual rate and assuming equal event rate across three arms, approximately 261 EFS events are expected to have occurred across all three arms (approximately 174 events for each two-arm comparison) by the EFS FA data cut-off date.

Data for Cohorts B1 and B2 will be reported at the time of PCD for Cohort A.

2.1. Modifications to the Analysis Plan Described in the Protocol

Changes	Protocol	SAP
Cohort A: Complete response as assessed by BICR	Specified	Removed
Cohort A: Duration of CR as assessed by BICR	Specified	Removed

The removal of the analysis of complete response (CR) and duration of CR (DoCR) by BICR described in the protocol is based on the infeasibility of these analyses without large amount of data imputations based on investigator assessment. As per protocol, on-study biopsy is not required at each disease assessment visit, therefore central pathology assessment of CR could not be performed at each visit, and independent oncology review of CR is not feasible.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

In addition, to adhere to the requirements from the FDA with reference to independent evaluation of EFS events and CR, the following summaries were added:

- BICR oncology evaluation of investigator-assessed EFS events (Section 6.2.2.3).
- BICR biopsy evaluation of investigator-assessed CR (Section 6.2.2.5).

2.2. Study Objectives, Endpoints, and Estimands

The treatment arms/cohorts to assess the objectives and endpoints are:

BCG Naïve Cohort (Cohort A)

- Arm A: PF-06801591 + BCG (IND and MNT)
- Arm B: PF-06801591 + BCG (IND only)
- Arm C: BCG only (IND and MNT)

where IND=induction and MNT=maintenance

BCG-Unresponsive Cohorts (Cohorts B1 and B2)

- Cohort B1; CIS with or without papillary disease: PF-06801591 only
- Cohort B2; Papillary disease only: PF-06801591 only

Table 2. Study Objectives and Endpoints (Cohort A)

Objectives	Endpoints
Primary	
• To demonstrate that PF-06801591 + BCG (induction and maintenance) is superior to BCG (induction and maintenance) in prolonging EFS in participants with high-risk non-muscular invasive bladder cancer (NMIBC).	• EFS as assessed by the investigator
Key Secondary	
• To demonstrate that PF-06801591 + BCG (induction) is superior to BCG (induction and maintenance) in prolonging EFS in participants with high-risk NMIBC.	• EFS as assessed by the investigator
• To demonstrate that PF-06801591 + BCG (induction and maintenance) is superior to BCG (induction and maintenance) in prolonging overall	• OS

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 2. Study Objectives and Endpoints (Cohort A)

Objectives	Endpoints
survival (OS) in participants with high-risk NMIBC.	
• To demonstrate that PF-06801591 + BCG (induction) is superior to BCG (induction and maintenance) in prolonging OS in participants with high-risk NMIBC.	• OS
Secondary	
• To evaluate the complete response (CR) rate of PF-06801591 + BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with carcinoma in situ (CIS) at randomization.	• CR as assessed by the investigator (in participants with CIS at randomization)
• To evaluate the duration of CR (DoCR) of PF-06801591 + BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with CIS at randomization	• DoCR for participants with CR as assessed by the investigator (in participants with CIS at randomization)
• To evaluate the time to recurrence of low grade disease of PF-06801591 + BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with high-risk NMIBC.	• Time to recurrence of low-grade disease as assessed by the investigator
• To evaluate the time to cystectomy of PF-06801591 + BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with high-risk NMIBC.	• Time to cystectomy
• To evaluate the disease-specific survival (DSS) of PF-06801591 + BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with high-risk NMIBC.	• DSS as assessed by the investigator
• To evaluate the overall safety profile of PF-06801591+BCG (induction and maintenance or induction) and BCG (induction and maintenance) in participants with high-risk NMIBC.	• Adverse events (AEs) as characterized by type, severity (as graded by National Cancer Institute Common Terminology Criteria for Adverse Events [NCI CTCAE] v5.0), timing, seriousness, and relationship to study therapy. Laboratory

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 2. Study Objectives and Endpoints (Cohort A)

Objectives	Endpoints
	abnormalities as characterized by type, severity (as graded by NCI CTCAE v5.0), and timing
• To assess the effects of PF-06801591+BCG (induction and maintenance or induction) and BCG (induction and maintenance) on patient-reported health-related quality of life in participants with high-risk NMIBC.	• Health-related quality of life as measured by: 1) EORTC QLQ-C30 (European Organization for Treatment of Cancer Quality of Life Questionnaire), 2) EORTC QLQ-NMIBC24, 3) PTAB (Patient Treatment Administration Burden Questionnaire)
• To characterize the PK of PF-06801591+BCG (induction and maintenance or induction).	• C_{trough} of PF-06801591 when in combination with BCG (induction and maintenance or induction); Arms A and B only.
• To evaluate the immunogenicity of PF-06801591+BCG (induction and maintenance or induction).	• ADAs; Nabs of PF-06801591 when in combination with BCG (induction and maintenance or induction); Arms A and B only.
• To evaluate PD-L1 expression in pre-treatment tumor tissue that may aid in the identification of a participant subpopulation most likely to benefit from treatment with PF-06801591 + BCG (induction and maintenance or induction).	• Tumor sample biomarker status based on PD-L1 expression (high or low)
Tertiary/Exploratory	
• To assess the effects of PF-06801591+BCG (induction and maintenance or induction) and BCG (induction and maintenance) on patient reported outcomes (PROs) in participants with high-risk NMIBC.	• Health state utilities as measured by the Euro Qol 5 Dimension (EQ-5D-5L) & visual analog scale (VAS) • Overall assessment of disease severity as measured by the Patient Global Impression of Severity (PGIS) • Overall assessment of change as measured by the Patient Global Impression of Change (PGIC) • Patient satisfaction as measured by the Treatment Satisfaction Questionnaire (TSQ)
• To explore the predictive and pharmacodynamic characteristics of	• Peripheral blood and additional tumor tissue biomarkers consisting of the levels

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 2. Study Objectives and Endpoints (Cohort A)

Objectives	Endpoints
peripheral blood and additional tumor tissue biomarkers that may be relevant to the mechanism of action of, or resistance to, treatment with PF-06801591 + BCG (induction and maintenance or induction), including but not limited to biomarkers related to anti-tumor immune response or target modulation.	of cells, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or proteins that may be related to anti-tumor immune response or disease progression

Due to the decision to close enrollment in Cohorts B1 and B2, the Cohort B1 and B2 study objectives are no longer required.

Table 3. Study Objectives and Endpoints (Cohorts B1 and B2)

Primary Objectives	Primary Endpoints
• To estimate the CR rate of PF-06801591 in participants with BCG-unresponsive CIS (Cohort B1 only)	• CR as assessed by the BICR
• To evaluate the EFS of PF-06801591 in participants with BCG-unresponsive NMIBC (Cohort B2 only)	• EFS as assessed by the investigator
Secondary Objectives	**Secondary Endpoints**
• To evaluate the duration of CR of PF-06801591 in participants with BCG-unresponsive CIS (Cohort B1 only).	• Duration of CR for participants with CR as assessed by the BICR
• To estimate the CR rate of PF-06801591 at 12 months in participants with BCG-unresponsive CIS (Cohort B1 only)	• CR at 12 months as assessed by the BICR
evaluate the EFS of PF-06801591 in participants with BCG unresponsive CIS (Cohort B1 only)	• EFS as assessed by the investigator
• To evaluate the time to cystectomy of PF 6801591 in participants with BCG-unresponsive NMIBC.(Cohorts B1 and B2).	• Time to cystectomy
• To evaluate OS of PF-06801591 in participants with BCG-unresponsive	• OS

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 3. Study Objectives and Endpoints (Cohorts B1 and B2)

NMIBC treated with PF-06801591 (Cohorts B1 and B2).	
• To evaluate the overall safety of PF-06801591 in participants with BCG-unresponsive NMIBC (Cohorts B1 and B2).	• AEs as characterized by type, severity (as graded by NCI CTCAE v5.0, timing, seriousness, and relationship to study therapy. Laboratory abnormalities as characterized by type, severity (as graded by NCI CTCAE v5.0), and timing
• To assess the effects of PF-06801591 on patient -reported health-related quality of life in participants with BCG-unresponsive NMIBC (Cohorts B1 and B2).	• Health-related quality of life as measured by: 1) EORTC QLQ-C30, 2) EORTC QLQ-NMIBC24, 3) PTAB
• To evaluate PD-L1 expression in pretreatment tumor tissue that may aid in the identification of a subpopulation of participants with BCG unresponsive NMIBC most likely to benefit from PF-06801591 (Cohorts B1 and B2).	• PD-L1 expression
• To characterize the PK and immunogenicity of PF-06801591 single agent in participants with BCG-unresponsive NMIBC (Cohorts B1 and B2).	• C_{trough}, C_{max} (Cohort B2 only), ADAs and Nabs following PF-06801591 single agent.
Tertiary/Exploratory Objectives	**Tertiary/Exploratory Endpoints**
• To assess the effects of PF-06801591 on patient reported outcomes (PROs) in participants with BCG unresponsive NMIBC (Cohorts B1 and B2).	• Health state utilities as measured by the EQ-5D-5L & VAS • Overall assessment of disease severity as measured by the PGIS • Overall assessment of change as measured by the PGIC • Patient satisfaction as measured by the TSQ
• To explore the predictive and pharmacodynamic characteristics of peripheral blood and additional tumor tissue biomarkers that may be relevant to the mechanism of action of, or resistance to, treatment with PF-06801591,	• Peripheral blood and additional tumor tissue biomarkers consisting of the levels of cells, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or proteins that may be related to anti-tumor immune response or disease progression.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 3. Study Objectives and Endpoints (Cohorts B1 and B2)

including but not limited to biomarkers related to anti-tumor immune response or target modulation (Cohorts B1 and B2).	

2.2.1. Primary estimand (Cohort A only)

Primary Estimand (EFS for Arm A vs Arm C): treatment effect, estimated based on data from all randomized participants, of Arm A on EFS compared to Arm C from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death regardless of tolerability, duration of study treatment, or initiation of subsequent anti-cancer therapy. The date of the event (event as defined in Table 10) is the date of disease assessment documenting recurrence of high-grade disease, progression of disease, persistence of CIS (applicable only to participants with CIS at randomization), or death, whichever occurs earlier. The date of persistence of CIS is the earliest date when persistence of CIS is observed, if CR is not observed after re-induction.

- Variable: EFS defined as the time from randomization until recurrence of high-grade disease, progression of disease, persistence of CIS or death due to any cause, whichever occurs first.
- Censoring: see Table 11.
- Population-level summary measure: hazard ratio for EFS including all randomized participants.

2.2.2. Secondary estimands (Cohort A only)

Key Secondary Estimand (EFS for Arm B vs Arm C): treatment effect, estimated based on data from all randomized participants, of Arm B on EFS compared to Arm C from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death regardless of tolerability, duration of study treatment, or initiation of subsequent anti-cancer therapy. The date of the event (event as defined in Table 10) is the date of disease assessment documenting recurrence of high-grade disease, progression of disease, persistence of CIS (applicable only to participants with CIS at randomization), or death, whichever occurs earlier. The date of persistence of CIS is the earliest date when persistence of CIS is observed, if CR is not observed after re-induction.

- Variable: EFS defined as the time from randomization until recurrence of high-grade disease, progression of disease, persistence of CIS or death due to any cause, whichever occurs first.
- Censoring: see Table 11.
- Population-level summary measure: hazard ratio for EFS including all randomized participants.

Key Secondary Estimand (OS): treatment effect, estimated based on data from all randomized participants, of each experimental arm (Arm A and Arm B) on OS compared to

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Arm C regardless of tolerability, duration of study treatment, initiation of subsequent anti-cancer therapy or participant's request to discontinue study procedures.

- Variable: OS
- Censoring: data for participants not known to have died are censored at the time of last contact.
- Population-level summary measure: hazard ratio for OS, including all randomized participants.

2.2.3. Additional estimands (Cohort A only)

Supportive Estimand 1 (EFS): treatment effect, estimated based on data from all randomized participants, of each experimental arm (Arm A and Arm B) on EFS compared to Arm C from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death, regardless of tolerability, duration of study treatment or initiation of subsequent anti-cancer therapy. The date of the event (event as defined in Table 10) is the date of disease assessment documenting recurrence of high-grade disease, progression of disease, date of randomization for participants with persistent CIS (applicable only to participants with CIS at randomization), or death, whichever occurs earlier.

- Variable: EFS defined as the time from randomization until recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause, whichever occurs first.
- Censoring: see Table 11
- Population-level summary measure: hazard ratio for EFS including all randomized participants.

Supportive Estimand 2 (EFS): treatment effect of each experimental arm (Arm A and Arm B) on EFS compared to Arm C from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death regardless of tolerability, duration of study treatment, or initiation of subsequent anti-cancer therapy. Data from randomized participants who do not meet per-protocol criteria as defined below are excluded. The date of the event (event as defined in Table 10) is the date of disease assessment documenting recurrence of high-grade disease, progression of disease, persistence of CIS (applicable only to participants with CIS at randomization), or death, whichever occurs earlier. The date of persistence of CIS is the earliest date when persistence of CIS is observed, if CR is not observed after re-induction.

- Variable: EFS defined as the time from randomization until recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause, whichever occurs first.
- Censoring: see Table 11
- Population-level summary measure: hazard ratio for EFS excluding randomized participants who did not receive at least 1 dose of study drug, did not meet inclusion criteria 2 or 3, or met exclusion criteria 1 or 2.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

2.3. Study Design

This is a phase 3, multinational, randomized, open-label, three parallel-arm study of PF-06801591, an anti-PD-1 Antibody, in combination with Bacillus Calmette-Guerin (BCG induction with or without BCG maintenance) versus BCG (induction and maintenance) in participants with high-risk, BCG naïve non-muscle invasive bladder cancer (Cohort A) or PF-06801591 as single agent in participants with BCG-unresponsive NMIBC (Cohorts B1 and B2).

In Cohort A, a total of approximately 999 participants (including a minimum of 250 participants with CIS) will be randomized in a 1:1:1 ratio to one of 3 treatment arms below.

- Arm A: PF-06801591 + BCG (induction and maintenance period)
- Arm B: PF-06801591 + BCG (induction period only)
- Arm C: BCG only (induction and maintenance period)

Randomization will be stratified by the presence of CIS (yes vs no) and geography (US vs Western Europe and Canada vs Rest of World [ROW]).

In Cohort B1, approximately 110 participants with BCG-unresponsive CIS (with or without concomitant recurrent Ta/T1 disease) will be enrolled.

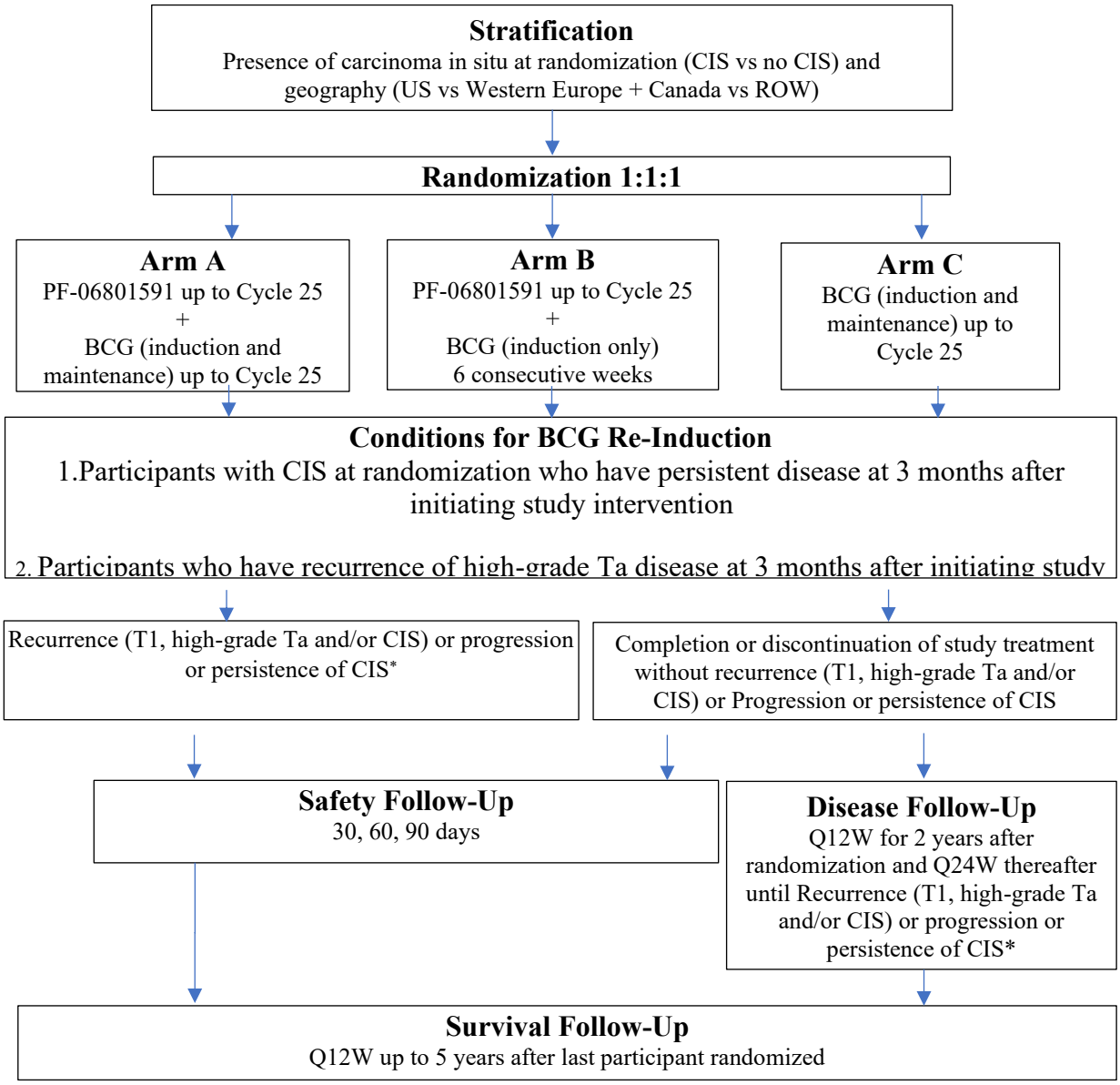
In Cohort B2, approximately 50 participants with BCG-unresponsive recurrent high-grade Ta/T1 only will be enrolled.

Participants discontinued from Cohort A are not permitted to enroll into Cohort B1 or Cohort B2.

As of 31 August 2022, enrollment in Cohorts B1 and B2 was closed by the sponsor for business strategy reasons. This decision was not made due to any safety concerns, new emerging data, or regulatory interaction. Participants that have already been enrolled into Cohorts B1 and B2 may continue treatment and study procedures and assessments per the Cohorts B1 and B2 Schedule of Activities in protocol amendment 5.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

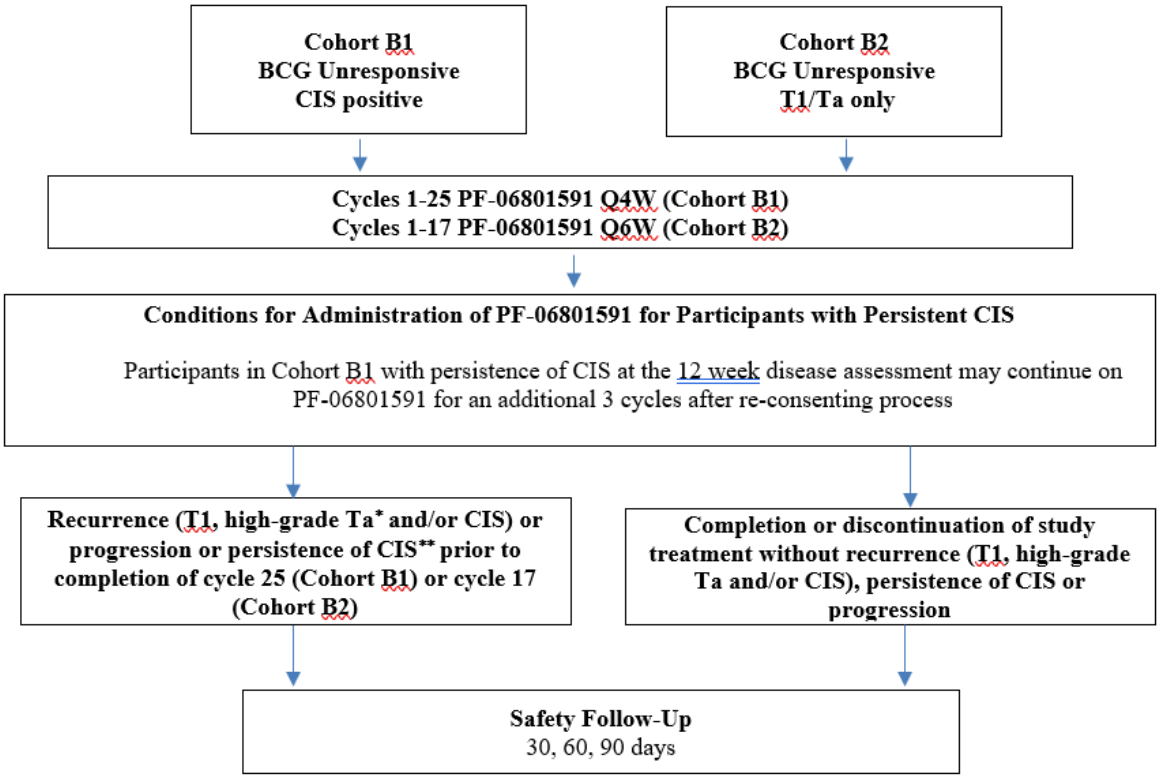
Figure 1. Study Design Schema (Cohort A)



* For participants with CIS at randomization
CIS: carcinoma in situ, BCG: Bacillus Calmette-Guerin; ROW: rest of world; US: United States

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Figure 2. Study Design Schema (Cohorts B1 and B2)



*See Section 4.1 for additional details
** For participants in Cohort B1 with CIS at registration

CIS: carcinoma in situ

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

Cohort A

- EFS for Arm A vs Arm C as assessed by the investigator

  EFS is defined as time from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause.

Cohort B1

- CR as assessed by the BICR (in participants with CIS at baseline as assessed by the BICR)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

CR is defined as histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy or negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or non-malignant tissue.

Cohort B2

- EFS as assessed by the investigator

  EFS is defined as time from date of initiation of study intervention (i.e. first dose) to the earliest of recurrence of high-grade disease, progression of disease, or death due to any cause.

3.2. Secondary Endpoints

3.2.1. Safety endpoints

- AEs as characterized by type, severity (as graded by NCI CTCAE v5.0), timing, seriousness, and relationship to study therapy.

  AEs will be graded by the investigator according to the CTCAE v5.0 and coded using the Medical Dictionary for Regulatory Activities (MedDRA).

- Laboratory abnormalities as characterized by type, severity (as graded by NCI CTCAE v5.0), and timing.

3.2.2. Efficacy endpoints

Cohort A

- Key Secondary: EFS for Arm B vs Arm C as assessed by the investigator

EFS is defined as time from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause.

- Key Secondary: OS

  OS is defined as the time from the date of randomization to the date of death due to any cause.

- EFS as assessed by the BICR

  EFS is defined as time from randomization to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause.

- CR as assessed by the investigator (in participants with CIS at randomization)

  CR is defined as histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy or negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or non-malignant tissue

- DoCR for participants with CR as assessed by the investigator (in participants with CIS at randomization)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

DoCR is defined as the time from the first documentation of CR to the date of an EFS event for participants with CR.

- Time to recurrence of low-grade disease as assessed by the investigator

  Time to recurrence of low-grade disease is defined as the time from randomization to the date of first documentation of recurrence of low-grade disease.

- Time to cystectomy

  Time to cystectomy is defined as time from randomization to cystectomy.

- DSS as assessed by the investigator

  DSS is defined as the time from randomization to death resulting from bladder cancer, as assessed by the investigator.

Cohorts B1 and B2

- CR as assessed by the investigator (in participants with CIS at baseline; Cohort B1 only)

  CR is defined as histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy or negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or non-malignant tissue.

- EFS as assessed by the investigator (Cohort B1 only)

  EFS is defined as time from date of initiation of study intervention (i.e. first dose) to the earliest of recurrence of high-grade disease, progression of disease, persistence of CIS, or death due to any cause.

3.2.3. Patient reported outcomes (Cohort A)

- Health-related quality of life as measured by: 1) EORTC QLQ-C30, 2) EORTC QLQ-NMIBC24, 3) PTAB

  The EORTC QLQ-C30 is a 30-question survey, which can be grouped into 5 functional domain subscales, including a physical functioning subscale, a role functioning subscale, an emotional functioning subscale, a cognitive functioning subscale and a social functioning subscale. Higher scores on the functional domains are indicative of higher levels of functioning. Oncology related symptoms assessed by the EORTC QLQ-C30 include fatigue (3 items), pain (2 items), nausea and vomiting (2 items), and dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial impact (1 item each). Higher scores are reflective of a greater presence of symptoms.

  The NMIBC24 has 24 items which can be grouped into 6 subscales: urinary symptoms (7 items), malaise (2 items), future worries (4 items), bloating/flatulence (2 items), sexual functioning (2 items), and male sexual issues (2 items). The NMIBC24 also assesses intravesical treatment, female sexual issues, sexual intimacy, risk of contaminating a partner, and sexual enjoyment (1 item each). Higher scores indicate greater impairment, with the exception of the sexual function and sexual enjoyment items, where higher scores indicate better function.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The PTAB questionnaire is a 2-item PRO designed to assess, from the participant perspective, any pain associated with the treatment administration and the burden of the amount of time required to complete the treatment administration procedures (1 item each).

3.2.4. Pharmacokinetic endpoints

- C_{trough} of PF-06801591 when in combination with BCG (induction and maintenance or induction); Arms A and B only in Cohort A.

  C_{trough} will be reported after single dose and at steady state.

Table 4. PK Parameters to be Determined for PF-06801591

Parameter	Definition	Method of Determination
C_{trough}	Predose/trough concentration	Observed directly from data

3.2.5. Immunogenicity endpoints

- Anti-drug antibodies (ADAs); neutralizing antibodies (Nabs) of PF-06801591 when in combination with BCG (induction and maintenance or induction). Arms A and B only in Cohort A.

3.2.6. Biomarker endpoints

- Tumor sample biomarker status based on PD-L1 expression (high or low).

Table 5. Biomarker Definition and Determination

Parameter	Definition	Method of Determination
PD-L1 expression	The number of PD-L1 positive cells and/or qualitative assessment of PD-L1 staining on tumor and immune cells in regions of interest that are defined by tumor cell morphology.	Pathologist, assisted by image analysis

3.3. Tertiary/Exploratory Endpoints

The following endpoints apply to Cohort A.

- Health state utilities as measured by the EQ-5D-5L & VAS

  The EQ-5D-5L is a 6-item patient-completed questionnaire designed to assess health status in terms of a single index value or utility score[1]. There are 2 components, a Health State Profile which has individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and a VAS in which participants rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for the creation of a single summary score. Overall scores range from 0 to 1, with lower scores representing a higher level of dysfunction.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- Overall assessment of disease severity as measured by the PGIS

  The PGIS is a single-item PRO designed to facilitate an anchor-based methodology for establishing meaningful within person change (MWPC) of other PRO-based endpoints such as the EORTC QLQ-C30 and NMIBC24. There are 4 response options, ranging from “none” to “severe”.

- Overall assessment of change as measured by the PGIC

  The PGIC is a single-item PRO designed to assess the participant’s overall impression of the degree of change they have experienced since the start of study treatment. There are 5 response options, ranging from “much worse” to “much better”.

- Patient satisfaction as measured by the TSQ

  The TSQ is a 3-item PRO designed to assess participant’s satisfaction with the study treatments. The TSQ assesses overall satisfaction with the study treatment, number of treatment administrations, and the form of treatment administration (injection into stomach and catheter into bladder) (1 item each).

- Peripheral blood and additional tumor tissue biomarkers consisting of the levels of cells, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or proteins that may be related to anti-tumor immune response or disease progression.

3.4. Baseline Variables

3.4.1. Study drug, study treatment and baseline definitions

In this study, ‘**study drug**’ refers to PF-06801591 or BCG and ‘**study treatment**’ (or ‘**treatment arm**’ or ‘**cohort**’) refers to one of the following:

BCG Naïve Cohort (Cohort A)

- Arm A: PF-06801591 + BCG (induction and maintenance period)
- Arm B: PF-06801591 + BCG (induction period only)
- Arm C: BCG only (induction and maintenance period)

BCG-Unresponsive Cohorts

- Cohort B1: CIS with or without papillary disease: PF-06801591 only
- Cohort B2: Papillary disease only: PF-06801591 only

Start and end dates of study treatment:

For Arm A and Arm B in Cohort A:

The date/time of first dose of study treatment in a combination arm is the earliest date/time of the first non-zero dose date/time for any of the study drugs in the combination.

The date/time of last dose of study treatment in a combination arm is the latest date/time of the last non-zero dose date/time for any of the study drugs in the combination.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

For Arm C (BCG only) in Cohort A and Cohorts B1 and B2 (PF-06801591 only):

The date/time of first dose of study treatment is the earliest date/time of non-zero dosing of the study drug.

The date/time of last dose of study treatment is the latest date/time of non-zero dosing of the study drug.

Definition of baseline:

'Start date' will be defined as date of randomization for Cohort A and date of first dose for Cohorts B1 and B2.

Definition of baseline for efficacy and PRO analyses

The last assessment prior to 'start date' will serve as the baseline assessment for efficacy and PRO analyses. For Cohort A only, if such a value is missing (since per protocol for some PRO endpoints the first assessment is planned to occur prior to dosing on Cycle 1 Day 1), the last assessment prior to the date of first dose of study treatment will be used as the baseline assessment except for analyses of disease assessment data where the baseline assessment would be considered as missing.

Definition of baseline for safety analyses

The last available assessment prior to the start of study treatment is defined as 'baseline' for safety analyses. If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 (day of first dose of study treatment) for baseline calculation. Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

Participants who start and discontinue treatment on the same day may have two different sets of data collected on study day 1 (one during study and one in the End of Treatment (EOT) visit). Data reported at the EOT visit are not eligible for baseline selection.

If a scheduled pre-dose assessment actually occurred post-dose, then the corresponding assessment will be treated and analyzed similar to an unscheduled post-dose assessment.

Baseline for RR and QT/QTc interval assessments will be derived from the visit where both RR and QT are not missing. QTcB and QTcF will be derived based on RR and QT.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Definition of baseline for immunogenicity analyses (Arms A and B in Cohort A and Cohorts B1 and B2)

The last available assessment prior to or at the start of treatment with PF-06801591(sample time ≤ 0 hr relative to PF-06801591 dose on Cycle 1 Day 1) is defined as 'baseline' for immunogenicity analyses. If an assessment is planned to be performed prior to the first dose of PF-06801591 in the protocol and the assessment is performed on the same day as the first dose of PF-06801591 but the assessment time point is not collected or is missing, it will be assumed that it was performed prior to PF-06801591 administration.

Definition of baseline for biomarker analyses

The last assessment prior to first dose of study treatment will serve as the baseline assessment for biomarker analyses. For biomarkers that are planned to be assessed on Cycle 1 Day 1, it will be assumed that the assessment was performed prior to study treatment administration, if the assessment time point is not collected or is missing.

3.4.2. Baseline characteristics

In Cohort A, randomization is stratified by the following, as recorded in the interactive response technology (IRT):

- Presence of carcinoma in situ at randomization (yes vs no);
- Geography (US vs Western Europe and Canada vs ROW).

The primary analyses for Cohort A of EFS and OS will be stratified by these randomization stratification factors.

Other baseline characteristics (including demographics, disease history, and prior anti-cancer therapies) are described in Section 6.5.1. These baseline characteristics are not planned to be included as stratification variables or covariates in statistical models unless otherwise specified in Section 6.

3.5. Safety Endpoints

3.5.1. Adverse events

Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day) (also refer to Appendix 1 for irAEs). The start day of new anti-cancer drug therapy after the first dose of study treatment is derived as outlined in Section 5.2.5.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Adverse Events of Special Interest (AESIs)

AESIs are immune-related adverse events (irAE) and injection site reactions related to PF-06801591. The criteria for classification of an AE as an irAE are described in Appendix 1.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis set prior to releasing the database and classifications will be documented per Pfizer's standard operating procedures.

Only participants who signed informed consent will be included in the analysis sets below.

Table 6. Analysis Sets

Population	Description
Enrolled	All participants who sign the informed consent document (ICD)
Randomly Assigned to Study treatment (Full Analysis Set [FAS]) (Cohort A)	All participants who are randomized in Cohort A. Participants will be classified according to the study treatment assigned at randomization.
Full Analysis Set [FAS] (Cohort B1)	All participants who receive at least 1 dose of study drug and are BCG unresponsive with CIS according to BICR assessment.
Full Analysis Set [FAS] (Cohort B2)	All participants who receive at least 1 dose of study drug and are BCG unresponsive with papillary disease only.
Per Protocol (PP) (Cohort A)	Participants randomized in Cohort A excluding those who did not receive at least one dose of study drug, did not meet inclusion criteria 2 or 3, or met exclusion criteria 1 or 2.
Safety	All participants who receive at least 1 dose of study drug. For Cohort A, participants will be classified according to the study treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case participants will be classified according to the first study treatment received.
Biomarker	The biomarker populations are a subset of the safety population and will include participants with at least 1 of the biomarkers evaluated at pre and/or post dose.
Immunogenicity	Subset of the safety population and will include participants who received at least one dose of PF-06801591 and have at least one ADA/Nab sample analyzed for anti-PF-06801591 antibodies.
PK	The PK concentration analysis population is a subset of the safety population and will include participants who have at least one post-1st dose concentration measurement above the lower limit of quantitation (LLQ) for PF-06801591. The PK parameter population is the same as the PK concentration analysis population.

Table 7 summarizes the use of the analysis sets for efficacy, safety, baseline characteristics, PRO, and exposure.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 7. Statistical Analyses by Analysis Set

Endpoints	Full Analysis Set	Per Protocol Analysis Set	Safety Analysis Set
Baseline Characteristics	✓		
Prior and Concomitant Therapies	✓		
Exposure			✓
Efficacy: Primary	✓	✓ (EFS for Arm A vs Arm C, Cohort A)	
Efficacy: Secondary	✓	✓ (EFS for Arm B vs Arm C, OS, Cohort A)	
Safety			✓
PRO	✓		

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and sample size determination

Cohort A

Approximately 999 participants (including a minimum of 250 participants with CIS) will be randomized in a 1:1:1 ratio to one of the 3 treatment arms (A, B, or C) with approximately 333 participants randomized to each treatment arm in Cohort A, stratified by the presence of CIS (yes vs no) and geography (US vs Western Europe and Canada vs ROW).

The primary objective will be tested at a significance level of 0.025 (1-sided) associated with H_{01} described below:

$$H_{01}: HR_{EFS(AvsC)} \geq 1 \text{ vs. } H_{11}: HR_{EFS(AvsC)} < 1$$

where $HR_{EFS(AvsC)}$ is the hazard ratio (HR) for EFS for Arm A vs Arm C.

In addition, the following statistical hypotheses will be tested to address the key secondary objectives:

$$H_{02}: HR_{EFS(BvsC)} \geq 1 \text{ vs. } H_{12}: HR_{EFS(BvsC)} < 1$$

$$H_{03}: HR_{OS(AvsC)} \geq 1 \text{ vs. } H_{13}: HR_{OS(AvsC)} < 1$$

$$H_{04}: HR_{OS(BvsC)} \geq 1 \text{ vs. } H_{14}: HR_{OS(BvsC)} < 1$$

where $HR_{EFS(BvsC)}$ is the HR for EFS for Arm B vs Arm C, $HR_{OS(AvsC)}$ is the HR for OS for Arm A vs Arm C, and $HR_{OS(BvsC)}$ is the HR for OS for Arm B vs Arm C.

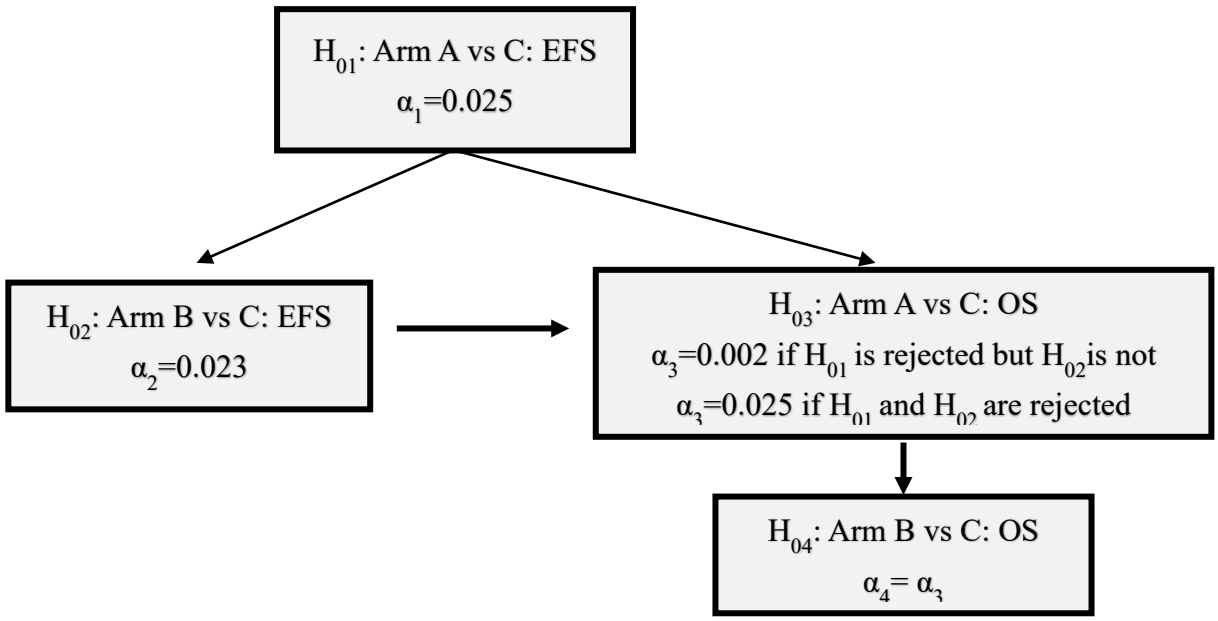
A graphical approach with group sequential testing outlined in Bretz et al. (2009) [2] will be used to strongly control the family-wise type I error rate at a 1-sided 0.025 level as shown in Figure 3. The initial 1-sided alpha allocated to H_{01} is 0.025. If H_{01} is statistically significant, H_{02} and H_{03} will be formally tested at initial 1-sided alpha of 0.023 and 0.002, respectively. If

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

H_{02} is statistically significant, alpha recycling will be applied to evaluate H_{03} at 1-sided 0.025 level. If H_{03} is statistically significant, all the alpha will be recycled to formally test H_{04}. One interim analysis on the OS endpoint is planned at the time of the EFS primary analysis. A Haybittle-Peto α-spending function will be used to determine the efficacy boundary.

Based on Protocol Amendment 5, the IA2 will be removed, and the FA for EFS based on a calendar-based data cut-off date will be conducted. Randomization to Cohort A was completed on 16 November 2021 and the EFS FA data cut-off date will be set to allow for approximately 3 years of follow-up after last randomized participant. The treatment period was completed on 16 November 2023. Based on the observed pooled blinded data of EFS accrual rate, approximately 261 EFS events are expected to have occurred across all three arms. Assuming equal event rate across three arms, approximately 174 events for each two-arm comparison are expected by the EFS FA data cut-off date.

Figure 3. Testing Strategy



The power for EFS is estimated based on the following assumptions:

- median EFS for participants receiving BCG (induction and maintenance) is 24 months[4,6,16.].
- treatment with PF-06801591 in combination with BCG (induction and maintenance) or BCG (induction only) is expected to increase the median EFS to 34.8 months, corresponding to a HR of 0.69 under the exponential model assumption.
- 20% drop-out rate for EFS within each arm.
- non-uniform participant accrual over 22 months.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- follow-up of approximately 33 months after the last participant is randomized.

For the primary endpoint, if the true HR is 0.69, 174 EFS events in Arm A and Arm C combined will provide 68.7% power to detect a difference using a 1-sided log-rank test at a significance level of 0.025.

Based on PA5,the FA for OS will be performed at the end of the study, which is approximately 5 years after the last participant has been randomized in Cohort A. Approximately 150 OS events are expected to have occurred across the three study arms assuming equal event rate across three arms (approximately 100 OS events for each two-arm comparison) by the OS FA data cut-off date.

If the primary objective is met, a gatekeeping approach will be used to allow further testing of the key secondary endpoints of EFS for Arm B vs. Arm C, OS for Arm A vs Arm C, and OS for Arm B vs Arm C as shown in Figure 3.

- For the key secondary endpoint of EFS for Arm B vs Arm C (H_{02}), if the true HR is 0.69 under the alternative hypothesis, 174 EFS events will provide 67.4% power to detect a difference using a 1-sided log-rank test at a significance level of 0.023.

- If both H_{01} and H_{02} are rejected, 100 OS events for Arm A vs C will provide 96.3% power to detect a HR of 0.472 using a 1-sided log-rank test at the 0.025 significance level and a 2-look group sequential design with Haybittle-Peto α-spending function to determine the efficacy boundary. The hazard ratio of 0.472 corresponds to an increase in the 5-year Kaplan-Meier rate for OS from 0.80[13] in Arm C to 0.90 in Arm A, under the exponential model assumption.

- If H_{01} is rejected but H_{02} is not rejected, 100 OS events for Arm A vs C will provide 80.8% power to detect a HR of 0.472 using a 1-sided log-rank test at the 0.002 significance level and a 2-look group sequential design with Haybittle-Peto α-spending function to determine the efficacy boundary.

- If H_{03} is rejected, H_{04} will be test with the same power as H_{03} under the same assumptions.

Cohorts B1 and B2

There will be no formal hypothesis testing for Cohorts B1 and B2.

Approximately 110 participants with BCG-unresponsive CIS will be enrolled in Cohort B1; this takes into account 10% of participants not meeting eligibility criteria. This sample size will allow for a 95% CI width for the CR rate to not exceed 0.21. The CR rate will be estimated in BCG-unresponsive CIS participants treated with PF-06801591.

Table 8 provides exact 95% CIs for CR rate based on different possible observed responses in the BCG unresponsive CIS Cohort (Cohort B1).

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 8. Exact 95% Confidence Interval for CR rate

Number of participants	Number of observed CRs	Observed CR rate	95% CI for True CR Rate
100	20	20%	(12.67%, 29.18%)
	25	25%	(16.88%, 34.66%)
	30	30%	(21.24%, 39.98%)
	35	35%	(25.73%, 45.18%)
	40	40%	(30.33%, 50.28%)
	45	45%	(35.03%, 55.27%)
	50	50%	(39.83%, 60.17%)
	55	55%	(44.73%, 64.97%)
	60	60%	(49.72%, 69.67%)

As of 31 August 2022, enrollment in Cohorts B1 and B2 was closed by the sponsor for business strategy reasons. This decision was not made due to any safety concerns, new emerging data, or regulatory interaction.

Due to the decision to close enrollment in Cohorts B1 and B2, the sample size for Cohorts B1 and B2 will not be achieved and the Cohort B1 and B2 objectives are no longer required.

5.1.2. Decision rules (Cohort A only)

Prior to protocol amendment 5, two interim (IA1(futility only) and IA2) and final (FA) analyses were planned for the primary endpoint of Cohort A.

IA1 for EFS was conducted after 658 participants were randomized to Cohort A for all 3 arms combined. As IA1 was for futility only, no alpha was spent at IA1.

Based on protocol amendment 5, the IA2 will be removed, and the FA for EFS based on a calendar-based data cut-off date will be conducted. Randomization to Cohort A was completed on 16 November 2021 and the EFS FA data cut-off date will be set to allow for approximately 3 years of follow-up after last randomized participant.

Cohort A of the study will be considered positive if the primary objective is met.

One interim and one final efficacy analyses are planned for OS endpoints. The IA for OS will be performed at the time of the final analysis for EFS. The FA for OS will be performed at the end of the study, which is approximately 5 years after the last participant has been randomized in Cohort A:

- IA: at the time of the final analysis for EFS; approximately 64 OS events (64% of the 100 OS events expected for each comparison) are expected to have occurred for each comparison.
- FA: this will occur at the end of the study, approximately 5 years after last participant randomized; approximately 100 OS events are expected to have occurred for each comparison.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 9. OS- Efficacy Boundaries

OS Comparison	Arm A vs Arm C		Arm B vs Arm C	
Analysis	**IA**	**Final Analysis**	**IA**	**Final Analysis**
Analysis cutoff trigger	EFS final analysis	5 years after last participant randomized	EFS final analysis	5 years after last participant randomized
Number of events (information fraction)[a]	64 (64%)	100 (100%)	64 (64%)	100 (100%)
p-value (z-value) for efficacy if H01 is rejected but H02 is not rejected	<0.0001 (<-3.719)	<0.00197 (<-2.883)	<0.0001 (<-3.719)	<0.00197 (<-2.883)
p-value (z-value) for efficacy if both H01 and H02 are rejected	<0.0001 (<-3.719)	<0.025 (<-1.96)	<0.0001 (<-3.719)	<0.025 (<-1.96)

a. For IA, the number of events is that expected under the alternative hypotheses (for EFS and OS) at the time of final EFS analysis; for FA, the number of events is that expected under the alternative hypothesis for OS at 5 years after last participant randomized.

The efficacy boundary for OS will be updated based on the actual number of observed OS events for each comparison using the pre-specified α-spending function. Therefore, the observed Z-test statistic at the IA for OS will be compared with the updated efficacy boundary. If OS is not met at the IA, the study for Cohort A will continue to the final analyses of OS, and the p-value that will be used to declare statistical significance at the FA for each comparison of OS will be based on the actual number of events documented at the cut-off date for the FA and the α already spent at the IA. Further details are provided in Section 7.

5.2. General Methods

As described in Section 3.4, in this study '**treatment arm**' or '**cohort**' refers to one of the following:

BCG Naïve Cohort (Cohort A)

- Arm A: PF-06801591 + BCG (induction and maintenance period)
- Arm B: PF-06801591 + BCG (induction period only)
- Arm C: BCG only (induction and maintenance period)

BCG-Unresponsive Cohort

- Cohort B1: CIS with or without papillary disease: PF-06801591 only
- Cohort B2: Papillary disease only: PF-06801591 only

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Endpoints will be summarized based on the analysis sets described in Table 6 by treatment arm in Cohort A and for Cohort B1 and Cohort B2, unless otherwise specified.

5.2.1. Data handling after the cut-off date

Data after the cut-off date may not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses, or imputations.

5.2.2. Pooling of centers

In order to provide overall estimates of treatment effects, data will be pooled across centers. The 'center' factor will not be considered in statistical models or for subgroup analyses due to the high number of participating centers in contrast to the anticipated small number of participants randomized/enrolled at each center.

5.2.3. Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics i.e., number of non-missing values and number of missing values [i.e., n (missing)], mean, median, standard deviation (SD), minimum, maximum and first and third quartile (Q1 and Q3).

Qualitative variables will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of proportions will include the missing category. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of participants still present at that visit, unless otherwise specified.

5.2.4. Definition of study day

Start day of study treatment is the day of the first dose of study treatment.

The study day for assessments occurring on or after the start of study treatment (e.g., adverse event onset, tumor measurement) will be calculated as:

Study day = Date of the assessment/event – start of study treatment + 1.

The study day for assessments occurring prior to the first dose of study treatment (e.g., baseline characteristics, medical history) will be negative and calculated as:

Study day = Date of the assessment/event – start of study treatment.

The study day will be displayed in all relevant data listings.

5.2.5. Definition of start of new anti-cancer drug therapy

Start date of new anti-cancer drug therapy is used to determine the end of the on-treatment period (see Section 5.2.7).

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The start date of new anti-cancer drug therapy is the earliest start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages that is after the first dose of study treatment. When start date of anti-cancer drug therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using only data from the 'Follow-up Cancer Therapy' eCRF pages.

5.2.6. Definition of start of new anti-cancer therapy

Start date of new anti-cancer therapy (drug, radiation, surgery) is used for censoring in sensitivity analyses for the primary efficacy endpoint of EFS for Arm A vs Arm C and the key secondary endpoint of EFS for Arm B vs Arm C (see Section 6.2.2.3).

The start date of new anti-cancer therapy is the earliest date after 'start date' amongst the following:

- Start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages
- Start date of radiation therapy recorded in 'Follow-up Radiation Treatment' eCRF pages.
- Surgery date, for lesions with a positive biopsy other than low grade Ta, or lesions with missing biopsy results, per the 'Oncology Biopsy" eCRF page, recorded in 'On Study & Follow-up Cancer Surgery' eCRF pages when 'Outcome of Procedure' = 'Resected' or 'Partially Resected'.

When start date of anti-cancer therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using 'Follow-up Cancer Therapy', 'Follow-up Radiation Treatment', and 'On Study & Follow-up Cancer Surgery' eCRF pages.

5.2.7. Definition of on-treatment period

Safety endpoints will be summarized based on the on-treatment period unless otherwise specified.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day) (also refer to Appendix 1 for irAEs).

Safety data collected outside the on-treatment period as described above will be listed and flagged in listings but will not be summarized.

5.2.8. Standard derivations and reporting conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Age [years]: (year of given informed consent – year of birth)

For reporting conventions, mean and median should generally be displayed one more decimal place than the raw data and standard deviation should be displayed to two more decimal places than the raw data. Percentages will be reported to one decimal place. The

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

rounding will be performed to closest integer / first decimal using the common mid-point between the two consecutive values. E.g., 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

5.2.9. Unscheduled visits

Generally, data collected at unscheduled visits will be included and analyzed for both safety and efficacy analyses in the same fashion as the data collected at scheduled visits except where otherwise noted in the sections that follow. Descriptive statistics (mean, SD, median, minimum, maximum, quartiles) by nominal visit or time point for safety endpoints such as laboratory measurements, electrocardiograms (ECG**s**) and vital signs will include only data from scheduled visits

5.2.10. Adequate baseline disease assessment

Adequate baseline is defined using the following criteria:

Cohort A, for the analysis of EFS per Investigator and per BICR

- All participants with CIS at baseline by the investigator are deemed to have adequate baseline disease.
- Participants without CIS at baseline by the investigator are deemed to have adequate baseline disease if all of the following criteria are met:
  - Baseline cystoscopy performed within 35 days prior to and including the date of randomization.
  - Baseline assessment for cystoscopy as collected on the “Cystoscopy” eCRF page must indicate “Yes” for “was Cystoscopy performed” and the cystoscopy result is “Negative”.
  - Baseline cytology as collected on the “Lab Urine Cytology” eCRF page is not “Positive”.

Cohort B1

- All participants with CIS by BICR at baseline are deemed to have adequate baseline disease.

Cohort B2

- Participants without CIS at baseline by the investigator are deemed to have adequate baseline disease if all of the following criteria are met:
  - Baseline cystoscopy performed within 35 days prior to and including the date of registration.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- Baseline assessment for cystoscopy as collected on the "Cystoscopy" eCRF page must indicate "Yes" for "was Cystoscopy performed" and the cystoscopy result is "Negative".
- Baseline cytology as collected on the "Lab Urine Cytology" eCRF page is not "Positive".

5.2.11. Adequate post-baseline disease assessment

An adequate post-baseline disease assessment is defined as an assessment recorded with progression of disease (all participants), recurrence of high-grade disease (all participants), recurrence of low-grade disease (participants without CIS at randomization (Cohort A/Cohort B2), CR (participants with CIS at randomization (Cohort A)/baseline (Cohort B1)), persistence of CIS disease (participants with CIS at randomization (Cohort A)/baseline (Cohort B1), or no evidence of disease (participants without CIS at randomization (Cohort A)/baseline (Cohort B2)) (Section 5.2.12). Time points where the response is not evaluable (NE) or no assessment was performed will not be used for determining the censoring date.

5.2.12. Disease Assessments

Disease will be assessed based on reported results from cytology, cystoscopy, biopsy, and imaging scans at different evaluation time points from the 'start date' until the first documentation of progression of disease, recurrence of high-grade disease, or persistence of CIS disease according to the following rules.

Complete Response, in participants with CIS at randomization (Cohort A)/baseline (Cohort B1), is defined as:

- histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy; or
- negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or non-malignant tissue.

The date of biopsy, cystoscopy, or cytology (whichever is earliest) will be used as the date of CR.

Persistence of CIS, in participants with CIS at randomization in Cohort A, is defined as persistence of CIS after induction or re-induction (if re-induction is administered). For participants in Cohort B1, persistence of CIS is defined as persistence of CIS at the 12-week disease assessment. This will require a cystoscopy or cytology positive result with biopsy positive for CIS. The date of positive biopsy, cystoscopy, or cytology (whichever positive result is earliest) will be used as the date of persistence of CIS.

Recurrence of high-grade disease is defined as re-appearance of high-grade disease (high-grade Ta, T1 or CIS) after 'start date' for the following 1) participants without CIS at baseline, 2) after CR for participants with CIS at baseline, or 3) before CR for participants with CIS and concurrent papillary disease at baseline in Cohort A or B1. This will require a cystoscopy or cytology positive result, with biopsy positive for T1, high-grade Ta and/or

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

CIS. The date of positive biopsy, cystoscopy, or cytology (whichever positive result is earliest) will be used as the date of recurrence of high-grade disease. If recurrence of high-grade disease occurs at the time of progression, progression should be the event. For participants with CIS at randomization, appearance of high-grade disease after a CR is considered recurrence of high-grade disease, while appearance of high-grade disease before a CR is progression.

Recurrence of low-grade disease is defined as re-appearance of low-grade disease (low-grade Ta) after 'start date'. This will require a cystoscopy or cytology positive result, with biopsy positive for low-grade Ta. The date of positive biopsy, cystoscopy, or cytology (whichever positive result is earliest) will be used as the date of recurrence of low-grade disease.

Progression of disease is defined as any of the following:

- Lamina propria invasion (e.g., increase from Ta to T1 or CIS to T1)
- Muscle invasive disease (stage ≥ T2)
- Lymph node positive disease (N+)
- Metastatic disease (M1)
- Appearance of high-grade Ta or T1 in participants with CIS only at baseline (Cohort A and Cohort B1) before achieving a CR.

  This will require a cystoscopy or cytology positive result with imaging positive and/or biopsy positive for disease progression as described above. The date of positive biopsy, imaging for new lesion, positive cystoscopy, or positive cytology (whichever result is earliest) will be used as the date of progression.

No Evidence of Disease: only applies to participants without CIS at baseline (Cohort A and Cohort B2) and is defined as disease that was completely resected prior to 'start date' and does not meet the criteria of progression of disease, recurrence of high-grade disease, or recurrence of low-grade disease while on study.

The date of imaging, biopsy, cystoscopy, or cytology (whichever result of the available assessments is earliest) that supports no evidence of disease will be used as the date of no evidence of disease.

5.3. Methods to Manage Missing Data

5.3.1. Missing data

Unless otherwise specified, all data will be evaluated as observed, and no imputation method for missing values will be used.

In all participant data listings imputed values will be presented and flagged as imputed.

Missing statistics, e.g. when they cannot be calculated, should be presented as 'ND' for not done, 'NR' for not reached or 'NA' for not applicable. For example, if N=1, the measure of variability cannot be computed and should be presented as 'ND' or 'NA'.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

5.3.1.1. Pharmacokinetic concentrations

Concentrations Below the Limit of Quantification

All concentrations assayed as below the level of quantification (BLQ) will be set to zero. The BLQ values will be excluded from calculations of geometric means and their CIs. A statement similar to 'All values reported as BLQ have been replaced with zero' should be included as a footnote to the appropriate tables and figures.

In listings, BLQ values will be reported as "<LLQ", where LLQ will be replaced with the value for the lower limit of quantification.

Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

1) A concentration has been reported as ND (i.e., not done) or NS (i.e., no sample);
2) A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular time point if more than 50% of the data are missing. No values will be imputed for missing data.

5.3.2. Handling of incomplete dates

5.3.2.1. Adverse events

Incomplete AE-related dates will be imputed as follows:

- If the AE onset date is missing completely, then the onset date will be replaced by the start of study treatment.
- If only the day part of the AE onset date is missing, but the month and year are equal to the start of study treatment, then the AE onset date will be replaced by the start of study treatment. For example, if the AE onset date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN/2015.
- If both the day and month of the AE onset date are missing but the onset year is equal to the start of study treatment, then the onset date will be replaced by the start of study treatment. For example, if AE onset date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed AE onset date will be 19/NOV/2014.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of participant's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete stop date will not be imputed. If stop date of AE is after the date of cut-off outcome of AE is ongoing at cut-off.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

5.3.2.2. Exposure

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 there is no End of Treatment eCRF page and no death date, the participant should be considered to be ongoing and use the cutoff date for the analysis as the last dosing date.
- If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment 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.

5.3.3. Imputation rules for date of last contact and efficacy assessments

5.3.3.1. Date of last contact

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

- All participant assessment dates (blood draws [laboratory, PK], vital signs, performance status, ECG, disease assessments, radiation, surgery)
- Start and end dates of anti-cancer therapies
- Start and stop dates of concomitant therapies including non-drug treatments or procedures
- Completion dates for PRO Questionnaires
- AE start and end dates
- Study drug start and end dates
- Randomization date
- Withdrawal of consent date
- Last date of contact where “Subject Remains in Follow-up” or “Subject No Longer Being Followed for Survival” collected on the “Survival Follow-up” eCRF
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up or death).

Only dates associated with actual examinations of the participant will be used in the derivation. Dates associated with a technical operation unrelated to participant status such as the date a blood sample was processed will not be used. Assessment dates after the cut-off date will not be applied to derive the last contact date.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

5.3.3.2. Death date

If there is a record for death, but the date is missing or is partial, it will be imputed based on the last contact date.

- If the date is missing, the death date will be imputed as the day after the date of last contact.
- 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 of last contact and the following:
  - 1st day of the month and year of death, if day of death is missing, OR
  - January 1st of the year of death, if both the day and month of death are missing.

5.3.3.3. Disease assessments

All investigation dates (e.g., cytology, cystoscopy, biopsy, CT/MRI/urogram scan) must be completed with day, month, and year.

If there are multiple disease assessment dates associated with an evaluation, i.e., if cytology, cystoscopy, biopsy, CT/MRI/urogram scan assessments occur over a series of days rather than the same day, the choice of date of assessment could impact the date of EFS event and/or date of complete response. If there are multiple assessment dates associated with an evaluation, the earliest of the assessment dates associated with the evaluation will be used as the date of assessment.

If one or more investigation dates for an evaluation are incomplete but other investigation dates are available, the incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the earliest of all investigation dates (e.g., cytology, cystoscopy, biopsy, or CT/MRI/urogram scan).

If all assessment dates for an evaluation have no day recorded, the 1st of the month is used.

If the month is not completed, for any of the investigations for an evaluation, the respective assessment will be considered to be at the date which is exactly between the previous and the following assessment. If both a previous and following assessments are not available, this assessment will not be used for any calculations.

5.3.3.4. Date of start of new anti-cancer therapy

Incomplete dates for start date of new anti-cancer therapy will be imputed as follows and may be used for determining censoring dates for efficacy analyses. Event date below refers to progression of disease, recurrence of high-grade disease or persistence of CIS date by investigator assessment. If the imputation results in an end date prior to the imputed start date, then the imputed start date should be set to the end date.

- The end date of new anti-cancer therapy will be included in the imputations for start date of new anti-cancer therapy. If the end date of new anti-cancer therapy is
  - completely missing then it will be ignored in the imputations below

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- partially missing with only year (YYYY) available then the imputations below will consider 31DECYYYY as the end date of the new anti-cancer therapy
- partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anti-cancer therapy

- For participants who have not discontinued study treatment at the analysis cut-off date, last dose of study treatment is set to the analysis cut-off date in the imputations below.
- If the start date of new anti-cancer therapy is completely or partially missing, then the imputed start date of new anti-cancer therapy is derived as follows:
  - Start date of new anti-cancer therapy is completely missing

    Imputed start date = min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]
  - Only year (YYYY) for start of anti-cancer therapy is available

    IF YYYY < Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy] THEN imputed start date = 31DECYYYY;

    ELSE IF YYYY = Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    THEN imputed start date = min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    ELSE IF YYYY > Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    THEN imputed start date = 01JANYYYY
  - Both Year (YYYY) and Month (MMM) for start of anti-cancer therapy are available

    IF

    YYYY = Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

    MMM < Month of min [max(Event date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

    THEN

    imputed start date = DAY (Last day of MMM) MMM YYYY ;

    ELSE IF

    YYYY = Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

    MMM = Month of min [max(Event date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

THEN

imputed start date = min [max(Event date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]);

ELSE IF

YYYY = Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM > Month of min [max(Event date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY;

ELSE IF

YYYY < Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = DAY (Last day of MMM) MMM YYYY;

ELSE IF

YYYY > Year of min [max(Event date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY.

5.3.4. Other Missing or Partial Dates

Imputation methods generally apply to partial dates as follows:

- If the day of the month is missing for a start date used in a calculation, the 1st of the month will be used to replace the missing date.
- If both the day and month are missing, the first day of the year is used.
- For stop dates, the last day of the month, or last day of the year is used if the day or day and month are missing, respectively.

These rules are used unless the calculations result in negative time durations (e.g., date of resolution cannot be prior to date of onset). In this case, the resolution and onset dates will be the same and the duration will be set to 1 day.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

6. ANALYSES AND SUMMARIES

Refer to Section 4 for definitions of analysis sets and Section 5.2 for general methodology.

For Cohort A, analyses and summaries will be by treatment arm.

Cohorts B1 and B2 will be reported separately from Cohort A.

6.1. Primary Endpoint(s)

6.1.1. Event-free survival as assessed by investigator (Cohort A only)

6.1.1.1. Primary analysis (for Arm A vs Arm C)

Refer to Section 2.2.1 for the primary estimand.

In this study, disease may be assessed through cytology, cystoscopy, biopsy, and imaging scans conducted at screening (required), every 12 weeks for 2 years following randomization, and every 24 weeks thereafter until first occurrence of progression of disease, recurrence of high-grade disease, or persistence of CIS (for participants with CIS at baseline) regardless of initiation of subsequent anti-cancer therapy.

EFS is defined as the time from randomization until progression of disease, recurrence of high-grade disease, persistence of CIS, or death due to any cause, whichever occurs first.

An event is as defined in Table 10. The censoring and event date options to be considered for the EFS analysis (Section 2.2.1) are presented in Table 11.

EFS (months) = [date of event or censoring– date of randomization +1]/30.4375

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 10. Possible Clinical Outcomes for Participants With or Without CIS at Randomization and EFS Evaluation

Population	Disease Assessment	EFS Event Y/N
Participants with CIS at randomization	Progressive disease [a] before achieving a CR for participants with CIS only at randomization	Y
	Progressive disease after achieving a CR	Y
	Recurrence of high-grade disease before achieving a CR for participants with CS and concurrent papillary disease at randomization	Y
	Recurrence of high-grade disease after achieving a CR	Y
	Persistence of CIS (non-CR) [b]	Y
	Recurrence of low-grade disease	N
Participants without CIS at randomization	Progressive disease	Y
	Recurrence of high-grade disease	Y
	Recurrence of low-grade disease	N

[a] Including appearance of new high-grade Ta or T1 disease for participants with CIS only at randomization.
[b] Persistence of CIS after induction, with CR after re-induction is not considered an event.

Table 11. Outcome and Event Dates for the Primary Analysis of EFS-Cohort A

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	Date of randomization [a]	Censored [a]
Event - After at most one missing or inadequate post-baseline disease assessment, OR - ≤ 24 weeks after the date of randomization	Date of event	Event
Event - After 2 or more missing or inadequate post-baseline disease assessments	Date of last adequate disease assessment documenting no event	Censored
No event	Date of last adequate disease assessment documenting no event	Censored

[a] However, if the participant dies ≤24 weeks after the date of randomization the death is an event with date on death date

The primary efficacy analysis will compare the EFS time based on the investigator assessment between Arm A and Arm C and will be performed using a 1-sided stratified log-rank test as described in Section 5.1. The following analyses will be based on the FAS.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio along with 95% CIs. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$,

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

where h(i,0;t) defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=control arm, 1= experimental arm) and β is the unknown regression parameter.

Ties will be handled using the Exact option in SAS (Ties=Exact option in SAS PROC PHREG).

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median EFS time with 2-sided 95% CIs. In particular, the EFS rate at 6, 12, 18, 24, 30, 36, 42, and 48 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot for EFS will be generated.

Frequency (number and percentage) of participants with each event type (progression of disease, recurrence of high-grade disease, persistence of CIS, or death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 12 following the hierarchy shown.

Table 12. EFS Censoring Reasons and Hierarchy – Cohort A

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment	No adequate baseline assessment[a]
2	Event after 2 or more missing or inadequate post-baseline disease assessments	Event after 2 or more missing assessments
3	No event and [withdrawal of consent date ≥ date of randomization]	Withdrawal of consent
4	No event and lost to follow-up in any disposition page	Lost-to-follow-up
5	No event and [EOS][b] present OR all EOT disposition pages of all drugs in the treatment arm say participant will not continue into any subsequent phase of the study] and no adequate post-baseline disease assessment	No adequate post-baseline disease assessment
6	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] However, if the participant dies ≤24 weeks after the date of randomization the death is an event with date on death date

[b] EOS here refers to completion of follow-up disposition page.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

For the participants with a PD event for EFS as assessed by the investigator, the number and percentage of participants with the following types of PD (defined hierarchically) will be presented:

- Progression to metastatic disease (defined as N+, M+)
- Progression to muscle invasive disease (defined as ≥ T2 disease)
- Stage Progression (defined as high-grade Ta or T1 disease)

The EFS time or censoring time and the reasons for censoring will also be presented in a participant listing.

Time of Follow-Up for EFS

A Kaplan-Meier plot for EFS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the EFS censoring and event indicators.

6.1.2. Complete response as assessed by BICR (Cohort B1 only)

6.1.2.1. Primary analysis

In this study, disease may be assessed through cytology, cystoscopy, biopsy, and imaging scans. Cytology, cystoscopy, and biopsy (as clinically indicated) will be required every 12 weeks for 2 years following date of first dose (after cycle 10 assessments will occur at cycle 14 and 16 followed by every 12 weeks) until first occurrence of progression of disease, recurrence of high-grade disease, or persistence of CIS (for participants with CIS at baseline) regardless of initiation of subsequent anti-cancer therapy. In addition, disease status will be assessed through imaging in the case of a positive cytology or cystoscopy and at screening, every 24 weeks for 2 years after initiation of study intervention until first occurrence of progression of disease, recurrence of high-grade disease, persistence of CIS (for participants with CIS at baseline) regardless of initiation of subsequent anti-cancer therapy, or EOT.

Complete Response (CR) is defined as histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy or negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or nonmalignant tissue at the 12-week assessment for participants with CIS at baseline by the BICR. CR will also be assessed by investigator.

A listing for CR will be provided.

6.1.3. Event-free survival as assessed by investigator (Cohort B2 only)

6.1.3.1. Primary analysis

In Cohort B2, disease may be assessed through cytology, cystoscopy, biopsy, and imaging scans. Cytology and cystoscopy are to be conducted at screening, every 12 weeks for 2 years following initiation of study intervention until first occurrence of progression of disease, recurrence of high-grade disease, regardless of initiation of subsequent anti-cancer therapy. Biopsies and imaging are performed at screening and as clinically indicated.

EFS is defined as the time from initiation of study intervention until recurrence of high-grade disease, progression of disease, or death due to any cause, whichever occurs first.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

An event is as defined in Table 13. The censoring and event date options to be considered for the EFS analysis are presented in Table 14.

Table 13. Possible Clinical Outcomes for Participants Without CIS at baseline and EFS Evaluation

Population	Disease Assessment	EFS Event Y/N
Participants without CIS at baseline	Progressive disease	Y
	Recurrence of high-grade disease	Y
	Recurrence of low-grade disease	N

Table 14. Outcome and Event Dates for the Primary Analysis of EFS – Cohorts B1 and B2

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	Date of initiation of study intervention [a]	Censored [a]
Event - After at most one missing or inadequate post-baseline disease assessment, OR - ≤ 24 weeks after the date of first dose	Date of event	Event
Event - After 2 or more missing or inadequate post-baseline disease assessments	Date of last adequate disease assessment documenting no event	Censored
No event	Date of last adequate disease assessment documenting no event	Censored

[a] However, if the participant dies ≤24 weeks after the date of first dose the death is an event with date on death date

EFS (months) = [date of event or censoring– date of first dose+1]/30.4375

Table 15. EFS Censoring Reasons and Hierarchy – Cohorts B1 and B2

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment	No adequate baseline assessment[a]
2	Event after 2 or more missing or inadequate post-baseline disease assessments	Event after 2 or more missing assessments
3	No event and [withdrawal of consent date ≥ date of first dose]	Withdrawal of consent
4	No event and lost to follow-up in any disposition page	Lost-to-follow-up
5	No event and [EOS][b] present OR all EOT disposition pages of all drugs in Cohort B2 say participant will not continue into any subsequent phase of the study] and no adequate post-baseline disease assessment	No adequate post-baseline disease assessment

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 15. EFS Censoring Reasons and Hierarchy – Cohorts B1 and B2

Hierarchy	Condition	Censoring Reason
6	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] However, if the participant dies ≤24 weeks after the date of first dose the death is an event with date on death
date
[b] EOS here refers to completion of follow-up disposition page.

A listing for EFS will be provided.

6.2. Secondary Endpoint(s)

6.2.1. Safety endpoints

Refer to Section 6.6.

6.2.2. Efficacy endpoints

The following analyses will be based on the FAS by treatment arm for Cohort A and for Cohorts B1 and B2, unless otherwise specified. Summaries will only be provided for Cohort A. For Cohorts B1 and B2, listings will be provided for efficacy endpoints as indicated.

For Cohort A, disease endpoints (CR, DoCR, time to recurrence of low-grade disease, and DSS) will be analyzed based on investigator assessment. EFS by BICR assessment will be analyzed as a secondary endpoint using the same methodology that is described in Section 6.1.1.1 now referring to BICR assessment instead of investigator assessment.

6.2.2.1. Event-free survival as assessed by the investigator (for Arm B vs Arm C in Cohort A)

Refer to Section 2.2.2 the key secondary estimand (EFS for Arm B vs Arm C).

The primary analysis of the key secondary endpoint of EFS for Arm B vs Arm C will repeat the primary analysis of the primary endpoint of EFS for Arm A vs Arm C (p-value, HR and 95% CIs) described in Section 6.1.1.1.

6.2.2.2. Overall survival (Cohort A only)

Overall survival (OS) is a key secondary endpoint for Cohort A.

The following analyses will be based on the FAS.

OS is defined as the time from the date of randomization to the date of death due to any cause. Participants last known to be alive will be censored at date of last contact.

OS (months) = [date of death or censoring– date of randomization +1]/30.4375

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The primary analysis of OS will compare the OS time between each of the experimental arms and the control arm and will be performed using a 1-sided stratified log-rank test as described in Section 5.1.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=control arm, 1= experimental arm) and β is the unknown regression parameter.

Ties will be handled using the Exact option in SAS (Ties=Exact option in SAS PROC PHREG).

In order to account for the group sequential design in this study, the repeated CI (RCI) method (Jennison and Turnbull, 2000)[9], will be used to construct the 2-sided RCIs for the hazard ratio at the interim and the final analyses of OS.

In addition, the unadjusted 95% CIs for the hazard ratio will also be reported at the interim and the final analyses for OS.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm for Cohort A along with a summary of associated statistics including the median OS time with 2-sided 95% CIs. In particular, the OS rate at 12, 24, 36, 48, 60, 72, and 84 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot for OS will be generated.

Frequency (number and percentage) of participants with an event (death) and censoring reasons will be presented by treatment arm.

Reasons for censoring will be summarized according to the categories in Table 16 following the hierarchy shown.

Table 16. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ 'start date']	Withdrawal of consent
2	No event and [lost to follow-up in any disposition page OR (data cut-off date – last contact date > 365 days)]	Lost to follow-up
3	No event and none of the conditions in the prior hierarchy are met	Alive

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The OS time or censoring time and the reasons for censoring will also be presented in a participant listing.

Time of Follow-Up for OS

A Kaplan-Meier plot for OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the OS censoring and event indicators.

6.2.2.3. Sensitivity analyses for event-free survival (Cohort A only)

The following analyses will be performed to explore the robustness of the primary analysis results. The sensitivity analyses will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.1.1.1 with the modifications below:

- EFS based on BICR results following the definitions in Table 10 and the censoring rules outlined in Table 11 and Table 12.
- EFS based on investigator assessment where the date of the event for participants with persistence of CIS is the date of randomization (refer to Supportive Estimand 1 Section 2.2.3)
- EFS based on investigator assessment following the definitions in Table 10 and the censoring rules outlined in Table 11 and Table 12 for participants in the per-protocol analysis set (refer to Supportive Estimand 2 Section 2.2.3)
- EFS based on investigator assessment following the definitions in Table 10 and the censoring rules outlined in Table 11 and Table 12 and using an unstratified analysis.
- EFS based on investigator assessment following the definitions in Table 10 and modifying the censoring rules in Table 11 (update footnote “a” to “If the participant dies ≤24 weeks after the date of randomization and did not initiate new anti-cancer therapy, the death is an event with date on death date”) and Table 12 to include initiation of subsequent anti-cancer therapies as a censoring reason after “no adequate baseline assessment”.
- EFS based on investigator assessment following the definitions in Table 10 and the censoring rules outlined in Table 11 and Table 12 and using CIS stratification factor derived based on the data entered by the investigator in the “Oncology Biopsy” eCRF at Screening.
- EFS based on investigator assessment following the definitions in Table 10 and modifying the censoring rules in Table 11 and Table 12 to not censor events after 2 or more missing or inadequate post-baseline disease assessments (ie. all events will be counted).
- If >10% of deaths are due to COVID-19, an analysis will be performed for EFS (primary definition excluding COVID-19 related deaths) by treating COVID-19 related deaths as a competing risk event.
- If there are >10% of participants with missed visits resulting in 2 or more missing assessments or inadequate post-baseline disease assessments due to COVID-19 before EFS events, censoring for event after 2 or more missing or inadequate disease assessment

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

will not occur per Table 11 and Table 12 and the EFS event will be counted (i.e. not censored per Table 11 and Table 12).

Sensitivity analysis for possibly informative censoring

To assess impact of possibly informative censoring due to potentially imbalanced dropouts on the primary analysis of EFS by the investigator, a reference-based multiple imputation method based on Bayes Gibbs sampling as outlined by Lu, Li, and Koch (2015)[12] will be implemented if imbalanced dropouts are observed between treatment arms. Imputations will be performed for participants who were censored due to lost to follow-up, withdrawal of consent, or no adequate post-baseline disease assessment as described in Table 12. This method assumes that the hazard for the participants in Arm A (or Arm B) who are censored due to the above censoring reasons lies between the hazard for the participants in Arm A (or Arm B) who continued and the hazard for the participants in Arm C, while the hazard for the participants in Arm C who are censored due to the above censoring reasons is the same as the hazard for the participants in Arm C who continued. Other sensitivity analysis methods for possibly informative censoring may also be considered, as applicable, and as appropriate.

Methods for evaluating the validity of model assumptions

The proportional hazards assumption will be checked visually by plotting log(-log(EFS)) versus log(time) within each randomization stratum.

Schoenfeld residuals for the stratified Cox's proportional regression model will be plotted to investigate graphically violations from the proportional hazards (PH) assumption; a non-zero slope is evidence of departure from PH. The PH assumption will be formally tested using Schoenfeld's residual test (Schoenfeld, 1980[15] Therneau & Grambsch, 2000[17]). Large departures from PH will be evidenced by a p-value <0.05.

If these show large departures from proportional hazards, then EFS based on central pathology assessment of biopsies and independent review of imaging scans will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013)[19].

Restricted Mean Survival Time (RMST)

The hazard ratio estimate from the Cox's PH model is routinely used to empirically quantify the between-arm difference under the assumption that the ratio of the two hazard functions is constant over time. When this assumption is plausible, such a ratio estimate may capture the relative difference between two survival curves. However, the clinical meaning of such a ratio estimate is difficult, if not impossible, to interpret when the underlying PH assumption is violated (i.e., the hazard ratio is not constant over time).

The RMST is a robust and clinically interpretable summary measure of the survival time distribution and is equivalent to the area under the Kaplan-Meier curve from 'start date' through a specific cut-off point. Unlike median survival time, it is estimable even under heavy censoring. There is a considerable body of methodological research (e.g., Royston and Parmar, 2011[14], Uno, Wei, et al., 2014[18]; Zhang, 2013[19]) about the use of RMST to estimate treatment effects as an alternative to the hazard ratio approach.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The RMST methodology is applicable independently of the PH assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results. However, when large departures from the PH assumption are observed, the log-rank test is underpowered to detect differences between the survival distributions for the treatment arms, and a test of the difference between the RMST for the experimental arm and the control arm may be more appropriate to determine superiority of the experimental arm compared to the control arm with respect to the time-to-event endpoint.

In particular, as it pertains to the **cut-off point (τ)** to evaluate the RMST, it is noted that the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can then be interpreted as the expected survival time restricted to the common follow-up time τ among all participants. To avoid arbitrary selection of the common cut-off τ for both treatment arms, τ is pre-specified as:

τ = minimum of (largest observed EFS time for the experimental arm, largest observed EFS time for the control arm).

The treatment effect between the experimental arm and the control arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated.

BICR vs Investigator Assessment

A summary of BICR evaluation of investigator-assessed EFS will be provided:

- Number and percentage of participants with a non-death EFS event as assessed by the investigator
- Number and percentage of participants with an investigator-assessed non-death EFS event confirmed by BICR.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

See subgroups as defined in Section 6.4.

Multivariable Cox's regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as:

$$h(t) = h(0;t)\, e^{Xb}$$

where h(0;t) defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

In the stepwise selection procedure, variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). This analysis will be performed using the stepwise selection method in SAS (Proc PHREG). Once this procedure stops, the factor ‘'treatment arm'' will be added to the last selected model in order to evaluate the effect of treatment on EFS time when adjusted for the selected explanatory variables. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% CIs. No interactions will be considered. Post-baseline factors will not be considered for the model.

6.2.2.4. Sensitivity analysis for OS (Cohort A only)

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results for OS. These analyses are regarded as purely exploratory. The sensitivity analyses will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.2.2.2 with the modifications below:

- PP analysis set;
- Unstratified;
- If >10% of deaths are due to COVID-19, an analysis will be performed for non-COVID-19 related deaths by treating COVID-19 related deaths as a competing risk event.

Methods for evaluating the validity of model assumptions

The same methodology described in Section 6.2.2.3 for EFS will be used for OS.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

The same methodology described in Section 6.2.2.3 for EFS will be used for OS.

6.2.2.5. Complete response (participants with CIS at randomization) (Cohort A only)

Complete Response (CR) is defined as histologic disappearance of malignancy on bladder biopsy with negative cytology and cystoscopy or negative cytology and positive cystoscopy with biopsy-proven low-grade Ta lesion or nonmalignant tissue for participants with CIS at randomization according to the CRF.

CR rate for participants with CIS at randomization will be calculated with the 2-sided 95% CI using the Clopper-Pearson method[5] (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option) for each treatment arm.

The difference in CR rates between each experimental arm (Arm A and Arm B) and the control arm (Arm C) will be tested with a 2-sided p-value using a stratified analysis stratified

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

by the randomization stratum of geographic region. The Mantel-Haenszel (MH) test and MH confidence limits will be obtained through PROC FREQ using the option of COMMONRISKDIFF(CL=MH Test=MH), which tests the null hypothesis that the common risk difference is 0.

BICR biopsy evaluation of investigator-assessed CR

For the participants with CIS at randomization by the investigator, the following will be presented:

- Number of participants achieved CR as assessed by the investigator
    - Number of participants achieved CR as assessed by the investigator which was confirmed by investigator biopsy assessment.
        - BICR agreed (investigator-assessed CR confirmed by BICR biopsy assessment).
        - BICR disagreed (investigator-assessed CR not confirmed by BICR biopsy assessment).
        - No BICR biopsy assessment (biopsy not available/evaluable for BICR assessment).

If there are >5% of participants who discontinued treatment due to COVID-19 prior to achieving CR, a sensitivity analysis based on cumulative incidence curve will be performed for CR to estimate the CR rate by treating COVID-19 related discontinuation as censoring.

6.2.2.6. Duration of complete response (Cohort A only)

Duration of complete response (DoCR) will be analyzed for participants with CIS at randomization.

DoCR is defined as the time from the first documentation of CR to the date of an EFS event for participants with CR.

For participants without an EFS event, DoCR will be censored according to the rules in Table 17

DoCR (months) = [date of event or censoring– first date of CR +1]/30.4375

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 17. Outcome and Event Dates for DoCR Analyses

Scenario	Date of event/censoring	Outcome
Event - After at most one missing or inadequate post-baseline disease assessment, OR - ≤ 24 weeks after the 'start date')	Date of event	Event
Event - After 2 or more missing or inadequate post-baseline disease assessments	Date of last adequate disease assessment documenting no event	Censored
No Event	Date of last adequate disease assessment documenting no event	Censored

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm for Cohort A together with a summary of associated statistics including the median DoCR time with 2-sided 95% CIs. In particular, the DoCR rates at 6, 12, 18, 24, 30, 36, 42, and 48 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of participants with each event type (progression of disease, recurrence of high-grade disease, or death) and censoring reasons will be presented by treatment arm for Cohort A. Reasons for censoring will be summarized according to the categories in Table 18 following the hierarchy shown.

Table 18. DoCR Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	Event after 2 or more missing or inadequate post-baseline disease assessments	Event after 2 or more missing assessments
2	No event and [withdrawal of consent date ≥ 'start date']	Withdrawal of consent
3	No event and lost to follow-up in any disposition page	Lost-to-follow-up
4	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

The DoCR time or censoring time and the reasons for censoring will also be presented in a participant listing.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

DoCR (Alternative Definition): Includes all randomized participants (Cohort A only)

Rather than limiting the analyses of DoCR to the subset of participants with CR, statistical inference on DoCR can be performed including all randomized participants with CIS at randomization based on the restricted mean methodology [(Huang et al, Annals of Internal Medicine, 2020)[7] and (Huang et al, JAMA Oncology, 2018)[8]].

DoCR is defined as EFS time minus P/D/R event-free time at cut-off time τ.

DoCR = EFS Time – (P/D/R) event-free time, at cutoff time τ.

where P/D/R is a composite endpoint of PD (EFS event), Death, or CR whichever occurs first. The cut-off time τ is intended to reflect the time point in the Kaplan-Meier curve beyond which there are no meaningful information on EFS or P/D/R event-free time.

Specifically,

- DoCR = Date of P/D or Censoring Date–- Date of CR, for responders whose EFS time $\leq$ cutoff time τ.
- DoCR = Randomization date + cutoff time τ – Date of CR, for responders whose EFS time > cutoff time τ.
- DoCR = 0, for non-responders before cutoff time τ.

where P/D stands for progression or death.

The mean DoCR is the area between the EFS curve and the P/D/R event-free curve in $(0, \tau)$.

The principle of selecting the truncation cutoff time τ for mean DoCR is that we can make a valid inference for the area under the curve with respect to the P/D/R and P/D events. There are 3 cases for the selection in a single arm:

1. If the estimated P/D/R event-free curve has not reached 0 on the basis of the observed data (that is, the largest observed time to P/D/R or censoring is not a P/D/R event time), the truncation time for mean DoCR can be as large as the largest observed time to P/D/R or censoring.

2. If the estimated P/D/R event-free curve has reached 0 but the estimated P/D event-free curve has not reached 0 on the basis of observed data (that is, the largest observed time to P/D/R or censoring is a P/D/R event time, but the largest observed time to P/D or censoring is not a P/D event time), then the truncation time for mean DoCR can be as large as the largest observed time to P/D or censoring.

3. If both estimated event-free curves have reached 0 on the basis of observed data (that is, the largest observed time to P/D or censoring is a P/D event time), there is no restriction in selecting the truncation time point for mean DoCR.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

After selecting the truncation cutoff time τ for each arm following the steps above, the truncation cutoff time in the 2-arm comparison is determined as follows:

1. If both the EFS curve and the P/D/R event-free curve crossed zero for both arms, use the maximum of the τ for each of the arms;

2. Else if both the EFS curve and the P/D/R event-free curve crossed zero for Arm A, then use the τ of Arm B;

3. Else if both the EFS curve and the P/D/R event-free curve crossed zero for Arm B, then use the τ of Arm A;

4. Else use the minimum of the τ for each of the arms

DoCR for each treatment arm will be displayed graphically and between treatment arms can also be compared using RMST methodology (see Section 6.2.2.3). The variance of the mean DoCR can be derived by bootstrapping or analytically based on counting process theory.

6.2.2.7. Event-free survival (Cohort B1)

In Cohort B1, cytology, cystoscopy, and biopsy (as clinically indicated) will be required every 12 weeks for 2 years following date of first dose until first occurrence of progression of disease, recurrence of high-grade disease, or persistence of CIS (for participants with CIS at baseline) regardless of initiation of subsequent anti-cancer therapy. In addition, disease status will be assessed through imaging in the case of a positive cytology or cystoscopy and at screening, every 24 weeks for 2 years after initiation of study intervention until first occurrence of progression of disease, recurrence of high-grade disease, persistence of CIS regardless of initiation of subsequent anti-cancer therapy, or EOT.

EFS is defined as the time from first dose until progression of disease, recurrence of high-grade disease, persistence of CIS, or death due to any cause, whichever occurs first.
The censoring and event date options to be considered for the EFS analysis are presented in Table 14. An event is as defined in Table 19.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 19. Possible Clinical Outcomes for Participants With CIS at baseline and EFS Evaluation

Population	Disease Assessment	EFS Event Y/N
Participants with CIS at baseline	Progressive disease [a] before achieving a CR for participants with CIS only at baseline	Y
	Progressive disease after achieving a CR	Y
	Recurrence of high-grade disease before achieving a CR for participants with CIS and concurrent papillary disease at baseline	Y
	Recurrence of high-grade disease after achieving a CR	Y
	Persistence of CIS (non-CR)	Y
	Recurrence of low-grade disease	N

[a] Including appearance of new high-grade Ta or T1 disease

EFS (months) = [date of event or censoring– date of first dose+1]/30.4375

An EFS listing only will be provided by investigator assessments.

6.2.2.8. Time to recurrence of low-grade disease (Cohort A only)

Time to recurrence of low-grade disease is defined as the time from randomization to the date of first documentation of recurrence of low-grade Ta disease, according to the "Oncology Biopsy" CRF page.

For participants without an event (recurrence of low-grade Ta), time to recurrence of low-grade disease will be censored according to the rules in Table 20.

Table 20. Outcome and Event Dates for the Analysis of Time to Recurrence of Low-Grade Disease

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	Date of randomization	Censored
Event - After at most one missing or inadequate post-baseline disease assessment, OR - ≤ 24 weeks after the date of randomization	Date of event	Event
Event - After 2 or more missing or inadequate post-baseline disease assessments	Date of last adequate disease assessment documenting no event	Censored
No event	Date of last adequate disease assessment documenting no event	Censored

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by randomization strata to calculate the hazard ratio for time to recurrence of low-grade disease between each experimental arm and Arm C.

Time to recurrence of low-grade disease will be analyzed using Kaplan-Meier methods. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time to recurrence of low-grade disease with 2-sided 95% CIs. Time to recurrence of low-grade disease rates at 6, 12, 18, 24, 30, 36, 42, and 48 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at specific time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (with back transformation to a CI on the untransformed scale). The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot for time to recurrence of low-grade disease will be generated.

Reasons for censoring will be summarized according to the categories in Table 21 following the hierarchy shown.

Table 21. Time to Recurrence of Low-Grade Disease Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment	No adequate baseline assessment
2	Event after 2 or more missing or inadequate post-baseline disease assessments	Event after 2 or more missing assessments
3	No event and [withdrawal of consent date ≥ date of randomization]	Withdrawal of consent
4	No event and lost to follow-up in any disposition page	Lost-to-follow-up
5	Death without an event	Death
6	No event and [EOS][a] present OR all EOT disposition pages of all drugs in a treatment arm say participant will not continue into any subsequent phase of the study] and no adequate post-baseline disease assessment	No adequate post-baseline disease assessment
7	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] EOS here refers to completion of follow-up disposition page.

A listing for time to recurrence of low-grade disease will be provided.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

6.2.2.9. Disease specific survival (Cohort A only)

Disease specific survival (DSS) is defined as the time from randomization to death resulting from bladder cancer, as assessed by the investigator.

DSS = [date of death due to bladder cancer or censoring – date of randomization +1]/30.4375

Participants last known to be alive will be censored at the date of last contact.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by randomization strata to calculate the hazard ratio for DSS between each experimental arm and Arm C.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median DSS time with 2-sided 95% CIs. DSS rates at 12, 24, 36, 48, 60, 72, and 84 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at specific time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (with back transformation to a CI on the untransformed scale). The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot for disease specific survival will be generated.

Reasons for censoring will be summarized according to the categories in Table 22 following the hierarchy shown.

Table 22. DSS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	Death due to reasons other than bladder cancer	Cause of death other than bladder cancer*
2	No event and [withdrawal of consent date ≥ date of randomization]	Withdrawal of consent
3	No event and [lost to follow-up in any disposition page OR data cut-off date – last contact date > 365 days]	Lost to follow-up
4	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

*Reasons other than disease under study on the Death Details CRF page

A listing for disease specific survival will be provided.

6.2.2.10. Time to cystectomy (Cohort A only)

Time to cystectomy is defined as time from 'start date' to cystectomy. Participants without a cystectomy will be censored at death date or last date known to be alive.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by randomization strata to calculate the hazard ratio for time to cystectomy between each experimental arm and Arm C.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time to cystectomy with 2-sided 95% CIs. Time to cystectomy rates at 6, 12, 18, 24, 30, 36, 42, and 48 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[3] and the CIs for the survival function estimates at specific time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[10] (with back transformation to a CI on the untransformed scale). The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot for time to cystectomy will be generated.

Reasons for censoring will be summarized according to the categories in Table 23 following the hierarchy shown.

Table 23. Time to Cystectomy Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and participant died	Death prior to event
2	No event and [withdrawal of consent date ≥ 'start date']	Withdrawal of consent
3	No event and [lost to follow-up in any disposition page OR data cut-off date – last contact date > 365 days]	Lost to follow-up
4	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

A listing for time to cystectomy will be provided.

6.2.3. Pharmacokinetic endpoints

The following pharmacokinetic analyses will be based on the PK analysis set by treatment arm for Cohort A.

C_{trough} for PF-06801591 will be listed and summarized by treatment arm (Cohort A), cycle and study day relative to PF-06801591 dosing, and the summarized data will be plotted using box whisker plots to assess attainment of steady state. EOT concentrations will be excluded from C_{trough} data summaries and plots, unless the EOT visit coincides with a scheduled predose PF-06801591 sample (Cycles 1, 2, 4, 6, 8, 10, or 13 of Cohorts A) and has been collected prior to dosing, in which case it will be included in the summary statistics for the planned PK sample time.

Presentation of pharmacokinetic data will include:

- Descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% CI) of serum C_{trough}/predose concentrations will be presented in tabular form by treatment arm (Cohort A)), cycle, and day. Zero values will

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

be excluded from the calculation of geometric means and its associated %CV. Data from samples collected at EOT will be presented in individual listings, but will be excluded from summary statistics, unless the EOT visit coincides with a scheduled PF-06801591 PK sample time.

- Box plots for C_{trough} for PF-06801591 will be generated by cycle for each treatment arm for Cohort A. The geometric mean and the median of the parameter in each treatment will be overlaid on the box plots. Data collected at EOT will be excluded, unless the EOT visit coincides with a scheduled PF-06801591 PK sample time.
- Individual serum C_{trough}/predose concentrations will be presented in spaghetti plots (linear scale) by treatment arm for Cohort A. Data collected at EOT will be excluded, unless the EOT visit coincides with a scheduled PF-06801591 PK sample time.

PK concentration data available for Cohort B1 and/or B2 will be provided in listings only.

6.2.4. Population pharmacokinetic endpoints

Pharmacokinetic and pharmacodynamic data from this study may be analyzed using modeling approaches and may also be pooled with data from other studies to investigate any association between PF-06801591 exposure and biomarkers or significant safety or efficacy endpoints. The results of these analyses, if performed, may be reported separately.

6.2.5. Biomarker endpoints

Summary statistics for levels of PD-L1 expression, and number and percentage of participants with tumors categorized with baseline PD-L1 expression level (high vs low) will be presented by arm for Cohort A.

Biomarker data, if available, for Cohort B1 and/or Cohort B2 will be provided in a listing only.

6.2.6. Endpoints for immunogenicity data of PF-06801591

All analyses described below will be performed on ADA and nAb data from participants who received PF-06801591. Data will be summarized by treatment in Cohort A (Arm A and Arm B) and combined across both treatment arms, unless otherwise specified.

Blood samples for PF-06801591 immunogenicity testing will be collected pre-dose (≤ 0 hr actual time) of the PF-06801591 injection on Day 1 relative to PF-06801591 dosing of Cycles 1, 2, 4, 6, 8, 10, 13 and EOT of Cohort A

Samples positive for ADA will be analyzed for titer and may be analyzed for nAb.

Participants will be characterized into different ADA and nAb categories based on the criteria defined in Table 24.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 24. Participants Characterized Based on Anti-Drug Antibody Results (ADA Status) and Neutralizing Antibody Results (nAb Status)

Term	Definition
ADA-positive sample	ADA sample titer ≥ 99
ADA-negative sample	ADA sample titer < 99
nAb-positive sample	nAb sample titer ≥ 4
nAb-negative sample	nAb sample titer < 4
Pre-existing ADA/nAb	Positive ADA/nAb at baseline (e.g. day 1 pre-dose)
Cross-reactivity	This often refers to immunogenicity testing against an endogenous antigen, biosimilar reference product or another biotherapeutic within the same therapeutic class.
Subject-level immunogenicity analysis population	
ADA evaluable population	All subjects with ≥ 1 post-treatment ADA result.
nAb evaluable population	ADA-positive subjects with ≥ 1 post-treatment nAb result, plus all ADA-negative subjects. An ADA-positive subject without any post-treatment nAb data is excluded from the analysis population.
Subject-level definitions	
Treatment-induced ADA	Baseline ADA titer is missing or negative and subject has ≥ 1 post-treatment positive ADA titer.
Treatment-boosted ADA	Baseline ADA titer is positive and subject has a ≥ 4-fold dilution increase in ADA titer from baseline in ≥ 1 post-treatment sample.
ADA-positive subject	A subject with ≥ 1 treatment-induced or treatment-boosted ADA response.
ADA-negative subject	An ADA evaluable subject without treatment-induced or treatment-boosted ADA response. Subject either has (1) all ADA-negative results throughout the study or (2) is ADA positive at baseline but did not become treatment-boosted post dose.
ADA incidence	The percent of ADA-positive subjects in a treatment group/cohort or study.
Treatment-induced nAb	Baseline nAb titer is missing or negative or ADA-negative and subject has ≥ 1 post-treatment positive nAb titer.
Treatment-boosted nAb	Baseline nAb titer is positive and subject has a ≥ 4-fold dilution increase in nAb titer from baseline in ≥ 1 post-treatment sample.
nAb-positive subject	An ADA-positive subject with ≥ 1 treatment-induced or treatment-boosted nAb response. For ADA-positive (treatment-boosted) subjects, subject is nAb positive only if the subject has ≥1 treatment-induced or treatment-boosted nAb response at the visit where the subject has a treatment-boosted ADA response. For visits where the subject did not show a boosted ADA response, the subject is classified as nAb-negative for the visit even if the subject has post-treatment positive nAb titer for that visit.
nAb-negative subject	nAb evaluable participant who is either (1) an ADA-negative subject or (2) an ADA-positive subject without treatment-induced or treatment-boosted nAb response (i.e. subject has all nAb-negative results throughout the study or subject is nAb positive at baseline but did not become treatment-boosted post dose). Note: in the event a subject is ADA-positive at baseline but did not show a boosted response post-treatment, subject is classified as ADA-negative and nAb-negative at the subject level even if the subject has post-treatment positive nAb titer, as all such ADA-negative subjects are nAb-negative regardless of nAb titer data.
nAb incidence	The percent of nAb-positive subjects in a treatment group/cohort or study.
Duration of ADA and nAb response (subject-level definitions):	
Transient ADA	An ADA-positive subject with (1) a treatment-induced or treatment-boosted ADA sample detected only at 1 sampling time (excluding the last time point) post-treatment, or (2) treatment-induced or treatment-boosted ADA samples detected at ≥ 2 time points where the first and last positive samples (irrespective of any negative samples in between) are separated by < 16 weeks, and the subject's last sample is ADA negative.
Persistent ADA	An ADA-positive subject with first and last positive ADA samples (treatment-induced or treatment-boosted) detected over a period of ≥ 16 weeks post-treatment, irrespective of any negative samples in between.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Table 24. Participants Characterized Based on Anti-Drug Antibody Results (ADA Status) and Neutralizing Antibody Results (nAb Status)

Term	Definition
Indeterminate ADA	An ADA-positive subject who is not persistent or transient.
Transient nAb	A nAb-positive subject with (1) a treatment-induced or treatment-boosted nAb sample detected only at 1 sampling time (excluding the last time point) post-treatment, or (2) treatment-induced or treatment-boosted nAb samples detected at ≥ 2 time points where the first and last positive samples (irrespective of any negative samples in between) are separated by < 16 weeks, and the subject's last sample is nAb negative or ADA negative.
Persistent nAb	A nAb-positive subject with first and last positive nAb samples (treatment-induced or treatment-boosted) detected over a period of ≥ 16 weeks post-treatment, irrespective of any negative samples in between.
Indeterminate nAb	A nAb-positive subject who is not persistent or transient.

Note: Duration of response (persistent, transient or indeterminate) definitions are only applicable to ADA (or nAb)-positive subjects.

All ADA and nAb data will be listed, and the number and percentage of participants in each ADA and nAb category will be summarized by treatment arm for Cohort A and combined across treatment arms. Incidence of ADA and nAb positive participants and time to first ADA and nAb detection will also be summarized by treatment arm for Cohort A and combined across treatment arms. Incidence may also be presented graphically and, if ≥ 10% in a treatment arm, immunogenicity titer data will also be summarized and box plots of PK concentrations by ADA and nAb status will be presented.

Where data are summarized by visit, results will be summarized as windowed visit (i.e., cumulatively from the day/time of prior visit to the day/time of current visit) to allow inclusion of results from unplanned visits. Data from samples collected at EOT will be presented in individual listings and will be included in categorical assessments and summaries.

Immunogenicity data available for Cohort B1 and/or Cohort B2 will be provided in listings only.

6.3. Other Endpoints

6.3.1. PRO endpoints (Cohort A only)

All PRO analyses will be performed based on the FAS.

6.3.1.1. PRO instruments

Patient-reported outcomes will be collected using the following instruments: European Organization for Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), non-muscle invasive bladder cancer symptoms (NMIBC24), health status in participants with NMIBC (EQ-5D-5L), disease severity (PGIS), patient global impression of change (PGIC), treatment administration burden (PTAB), and treatment satisfaction questionnaire (TSQ).

6.3.1.2. PRO scoring

The EORTC QLQ-C30, NMIBC24, and EQ-5D-5L will be scored according to their respective user guides. The current version at the time of analysis, of UK weights, will be

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

used to construct the index values of the EQ-5D-5L. For the EORTC QLQ-C30, EORTC QLQ NMIBC24, and EQ-5D-5L, missing items will be handled per the respective scoring manuals of each questionnaire. For the PGIS, PGIC, TSQ, and PTAB there will be no adjustments for missing data.

PGIS, PGIC, PTAB, and TSQ questionnaire items will be analyzed separately, items within each questionnaire will not be combined or scored.

6.3.1.3. Instrument completion rates

For each treatment arm in Cohort A, at each time point the number and percentage of participants who complete each of the EORTC QLQ-C30, NMIBC24, EQ-5D-5L, PGIS, PGIC, PTAB, and TSQ will be summarized.

6.3.1.4. Descriptive Summary

Separate tables for each questionnaire for Cohort A will be used to summarize the mean, SD, median, range and 95% CI of absolute scores and change from baseline of the total, all subscales of the EORTC QLQ-C30, NMIBC24, EQ-5D-5L (index and VAS) at each time point. A line chart depicting the means along with 95% CI bars over time will be provided for each scale in each Cohort A treatment arm.

For the PTAB, PGIS, PGIC, and TSQ, summary statistics (number and percentage) will be provided for each item, by treatment arm in Cohort A at each time point.

A figure will be provided for the cumulative patient disposition by treatment arm in Cohort A per PRO assessment window for all scheduled assessments using the following categories:

- PRO assessment expected;
- PRO assessment not expected due to lack of efficacy;
- PRO assessment not expected due to death;
- PRO assessment not expected due to other reasons.

6.3.1.5. Health Status

For the EQ-5D-5L, for each treatment arm in Cohort A, the number and percentage of participants reported having “none”, “slight”, “moderate”, “severe”, or “extreme/unable” problems at each time point will be calculated.

6.3.1.6 Random Coefficient Models

Random coefficient models will be carried out in Cohort A for the EORTC QLQ-C30 and NMIBC24 (all domains, subscales and symptoms), EQ-5D-5L (Index), and EQ-VAS using PROC MIXED. Outcomes are post-baseline scores and change scores and the predictors are the corresponding baseline PRO score, treatment, time (treated as a continuous variable), and treatment-by-time interaction. Intercept and time are considered as random effects particular to each participant. All available data for each participant prior to the end of therapy should be used in the analyses. All parameter estimates should be obtained using restricted

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

maximum likelihood. The unstructured covariance structure should be used to define covariance between random effects (using option "Type=UN" as a part of the RANDOM statement in PROC MIXED). For the degrees-of-freedom calculations the Kenward and Roger algorithm should be used (using option ddfm = kr") as a part of the MODEL statement in PROC MIXED).

6.3.2. Exploratory biomarker endpoints (Cohort A only)

Peripheral blood and additional tumor tissue biomarkers consisting of the levels of cells, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or proteins that may be related to anti-tumor immune response or disease progression.

Summary statistics for pre-treatment biomarkers will be presented by treatment arm for Cohort A. Summary statistics for on-treatment biomarkers will additionally include the ratio to baseline for continuous biomarkers, and a contingency table for non-continuous biomarkers, as appropriate.

6.4. Subset Analyses (Cohort A only)

Subset analyses, if applicable, will be performed for OS, EFS, and CR per investigator assessment as noted below, based on the FAS for the subgroups defined below.

The following subgroups will be defined and used for analyses:

- Randomization stratification factor(s)
  - Presence of CIS per IRT (EFS per investigator, and OS)
    - Yes (Reference)
    - No
  - Geography
    - US (Reference)
    - Western Europe and Canada
    - ROW
- Presence of CIS per CRF (EFS per investigator and OS only)
    - Yes (Reference)
    - No
- Age
  - Age < 65 years (Reference)
  - Age ≥ 65 years
- Sex
  - Male (Reference)
  - Female

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- Race
  - White (Reference)
  - Asian
  - Black or African American
  - Other
- Ethnicity
  - Hispanic or Latino or of Spanish origin
  - Not Hispanic or Latino or of Spanish origin (Reference)
- TICE Strain
  - TICE (Reference)
  - No TICE
- BCG strain at induction
  - TICE (Reference)
  - RIVM
  - TOKYO172
  - BCG-1
  - CHINA D2PB302
- T1 disease at baseline per CRF (EFS per investigator and OS only)
  - Yes
  - No (Reference)

Subset analyses for EFS and OS will use the censoring rules described in Sections 6.1.1.1, Section 6.2.2.1 and Section 6.2.2.2.

Treatment arms will be compared for each endpoint using a 2-sided unstratified log-rank test for each subgroup level and the unstratified HR and its corresponding 95% CI will be computed per subgroup level.

Subset analyses for difference in CR rates will be performed and the 2-sided 95% CI will be provided.

All the subset analyses will be unstratified and exploratory; no adjustment for multiplicity will be performed. In the case of a low number of participants within a category (<5% of the FAS) and low number of events within the category (< 10 events) the categories will be pooled or in cases where no meaningful pooling can be performed, the category may not be summarized.

To assess the heterogeneity of treatment effects for EFS and OS across the subgroup levels, a Cox's regression models will be fitted with EFS and OS as the dependent variable,

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

respectively, and subgroup, treatment, and the treatment-by-subgroup interaction as explanatory variables.

- Model: treatment + subgroup + treatment×subgroup-variable

A p-value for the interaction test (Wald's test) will be provided together with the HR and corresponding 95% CI for the interaction model parameter.

The HR for EFS and OS and corresponding 95% CIs for each subgroup will be presented in a forest plot.

The difference in CR rates and corresponding 95% CIs for each subgroup will be presented in a forest plot.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline summaries

The following analyses will be based on the FAS overall and separately by treatment arm in Cohort A. Baseline data for Cohorts B1 and B2 will be provided in listings only.

6.5.1.1. Demographic and baseline characteristics

Demographic characteristics will be summarized by treatment arm and overall in Cohort A using the following information from the 'Demography' and 'ECOG Performance Status' eCRF pages.

- Demographic characteristics
  - Sex: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiracial, Not Reported
  - Ethnic origin:
    - Hispanic or Latino or of Spanish origin
    - Not Hispanic or Latino or of Spanish origin
    - Not Reported
  - Age (years): summary statistics
  - Age categories:
    - < 65 years, ≥ 65 years
    - 65-<75, 75-<85, ≥ 85 years
  - Geographic Region (as applicable):
    - North America
    - Latin America
    - Western Europe
    - Eastern Europe

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

  - Middle East
  - Australasia
  - Asia
  - Africa
- Eastern Cooperative Oncology Group (ECOG) Performance Status at baseline: 0, 1, 2, 3, and 4

Center codes will be used for the determination of the participant's geographic region.

The listing of demographics and baseline characteristics will include the following information: participant identifier, treatment arm/cohort, age, sex, race, ethnicity, and ECOG performance status at baseline.

6.5.1.2. Medical history

Medical history will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized from the 'Significant Medical History' eCRF page. Medical history will be summarized as the numbers and percentages of participants by MedDRA PT as event category and MedDRA primary SOC as summary category. Each participant will be counted only once within each PT or SOC.

Medical history will be displayed in terms of frequency tables: ordered by primary SOC and PT in alphabetical order.

6.5.1.3. Disease characteristics

Information on disease characteristics collected on 'Primary Diagnosis', 'Primary Cancer Therapy', 'Substance Use', and disease assessment eCRF pages will be summarized overall and by treatment arm in Cohort A and listed for Cohort B1 and B2. Summary statistics for Cohort A and a listing only for Cohorts B1 and B2 will be presented for the following, as applicable.

From the 'Primary Diagnosis' and 'Primary Cancer Therapy' eCRF page:

- Primary diagnosis (summarize all categories (Urothelial (transitional cell) carcinoma, urothelial (transitional cell) carcinoma with squamous differentiation, urothelial (transitional cell) carcinoma with glandular differentiation, urothelial (transitional cell) carcinoma with variant histology (specify), Other) collected in the 'Primary Diagnosis' eCRF page)
- Disease Recurrence Type (Recurrence, Newly Diagnosed) (Cohort A only)
- Time since initial diagnosis to 'start date' (months), defined as ('start date' – date of initial diagnosis)/30.4375
- Reason participant did not undergo cystectomy (Participant Refused, Participant Ineligible) (Cohorts B1 and B2 only)

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Separately from the disease status eCRF pages based on investigator assessment ('Cystoscopy', and 'Oncology Biopsy') at baseline, except for Cohort B1 where the BICR biopsy data at baseline will be used:

- CIS per CRF: Yes, No
- Number of tumors identified by cystoscopy: Single, Multiple, Unknown
- Size of largest tumor identified by cystoscopy: <3cm, ≥3cm, Unknown
- Highest grade of tumors identified by biopsy: High, Low
- Worst T stage: CIS, Ta, T1, T2, T3, T4

From the 'Substance Use' eCRF page:

- Smoking history (never smoker vs current vs former smoker)

Listing of disease history will be provided for Cohort A and Cohorts B1 and B2 with all relevant data (as collected on the 'Primary Diagnosis' and 'Primary Cancer Therapy' eCRF pages) and derived variables as above.

6.5.1.4. Prior anti-cancer therapies

The prior anti-cancer therapies are collected under the 'Prior Cancer Therapy' and 'Prior anti-cancer Surgery' eCRF pages.

The number and percentage of participants in each of the following anti-cancer therapy categories will be tabulated:

- Participants with at least one type of prior anti-cancer therapy
- Participants with at least one prior anti-cancer drug therapy
- Participants with at least one prior anti-cancer surgery

Prior anti-cancer drug therapy will be summarized as follows based on the number and percentage of participants with the following:

- At least one prior anti-cancer drug therapy
- Number of prior anti-cancer drug therapy regimens: missing, 1, 2, 3, ≥4

The following will be provided in a listing:

- Prior BCG strain (Cohorts B1 and B2)
- Number of prior BCG doses (Cohorts B1 and B2)
- Prior BCG failure: persistence of CIS, recurrence (Cohorts B1 and B2)

The prior anti-cancer drugs will also be summarized based on the number and percentage of participants by the drug class and preferred term. A participant will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. The summary will be sorted on decreasing frequency of drug

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used.

Prior anti-cancer drug therapies will be included in the listing that follows with a flag to identify prior drug therapies. These will include the participant identification number, and all the relevant collected data-fields on the "Prior Cancer Therapy" eCRF page.

- Listing of anti-cancer drug therapies

6.5.2. Study conduct and participant disposition

The following analyses will be performed based on the FAS overall (screening only) and separately by treatment arm in Cohort A. Data for Cohorts B1 and B2 will be provided in listings only.

6.5.2.1. Participant disposition

The percentages below will be calculated based on the number of participants screened. The following will be summarized:

- Total number of participants screened overall
- n (%) of participants who discontinued prior to 'start date' overall and by reasons for discontinuation.

The percentages below will be calculated based on the number of participants in the FAS. Discontinuations will be presented overall and by reason for discontinuation:

- n (%) in each of the analysis sets defined in Section 4
- n (%) discontinued prior to receiving any study drug
- PF-06801591 disposition
  - n (%) discontinued treatment with PF-06801591 and by reason
  - n (%) ongoing treatment with PF-06801591
  - n (%) completed treatment with PF-06801591
- BCG disposition, separately for induction, re-induction and maintenance periods:
  - n (%) entering each period
  - n (%) discontinued treatment with BCG and by reason
  - n (%) ongoing treatment with BCG
  - n (%) completed treatment with BCG
- Follow-up
  - n (%) entered follow-up
  - n (%) discontinued follow-up and by reason
  - n (%) ongoing

A listing of participant disposition will be provided.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

A listing of participant discontinuations related to COVID-19 will be presented.

For Cohort A, the results of the randomization algorithm (according to IRT) will be summarized as follows:

- Number and percentage of randomized participants overall, by region, by country, and by center
- Number and percentage of randomized participants by randomization strata (IRT)
- Cross tabulation: stratum by IRT vs. stratum by eCRF where CIS is derived based on the data entered by the investigator on the "Oncology Biopsy" eCRF at Screening.

6.5.2.2. Protocol deviations

All protocol violations that impact the safety of the participants and/or the conduct of a study and/or its evaluation will be reported. These include:

- Participants who are dosed on the study despite not satisfying the inclusion criteria
- Participants who develop withdrawal criteria whilst on the study but are not withdrawn
- Participants who receive the wrong treatment
- Participants who receive an excluded concomitant medication
- Deviations from Good Clinical Practice (GCP).

A listing of protocol deviations will be provided.

A listing of protocol deviations related to COVID-19 will be provided. The identification of these and other CSR-reportable deviations will be based on the inclusion/exclusion criteria or other criteria presented in the protocol.

A listing of participants impacted by COVID-19 will be provided to describe how the individual's participation was altered in terms of investigator disease assessments, including a row for every follow-up visit, as well as extra rows for missed follow-up visits.

6.5.3. Study treatment compliance and exposure

The following analyses will be based on the safety analysis set by treatment arm in Cohort A for the following study drugs (administered alone or in combination):

- **BCG** suggested administration (full dose per BCG labeling guidelines):
  - Induction period: One dose every week for 6 consecutive weeks (i.e., C1D1, C1D8, C1D15, C1D22, C2D1, and C2D8).
  - Re-induction period, up to 6 consecutive weeks (following induction period, for participants with CIS at randomization who have persistent disease and participants with recurrence of high-grade Ta disease); suggested schedule is C4D1, C4D8, C4D15, C4D22, C5D1, and C5D8. Maintenance will then begin on C7D1. If TURBT

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

is performed, schedule for BCG re-induction should be modified according to the product label.

- Maintenance period (Arm A and Arm C): D1, D8, D15 during Cycles C4, C7, C13, C19 and C25. For participants that have a re-induction period, the maintenance period will begin at C7D1.

- **PF-06801591** administration: SC Q4W (last dose at Cycle 25) at a dose of 300 mg for Cohorts A and B1 or SC Q6W (last dose at Cycle 17) at 600 mg for Cohort B2. Doses of PF-06801591 may be recorded in the eCRF as 300 mg, 600 mg (Cohort B2 only), 0 mg, or 'partial dose', where a 'partial dose' is captured as "UNK" on the PF-06801591 CRF dosing page. A 'partial dose' should not be considered a 0 mg dose and must be included in the derivations for first and last dose of study drug.

6.5.3.1. Exposure to study drug

Exposure to each study drug will be summarized for Cohort A as follows.

Duration of exposure to BCG (weeks) =

- Overall: (last dose date of BCG – first dose date of BCG +7)/7.

Duration of exposure to PF-06801591 (weeks) =

For Cohorts A:

(last dose date of PF-06801591 – first dose date of PF-06801591 +28)/7

The number and percentage of participants with 1, 2, 3, 4, 5, 6 or > 6 non-zero doses of BCG during induction, 1, 2, 3, 4, 5, 6 or > 6 non-zero doses of BCG during re-induction, and 1 through 15 and >15 non-zero doses of BCG during maintenance will be summarized. In addition, the median and range of number of non-zero doses of BCG received by a participant in each period will be provided.

The number and percentage of participants with each of 1 through 25 or > 25 (Cohorts A) non-zero doses of PF-06801591 will be summarized. In addition, the median and range of non-zero doses of PF-06801591 received by a participant will be provided.

The summary of exposure to study drug will include:

- Duration of exposure to PF-06801591 (overall)
- Duration of exposure to BCG (overall)
- Doses received for PF-06801591 (overall)
- Doses received for BCG (overall and by period)

A listing of exposure to BCG and PF-06801591 will be provided with relevant information for Cohort A and Cohorts B1 and B2, as applicable.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

6.5.3.2. Dose reductions

Applicable to BCG. Dose reduction is defined as actual non-zero dose < 90% of the planned dose.

The number and percentage of participants with at least one dose reduction as well as a breakdown of the number of dose reductions (1, 2, 3, 4, 5, ≥6) will be summarized.

Applicable to PF-06801591. The dose for PF-06801591 can be entered as 0, 300 mg, 600 mg (Cohort B2 only), or 'partial dose'. Where a 'partial dose' is captured as "UNK" on the PF-06801591 CRF dosing page for all cohorts. For Cohort B2, 300mg dose is also a 'partial dose'. The number and percentage of participants with at least one 'partial dose' dose as well as a breakdown of the number of partial doses (1, 2, 3, 4, 5, ≥6) will be summarized for Cohort A.

6.5.3.3. BCG retention interruption

An interruption in BCG retention based on void time will be assessed. A void time <= 30 minutes, >30 minutes and ≤ 1 hour, and >1 hour and < 2 hours will be summarized by number and percentage of participants. In addition, the number and percentage of participants with at least one retention interruption as well as a breakdown of the number of retention interruptions (1, 2, 3, ≥4) by time category and overall will be summarized.

6.5.3.4. Dose delays

Applicable to PF-06801591 and BCG for Cohort A.

Dose Delay is the difference between the actual time between two consecutive non-zero doses and the planned time between the same two consecutive non-zero doses. 'Unknown' dose amounts for PF-06801591 are considered a non-zero dose.

For PF-06801591:

For Cohort A: Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – 28

For BCG:

Dosing in each period will be determined by the CRF disposition pages for when each period started and ended based on when participants started the next period.

- Induction: Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – 7,
- Re-induction: Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – 7,
- Maintenance:

  Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) –7, for doses within a cycle (i.e. between Days 1, 8, and 15 of Cycle 4, Cycle 7, Cycle 13, Cycle 19 and Cycle 25).

  Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – 70, between the last dose of the cycle prior to Cycle 7 and next dose (i.e. between Cycle 4 Day 15 to

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Cycle 7 Day 1)

Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) –154, between the last dose of the previous cycle (Cycle 7 and after) and next dose (i.e. between maintenance cycles 7 to 25 from Day 15 to Day 1 (i.e. Cycle 7 Day 15 to Cycle 13 Day 1)

For Cohort A, dose delays will be grouped into the following categories, Cycles 1-3 and Cycle 4-25, for both BCG and PF-06801591.

Cycles 1-3 (Cohort A only)

- No delay (0-3 days window)
- 4 or more days delay

Cycles 4-25 (Cohort A)

- No delay (0-7 days window)
- 8 or more days delay

For example, for PF-06801591, administered on a 4-week schedule, if a participant receives PF-06801591 on Cycle 1 Day 1 (study day 1) and the next PF-06801591 administration is on study day 33, this is considered as a 4-day delay.

The number and percentage of participants with delayed study drug administration and maximum length of delay, i.e., the worst case of delay if participants have multiple dose delays will be summarized.

6.5.4. Concomitant medications and non-drug treatments

The following analyses will be based on the safety analysis set by treatment arm in Cohort A.

Concomitant medications are medications, other than study drugs, which started prior to first dose date of study treatment and continued during the on-treatment period as well as those started during the on-treatment period. Concomitant medications will be summarized from the 'Concomitant Medications' eCRF page.

Summary of concomitant medications will include the number and percentage of participants by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. A participant will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under 'Unavailable ATC classification' category.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

A listing of concomitant medications will be created with the relevant information collected on the 'Concomitant Medications' eCRF page.

6.5.5. Subsequent anti-cancer therapies

The following analyses will be based on the FAS by treatment arm in Cohort A.
Anti-cancer drug treatment will be provided in a data listing with data retrieved from 'Follow-up Cancer Therapy' eCRF page.
Number and percentage of participants with any anti-cancer therapy will be tabulated overall and by type of therapy based on the data collected from the 'Follow-up Cancer Therapy', 'Follow-up Radiation Treatment' and 'On Study & Follow-up Cancer Surgery' eCRF pages for Cohort A only.

A listing of anti-cancer therapies will be provided for Cohort A and Cohorts B1 and B2.

6.6. Safety Summaries and Analyses

The Safety Analysis Set will be the primary analysis set for safety evaluations. Summaries of AEs and other safety parameters will be based on the safety analysis set by treatment arm in Cohort A and by Cohorts B1 and B2.

6.6.1. Adverse events

TEAEs are those events with onset dates occurring during the on-treatment period as defined in Section 3.5.1.

All analyses described will be based on TEAEs (started during the on-treatment period) if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). AEs outside the on-treatment period will be flagged in the listings.

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (i.e., no answer to the question 'Relationship with study treatment'). Related AEs are those related to any study drug (i.e., at least one of the study drugs).
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious = Yes).
- **Adverse Events Leading to Interruption of Study Treatment**: adverse events leading to interruption of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug interrupted).
- **Adverse Events Leading to Discontinuation of Study Drug:** adverse events leading to discontinuation of study drug (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- **Adverse Events Leading to Death:** adverse event leading to death (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

- **Immune-related Adverse Events (irAE):** irAEs (as identified according to the methodology outlined in Appendix 1 for a pre-specified search list of MedDRA PTs documented in the Safety Review Plan [SRP] and finalized for analysis of this study's data prior to database lock)

- **Injection site reaction (ISR) related to PF-06801591**: adverse events with relationship to PF-06801591 (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (i.e., no answer to the question 'Relationship with study treatment'), and with high level term (HLT) of Injection site reactions.

Unless otherwise specified, AEs will be summarized by number and percentage of participants with the AE in the category of interest as described above, by treatment arm in Cohort A and by Cohorts B1 and B2, primary SOC, and PT in decreasing frequency based on the frequencies observed for Arm A (for Cohort A) and Cohort B1 (for Cohorts B1 and B2).

Each participant will be counted only once within each SOC or PT. If a participant experiences more than one AE within a SOC or PT for the same summary period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

6.6.1.1. All adverse events

Adverse events will be summarized by worst severity (according to CTCAE v5.0) per participant, using the latest version of MedDRA PT as event category and MedDRA primary SOC body term as Body System category.

In case a participant has AEs with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The following tables will be created:

- The overall summary of AEs table will include the frequency (number and percentage) of participants with each of the following by treatment arm in Cohort A and by Cohorts B1 and B2, as applicable:
  - TEAEs
  - TEAEs, Grade ≥ 3
  - Related TEAEs
  - Related TEAEs, Grade ≥ 3
  - TEAEs leading to interruption of PF-06801591
  - TEAEs leading to interruption of BCG
  - TEAEs leading to discontinuation of PF-06801591
  - TEAEs leading to discontinuation of BCG
  - TEAEs leading to discontinuation of all study drugs

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

  - Related TEAEs leading to discontinuation of PF-06801591
  - Related TEAEs leading to discontinuation of BCG
  - Related TEAEs leading to discontinuation of all study drugs
  - Serious TEAEs
  - Related Serious TEAEs
  - TEAEs leading to death
  - Related TEAEs leading to death
  - irAEs
  - ISRs related to PF-06801591

- TEAEs by SOC and PT and worst grade
- Related TEAEs by SOC and PT and worst grade
- TEAEs leading to death by SOC and PT (Cohort A only)
- Related TEAEs leading to death by SOC and PT (Cohort A only)
- COVID-19 TEAEs by SOC and PT, as noted on the 'Adverse Event' eCRF.

A listing of all AEs will be provided with relevant information (e.g., SOC/PT, start and stop dates, causality, study drug action, medication error, SAE, and concomitant medications).

6.6.1.2. Adverse events leading to discontinuation of study drug

The frequency (number and percentage) of participants with each of the following will be presented for TEAEs leading to discontinuation of each study drug, by treatment arm in Cohort A and by Cohorts B1 and B2, as applicable:

- TEAEs leading to discontinuation of PF-06801591 by SOC and PT
- Related TEAEs leading to discontinuation of PF-06801591 by SOC and PT
- TEAEs leading to discontinuation of BCG by SOC and PT (Cohort A only)
- Related TEAEs leading to discontinuation of BCG by SOC and PT (Cohort A only)
- TEAEs leading to discontinuation of all study drugs by SOC and PT (Cohort A only)
- Related TEAEs leading to discontinuation of all study drugs by SOC and PT (Cohort A only)

The listing of all AEs leading to discontinuation of study drug will also be provided with the relevant information.

6.6.1.3. Adverse events leading to interruption of study treatment

AEs leading to interruption will be defined as AEs identified in the AE eCRF page with an action taken with study treatment of 'drug interrupted'.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The frequency (number and percentage) of participants with each of the following will be presented for TEAEs leading to interruption of each study drug by treatment arm in Cohort A:

- TEAEs leading to interruption of PF-06801591 by SOC and PT
- TEAEs leading to interruption of BCG by SOC and PT

The listing of all AEs leading to interruption of study treatment will also be provided with the relevant information for Cohort A and Cohorts B1 and B2.

6.6.2. Deaths

The frequency (number and percentage) of participants in the safety analysis set who died and who died within 30 days after last dose of study treatment as well as the reason for death, will be tabulated based on information from the 'Notice of Death' eCRF, by treatment arm in Cohort A and by Cohorts B1 and B2.

- All deaths
- Deaths within 30 days after last dose of study treatment
- Deaths due to COVID-19 (if >10% are due to COVID-19)
- Reason for Death
  - Disease under study
  - Study treatment toxicity
  - Unknown
  - Other.

In addition, date and cause of death will be provided in individual participant data listing together with selected dosing information (study treatment received, date of first / last administration, dose) and will include the following information:

- AEs with fatal outcome (list PTs of AEs with outcome=Fatal, as well as AEs of Grade 5),
- Flag for death within 30 days of last dose of study treatment.
- Flag for deaths related to COVID-19.

6.6.3. Serious adverse events

The frequency (number and percentage) of participants with each of the following will be presented for treatment-emergent SAEs by treatment arm in Cohort A and by Cohorts B1 and B2:

- SAEs by SOC and PT
- Related SAEs by SOC and PT

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The listings of all SAEs will also be provided with the relevant information with a flag for SAEs with onset outside of the on-treatment period for Cohort A and Cohorts B1 and B2.

6.6.4. Other significant adverse events

Adverse events of special interest in this study are immune-related adverse events and injection site reactions.

6.6.4.1. Immune related adverse events

The criteria for classification of an adverse event as an irAE is described in Appendix 1. The frequency (number and percentage) of participants with each of the following will be presented for irAEs, by treatment arm in Cohort A and by Cohorts B1 and B2, if applicable:

- irAEs leading to death, by Cluster and PT (Cohort A only)
- irAEs, by Cluster, PT, and worst Grade
- irAEs, Any Grade and Grade ≥ 3, by Cluster and PT (Cohort A only)
- irAEs leading to discontinuation of PF-06801591, by Cluster and PT
- irAEs leading to discontinuation of BCG, by Cluster and PT (Cohort A only)
- irAEs leading to discontinuation of all study drugs, by Cluster and PT (Cohort A only)
- Serious irAEs, by Cluster and PT

The listing of all irAEs will also be provided with the relevant information with a flag for irAEs with onset outside of the on-treatment period for irAE (i.e., irAEs not meeting the programmatic criteria in Step 2 as outlined in Table 25 of Appendix 1

The frequency (number and percentage) of participants with potential irAEs (i.e., AEs meeting the programmatic criteria Step 1 and Step 2 outlined in Table 25 of Appendix 1) will also be presented by treatment arm in Cohort A and by Cohorts B1 and B2.

6.6.4.2. Injection site reactions related to PF-06801591

The frequency (number and percentage) of participants with the following will be presented by treatment arm in Cohort A and by Cohorts B1 and B2

- ISRs related to PF-06801591 by HLT, PT, and worst grade

6.6.5. Laboratory data

Summaries of laboratory data will be based on the safety analysis set by treatment arm in Cohort A. For Cohorts B1 and B2 only listings will be provided.

6.6.5.1. Hematology and chemistry parameters

Laboratory results will be classified according to the CTCAE v5.0. Non-numerical qualifiers (with the exception of fasting flags) will not be taken into consideration in the derivation of

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

CTCAE criteria. Additional laboratory results that are not graded per CTCAE v5.0 will be presented according to the categories: below normal limit, within normal limits and above normal limit (according to the laboratory normal ranges).

Quantitative data will be summarized using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time (unscheduled measurements would therefore not be included in these summaries as described in Section 5.2.9). End of Treatment visit laboratory results will be summarized separately. The changes computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (i.e., Low, Normal, High).

Abnormalities classified according to CTCAE v5.0 will be described using the worst grade. For those parameters which are graded with two toxicities, the toxicities will be summarized separately.

Low direction toxicity grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity, and vice versa.

For **WBC differential counts** (total neutrophil [including bands], lymphocyte, monocyte, eosinophil, and basophil counts), the absolute value will be used when reported. When only percentages are available (this is mainly important for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count = (WBC count) × (Differential %value / 100)

If the range for the differential absolute count is not available (only range for value in % is available) then Grade 1 will be attributed to as follows:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 800/mm^3
- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 1500/mm^3

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium (CALCIO). Corrected Calcium is calculated from Albumin and Calcium as follows

Corrected calcium (mmol/L) = measured total Calcium (mmol/L) + 0.02 (40 – serum albumin [g/L])

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity. The ratios of test result over upper limit of normal (ULN) will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of participants with each of the following during the on-treatment period will be summarized by treatment arm in Cohort A and by Cohorts B1 and B2:

- ALT ≥ 3×ULN, ALT ≥ 5×ULN, ALT ≥ 10×ULN, ALT ≥ 20×ULN
- AST ≥ 3×ULN, AST ≥ 5×ULN, AST ≥ 10×ULN, AST ≥ 20×ULN
- (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN
- TBILI ≥ 2×ULN
- Concurrent ALT ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent AST ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and ALP > 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing)

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, i.e., a participant with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN. Liver function elevation and possible Hy's Law cases will be summarized using frequency counts and percentages.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different treatment arms, by graphically displaying

- peak serum ALT(/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN.
- peak serum AST(/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN.

In addition, a listing of all TBILI, ALT, AST and ALP values for participants with concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing) will be provided.

Parameters graded per CTCAE v5.0:

The laboratory toxicities will be tabulated using descriptive statistics (number of participants and percentages) during the on-treatment period. The denominator to calculate percentages

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

for each laboratory parameter is the number of participants evaluable for CTCAE grading (i.e., those participants for whom a Grade 0, 1, 2, 3 or 4 can be derived).

- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade (Grade 0, 1, 2, 3, or 4) for the summary.
- The number and percentage of participants with newly occurring or worsening laboratory abnormalities during the on-treatment period will be summarized by worst grade on-treatment (Grade 1, 2, 3, 4, Grade 3/4 and any grade (Grades 1-4)).

The above analyses apply to hematology and chemistry evaluations which can be graded per CTCAE v5.0.

Parameters with CTCAE grades not available:

Hematology and chemistry evaluations which cannot be graded per CTCAE criteria will be summarized in a shift table by frequency (number and percentage) of participants.

6.6.5.2. Other laboratory parameters

The listings of laboratory results will be provided for all laboratory parameters. The listings will be sorted by parameters and assessment dates or visits for each participant. Laboratory values that are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges. A listing of CTCAE grading will also be generated for those laboratory tests.

6.6.6. Electrocardiogram

QT intervals will be corrected for HR (QTc) using standard correction factors [ie, Fridericia's (default correction), Bazett's, and possibly a study specific factor, as appropriate.

QTcB and QTcF will be derived based on RR and QT (see below).

Selecting Primary QT Correction for Heart Rate

The analysis of QT data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected value, denoted QTc, which is independent of heart rate. This QTc interval is intended to represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis we will use some of the correction methods described below. The QT interval corrected for heart rate by the Bazett's formula, QTcB, is defined as

$$QTcB = \frac{QT}{\sqrt{RR}},$$

the QT interval corrected for heart rate by the Fridericia's formula, QTcF, is defined as

$$QTcF = \frac{QT}{\sqrt[3]{RR}},$$

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

where RR represents the RR interval of the ECG, in seconds, and can be estimated as 60/Heart Rate.

Data will be summarized using QTcF and QTcB. Additional concentration-corrected QT analyses may be performed. Data may be pooled with other study results and/or explored further with PK/PD models.

ECG Summaries

The following analyses will be performed for each applicable ECG parameters (RR, PR, QRS, QT, heart rate, and QTc) by treatment arm in Cohort A during the on-treatment period. The denominator to calculate percentages for each category is the number of participants evaluable for the category.

- Frequency (number and percentage) of participants with notable ECG values according to the following categories:
  - QT increase from baseline >30 ms, >60 ms
  - QTc increase from baseline >30 ms, >60 ms
  - QT > 450 ms, > 480 ms, > 500 ms
  - QTc > 450 ms, > 480 ms, > 500 ms
  - Heart rate ≤ 50 bpm and decrease from baseline ≥ 20 bpm
  - Heart rate ≥ 120 bpm and increase from baseline ≥ 20 bpm
  - PR ≥ 220 ms and increase from baseline ≥ 20 ms
  - QRS ≥ 120 ms

Participants with notable ECG interval values and qualitative ECG abnormalities will be listed for each participant and time point and the corresponding notable values and abnormality findings will be included in the listings.

For Cohorts B1 and B2, only ECG listings will be provided.

7. INTERIM ANALYSES (COHORT A ONLY)

7.1. Introduction

The goals of the interim analyses are to allow early stopping of treatment arm(s) for futility or efficacy.

Prior to protocol amendment 5, two interim (IA1(futility only) and IA2) and final (FA) analyses were planned for the primary endpoint of Cohort A. IA1 for EFS was conducted after 658 participants were randomized to Cohort A for all 3 arms combined. As IA1 was for futility only, no alpha was spent at IA1.

Based on protocol amendment 5, the IA2 will be removed, and the FA for EFS based on a calendar-based data cut-off date will be conducted. Randomization to Cohort A was completed on 16 November 2021 and the EFS FA data cut-off date will be set to allow for approximately 3 years of follow-up after last randomized participant.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

The interim efficacy analysis of OS will be performed as described in Sections 5.1.1 and 5.1.2 using the methodology described in Section 6.2.2.2.

As this is an open-label randomized study, the treatment is unblinded on participant level. However, the aggregate/cumulative data summaries by treatment arm will be unavailable to the study team and the external investigators until the database snapshot for the primary analysis.

Unblinded results from the interim futility analysis for EFS will not be communicated to the Sponsor's clinical team or to any party involved in the study conduct (apart from the independent statistician and E-DMC members) until the E-DMC has determined that either (i) EFS analysis has crossed the pre-specified boundary for efficacy or (ii) the study needs to be terminated due to any cause, including futility or safety reasons. Further details will be described in the E-DMC charter.

At the time of final EFS analyses, all participants will continue to be followed for OS until the final OS analysis (approximately 5 years after the last participant is randomized).

Based on PA5, since EDMC review of interim efficacy data (IA2) is no longer applicable, the study team will be unblinded to aggregate/cumulative data summaries by treatment arm at the time of FA for EFS. No additional EDMC activities are foreseen after the FA for EFS.

7.2. Interim Analyses and Summaries (OS)

At each OS efficacy analysis time point, the critical boundaries for the group sequential test will be derived from the predefined spending function(s) as described in Section 5.1. The calculations of boundaries will be performed using EAST.

Let $u(t_1)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_1 and Z_F for efficacy at the interim and the final analysis, respectively.

The critical value $u(t_1)$ for the interim analysis of OS is determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1) \ ,$$

where P_0 denotes the probability under the null hypothesis and $\alpha(t_1)$ denotes the α spent based on the predefined spending function at information fraction t_1 (t_1 is calculated as the ratio of the number of OS events observed at the time of the cut-off for the interim analysis and the total number of OS events targeted for the final analysis).

The boundary for the final efficacy analysis will be calculated such that

$\alpha(t_1) + P_0(Z_1 < u(t_1), Z_F \geq u_F) = \alpha'$ where α'=0.002 or 0.025 for the analysis of OS determined according to the testing strategy described in Section 5.1.1.

As described in Section 5.1.2, if the number of OS events in the final analysis deviates from the target number of OS events, the final analysis criteria will be determined as above taking

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

into account the actual α spent at the interim analysis and the actual correlation between the two test statistics Z_I and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.025.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

8. REFERENCES

1. Anonymous. EuroQol- a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990; 16(3):199-208.

2. Bretz F, Maurer W, Brannath W and Posch. A graphical approach to sequentially rejective multiple test procedures. Statist. Med. 2009; 28:586–604.

3. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics. 38: 29-41, 1982.

4. Cambier S, Sylvester RJ, Collette L, et al. EORTC nomograms and risk groups for predicting recurrence, progression, and disease-specific and overall survival in nonmuscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1-3 years of maintenance Bacillus Calmette-Guerin. Eur Urol 2016; 69(1):60-9.

5. Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika; 26, 404-413.

6. Fernandez-Gomez J, Madero R, Solsona E, et al. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. J Urol 2009; 182(5):2195-203.

7. Huang B, Tian L, McCaw ZR, et al. Analysis of response data for assessing treatment effects in comparative clinical studies. Annals of Internal Medicine. 2020;173(5):368-74. https://doi.org/10.7326/M20-0104.

8. Huang B, Tian L, Talukder E, Rothenberg M, Kim DH, Wei LJ. Evaluating treatment effect based on duration of response for a comparative oncology study. JAMA oncology. 2018 Jun 1;4(6):874-6.

9. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.

10. Kalbfleisch JD, Prentice, RL. Statistical Analysis of Failure Time Data, 2nd Edition. Hoboken, Wiley Interscience, 2002.

11. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53: 457-81, 1958.

12. Lu K, Li D, and Koch G. Comparison between 2 controlled multiple imputation methods for sensitivity analyses of time-to-event data with possibly informative censoring. Statistics in Biopharmaceutical Research, 2015;Vol. 7, 199-213.

13. Oddens J, Brausi M, Sylvester R, et al. Final results of an EORTC-GU cancers group randomized study of maintenance bacillus Calmette-Guérin in intermediate- and high-risk Ta, T1 papillary carcinoma of the urinary bladder: one-third dose versus full dose and 1 year versus 3 years of maintenance. Eur Urol 2013; 63(3):462-72.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

14. Royston, P. and Parmar, M.K. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine*, 2011; *30*(19), pp.2409-2421.

15. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 1980; 67:145-53.

16. Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol 2006; 49(3):466-5; discussion 475-7.

17. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health 2000.

18. Uno H, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol 2014; 32.

19. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

9. APPENDICES

Appendix 1. Immune-Related Adverse Events

The MedDRA PTs and clusters for irAEs are defined in the SRP for PF-06801591.

Immune-related AEs (irAEs) will be identified using a combination of programmatic checks and medical review as outlined in Table 25. This case definition is hierarchical, i.e., each step is only checked for participants and events that have already met the prior step.

Table 25. Case Definition for irAEs

Step	Selection Criteria	
1	Adverse Event (AE) selected based on a list of pre-specified MeDRA PTs within clusters. These are included in the SRP. If AE matches the list, then it is included in the next step.	
2	AE onset on or after the first dose of study drug and on or before 90 days after last dose of study treatment.	This is regardless of start of new anti-cancer drug therapy and regardless of TEAE classifications.
3	AE treated with corticosteroids or other immunosuppressant therapy. For endocrinopathies only: AE required hormone replacement.	Look in the conmed pages for AEs where concomitant medications match any of the following A) conmed ATC code is in (H02A, H02B, D07, A01AC, S01BA, S01BB, L04AA, L04AB, L04AC, L04AD, L04AX, A07EA) and AE PT is in any of the irAE clusters. B) conmed ATC code is in (H03A, H03B) and AE PT is in one of the irAE clusters associated with "Immune-mediated endocrinopathies: Thyroid disorders" C) conmed ATC code is A10A and AE PT is in the irAE cluster associated with "Immune-mediated endocrinopathies: Type I Diabetes Mellitus"
4	4A) AE has no clear alternative etiology other than the immune-mediated etiology. OR 4B) AE has a histopathology / biopsy consistent with immune-mediated event.	Steps 4A) and 4B) are assessed by the clinical team based on medical review of the data available in the CRF and SAE report (if applicable). Steps 4A) and 4B) are not completed programmatically.

The data set associated with irAEs may be further refined based on medical review. The final data set including any changes based on medical review will be the basis of the irAE analyses.

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Appendix 2. List of Abbreviations

Abbreviation	Term
ADA	anti-drug antibodies
AE	adverse event
AESI	adverse events of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical
BCG	bacille calmette guérin
BLQ	below the level of quantification
bpm	beats per minute
CD8	cluster of differentiation-8
CI	confidence interval
CIS	carcinoma in situ
CR	complete response
CRF	case report form
CSR	clinical study report
CT	computed tomography
CTCAE	common terminology criteria for adverse events
C_{max}	maximum observed concentration
C_{trough}	concentration at predose/trough
CV	coefficient of variation
DNA	deoxyribonucleic acid
DoCR	duration of complete response
DSS	disease-specific survival
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
eDISH	evaluation of drug-induced serious hepatotoxicity
EFS	event-free survival
EORTC	European Organisation for Research and Treatment of Cancer
EORTC-GU	EORTC-Genito-Urinary
EOT	end of treatment
EQ-5D-5L	EuroQol 5 Dimensions, 5-Level
FA	final analyses
FAS	Full analysis set
FFPE	formalin-fixed paraffin-embedded
GCP	Good clinical practice

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Abbreviation	Term
HR	hazard ratio
IA	interim analysis
ICD	informed consent document
IgA	immunoglobulin A
IgG	immunoglobulin G
irAE	immune-related adverse event
IRT	interactive response technology
LLQ	lower limit of quantitation
LLN	Lower limit of normal
MNT	maintenance
MRI	magnetic resonance imaging
MWPC	meaningful within person change
N+	lymph node positive disease
NAb	neutralizing antibodies
NCI	National Cancer Institute
ND	not done
NMIBC	non-muscle invasive bladder cancer
NR	Not reached
NS	no sample
OS	overall survival
PD	progressive disease
PGIC	patient global impression of change
PGIS	patient global impression of severity
PH	proportional hazards
PK	pharmacokinetic
PRO	patient reported outcome
PT	preferred term
PTAB	Patient Treatment Administration Burden
Q1	First quartile
Q3	Third quartile
Q12W	Once every 12 weeks
Q24W	Once every 24 weeks
QLQ-C30	Quality of Life Questionnaire–30 (items)
QLQ NMIBC24	Quality of Life Questionnaire Non-Muscle Invasive Bladder Cancer 24 (items)
QT	time from the beginning of the QRS complex to the end of the T wave
QTc	QT interval corrected for heart rate
QTcF	QT corrected for heart rate using Fridericia's formula
QTcB	QT corrected for heart rate using Bazett’s formula
RCI	repeated confidence interval

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)

Abbreviation	Term
RNA	ribonucleic acid
RMST	restricted mean survival time
ROW	rest of world
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
SD	standard deviation
SOC	system organ class
SRP	safety review plan
T1	stage of cancer in which the cancer cells are only growing in the most superficial layer of tissues and have not grown into deeper tissues; in bladder cancer, T1 is defined as an invasion into the lamina propria without invasion into the muscularis propria
Ta	stage of bladder cancer defined as a non-invasive papillary carcinoma
TBILI	total bilirubin
TEAE	treatment-emergent adverse event
TSQ	Treatment Satisfaction Questionnaire
TURBT	transurethral resection of the bladder tumor
UK	United Kingdom
ULN	upper limit of normal
US	United States
VAS	visual analog scale
WBC	white blood cell

090177e1a25f065f\Approved\Approved On: 11-Dec-2024 10:57 (GMT)