- Objective: To compare overall survival (OS) for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo. - Hypothesis (H2): The combination of pembrolizumab + lenvatinib has superior OS	- OS, defined as the time from randomization to death due to any cause*
Secondary Objectives	Secondary Endpoints
- Objective: To compare objective response rate (ORR) as assessed by BICR according to RECIST 1.1, as defined in Section 4.2.1.1.1, for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo - Hypothesis (H3): The combination of pembrolizumab + lenvatinib has superior ORR per RECIST 1.1, as defined in Section 4.2.1.1.1, based on BICR	- Objective response (OR), defined as a confirmed complete response (CR) or partial response (PR)
- Objective: To evaluate the safety and tolerability for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	- Adverse events (AEs) and study intervention discontinuation due to AEs
- Objective: To compare the mean change from baseline in the global health status/quality of life (QoL), cough, chest pain, dyspnea, and physical functioning for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	- Change from baseline for the following patient-reported outcomes (PROs) scales/items: global health status/QoL (EORTC QLQ-C30 items 29 and 30), cough (EORTC QLQ-LC13 item 31), chest pain (EORTC QLQ-LC13 item 40), dyspnea (EORTC QLQ-C30 item 8), and physical functioning (EORTC QLQ-C30 items 1-5)

1.3 Schedule of Activities (SoA)

1.3.1 Initial Treatment Phase

NOTE: As of Amendment 007-06, lenvatinib and matching placebo have been removed from the study. Participants who remain on treatment will receive open-label pembrolizumab monotherapy as per protocol. Administration of lenvatinib and placebo, collection of ePRO data, and collection of most of the blood biomarkers have been removed from the SoA. Other study procedures specific to lenvatinib/matching have also been removed.

Study Period	**Screening**	**Treatment Cycle = 21 Days**									**EOT**	**Posttreatment**			**Notes**
												Safety Follow-up	**Follow-up**	**Survival Follow-up**	**All procedures are to be performed before study intervention administration unless otherwise indicated. Refer to Section 8.11 for visit details.**
Visit Timing/ Cycle Number	**-28 to -1**	**1**			**2**		**3**	**4**	**5**	**6 to 35**	**At DC**	**30 Days post last dose**	**Every 12 weeks**	**Every 12 weeks**	
Cycle Day (±3 days unless otherwise specified)		**1**	**8**	**15**	**1**	**15**	**1**	**1**	**1**	**1**	**–**				
Scheduling Window (Days)		**+ 3**	± 3	±3	± 3	± 3	± 3	± 3	± 3	± 3		**+ 7**	**± 7**	**± 14**	
Administrative and General Procedures															
Informed consent	X														ICF must be signed before any protocol-specific screening procedures are performed. Additional consent is required at PD.
Inclusion/exclusion criteria	X														
Participant identification card	X	X													Update at C1D1.
Demographic and medical history	X														

Study Period	**Screening**	**Treatment Cycle = 21 Days**									**EOT**	**Posttreatment**			**Notes**
												Safety Follow-up	**Follow-up**	**Survival Follow-up**	**All procedures are to be performed before study intervention administration unless otherwise indicated. Refer to Section 8.11 for visit details.**
Visit Timing/ Cycle Number	**-28 to -1**	**1**			**2**		**3**	**4**	**5**	**6 to 35**	**At DC**	**30 Days post last dose**	**Every 12 weeks**	**Every 12 weeks**	
Cycle Day (±3 days unless otherwise specified)		**1**	**8**	**15**	**1**	**15**	**1**	**1**	**1**	**1**	**–**				
Scheduling Window (Days)		**+ 3**	± 3	±3	± 3	± 3	± 3	± 3	± 3	± 3		+ 7	**± 7**	**± 14**	
Prior/concomitant medications	X	←								→					Prior concomitant medications received within 30 days before the first dose of study intervention through 30 days after the last dose of study intervention (or 90 days if used to treat an SAE) will be recorded.
Treatment randomization		X													Dose within 3 days of randomization.

In adult participants with treatment-naïve, metastatic NSCLC expressing PD-L1 (TPS ≥1%):

Objectives	Endpoints
Primary	
• Objective: To compare progression-free survival (PFS) as assessed by blinded independent central review (BICR) according to Response Evaluation Criteria in Solid Tumors, version 1.1 (RECIST 1.1), adjusted to follow a maximum of 10 target lesions and a maximum of 5 target lesions per organ, for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo – Hypothesis (H1): The combination of pembrolizumab + lenvatinib has superior PFS per RECIST 1.1, as defined in Section 4.2.1.1.1, based on BICR	• PFS, defined as the time from randomization to the first documented progressive disease (PD) or death due to any cause, whichever occurs first*
• Objective: To compare overall survival (OS) for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo – Hypothesis (H2): The combination of pembrolizumab + lenvatinib has superior OS	• OS, defined as the time from randomization to death due to any cause*
Secondary	
• Objective: To compare objective response rate (ORR) as assessed by BICR according to RECIST 1.1, as defined in Section 4.2.1.1.1, for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo – Hypothesis (H3): The combination of pembrolizumab + lenvatinib has superior ORR per RECIST 1.1, as defined in Section 4.2.1.1.1, based on BICR	• Objective response (OR), defined as a confirmed complete response (CR) or partial response (PR)

Extension Study in China

NOTE: As of Amendment 007-06, enrollment in the Extension Study in China has been stopped; the planned enrollment goal of 120 participants was not reached.

Lenvatinib and matching placebo are removed from the extension study and it will proceed identical to the global study as per this protocol amendment (eg, study endpoints, primary and secondary objectives, study procedures), with the exception of an additional supplemental statistical analysis plan (sSAP), which will provide details of analyses specific to Chinese participants.

Specific procedures to be performed during the study, as well as their prescribed times and associated visit windows, are outlined in the SoA in Section 1.3. Details of each procedure are provided in Section 8.

4.2 Scientific Rationale for Study Design

This Phase 3 study is being conducted to evaluate the efficacy and safety of pembrolizumab in metastatic NSCLC when administered in combination with lenvatinib. Lenvatinib (20 mg QD) will be evaluated in combination with the standard dose of pembrolizumab (200 mg Q3W).

4.2.1 Rationale for Endpoints

4.2.1.1 Efficacy Endpoints

This study will use PFS based on RECIST 1.1 criteria, adjusted to follow a maximum of 10 target lesions and a maximum of 5 target lesions per organ, as assessed by BICR as a primary endpoint and OR as a secondary efficacy endpoint. PFS and OR are acceptable measures of clinical benefit for a late-stage study that demonstrates superiority of a new antineoplastic therapy, especially if the magnitude of the effect is large and the therapy has an acceptable risk/benefit profile. The use of BICR and RECIST 1.1 to assess PFS and OR are typically considered acceptable by regulatory authorities. Images will be read by a central imaging vendor blinded to treatment assignment to minimize bias in the response assessments. In addition, the final determination of radiologic progression will be based on the central assessment of progression, rather than a local site investigator/radiology assessment. Real time determination of radiologic progression as determined by central review will be communicated to the site.

The primary efficacy endpoint OS has been recognized as the gold standard for the demonstration of superiority of a new antineoplastic therapy in randomized clinical studies.

For additional details about assessing efficacy endpoints using RECIST 1.1 and iRECIST, see Appendix 6.

4.2.1.1.1 RECIST 1.1

RECIST 1.1 will be used by the BICR when assessing images for efficacy measures and by the local site when determining eligibility (Section 8.2.1.5). Although traditional RECIST 1.1 references a maximum of 5 target lesions in total and 2 per organ, this protocol has implemented an adjustment to RECIST 1.1 to follow a maximum of 10 target lesions in total and 5 per organ.

4.2.1.1.2 Adjusted RECIST 1.1 for Immune-based Therapeutics iRECIST

RECIST 1.1 will be adapted to account for the unique tumor response characteristics seen following treatment with pembrolizumab (Section 8.2.1.6). Immunotherapeutic agents such as pembrolizumab may produce antitumor effects by potentiating endogenous cancer-specific immune responses. The response patterns seen with such an approach may extend beyond the typical time course of responses seen with cytotoxic agents, and patients treated with pembrolizumab may manifest a clinical response after an initial increase in tumor burden or even the appearance of new lesions. Thus, standard RECIST 1.1 may not fully capture the treatment benefits from immunotherapeutic agents such as pembrolizumab. Based on an analysis of participants with melanoma enrolled in KEYNOTE-001 (KN001), 7% of evaluable participants experienced delayed or early tumor pseudoprogression. Of note, participants who had PD by RECIST 1.1 but not by the immune-related response criteria [Wolchok, J. D., et al 2009] had longer OS than participants with PD by both criteria [Hodi, F. S., et al 2014]. Additionally, the data suggest that RECIST 1.1 may underestimate the benefit of pembrolizumab in approximately 15% of participants. These findings support the need to apply an adjustment to RECIST 1.1 that takes into account the unique patterns of atypical responses in immunotherapy and enables treatment beyond initial radiographic progression, if the participant is clinically stable.

iRECIST assessment has been developed and published by the RECIST Working Group, with input from leading experts from industry and academia, along with participation from the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA) [Seymour, L., et al 2017]. The unidimensional measurement of target lesions, qualitative assessment of nontarget lesions, and response categories are identical to RECIST 1.1, until progression is seen by RECIST 1.1. However, if a participant is clinically stable, additional imaging may be performed to confirm radiographic progression. iRECIST will be used by investigators to assess tumor response and progression and make treatment decisions as well as for exploratory efficacy analyses where specified.

4.2.1.2 Safety Endpoints

Safety parameters commonly used for evaluating investigational systemic anticancer treatments are included as safety endpoints including, but not limited to, the incidence of, causality, and outcome of AEs/serious AEs (SAEs); and changes in vital signs and laboratory values. Adverse events will be assessed as defined by the NCI Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0.

4.2.1.3 Patient-reported Outcomes

NOTE: As of Amendment 007-06, ePRO data are no longer being collected.

Symptomatic improvement is considered a clinical benefit and accepted by health authorities. As part of the analyses for this study, participants will provide information regarding their health-related quality of life (HRQoL) via the following assessment tools: European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30, EORTC QLQ-LC13, and EuroQoL EQ-5D-5L questionnaires. The ePROs are not pure efficacy or safety endpoints because they are affected by both PD and treatment tolerability.

4.2.1.3.1 EORTC QLQ-C30

EORTC QLQ-C30 is the most widely used cancer-specific, HRQoL instrument, which contains 30 items and measures 5 functional dimensions (physical, role, emotional, cognitive, and social), 3 symptom items (fatigue, nausea/vomiting, and pain), 6 single items (dyspnea, sleep disturbance, appetite loss, constipation, diarrhea, and financial impact), and a global health and QoL scale [Aaronson, N. K., et al 1993]. The EORTC QLQ-C30 is a psychometrically and clinically validated instrument appropriate for assessing QoL in oncology studies [Aaronson, N. K., et al 1993].

4.2.1.3.2 EORTC QLQ-LC13

The EORTC Quality of Life Questionnaire and Lung Cancer Module 13 (QLQ-LC13), a supplemental lung cancer-specific module used in combination with QLQ-C30, comprises multi-item and single-item measures of lung cancer-associated symptoms (cough, hemoptysis, dyspnea, and site-specific pain) and treatment-related symptoms (sore mouth, dysphagia, peripheral neuropathy, and alopecia) [Bergman, B., et al 1994]. It is scored on a 4-point scale (1 = not at all, 2 = a little, 3 = quite a bit, 4 = very much) and has been translated and validated into more than 60 languages.

The EORTC QLQ-C30 and QLQ-LC13 are the most frequently used PRO measures in lung cancer clinical studies. The reliability, validity, and practicality of these instruments have been reported [Bergman, B., et al 1994] [Aaronson, N. K., et al 1993].

4.2.1.3.3 EuroQoL EQ-5D-5L

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome and will provide data to develop health utilities for use in health economic analyses [Rabin, R. 2001]. The 5 health state dimensions in the EQ-5D-5L include the following: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is rated on a 5 point scale from 1 (no problem) to 5 (unable to/extreme problems). The EQ-5D-5L also includes a graded (0 to 100) vertical visual analog scale on which the participant rates his or her general state of health at the time of the assessment. This instrument has been used extensively in cancer studies and published results from these studies support its validity and reliability [Pickard, A. S., et al 2007].

4.2.1.4 Pharmacokinetic Endpoints

Prior to Amendment 05, blood samples were obtained to measure the PK of pembrolizumab and lenvatinib. The Sponsor has determined sufficient pembrolizumab/lenvatinib PK data in NSCLC participants have been collected. As such, no further PK samples will be collected.

Standard PK parameters of clearance (CL) and volume of distribution (V) at steady state are planned to be calculated for lenvatinib when coadministered with pembrolizumab, using the accepted mixed-effects modeling approach. PK data from this study may be combined with data from other studies and analyzed using standard population PK techniques to further characterize basic PK parameters, evaluate the effect of extrinsic and intrinsic factors in support of the proposed dosing regimen, and evaluate safety in the proposed participant population.

4.2.1.5 Planned Exploratory Biomarker Research

Cancer immunotherapies represent an important and novel class of antitumor agents. However, the mechanism of action of these exciting new therapies, including novel combinations with antiangiogenesis therapy, is not completely understood and much remains to be learned regarding how best to leverage these new drugs in treating patients. Thus, to aid future patients, it is important to investigate the determinants of response or resistance to cancer immunotherapy and other treatments administered, as well as determinants of AEs in the course of our clinical studies. These efforts may identify novel predictive/PD biomarkers and generate information that may better guide single agent and combination therapy with immuno-oncology drugs. To identify novel biomarkers, biospecimens (ie, blood components, tumor material) will be collected to support analyses of cellular components (eg, protein, deoxyribonucleic acid [DNA], ribonucleic acid [RNA], metabolites) and other circulating molecules. Investigations may include but are not limited to:

Germline (blood) genetic analyses (eg, SNP analyses, whole exome sequencing, whole genome sequencing)

This research may evaluate whether genetic variation within a clinical study population correlates with response to the treatment(s) under evaluation. If genetic variation is found to predict efficacy or AEs, the data might inform optimal use of therapies in the patient population. Furthermore, it is important to evaluate germline DNA variation across the genome in order to interpret tumor-specific DNA mutations.

Genetic (DNA) analyses from tumor

The application of new technologies, such as next generation sequencing, has provided scientists the opportunity to identify tumor-specific DNA changes (ie, mutations, methylation status, microsatellite instability) contributing towards the development/progression of cancer and/or driving response to therapy. Key molecular changes of interest to immuno-oncology drug development include the mutational burden of tumors and the clonality of T-cells in the tumor microenvironment. Increased mutational burden (sometimes referred to as a 'hyper-mutated' state) may generate neo-antigen presentation in the tumor microenvironment. To

conduct this type of research, it is important to identify tumor-specific mutations that occur across all genes in the tumor genome. Evaluation of molecular targets and signaling pathways including angiogenesis- or and growth factor related signaling pathways related to pembrolizumab and lenvatinib may also be explored. Thus, genome-wide approaches may be used for this effort. Note that in order to understand tumor-specific mutations, it is necessary to compare the tumor genome with the germline genome. Circulating tumor DNA and/or RNA may also be evaluated from blood samples.

Tumor and blood RNA analyses

Both genome-wide and targeted messenger RNA (mRNA) expression profiling and sequencing in tumor tissue and in blood may be performed to define gene signatures that correlate to clinical response to treatment with pembrolizumab or other immunotherapies. Pembrolizumab induces a response in tumors that likely reflects an inflamed/immune phenotype. Specific immune-related gene sets (ie, those capturing interferon-gamma transcriptional pathways) may be evaluated and new signatures may be identified. Individual genes related to the immune system and growth factor signaling pathways (eg, VEGF and FGF) may also be evaluated. MicroRNA profiling may also be pursued as well as exosomal profiling.

Proteomics and immunohistochemistry (IHC) using blood or tumor

Tumor and blood samples from this study may undergo proteomic analyses (eg, PD-L1 IHC). PD-L1 protein level in tumor sections, assessed by IHC, has been shown to correlate with response to pembrolizumab in patients with NSCLC, and an in vitro diagnostic (IVD) device has been developed for use with pembrolizumab in NSCLC. Preliminary data indicates that this association may also be true in additional cancer types (ie, triple negative breast cancer, head and neck, and gastric). Additional tumor or blood-derived proteins may also correlate with response to pembrolizumab and lenvatinib combination therapy. Therefore, tumor tissue may be subjected to proteomic analyses using a variety of platforms that could include but are not limited to immunoassays and liquid chromatography/mass spectrometry. This approach could identify novel protein biomarkers that could aid in patient selection for pembrolizumab (MK-3475) and lenvatinib combination therapy.

Other blood-derived biomarkers

In addition to expression on the tumor tissue, PD-L1, circulating cytokines and angiogenic factors, and other tumor derived proteins can be shed from tumor and released into the blood. Assays such as enzyme-linked immunoassay (ELISA) measure such proteins in serum. Correlation of expression with response to pembrolizumab and lenvatinib combination therapy may identify new approaches for predictive biomarkers in blood, representing a major advance from today's reliance on assessing tumor biomarkers. This research would serve to develop such assays for future clinical use.

Other molecular changes of interest include the subtype of T-cells in the tumor microenvironment. The T-cell repertoire from tumor tissue and blood components may be evaluated.

study. Eight of the participants had RCC, 2 had NSCLC, 2 had endometrial cancer, and 1 had melanoma. There were 2 DLTs at the dose of lenvatinib 24 mg/day + pembrolizumab 200 mg IV Q3W (1 participant had Grade 3 arthralgia and another had Grade 3 fatigue); hence, this was defined as the toxic dose. No DLTs were reported in the next 10 participants (expansion part), all of whom received the lenvatinib 20 mg/day + pembrolizumab 200 mg Q3W dose.

Based on review of all of the clinical data from these 13 participants, the MTD and RP2D were determined to be 20 mg of lenvatinib daily in combination with a fixed dose of 200 mg of pembrolizumab given Q3W. Based on the promising antitumor efficacy and tolerable safety profile seen in both the endometrial carcinoma and RCC expansion cohorts from Study 111/KEYNOTE-146 [Makker, V., et al 2018], 2 Phase 3 studies have been initiated for both of these tumor types, Study E7080-G000-309/KEYNOTE-775 and Study E7080-G000-307/KEYNOTE-581.

4.3.2.2 Pembrolizumab Dosing

The rationale for the use of a fixed dose of pembrolizumab in participants with solid tumors was based on the following:

- Similar efficacy and safety of pembrolizumab when dosed at 2 mg/kg Q3W, 10 mg/kg Q3W, or 10 mg/kg every 2 weeks in participants with melanoma and NSCLC

- Flat exposure-response relationships of pembrolizumab for both efficacy and safety in the dose range of 2 mg/kg Q3W to 10 mg/kg Q3W

- The lack of clinically relevant effect of tumor burden or indication on distribution behavior of pembrolizumab (as assessed by the population pharmacokinetics [PopPK] model)

- Pharmacology data showing full target saturation in both systemic circulation (inferred from PK data) and tumor (inferred from physiologically based PK analysis) at 200 mg Q3W

The choice of pembrolizumab 200 mg Q3W as an appropriate fixed dose was based on simulations performed using the PopPK model of pembrolizumab, which showed the following for a fixed dose of 200 mg Q3W:

- Provide similar control of PK variability as weight-based dosing, with considerable overlap in the distribution of exposures from the 200 mg Q3W fixed dose and 2 mg/kg Q3W dose

- Maintain individual participant exposures in the exposure range associated with maximal efficacy response that was established in participants with melanoma and NSCLC

6.4 Study Intervention Compliance

Interruptions from the protocol-specified treatment plan for >12 weeks (pembrolizumab) require consultation between the investigator and the Sponsor and written documentation of the collaborative decision on participant management.

6.5 Concomitant Therapy

Medications or vaccinations specifically prohibited in the exclusion criteria are not allowed during the treatment period. If there is a clinical indication for any medication or vaccination specifically prohibited, discontinuation from study intervention may be required. The investigator is to discuss prohibited medication/vaccination with the Sponsor's Clinical Director. The final decision on any supportive therapy or vaccination rests with the investigator and/or the participant's primary physician, but the decision to continue the participant on study intervention requires the mutual agreement of the investigator, the Sponsor, and the participant.

All prior medications (including over-the-counter medications) administered within 30 days before the first dose of study drug and any concomitant therapy administered to the participant during the course of the study (starting at the date of informed consent) until 30 days after the final dose (or 90 days if used to treat an SAE) of study drug will be recorded. Additionally, all diagnostic, therapeutic, or surgical procedures relating to malignancy should be recorded. Any medication that is considered necessary for the participant's health and that is not expected to interfere with the evaluation of or interact with the study interventions may be continued during the study.

6.5.1 Allowed Concomitant Medications

Treatment of complications or AEs, or therapy to ameliorate symptoms (including blood products, blood transfusions, fluid transfusions, antibiotics, and antidiarrheal drugs), may be given at the discretion of the investigator, unless it is expected to interfere with the evaluation of (or to interact with) the study intervention. Antiemetic or any other prophylaxis should be considered in accordance with institutional guidelines.

The following concomitant medications are also allowed:

- Hormone replacement therapy (HRT)
- Thyroid hormone suppressive therapy
- Anticoagulants, including low molecular weight heparin, warfarin, and anti-Xa agents
- Anti-inflammatory agents

Based on existing clinical study data, most irAEs were reversible and could be managed with interruptions of pembrolizumab, administration of corticosteroids and/or other supportive care. For suspected irAEs, ensure adequate evaluation to confirm etiology or exclude other causes. Additional procedures or tests such as bronchoscopy, endoscopy, skin biopsy may be included as part of the evaluation. Based on the severity of irAEs, withhold or permanently discontinue pembrolizumab and administer corticosteroids.

Modification and toxicity management guidelines for irAEs associated with pembrolizumab monotherapy, coformulations, or IO combinations are provided in Table 3.

irAEs	Toxicity Grade (CTCAEv4.0)	Action With Pembrolizumab Monotherapy, Coformulations or IO Combinations	Corticosteroid and/or Other therapies	Monitoring and Follow-up
Diarrhea / Colitis	Grade 2 or 3	Withhold	Administer corticosteroids (initial dose of 1-2 mg/kg prednisone or equivalent) followed by taper	Monitor participants for signs and symptoms of enterocolitis (ie, diarrhea, abdominal pain, blood or mucus in stool with or without fever) and of bowel perforation (ie, peritoneal signs and ileus) Participants with ≥Grade 2 diarrhea suspecting colitis should consider GI consultation and performing endoscopy to rule out colitis Participants with diarrhea/colitis should be advised to drink liberal quantities of clear fluids. If sufficient oral fluid intake is not feasible, fluid and electrolytes should be substituted via IV infusion.
	Recurrent Grade 3 or Grade 4	Permanently discontinue		
AST / ALT Elevation or Increased Bilirubin	Grade 2	Withhold	Administer corticosteroids (initial dose of 0.5-1 mg/kg prednisone or equivalent) followed by taper	Monitor with liver function tests (consider weekly or more frequently until liver enzyme value returned to baseline or is stable)
	Grade 3 or 4	Permanently discontinue	Administer corticosteroids (initial dose of 1-2 mg/kg prednisone or equivalent) followed by taper	
T1DM or Hyperglycemia	New onset T1DM or Grade 3 or 4 hyperglycemia associated with evidence of β-cell failure	Withhold [a]	Initiate insulin replacement therapy for participants with T1DM Administer antihyperglycemic in participants with hyperglycemia	Monitor participants for hyperglycemia or other signs and symptoms of diabetes

- Procedures conducted as part of the participant's routine clinical management (eg, blood count) and obtained before signing of ICF may be utilized for screening or baseline purposes provided the procedure met the protocol-specified criteria and were performed within the time frame defined in the SoA.

- Additional evaluations/testing may be deemed necessary by the investigator and or the Sponsor for reasons related to participant safety. In some cases, such evaluation/testing may be potentially sensitive in nature (eg, HIV, hepatitis C), and thus local regulations may require that additional informed consent be obtained from the participant. In these cases, such evaluations/testing will be performed in accordance with those regulations.

Repeat or unscheduled samples may be taken for safety reasons or for technical issues with the samples.

8.1 Administrative and General Procedures

8.1.1 Informed Consent

The investigator or medically qualified designee (consistent with local requirements) must obtain documented informed consent from each potential participant or their legally acceptable representative prior to participating in this clinical study or future biomedical research. If there are changes to the participant's status during the study (eg, health or age of majority requirements), the investigator or medically qualified designee must ensure the appropriate documented informed consent is in place.

8.1.1.1 General Informed Consent

Informed consent given by the participant or their legally acceptable representative must be documented on a consent form. The form must include the trial protocol number, trial protocol title, dated signature, and /agreement of the participant (or his/her legally acceptable representative) and of the person conducting the consent discussion.

A copy of the signed and dated informed consent form should be given to the participant (or their legally acceptable representative) before participation in the study.

The initial ICF, any subsequent revised written ICF, and any written information provided to the participant must receive the IRB/IEC's approval/favorable opinion in advance of use. The participant or his/her legally acceptable representative should be informed in a timely manner if new information becomes available that may be relevant to the participant's willingness to continue participation in the study. The communication of this information will be provided and documented via a revised consent form or addendum to the original consent form that captures the participant's or the participant's legally acceptable representative's dated signature.

The process for scan collection and transmission to the central imaging vendor can be found in the Site Imaging Manual. Tumor scans are strongly preferred to be acquired by computed tomography (CT). For the abdomen and pelvis, contrast-enhanced MRI may be used when CT with iodinated contrast is contraindicated, or when mandated by local practice. The same scan technique regarding modality, ideally the same scanner, and the use of contrast should be used in a participant throughout the study to optimize the reproducibility of the assessment of existing and new tumor burden and improve the accuracy of the assessment of response or progression based on scans. Note: for the purposes of assessing tumor scans, the term "investigator" refers to the local investigator at the site and/or the radiological reviewer located at the site or at an offsite facility.

MRI is preferred for brain scans; however, CT scans will be acceptable, if MRI is medically contraindicated.

Bone scans may be performed to evaluate bone metastases. Any supplemental scans performed to support a positive or negative bone scan, such as plain X-rays acquired for correlation, should be submitted to the central imaging vendor.

Participant eligibility will be determined using local assessment (investigator assessment) based on RECIST 1.1. All scheduled scans for each participant will be submitted to the central imaging vendor. In addition, unscheduled scans to determine PD and scans obtained for other reasons, but demonstrate radiologic progression, are to be submitted to the central imaging vendor.

When the investigator identifies disease progression per RECIST 1.1, but elects to implement iRECIST, the investigator will assess for confirmation of progression by iRECIST at subsequent time points. Images should continue to be submitted to the central imaging vendor.

8.2.1.1 Initial Tumor Scans

Initial tumor scans at screening must be performed within 28 days before the date of randomization. The site study team must review screening scans to confirm the participant has measurable disease per RECIST 1.1.

The screening scans must be submitted to the central imaging vendor for retrospective eligibility review.

Tumor scans performed as part of routine clinical management is acceptable for use as screening tumor scans if it is of diagnostic quality and performed within 28 days before the date of randomization and can be assessed by the central imaging vendor.

If brain scan is performed to document the stability of existing metastases, an MRI scan should be used if possible. If MRI is medically contraindicated, CT with contrast is an acceptable alternative.

It is best practice and strongly recommended that the ePROs are administered to randomized participants before drug administration, AE evaluation, and disease status notification. If the participant does not complete the ePROs at a scheduled time point, the MISS_MODE form must be completed to capture the reason the assessment was not performed.

8.3 Safety Assessments

Details regarding specific safety procedures/assessments to be performed in this study are provided. Planned time points for all safety assessments are provided in the SoA.

8.3.1 Physical Examinations

The investigator or qualified designee will perform a complete physical examination during the screening period. Clinically significant abnormal findings should be recorded as medical history. The time points for full physical examination are described in the SoA (Section 1.3). After the first dose of study treatment, new clinically significant abnormal findings should be recorded as AEs.

For cycles that do not require a full physical examination per the SoA (Section 1.3), the investigator or qualified designee will perform a directed physical examination as clinically indicated before study treatment administration. New clinically significant abnormal findings should be recorded as AEs.

Investigators should pay special attention to clinical signs related to previous serious illnesses.

8.3.2 Vital Signs

NOTE: As of Amendment 007-06, lenvatinib and matching placebo have been removed from the study. Repeat BP measurements for monitoring participants receiving lenvatinib are no longer required. This section has been updated accordingly.

The investigator or qualified designee will take vital signs at screening, before the administration of each dose of study treatment and during the follow-up period, as specified in the SoA (Section 1.3). Vital signs include temperature, heart rate, respiratory rate, weight, and blood pressure. Height will be measured at Visit 1 only.

8.3.3 Electrocardiograms

NOTE: As of Amendment 007-06, lenvatinib and matching placebo have been removed from the study. Additional ECG measurements for monitoring participants receiving lenvatinib are no longer required. This section has been updated accordingly.

Electrocardiograms (ECGs) will be obtained as designated in the SoA (Section 1.3). Complete, standardized, 12-lead ECG recordings that permit all 12 leads to be displayed on a single page with an accompanying lead II rhythm strip below the customary 3 × 4 lead format are to be used. In addition to a rhythm strip, a minimum of 3 full complexes should be recorded from each lead simultaneously. Participants must be in the recumbent position for a

period of 5 minutes prior to the ECG. The Fridericia correction method for calculating QTc will be used.

An ECG abnormality may meet the criteria of an AE as described in this protocol (see Appendix 3) and the eCRF Entry Guidelines. In these instances, the AE corresponding to the ECG abnormality will be recorded on the appropriate case report form (CRF).

8.3.4 Echocardiograms or Multigated Acquisition Scans

A MUGA scan (using a technetium-based tracer) or an ECHO will be performed to assess LVEF as designated in the SoA (Section 1.3). Additional assessments may be performed as clinically indicated.

MUGA scans or ECHOs should be performed locally in accordance with the institution's standard practice. MUGA scans are the preferred modality. However, whichever modality is used for an individual participant at baseline should be repeated for all subsequent LVEF assessments for that participant.

LVEF, as assessed by the institution, will be entered in the electronic case report form (eCRF). Investigator assessment will be based on institutional reports.

8.3.5 Clinical Safety Laboratory Assessments

8.3.5.1 Laboratory Safety Evaluations (Hematology, Chemistry and Urinalysis)

- Refer to Appendix 2 for the list of clinical laboratory tests to be performed and to the SoA for the timing and frequency.

- The investigator or medically qualified designee (consistent with local requirements) must review the laboratory report, document this review, and record any clinically relevant changes occurring during the study in the AE section of the case report form (CRF). The laboratory reports must be filed with the source documents. Clinically significant abnormal laboratory findings are those which are not associated with the underlying disease, unless judged by the investigator to be more severe than expected for the participant's condition.

- CBC with differential and clinical chemistry results must be reviewed before administration of study intervention. Electrolytes such as potassium, calcium, and magnesium should be monitored and abnormalities, when considered clinically significant, should be corrected in all participants before starting study intervention.

- All protocol-required laboratory assessments, as defined in Appendix 2, must be conducted in accordance with the laboratory manual and the SoA.

- If laboratory values from nonprotocol specified laboratory assessments performed at the institution's local laboratory require a change in study participant management or

are considered clinically significant by the investigator (eg, SAE or AE or dose modification), then the results must be recorded in the appropriate CRF (eg, SLAB).

- For any laboratory tests with values considered clinically significantly abnormal during participation in the study or within 30 days after the last dose of study intervention, every attempt should be made to perform repeat assessments until the values return to normal or baseline or if a new baseline is established as determined by the investigator.

8.3.5.2 Pregnancy Test

All women who are being considered for participation in the study, and who are not surgically sterilized or postmenopausal, must be tested for pregnancy within 24 hours of the first dose of study intervention. If a urine test is positive or not evaluable, a serum test will be required. Participants must be excluded/discontinued from the study in the event of a positive test result. Repeated pregnancy test (such as monthly testing) may be conducted if required by local regulations.

8.4 Adverse Events (AEs), Serious Adverse Events (SAEs), and Other Reportable Safety Events

The definitions of an AE or SAE, as well as the method of recording, evaluating, and assessing causality of AE and SAE and the procedures for completing and transmitting AE, SAE, and other reportable safety event reports can be found in Appendix 3.
Progression of the cancer under study is not considered an AE as described in Section 8.4.6 and Appendix 3.

Adverse events, SAEs, and other reportable safety events will be reported by the participant (or, when appropriate, by a caregiver, surrogate, or the participant's legally authorized representative).

The investigator and any designees are responsible for detecting, documenting, and reporting events that meet the definition of an AE or SAE as well as other reportable safety events. Investigators remain responsible for following up AEs, SAEs, and other reportable safety events for outcome according to Section 8.4.3.

The investigator, who is a qualified physician, will assess events that meet the definition of an AE or SAE as well as other reportable safety events with respect to seriousness, intensity/toxicity and causality.

8.4.1 Time Period and Frequency for Collecting AE, SAE, and Other Reportable Safety Event Information

All AEs, SAEs, and other reportable safety events that occur after the participant provides documented informed consent but before intervention allocation/randomization must be reported by the investigator if the participant is receiving placebo run-in or other run-in treatment, if the event cause the participant to be excluded from the study, or is the result of a

protocol-specified intervention, including but not limited to washout or discontinuation of usual therapy, diet, or a procedure.

- All AEs from the time of intervention allocation/randomization through 30 days following cessation of study intervention must be reported by the investigator.

- All AEs meeting serious criteria, from the time of intervention allocation/randomization through 90 days following cessation of study intervention or 30 days following cessation of study intervention if the participant initiates new anticancer therapy, whichever is earlier, must be reported by the investigator.

- All pregnancies and exposure during breastfeeding, from the time of treatment allocation/randomization through 120 days following pembrolizumab/placebo or 30 days following cessation of lenvatinib/placebo, whichever occurs last, must be reported by the investigator. If the participant initiates new anticancer therapy following discontinuation of study intervention, the time period for reporting pregnancies and exposure during breastfeeding is reduced to 30 days following cessation of study intervention.

  Exception: A positive pregnancy test at the time of initial screening is not a reportable event unless the participant has received study intervention.

- Additionally, any SAE brought to the attention of an investigator at any time outside the time period specified above must be reported immediately to the Sponsor if the event is considered related to study intervention.

Investigators are not obligated to actively seek AEs or SAEs or other reportable safety events in former study participants. However, if the investigator learns of any SAE, including a death, at any time after a participant has been discharged from the study, and he/she considers the event to be reasonably related to the study intervention or study participation, the investigator must promptly notify the Sponsor.

All initial and follow-up AEs, SAEs, and other reportable safety events will be recorded and reported to the Sponsor or designee within the time frames as indicated in Table 6.

8.4.2 Method of Detecting AEs, SAEs, and Other Reportable Safety Events

Care will be taken not to introduce bias when detecting AEs and/or SAEs and other reportable safety events. Open-ended and nonleading verbal questioning of the participant is the preferred method to inquire about AE occurrence.

8.4.3 Follow-up of AE, SAE, and Other Reportable Safety Event Information

After the initial AE/SAE report, the investigator is required to proactively follow each participant at subsequent visits/contacts. All AEs, SAEs, and other reportable safety events including pregnancy and exposure during breastfeeding, events of clinical interest (ECIs), cancer, and overdose will be followed until resolution, stabilization, until the event is otherwise explained, or the participant is lost to follow-up (as defined in Section 7.3). In addition, the investigator will make every attempt to follow all nonserious AEs that occur in randomized participants for outcome. Further information on follow-up procedures is given in Appendix 3.

8.4.4 Regulatory Reporting Requirements for SAE

Prompt notification (within 24 hours) by the investigator to the Sponsor of SAE is essential so that legal obligations and ethical responsibilities towards the safety of participants and the safety of a study intervention under clinical investigation are met.

The Sponsor has a legal responsibility to notify both the local regulatory authority and other regulatory agencies about the safety of a study intervention under clinical investigation. The Sponsor will comply with country-specific regulatory requirements and global laws and regulations relating to safety reporting to regulatory authorities, IRB/IECs, and investigators.

Investigator safety reports must be prepared for suspected unexpected serious adverse reactions (SUSARs) according to local regulatory requirements and Sponsor policy and forwarded to investigators as necessary.

An investigator who receives an investigator safety report describing an SAE or other specific safety information (eg, summary or listing of SAE) from the Sponsor will file it along with the IB and will notify the IRB/IEC, if appropriate according to local requirements.

8.4.5 Pregnancy and Exposure During Breastfeeding

Although pregnancy and infant exposure during breastfeeding are not considered AEs, any pregnancy or infant exposure during breastfeeding in a participant (spontaneously reported to the investigator or their designee) that occurs during the study are reportable to the Sponsor.

All reported pregnancies must be followed to the completion/termination of the pregnancy. Pregnancy outcomes of spontaneous abortion, missed abortion, benign hydatidiform mole, blighted ovum, fetal death, intrauterine death, miscarriage, and stillbirth must be reported as serious events (Important Medical Events). If the pregnancy continues to term, the outcome (health of infant) must also be reported.

8.4.6 Disease-related Events and/or Disease-related Outcomes Not Qualifying as AEs or SAEs

Efficacy endpoints as outlined in this section will not be reported to the Sponsor, as described in Section 8.4.1.

Specifically, the suspected/actual events covered in this exception include any event that is progression of the cancer under study.

The Sponsor will monitor unblinded aggregated efficacy endpoint events and safety data to ensure the safety of the participants in the study.

8.4.7 Events of Clinical Interest (ECIs)

Selected nonserious and SAEs are also known as ECIs and must be reported to the Sponsor.

Events of clinical interest for this study include:

1. An overdose of pembrolizumab, as defined in Section 8.5, that is not associated with clinical symptoms or abnormal laboratory results.

2. Any dose over the prescribed dose of lenvatinib associated with an AE.

3. An elevated AST or ALT lab value that is greater than or equal to 3X the upper limit of normal and an elevated total bilirubin lab value that is greater than or equal to 2X the upper limit of normal and, at the same time, an alkaline phosphatase lab value that is less than 2X the upper limit of normal, as determined by way of protocol-specified laboratory testing or unscheduled laboratory testing.*

   *Note: These criteria are based upon available regulatory guidance documents. The purpose of the criteria is to specify a threshold of abnormal hepatic tests that may require an additional evaluation for an underlying etiology. The study site guidance for assessment and follow-up of these criteria can be found in the Investigator Study File Binder (or equivalent).

8.5 Treatment of Overdose

NOTE: After amendment 007-06, lenvatinib and matching placebo were discontinued. This section was left unchanged for reference.

For purposes of this study, an overdose will be defined as any dose exceeding the prescribed dose as follows:

- Pembrolizumab: ≥5 times the protocol-specified dose.

- Lenvatinib: any dose above the protocol-specified dose if associated with an adverse event.

There is no specific antidote for an overdose of lenvatinib. Due to its high degree of plasma protein binding, lenvatinib is not expected to be dialyzable. Adverse reactions in patients receiving single doses of lenvatinib as high as 40 mg were similar to those in clinical studies at the recommended dose for differentiated thyroid cancer and RCC.

No specific information is available on the treatment of overdose of pembrolizumab or lenvatinib. In the event of overdose, the participant should be observed closely for signs of toxicity. Appropriate supportive treatment should be provided if clinically indicated.

All reports of pembrolizumab overdose with and without an AE and all reports of lenvatinib overdose with an AE must be reported by the investigator within 24 hours to the Sponsor either by electronic media or paper.

Reports of pembrolizumab overdose without any associated clinical symptoms or abnormal laboratory results, should be reported using the terminology “accidental or intentional overdose without adverse effect.”

8.6 Pharmacokinetics

As of Protocol Amendment 05, the Sponsor has determined that there is sufficient pembrolizumab/lenvatinib PK data in NSCLC participants. Consequently, no further PK samples for pembrolizumab and lenvatinib or ADA samples for pembrolizumab will be collected.

8.7 Pharmacodynamics

Pharmacodynamic parameters will not be evaluated in this study.

8.8 Planned Genetic Analysis Sample Collection

Samples are to be collected for planned analysis of associations between genetic variants in germline/tumor DNA and drug response. If a documented law or regulation prohibits (or local IRB/IEC does not approve) sample collection for these purposes, then such samples are not be collected at the corresponding sites.

8.9 Biomarkers

To identify novel biomarkers, the following biospecimens to support exploratory analyses of cellular components (eg, protein, RNA, DNA, metabolites) and other circulating molecules will be collected from all participants as specified in the SoA:

- Blood for genetic analysis
- Newly obtained/archival tissue sample for PD-L1 analysis

Sample collection, storage, and shipment instructions for the exploratory biomarker specimens will be provided in the laboratory manual.

9 STATISTICAL ANALYSIS PLAN

NOTE: Data from an interim safety and futility eDMC for LEAP-007 (data cutoff: 19-MAY-2021) indicated that the study met the prespecified nonbinding futility criteria for OS for the combination of lenvatinib plus pembrolizumab compared with placebo plus pembrolizumab. This futility analysis was requested at a previous eDMC meeting and conducted as described in the sSAP. Based upon these data and the recommendation of the eDMC, Amendment 007-06 was implemented to unblind the study and remove lenvatinib and matching placebo from the treatment arms. The prespecified interim and final analyses of the study described in the statistical analysis plan (SAP) will not be performed. Selected analyses of safety and ePRO endpoints will be performed at the end of the study; there will be no further analyses of efficacy endpoints.

This section outlines the statistical analysis strategy and procedures for the study. **As of Amendment 007-06, the study has been unblinded.** Changes made to primary and/or key secondary hypotheses, or the statistical methods related to those hypotheses, that occurred prior to study unblinding were documented in previous protocol amendment(s) (consistent with International Council for Harmonisation [ICH] of Technical Requirements for Pharmaceuticals for Human Use Guideline E9). Changes to exploratory or other nonconfirmatory analyses made after the protocol has been finalized, but before unblinding/final database lock, will be documented in a supplemental statistical analysis plan (sSAP) and referenced in the clinical study report (CSR) for the study. Post hoc exploratory analyses will be clearly identified in the CSR. Separate analysis plans (ie, separate documents from the sSAP) will be developed to detail PK and biomarker analyses. The ePRO analysis plan will be included in the sSAP.

9.1 Statistical Analysis Plan Summary

Key elements of the SAP are summarized below. The comprehensive plan is provided in Sections 9.2 through 9.12. **As of Amendment 007-06, the prespecified interim and final analyses of the study described in SAP will not be performed. Selected analyses of safety and ePRO endpoints will be performed at the end of the study; there will be no further analyses of efficacy endpoints. The SAP summary has been updated accordingly.**

Study Design Overview	Phase 3 study of pembrolizumab plus lenvatinib vs pembrolizumab plus placebo for 1L treatment of metastatic NSCLC in participants whose tumors express PD-L1 (TPS ≥1%)
Treatment Assignment	Approximately 620 participants will be randomized in a 1:1 ratio between 2 treatment arms: (1) pembrolizumab + lenvatinib and (2) pembrolizumab + placebo. Stratification factors are as follows: • Geographic region (East Asia vs non-East Asia) • ECOG PS (0 vs 1) • TPS (1%-49% vs ≥50%) **As of Amendment 007-06, the study is unblinded and all ongoing participants will continue treatment with open-label pembrolizumab monotherapy.**

9.4.1.3 Exploratory Endpoints

***DOR*:** the time from the earliest date of qualifying response until earliest date of PD or death from any cause, whichever comes first.

9.4.2 Safety Endpoints

Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, SAEs, fatal AEs, laboratory tests, and vital signs. Furthermore, specific events will be collected and designated as ECIs as described in Section 8.4.7.

9.4.3 Patient-reported Outcome Endpoints

The following secondary PRO endpoints will be evaluated as described in Section 4.2.1.3 and analyzed as described in Section 9.6.3:

Change from baseline in

- Global Health Status/QoL scale (QLQ-C30 items 29-30)
- Single-item symptom scales: cough (QLQ-LC13 item 31), chest pain (QLQ-LC13 item 40), and dyspnea (QLQ-C30 item 8)
- Physical functioning scale (QLQ-C30 items 1-5)

Time to true deterioration (TTD) in

- Global Health Status/QoL scale (QLQ-C30 items 29-30)
- Single-item symptom scales: cough (QLQ-LC13 item 31), chest pain (QLQ-LC13 item 40), and dyspnea (QLQ-C30 item 8)
- Physical functioning scale (QLQ-C30 items 1-5)
- Composite symptom endpoint: cough (QLQ-LC13 item 31), chest pain (QLQ-LC13 item 40), or dyspnea (QLQ-C30 item 8)

The TTD in global health status/QoL, cough, chest pain, dyspnea, and physical functioning is defined as the time from baseline to the first onset of a 10 or more points deterioration from baseline with confirmation by the subsequent visit of a 10 or more points deterioration from baseline. The TTD in the composite endpoint of cough, chest pain, or dyspnea is defined as the time to first onset of 10 or more points deterioration from baseline in any one of 3 scale items with confirmation by the subsequent visit of 10 points or more deterioration from baseline in the same scale as the first onset. The EQ-5D-5L will be evaluated as an exploratory endpoint. These analyses and other supportive PRO analyses will be described in the sSAP.

9.5 Analysis Populations

9.5.1 Efficacy Analysis Populations

The analyses of the primary efficacy endpoints are based on the ITT population. All randomized participants will be included in this population. Participants will be analyzed in the treatment group to which they are randomized. Details on the approach to handling missing data are provided in Section 9.6.

Extension Study in China

NOTE: As of Amendment 007-06, enrollment in the Extension Study in China has been stopped. This section has been updated accordingly.

The participants in China who were randomized in the extension study will not be included in the above global study primary efficacy analysis population. The ITT participants in China, including all participants in China randomized in the global study and the extension study, will be analyzed.

9.5.2 Safety Analysis Populations

Safety analyses will be conducted in the APaT population, which consists of all randomized participants who received at least 1 dose of study intervention. Participants will be included in the treatment group corresponding to the study intervention they actually received for the analysis of safety data using the APaT population. This will be the treatment group to which they were randomized, except for participants who take incorrect study intervention for the entire treatment period; such participants will be included in the treatment group corresponding to the study intervention actually received. Any participant who receives the incorrect study intervention for 1 cycle, but receives the randomized treatment for all other cycles, will be analyzed according to the randomized treatment group, and a narrative will be provided for any events that occur during the cycle for which the participant is incorrectly dosed.

At least 1 laboratory, vital sign, or ECG measurement obtained after at least 1 dose of study intervention is required for inclusion in the analysis of the respective safety parameter. To assess change from baseline, a baseline measurement is also required.

Extension Study in China

The participants in China who were randomized and treated in the extension study will not be included in the above global study primary safety analysis population. The APaT participants in China, including all randomized participants in China (in the global study and extension study) who received at least 1 dose of study treatment, will be analyzed.

9.5.3 Patient-reported Outcome Analysis Population

The analyses of PRO endpoints will be based on the PRO full analysis set (FAS) population following the ITT principle and ICH E9 guidelines. The PRO FAS population consists of all

randomized participants who have received at least 1 dose of study intervention and have completed at least 1 PRO assessment.

9.5.4 Pharmacokinetic Analysis Population

The population of PK analysis set includes all the participants who have received at least 1 dose of study intervention with documented dosing history in the lenvatinib + pembrolizumab arm and have measurable plasma levels of lenvatinib or serum levels of pembrolizumab.

9.6 Statistical Methods

NOTE: As of Amendment 007-06, the prespecified interim and final analyses of the study will not be performed. Selected analyses of safety and ePRO endpoints will be performed at end of study; there will be no further analyses of efficacy endpoints. The subsections below are retained for reference.

9.6.1 Statistical Methods for Efficacy Analyses

This section describes the statistical methods that address the primary and secondary efficacy objectives. Methods related to exploratory objectives will be described in the sSAP. Efficacy results that will be deemed to be statistically significant after consideration of the Type I error control strategy are described in Section 9.8. Nominal *p* values will be computed for other efficacy analyses, but should be interpreted with caution because of potential issues of multiplicity.

The stratification factors used for randomization (see Section 6.3.2) will be applied to all stratified analyses, in particular, the stratified log-rank test, stratified Cox model, and stratified Miettinen and Nurminen method [Miettinen, O. 1985]. If there are small strata, for the purpose of analysis, strata will be combined to ensure sufficient number of participants, responses and events in each stratum. Details regarding the pooling strategy will be prespecified in the sSAP prior to the database lock for the first analysis when each applicable endpoint will be analyzed, and decisions regarding the pooling will be based on a blinded review of response and event counts by stratum.

The efficacy analyses for OR, DOR and PFS will include responses and documented progression events that occur prior to Second-course treatment.

A summary of the primary analysis strategy for the key efficacy endpoints is provided in Table 7.

PDs/deaths are counted as events regardless of missed study visits or initiation of new anticancer therapy). The second sensitivity analysis considers discontinuation of treatment or initiation of an anticancer treatment after discontinuation of study-specified treatments, whichever occurs later, to be a PD event for participants without documented PD or death. If a participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied. The censoring rules for primary and sensitivity analyses are summarized in Table 8.

Table 8 Censoring Rules for Primary and Sensitivity Analyses of Progression-free Survival

Situation	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
No PD and no death; new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment	Progressed at treatment discontinuation for reasons other than CR; otherwise, censored at last disease assessment if still on study intervention or completed study intervention
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment	Censored at last disease assessment	Progressed at date of new anticancer treatment
PD or death documented after ≤1 missed disease assessment and before new anticancer therapy	Progressed at date of documented PD or death	Progressed at date of documented PD or death	Progressed at date of documented PD or death
PD or death documented immediately after ≥2 consecutive missed disease assessments or after new anticancer therapy	Censored at last disease assessment before the earlier date of ≥2 consecutive missed disease assessment and new anticancer therapy	Progressed at date of documented PD or death	Progressed at date of documented PD or death
Abbreviation: CR = complete response; PD = progressive disease.			

9.6.1.2 Overall Survival

The nonparametric Kaplan-Meier method will be used to estimate the survival curves. The treatment difference in survival will be assessed by the stratified log-rank test (based on the stratification factors defined in Section 6.3.2). A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (ie, the HR). The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported. The stratification factors used for randomization (Section 6.3.2) will be applied to both the stratified log-rank test and the stratified Cox model. Participants without documented death at the time of analysis will be censored at the

class terms) and events that meet predefined limits of change in laboratory, vital signs, and ECG parameters are either prespecified as "Tier 1" endpoints or will be classified as belonging to "Tier 2" or "Tier 3," based on observed proportions of participants with an event.

Tier 1 Events

Safety parameters or AEs of special interest that are identified a priori constitute "Tier 1" safety endpoints that will be subject to inferential testing for statistical significance. AEs that are immune-mediated or potentially immune-mediated are well documented and will be evaluated separately; however, these events have been characterized consistently throughout the pembrolizumab clinical development program, and determination of statistical significance is not expected to add value to the safety evaluation. Similarly, the combination of pembrolizumab and lenvatinib has not been associated with any new safety signals. Finally, there are no known AEs associated with participants with NSCLC for which determination of a *p* value is expected to impact the safety assessment. Therefore, there are no Tier 1 events for this protocol.

Tier 2 Events

Tier 2 parameters will be assessed via point estimates with 95% CIs provided for differences in the proportion of participants with events using the Miettinen and Nurminen method, an unconditional, asymptotic method [Miettinen, O. 1985].

Membership in Tier 2 requires that at least 10% of participants in any treatment group exhibit the event; all other AEs and predefined limits of change will belong to Tier 3. The threshold of at least 10% of participants was chosen for Tier 2 events because the population randomized in this study are in critical conditions and usually experience various AEs of similar types regardless of treatment; events reported less frequently than 10% of participants would obscure the assessment of the overall safety profile and add little to the interpretation of potentially meaningful treatment differences. In addition, Grade 3 to 5 AEs (≥5% of participants in one of the treatment groups) and SAEs (≥5% of participants in one of the treatment groups) will be considered Tier 2 endpoints. Because many 95% CIs may be provided without adjustment for multiplicity, the CIs should be regarded as a helpful descriptive measure to be used in safety review, not a formal method for assessing the statistical significance of the between-group differences.

Tier 3 Events

Safety endpoints that are not Tier 1 or 2 events are considered Tier 3 events. Only point estimates by treatment group are provided for Tier 3 safety parameters.

Continuous Safety Measures

For continuous measures such as changes from baseline in laboratory, vital signs, and ECG parameters, summary statistics for baseline, on-treatment, and change from baseline values will be provided by treatment group in table format.

9.6.4 Demographic and Baseline Characteristics

The comparability of the treatment groups for each relevant demographic and baseline characteristic will be assessed by the use of tables and/or graphs. No statistical hypothesis tests will be performed on these characteristics. The number and percentage of participants screened and randomized and the primary reason for screening failure and discontinuation will be displayed. Demographic variables, baseline characteristics, primary and secondary diagnoses, and prior and concomitant therapies will be summarized by treatment, either by descriptive statistics or categorical tables.

9.7 Interim Analyses

NOTE: As of Amendment 007-06, no interim analyses of the study will be performed. This section is no longer applicable.

9.7.1 Safety Interim Analyses

The DMC conducted regular safety monitoring. The timing of the safety monitoring was specified in the DMC charter. No further analysis is warranted.

9.8 Multiplicity

NOTE: As of Amendment 007-06, no interim or final analyses of the study will be performed. This section is no longer applicable.

9.9 Sample Size and Power Calculations

NOTE: As of Amendment 007-06, no interim or final analyses of the study will be performed. This section is retained for reference.

The study will randomize 620 participants in a 1:1 ratio into the pembrolizumab + lenvatinib and pembrolizumab + placebo arms. PFS and OS are primary endpoints for the study, with OR as the secondary endpoint.

For the PFS endpoint, based on a target number of 416 events at the final PFS analysis, the study has approximately 86.5% power to detect an HR of 0.7 at the initially allocated α=0.0055 (1-sided).

For the OS endpoint, based on a target number of 388 events and one interim analysis at approximately 78% of the target number of events, the study has approximately 90% power to detect an HR of 0.71 at the initially allocated α=0.0195 (1-sided).

Based on KEYNOTE-042 data, the above sample size and power calculations for PFS and OS assume the following:

PFS follows a piecewise exponential distribution, with a median of 5 months up to 6.5 months and then a median of 12 months thereafter for the control group.

10.3 Appendix 3: Adverse Events: Definitions and Procedures for Recording, Evaluating, Follow-up, and Reporting

10.3.1 Definition of AE

AE definition

- An AE is any untoward medical occurrence in a clinical study participant, temporally associated with the use of study intervention, whether or not considered related to the study intervention.

- NOTE: An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of a study intervention.

- NOTE: For purposes of AE definition, study intervention (also referred to as Sponsor's product) includes any pharmaceutical product, biological product, vaccine, diagnostic agent, or protocol specified procedure whether investigational or marketed (including placebo, active comparator product, or run-in intervention), manufactured by, licensed by, provided by, or distributed by the Sponsor for human use in this study.

Events meeting the AE definition

- Any abnormal laboratory test results (hematology, clinical chemistry, or urinalysis) or other safety assessments (eg, ECG, radiological scans, vital signs measurements), including those that worsen from baseline, considered clinically significant in the medical and scientific judgment of the investigator.

- Exacerbation of a chronic or intermittent pre-existing condition including either an increase in frequency and/or intensity of the condition.

- New conditions detected or diagnosed after study intervention administration even though it may have been present before the start of the study.

  Note: Congenital disorders (eg, present from birth) not detected or diagnosed prior to study intervention administration do not qualify for reporting as AE.

- Signs, symptoms, or the clinical sequelae of a suspected drug-drug interaction.

- Signs, symptoms, or the clinical sequelae of a suspected overdose of either study intervention or a concomitant medication.

- For all reports of overdose (whether accidental or intentional) with an associated AE, the AE term should reflect the clinical symptoms or abnormal test result. An overdose without any associated clinical symptoms or abnormal laboratory results is reported using the terminology "accidental or intentional overdose without adverse effect."

Events NOT meeting the AE definition

- Medical or surgical procedure (eg, endoscopy, appendectomy): the condition that leads to the procedure is the AE.
- Situations in which an untoward medical occurrence did not occur (social and/or convenience admission to a hospital).
- Anticipated day-to-day fluctuations of pre-existing disease(s) or condition(s) present or detected at the start of the study that do not worsen.
- Surgery planned prior to informed consent to treat a pre-existing condition that has not worsened.
- Refer to Section 8.4.6 for protocol-specific exceptions.

10.3.2 Definition of SAE

If an event is not an AE per definition above, then it cannot be an SAE even if serious conditions are met.

An SAE is defined as any untoward medical occurrence that, at any dose:

a. Results in death

b. Is life-threatening

- The term "life-threatening" in the definition of "serious" refers to an event in which the participant was at risk of death at the time of the event. It does not refer to an event, which hypothetically might have caused death, if it were more severe.

c. Requires inpatient hospitalization or prolongation of existing hospitalization

- Hospitalization is defined as an inpatient admission, regardless of length of stay, even if the hospitalization is a precautionary measure for continued observation. (Note: Hospitalization for an elective procedure to treat a pre-existing condition that has not worsened is not an SAE. A pre-existing condition is a clinical condition that is diagnosed prior to the use of an MSD product and is documented in the participant's medical history.

d. Results in persistent or significant disability/incapacity

- The term disability means a substantial disruption of a person's ability to conduct normal life functions.
- This definition is not intended to include experiences of relatively minor medical significance such as uncomplicated headache, nausea, vomiting, diarrhea, influenza,

- There may be instances when copies of medical records for certain cases are requested by the Sponsor. In this case, all participant identifiers, with the exception of the participant number, will be blinded on the copies of the medical records before submission to the Sponsor.

- The investigator will attempt to establish a diagnosis of the event based on signs, symptoms, and/or other clinical information. In such cases, the diagnosis (not the individual signs/symptoms) will be documented as the AE/SAE.

Assessment of intensity/toxicity

- An event is defined as "serious" when it meets at least 1 of the predefined outcomes as described in the definition of an SAE, not when it is rated as severe.

1. The investigator will make an assessment of intensity for each AE and SAE (and other reportable safety event) according to the NCI Common Terminology for Adverse Events (CTCAE), version 4. Any AE that changes CTCAE grade over the course of a given episode will have each change of grade recorded on the AE CRFs/worksheets.

    - Grade 1: Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.

    - Grade 2: Moderate; minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living (ADL).

    - Grade 3: Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL.

    - Grade 4: Life threatening consequences; urgent intervention indicated.

    - Grade 5: Death related to AE.

Assessment of causality

1. Did the Sponsor's product cause the AE?

2. The determination of the likelihood that the Sponsor's product caused the AE will be provided by an investigator who is a qualified physician. The investigator's signed/dated initials on the source document or worksheet that supports the causality noted on the AE form, ensures that a medically qualified assessment of causality was done. This initialed document must be retained for the required regulatory time frame. The criteria below are intended as reference guidelines to assist the investigator in assessing the likelihood of a relationship between the test product and the AE based upon the available information.

SAE reporting to the Sponsor via paper CRF

- If the EDC tool is not operational, facsimile transmission or secure e-mail of the SAE paper CRF is the preferred method to transmit this information to the Sponsor.

- In rare circumstances and in the absence of facsimile equipment, notification by telephone is acceptable with a copy of the SAE data collection tool sent by overnight mail or courier service.

- Initial notification via telephone does not replace the need for the investigator to complete and sign the SAE CRF pages within the designated reporting time frames.

- Contacts and instructions for SAE reporting and paper reporting procedures can be found in the Investigator Study File Binder (or equivalent).

10.5 Appendix 5: Contraceptive Guidance and Pregnancy Testing

10.5.1 Definitions

Women of Childbearing Potential (WOCBP)

A woman is considered fertile following menarche and until becoming postmenopausal unless permanently sterile (see below).

If fertility is unclear (eg, amenorrhea in adolescents or athletes) and a menstrual cycle cannot be confirmed before first dose of study intervention, additional evaluation should be considered.

Women in the following categories are not considered WOCBP:

- Premenarchal
- Premenopausal female with 1 of the following:
  - Documented hysterectomy
  - Documented bilateral salpingectomy
  - Documented bilateral oophorectomy

For individuals with permanent infertility due to an alternate medical cause other than the above (eg, Mullerian agenesis, androgen insensitivity), investigator discretion should be applied to determining study entry.

Note: Documentation can come from the site personnel's review of the participant's medical records, medical examination, or medical history interview.

- Postmenopausal female
  - A postmenopausal state is defined as no menses for 12 months without an alternative medical cause.
    - A high follicle-stimulating hormone (FSH) level in the postmenopausal range may be used to confirm a postmenopausal state in women not using hormonal contraception or HRT. However, in the absence of 12 months of amenorrhea, confirmation with 2 FSH measurements in the postmenopausal range is required.
    - Females on HRT and whose menopausal status is in doubt will be required to use one of the nonhormonal highly effective contraception methods if they wish to continue their HRT during the study. Otherwise, they must discontinue HRT to allow confirmation of postmenopausal status before study enrollment.

10.5.2 Contraception Requirements

Contraceptives allowed during the study include[a]:
Highly Effective Contraceptive Methods That Have Low User Dependency *Failure rate of <1% per year when used consistently and correctly.*
• Progestogen-only subdermal contraceptive implant[b] • Intrauterine hormone-releasing system (IUS)[c] • Intrauterine device (IUD) • Bilateral tubal occlusion
• Azoospermic partner (Vasectomized or secondary to medical cause) • This is a highly effective contraception method provided that the partner is the sole male sexual partner of the WOCBP and the absence of sperm has been confirmed. If not, an additional highly effective method of contraception should be used. A spermatogenesis cycle is approximately 90 days. Note: Documentation of azoospermia can come from the site personnel's review of the participant's medical records, medical examination, or medical history interview.
Sexual Abstinence • Sexual abstinence is considered a highly effective method only if defined as refraining from heterosexual intercourse during the entire period of risk associated with the study intervention. The reliability of sexual abstinence needs to be evaluated in relation to the duration of the study and the preferred and usual lifestyle of the participant.
a. Contraceptive use by men or women should be consistent with local regulations regarding the use of contraceptive methods for participants of clinical studies. b. If locally required, in accordance with Clinical Trial Facilitation Group (CTFG) guidelines, acceptable contraceptive implants are limited to those which inhibit ovulation. c. IUS is a progestin-releasing IUD. Note: The following are not acceptable methods of contraception: – Periodic abstinence (calendar, symptothermal, postovulation methods), withdrawal (coitus interruptus), spermicides only, and lactational amenorrhea method (LAM). – Male condom with cap, diaphragm, or sponge with spermicide. – Male and female condom should not be used together (due to risk of failure with friction).

10.5.3 Pregnancy Testing

WOCBP should only be included after a negative highly sensitive urine or serum pregnancy test. Refer to Section 8.3.5.2 and Appendix 2 for further details.

Abbreviation	Expanded Term
EOC	Executive Oversight Committee
EORTC	European Organization for Research and Treatment of Cancer
ePRO	electronic patient-reported outcome
EQ-5D-5L	EuroQoL-5D-5L
FAS	Full Analysis Set
FDA	Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
FSH	follicle-stimulating hormone
GCP	Good Clinical Practice
HBsAg	hepatitis B surface antigen
HBV	hepatitis B virus
HCV	hepatitis C virus
HIV	human immunodeficiency virus
HR	hazard ratio
HRQoL	health-related quality of life
HRT	hormone replacement therapy
IA	interim analysis
IB	Investigator's Brochure
ICF	informed consent form
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
iCPD	confirmed PD
iCR	confirmed CR
IEC	Independent Ethics Committee
irAE	immune-related adverse event
iPR	confirmed PR
IRB	Institutional Review Board
iRECIST	adjusted RECIST 1.1 for Immune-based Therapeutics
iSD	confirmed SD
ITT	Intention to Treat

- TTD as measured by each of the EORTC QLQ-C30 global health status/QoL, physical functioning, and dyspnea scores, and EORTC QLQ-LC13 cough and chest pain scores.

- TTD as measured by a composite symptom endpoint: cough (QLQ-LC13 item 31), chest pain (QLQ-LC13 item 40), or dyspnea (QLQ-C30 item 8)

Based on prior literature (Osoba D et al., 1998; King et al., 1996, Maringwa et al., 2011), a ≥10 point or greater worsening from baseline for each scale represents a minimally important difference (MID) that represents a clinically relevant deterioration. TTD in the individual scores specified above is defined as the time to first onset of ≥10 (out of 100) point deterioration from baseline in a given scale/subscale/item and confirmed by a second adjacent ≥10 point deterioration from baseline. TTD in the composite endpoint is defined as the time to first onset of a ≥10 point deterioration (i.e., increase in score) from baseline in any one of the 3 lung cancer symptom scale items (EORTC-QLQ-C30 [dyspnea] and EORTC-QLQ-LC13 [cough and chest pain]) with confirmation by the subsequent visit of a ≥10 point deterioration (i.e., increase in score) from baseline in the same scale as the first onset. Changes from baseline in EORTC QLQ-C30 scores will also be interpreted according to recent subscale-specific guidelines, which indicate that clinically meaningful differences vary by scale (Cocks et al., 2012)

Exploratory

- Change from baseline in the additional scales of the EORTC QLQ-C30, EORTC QLQ-LC13, and EuroQoL EQ-5D visual analogue scale (VAS) scores. Specifically:

    - Each EORTC QLQ-C30 functional scales: role functioning, emotional functioning, cognitive functioning, and social functioning

    - Each of three EORTC QLQ-C30 symptom scales (fatigue, nausea/vomiting, and pain), and single item measures (sleep disturbance, appetite loss, constipation, diarrhea, and financial difficulties)

    - Each EORTC QLQ-LC13 item: site-specific pain (pain in arm or shoulder, and pain in other parts), sore mouth, dysphagia, peripheral neuropathy, alopecia, and hemoptysis

    - The EQ-5D VAS

- Proportions of improvement / stability / stability + improvement / deterioration in EORTC QLQ-C30 global health status / QoL, physical functioning, and dyspnea scores, and EORTC QLQ-LC13 cough and chest pain scores where

    - Improvement

        - Improvement for the global health status/QoL and physical functioning scores is defined by an increase of 10 points or more in score from baseline at any time during the study and confirmed by an increase of 10 points or more in score at the next consecutive visit.

Table 1 Analysis Methods for Key Efficacy Endpoints

Endpoint/Variable	Statistical Method	Analysis Population	Missing Data Approach
Primary Analyses:			
PFS (RECIST 1.1) by BICR	Testing: stratified log-rank test Estimation: Stratified Cox model with Efron's tie-handling method	ITT	Censored according to rules in Table 2
OS	Testing: stratified log-rank test Estimation: Stratified Cox model with Efron's tie-handling method	ITT	Censored at last known alive date
Secondary Analyses:			
ORR (RECIST 1.1) by BICR	Testing and estimation: stratified Miettinen and Nurminen method	ITT	Participants with missing data are considered nonresponders
Abbreviations: BICR = blinded independent central review; ITT = intention-to-treat; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors.			

The strategy to address multiplicity issues with regard to multiple endpoints and interim analyses is described in Section 3.7, Interim Analyses, and in Section 3.8, Multiplicity.

3.6.1.1. Progression-free Survival

The nonparametric Kaplan-Meier method will be used to estimate the PFS curve in each treatment group. The hypotheses of treatment difference in PFS will be tested by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to estimate the magnitude of the treatment difference (ie, HR) between the treatment arms. The HR and its 95% CI from the stratified Cox model with Efron's method of tie handling and with a single treatment covariate will be reported. The stratification factors used for randomization (Section 6.3.2 in the protocol) will be applied to both the stratified log-rank test and the stratified Cox model.

Since PD is assessed periodically, PD can occur any time in the time interval between the last assessment where PD was not documented and the assessment when PD is documented. For the primary analysis, for the participants who have PD, the true date of PD will be approximated by the date of the first assessment at which PD is objectively documented per RECIST 1.1 by the BICR vendor, regardless of discontinuation of study drug.

To evaluate the robustness of the PFS endpoint per RECIST 1.1 via BICR by the imaging vendor, 1 primary and 2 sensitivity analyses with a different set of censoring rules will be performed. For the primary analysis, if the events (PD or death) are immediately after more than 1 consecutive missed disease assessment, the data are censored at the last disease assessment before missing visits. Also data after new anticancer therapy are censored at the last disease assessment before the initiation of new anticancer therapy. The first sensitivity analysis follows ITT principles (ie, PDs/deaths are counted as events regardless of missed study visits or initiation

of new anticancer therapy). The second sensitivity analysis considers discontinuation of treatment or initiation of an anticancer treatment subsequent to discontinuation of study-specified treatments, whichever occurs later, to be a PD event for participants without documented PD or death. If a participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied. The censoring rules for primary and sensitivity analyses are summarized in Table 2.

Table 2 Censoring Rules for Primary and Sensitivity Analyses of Progression-free Survival

Situation	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
No PD and no death; new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment	Progressed at treatment discontinuation for reasons other than CR; otherwise, censored at last disease assessment if still on study intervention or completed study intervention
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment	Censored at last disease assessment	Progressed at date of new anticancer treatment
PD or death documented after ≤1 missed disease assessment and before new anticancer therapy	Progressed at date of documented PD or death	Progressed at date of documented PD or death	Progressed at date of documented PD or death
PD or death documented immediately after ≥2 consecutive missed disease assessments or after new anticancer therapy	Censored at last disease assessment before the earlier date of ≥2 consecutive missed disease assessment and new anticancer therapy	Progressed at date of documented PD or death	Progressed at date of documented PD or death
Abbreviation: CR = complete response; PD = progressive disease.			

In case the proportional hazards assumption is not valid, Restricted Mean Survival Time (RMST) method may be conducted for PFS to account for the possible non-proportional hazards effect as a sensitivity analysis.

3.6.1.2. Overall Survival

The nonparametric Kaplan-Meier method will be used to estimate the survival curves. The treatment difference in survival will be assessed by the stratified log-rank test (based on the stratification factors defined in Section 6.3.2 in the protocol). A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (ie, the HR). The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported. The stratification factors used for randomization (Section 6.3.2 in the protocol) will be applied to both the stratified log-rank test and the stratified

The analysis of safety results will follow a tiered approach (Table *4*). The tiers differ with respect to the analyses that will be performed. AEs (specific terms as well as system organ class terms) and events that meet predefined limits of change in laboratory, vital signs, and ECG parameters are either prespecified as "Tier 1" endpoints or will be classified as belonging to "Tier 2" or "Tier 3," based on observed proportions of participants with an event.

Tier 1 Events

Safety parameters or AEs of special interest that are identified a priori constitute "Tier 1" safety endpoints that will be subject to inferential testing for statistical significance. AEs that are immune-mediated or potentially immune-mediated are well documented and will be evaluated separately; however, these events have been characterized consistently throughout the pembrolizumab clinical development program, and determination of statistical significance is not expected to add value to the safety evaluation. Similarly, the combination of pembrolizumab and lenvatinib has not been associated with any new safety signals. Finally, there are no known AEs associated with participants with NSCLC for which determination of a *p* value is expected to impact the safety assessment. Therefore, there are no Tier 1 events for this protocol.

Tier 2 Events

Tier 2 parameters will be assessed via point estimates with 95% CIs provided for differences in the proportion of participants with events using the Miettinen and Nurminen method, an unconditional, asymptotic method [Miettinen, O. 1985].

Membership in Tier 2 requires that at least 10% of participants in any treatment group exhibit the event; all other AEs and predefined limits of change will belong to Tier 3. The threshold of at least 10% of participants was chosen for Tier 2 events because the population randomized in this study are in critical conditions and usually experience various AEs of similar types regardless of treatment; events reported less frequently than 10% of participants would obscure the assessment of the overall safety profile and add little to the interpretation of potentially meaningful treatment differences. In addition, Grade 3 to 5 AEs (≥5% of participants in one of the treatment groups) and SAEs (≥5% of participants in one of the treatment groups) will be considered Tier 2 endpoints. Because many 95% CIs may be provided without adjustment for multiplicity, the CIs should be regarded as a helpful descriptive measure to be used in safety review, not a formal method for assessing the statistical significance of the between-group differences.

Tier 3 Events

Safety endpoints that are not Tier 1 or 2 events are considered Tier 3 events. Only point estimates by treatment group are provided for Tier 3 safety parameters.

Continuous Safety Measures

For continuous measures such as changes from baseline in laboratory, vital signs, and ECG parameters, summary statistics for baseline, on-treatment, and change from baseline values will be provided by treatment group in table format.

The scoring method for each subscale of the EORTC-QLQ-C30 will be performed according to the scoring manual (Scott, et al., 2008). According to the EORTC QLQ-C30 manual, if items $I_1, I_2, \cdots, I_n$ are included in a scale, the linear transformation procedure is as follows:

1. Compute the raw score:

$$RS = (I_1 + I_1 + \cdots + I_n)/n$$

2. Linear transformation to obtain the score *S*:

$$\text{Functional scales: } S = \left(1 - \frac{RS-1}{Range}\right) \times 100$$

$$\text{Symptom scales/items: } S = \left(1 - \frac{RS-1}{Range}\right) \times 100$$

$$\text{Global health status/quality of life scale: } S = \left(1 - \frac{RS-1}{Range}\right) \times 100$$

The range is defined as the difference between the maximum possible value of *RS* and the minimum possible value for RS. If more than half of the items within one scale are missing, then the scale is considered missing; otherwise, the score will be calculated as the average score of those available items.

EORTC QLQ-LC13:

The lung cancer questionnaire module is a validated self-reported PRO questionnaire that is intended to be used in conjunction with the QLQ-C30 and includes 13 questions that share the four level ordinal response with the QLQ-C30 and can be summarized and scored as described above (1= “not at all”; 2= “a little”; 3= “quite a bit”; 4= “very much”). The LC13 comprises both multi-item and single-item measures of lung cancer-associated symptoms (i.e. coughing, hemoptysis, dyspnea, and pain) and treatment related symptoms (i.e., sore mouth, dysphagia, peripheral neuropathy, and alopecia). A linear transformation will be applied to standardize the scores between 0 (least severe symptom) and 100 (most severe symptom) as described above for the EORTC QLQ-C30 symptom scales/items scoring.

EQ-5D:

The EQ-5D-5L is primarily designed for self-completion and consists of 2 parts: a descriptive system and the VAS. The EQ-5D-5L descriptive system includes 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The 5 dimensions each have 5 levels: no problems, slight, moderate, severe problems, and extreme problems. The responses patterns on the 5 dimensions are scored using country-specific population weights to provide an aggregate index score anchored at 0 (death) and 1 (perfect health); depending on the algorithm used some states may be considered worse than death.

The EQ-5D VAS records the respondent’s self-rated health on a vertical, visual analogue scale (0-100), with endpoints labeled ‘the best health you can imagine’ and ‘the worst health you can

categorical variable so that no restriction is imposed on the trajectory of the means over time. The cLDA model is specified as follows:

$$E(Y_{ijt}) = \gamma_0 + \gamma_{jt} I(t > 0) + \beta X_i, j = 1,2,,3,..,n;\ t = 0,1,2,3,..k$$

where Y_{ijt} is the PRO score for participant i, with treatment assignment j at visit t; γ_0 is the baseline mean for all treatment groups, γ_{jt} is the mean change from baseline for treatment group j at time t; X_i is the stratification factor (binary) vector for this participant, and β is the coefficient vector for stratification factors. An unstructured covariance matrix will be used to model the correlation among repeated measurements. If the unstructured covariance model fails to converge with the default algorithm, then Fisher scoring algorithm or other appropriate methods can be used to provide initial values of the covariance parameters. In the rare event that none of the above methods yield convergence, a structured covariance such as Toeplitz can be used to model the correlation among repeated measurements. In this case, the asymptotically unbiased sandwich variance estimator will be used. The cLDA model implicitly treats missing data as missing at random (MAR).

Line plots for the empirical mean change from baseline in the EORTC QLQ-C30 global health status/QoL, physical functioning and dyspnea scores and EORTC QLQ-LC13 cough and chest pain scores (as specified in section 3.4.3) will be provided across all time points as a supportive analysis.

In addition, the model-based LS mean change from baseline to the specified post-baseline time point together with 95% CI will be plotted in bar charts for the EORTC QLQ-C30 global health status/QoL, physical functioning and symptom scales, and EORTC QLQ-LC13 symptom scales (as specified in section 3.4.3).

3.6.3.4. Time-to-True Deterioration (TTD) Analysis

The non-parametric Kaplan-Meier method will be used to estimate the deterioration curve in each group. The estimate of median time to deterioration and its 95% confidence interval will be obtained from the Kaplan-Meier survival estimates. The treatment difference in TTD will be assessed by the stratified log-rank test, and nominal two-sided p-values will be reported. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (hazard ratio) between treatment arms. Stratification factors used for randomization (See Section 6.3.2 in the protocol) will be applied to the analysis will be used in the stratified Cox PH model.

The approach for the time-to-deterioration analysis will be based on the assumption of non-informative censoring. The participants who do not have deterioration on the last date of evaluation will be censored. Table 6 provides censoring rule for TTD analysis.

Table 7 Analysis Strategy for Key PRO Endpoints

Endpoint/Variable	Statistical Method	Analysis Population	Missing Data Approach
Mean change from baseline in – EORTC QLQ-C30 • Global health status/QoL • Physical functioning • dyspnea – EORTC QLQ-LC13 • Cough • Chest pain	cLDA model	FAS	Model-based.= cLDA
TTD in – EORTC QLQ-C30 • Global health status/QoL • Physical functioning • dyspnea – EORTC QLQ-LC13 • Cough • Chest pain	Stratified log-rank test and HR estimation using stratified Cox model with Efron's tie handling method	FAS	Censored according to rules in Table *6*.
Overall improvement and overall improvement/stability in – EORTC QLQ-C30 • Global health status/QoL • Physical functioning • dyspnea – EORTC QLQ-LC13 • Cough • Chest pain	Stratified Miettinen and Nurminen method	FAS	Participants with missing data are considered not achieving improvement/stability.
Abbreviations: cLDA = constrained longitudinal data analysis, FAS = full analysis set, QoL = quality of life.			

3.6.4. Demographic and Baseline Characteristics

The comparability of the treatment groups for each relevant demographic and baseline characteristic will be assessed by the use of tables and/or graphs. No statistical hypothesis tests will be performed on these characteristics. The number and percentage of participants screened and randomized and the primary reason for screening failure and discontinuation will be displayed. Demographic variables, baseline characteristics, primary and secondary diagnoses, and prior and concomitant therapies will be summarized by treatment, either by descriptive statistics or categorical tables.

3.7. Interim Analyses

Blinding to treatment assignment will be maintained at all investigational sites. The results of interim analyses will not be shared with the investigators before the completion of the study. Participant-level unblinding will be restricted to an external unblinded statistician and scientific

programmer performing the interim analysis, who will have no other responsibilities associated with the study.

An external DMC will serve as the primary reviewer of the results of the interim analyses of the study and will make recommendations for discontinuation of the study or protocol modifications to the study Executive Oversight Committee (EOC). If the DMC recommends modifications to the design of the protocol or discontinuation of the study, the EOC (and potentially other limited Sponsor personnel) may be unblinded to results at the treatment level to act on these recommendations. The extent to which individuals are unblinded with respect to results of interim analyses will be documented by the unblinded statistician. Additional logistical details will be provided in the DMC charter. Key aspects of the interim analyses are described below.

Treatment-level results from the interim analysis will be provided to the DMC by the unblinded statistician. The unblinded statistician will not be involved in any discussions regarding modifications to the protocol, statistical methods, identification of protocol deviations, or data validation efforts after the interim analyses.

If the study is positive at an interim analysis for both primary endpoints, additional analyses, including but not limited to the protocol-specified final analysis, may be carried out for exploratory purpose or upon regulatory request.

3.7.1. Efficacy Interim Analysis

One interim analysis is planned in addition to the final analysis for this study. For the interim and final analyses, all the randomized participants will be included. Results of the interim analysis will be reviewed by the DMC. Details on the boundaries for establishing statistical significance with regard to efficacy are discussed further in Section 3.8.

The analyses planned, endpoints evaluated, and drivers of timing are summarized in Table *8*.

Table 9 Efficacy Boundaries and Properties for Progression-free Survival Analyses

Analysis	Value	α=0.0055	α=0.025
IA: 100%[a] N: 620 Events: 416 Month: 30.5[f]	Z	2.5427	1.9600
	p (1-sided)[b]	0.0055	0.0250
	~HR at bound[c]	0.7793	0.8252
	P(Cross) if HR=1[d]	0.0055	0.0250
	P(Cross) if HR=0.7[e]	0.8650	0.9536

Abbreviations: HR = hazard ratio; IA = interim analysis.
The number of events and timings are estimated.
[a] Percentage of total planned events at the interim analysis.
[b] The nominal α for testing.
[c] The approximate HR required to reach an efficacy bound
[d] The probability of crossing a bound under the null hypothesis
[e] The probability of crossing a bound under the alternative hypothesis
[f] The approximate number of months since first participant randomized

3.8.3. Overall Survival

The study will test OS at IA and FA. Following the multiplicity strategy as outlined in Figure *1*, the OS hypothesis may be tested at α=0.0195 (initially allocated α) or α=0.025 (if the null hypotheses for both PFS and ORR are rejected). Table 10 shows the boundary properties for each of these α-levels for OS at IA and FA, which were derived using a Lan-DeMets and O'Brien-Fleming spending function. Note that if the PFS null hypothesis is rejected at IA, OS interim and final analysis test may be compared with its updated bounds considering the α reallocation from the PFS hypothesis.

Based on prior literature (Osoba D et al., 1998; King et al., 1996, Marringwa et al., 2011), a ≥10 point or greater worsening from baseline for each scale represents a minimally important difference (MID) that represents a clinically relevant deterioration. TTD in the individual scores specified above is defined as the time to first onset of ≥10 (out of 100) point deterioration from baseline in a given scale/subscale/item and confirmed by a second adjacent ≥10 point deterioration from baseline. TTD in the composite endpoint is defined as the time to first onset of a ≥10 point deterioration (i.e., increase in score) from baseline in any one of the 3 lung cancer symptom scale items (EORTC-QLQ-C30 [dyspnea] and EORTC-QLQ-LC13 [cough and chest pain]) with confirmation by the subsequent visit of a ≥10 point deterioration (i.e., increase in score) from baseline in the same scale as the first onset. Changes from baseline in EORTC QLQ-C30 scores will also be interpreted according to recent subscale-specific guidelines, which indicate that clinically meaningful differences vary by scale (Cocks et al., 2012)

4.5. Analysis Populations

4.5.1. Efficacy Analysis Populations

Efficacy analysis will be carried out in the intention-to-treat (ITT) China subpopulation. This population will include all Chinese participants who are randomized in the global study and all participants who are randomized in the extension portion.

Participants from China enrolled in the extension portion of this study after completion of the global enrollment will not be included in the primary efficacy analysis population for the global study.

4.5.2. Safety Analysis Populations

Safety analysis will be carried out in the All Patients as Treated (APaT) China subpopulation, i.e., all randomized Chinese participants (in the global study and extension portion) who received at least 1 dose of study treatment.

Participants from China randomized and treated in the extension portion of this study after completion of the global enrollment will not be included in the primary safety analysis population for the global study.

4.5.3. Patient-reported Outcome Analysis Population

The PRO analyses are based on the PRO Full Analysis Set (PRO FAS) China subpopulation, defined as all randomized participants who have at least one PRO assessment available for the specific endpoint and have received at least one dose of the study intervention. Participants will be analyzed in the treatment group to which they are randomized. The PRO FAS China subpopulation will include all Chinese participants (in the global study and extension portion) in this population.

Participants from China enrolled in the extension portion of this study after completion of the global enrollment will not be included in the primary PRO analysis population for the global study.

4.5.4. Pharmacokinetic Analysis Population

The analysis of PK data will be carried out in the population PK analysis set China subpopulation. The Population PK Analysis set will include all participants who have received at least 1 dose of study intervention with documented dosing history in the lenvatinib + pembrolizumab arm and have measurable plasma levels of lenvatinib or serum levels of pembrolizumab. Population PK Analysis Set China subpopulation will include all Chinese participants (in the global study and extension portion) in this population.

Participants from China enrolled in the extension portion of this study after completion of the global enrollment will not be included in the primary PK analysis population for the global study.

4.6. Statistical Methods

Regarding the analysis for extension, no formal hypothesis testing is planned. No multiplicity adjustment will be applied to the analysis for extension.

4.6.1. Statistical Methods for Efficacy Analyses

Analyses regarding objective response and disease progression will include all events in all participants excluding events that occur in the Second Course Treatment.

4.6.1.1. Progression-free Survival

Analysis of PFS for extension is the same to that for the global study if applicable.

The nonparametric Kaplan-Meier method will be used to estimate the PFS curve in each treatment group. The treatment difference in PFS will be assessed by the stratified log-rank test (based on the stratification factors defined in Section 6.3.2 in the protocol). . A stratified Cox proportional hazard model with Efron's method of tie handling will be used to estimate the magnitude of the treatment difference (ie, HR) between the treatment arms. The HR and its 95% CI from the stratified Cox model with Efron's method of tie handling and with a single treatment covariate will be reported. The same stratification factors used in the global study will be used (see section 6.3.2 in the protocol). For the Chinese subgroup analyses, the stratified method will only be used if applicable. The factor of geographic region (East Asia vs. non-East Asia) will not be included in the Chinese subgroup stratified analyses. An analysis using the restricted mean survival time method may be conducted for PFS to account for the possible nonproportional hazards effect.

Consistency of efficacy will be evaluated using the percentage of risk reduction preserved in the Chinese subpopulation from the empirical risk reduction from the global primary efficacy analyses (based on point estimates). Sample size is designed to provide about 80% chance of observing the point estimate of Chinese subpopulation preserves ≥ approximately 50% of empirical risk

4.6.2. Statistical Methods for Safety Analyses

Safety analyses for extension are the same to that for the global study as described in Section 3.6.2.

4.6.3. Statistical Methods for PRO Analysis

PRO Analysis are the same for extension to that for the global study as described in Section 3.6.3.

4.6.4. Summaries of Baseline Characteristics, Demographics, and Other Analyses

They are the same for extension to that for the global study as described in Section 3.6.4.

4.7. Interim Analysis and Final analysis

The primary analysis for PFS will be conducted in the Chinese subpopulation when approximately 75 PFS events have been collected. OS will also be analyzed.

The primary analysis for OS will be conducted in the Chinese subpopulation when approximately 75 OS events have been collected.

4.8. Multiplicity

No multiplicity adjustment will be applied to the analysis of China.

4.9. Sample Size and Power Calculations

After the completion of global study enrollment, the extension portion will continue to enroll participants and randomize eligible participants until the sample size for the overall randomized Chinese subpopulation reaches approximately 120. Participants from China enrolled in the extension portion of this study after completion of the global enrollment will not be included in the primary efficacy analysis population for the global study.

The extension portion will complete after approximately 75 OS events have been observed between the two arms in the Chinese subpopulation. With 75 OS events and a true hazard ratio of 0.7, the extension portion has ~80% chance to observe a point estimate of OS that preserves ≥ approximately 50% of the empirical risk reduction from the global analysis in the Chinese subpopulation. With 75 PFS events and a true hazard ratio of 0.7, the extension portion has ~80% chance to observe a point estimate of PFS that preserves ≥ approximately 50% of the empirical risk reduction from the global analysis in the Chinese subpopulation.

The above calculations for the consistency evaluation in OS and PFS are based on the same assumptions in the global study for sample size and power evaluation as specified in Section 3.9.

Abbreviation	Expanded Term
cLDA	constrained longitudinal data analysis
CR	complete response
CRF	case report form
CTCAE	Common Terminology Criteria for Adverse Events
DMC	data monitoring committee
DOR	duration of response
ECG	electrocardiogram
ECI	event of clinical interest
EOC	Executive Oversight Committee
EORTC	European Organisation for Research and Treatment of Cancer
ePRO	electronic patient-reported outcome
EQ-5D	EuroQoL-5D-5L
FAS	Full Analysis Set
HR	hazard ratio
HRQoL	health-related quality of life
HSD	Hwang-Shih-DeCani
IA	interim analysis
ICH	International Conference on Harmonization
iRECIST	Modified RECIST 1.1 for Immune-based Therapeutics
ITT	Intention to Treat
NPH	Non-proportional Hazards
PH	Proportional Hazards
NSCLC	non-small cell lung cancer
OR	objective response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PD-L1	programmed cell death ligand 1
PFS	progression-free survival
PH	proportional hazards

- Objective: To compare the time to true deterioration (TTD) in global health status/QoL, cough, chest pain, dyspnea, and physical functioning for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	- TTD, defined as the time from baseline to the first onset of a ≥10-point deterioration from baseline with confirmation by the subsequent visit of a ≥10-point deterioration from baseline in global health status/QoL (EORTC QLQ-C30 items 29 and 30), cough (EORTC QLQ-LC13 item 31), chest pain (EORTC QLQ-LC13 item 40), dyspnea (EORTC QLQ-C30 item 8), and physical functioning (EORTC QLQ-C30 items 1-5) - TTD in the composite endpoint (combination of cough [QLQ-LC13 item 31], chest pain [QLQ-LC13 item 40], or dyspnea [QLQ-C30 item 8]) defined as the time to first onset of a ≥10-point deterioration from baseline in any one of 3 scale items with confirmation by the subsequent visit of a ≥10-point deterioration from baseline in the same scale as the first onset

*This study will be considered to have met its success criteria if the combination of pembrolizumab + lenvatinib is superior to pembrolizumab + matching placebo in PFS or OS.

Overall Design:

Study Phase	Phase 3
Primary Purpose	Treatment
Indication	First-line treatment of metastatic NSCLC with PD-L1 expression (TPS ≥1%)
Population	Adult participants with treatment-naïve, metastatic NSCLC (TPS ≥1%)
Study Type	Interventional
Intervention Model	Parallel This is a multi-site study.

Study Period	**Screening**	**Treatment Cycle = 21 Days**									**EOT**	**Posttreatment**			**Notes**
												Safety Follow-up	**Follow-up**	**Survival Follow-up**	**All procedures are to be performed before study intervention administration unless otherwise indicated. Refer to Section 8.11 for visit details.**
Visit Timing/ Cycle Number	**-28 to -1**	**1**			**2**		**3**	**4**	**5**	**6 to 35**	**At DC**	**30 Days post last dose**	**Every 12 weeks**	**Every 12 weeks**	
Cycle Day (±3 days unless otherwise specified)		**1**	**8**	**15**	**1**	**15**	**1**	**1**	**1**	**1**	**–**				
Scheduling Window (Days)		**+ 3**	± 3	±3	± 3	± 3	± 3	± 3	± 3	± 3		**+ 7**	**± 7**	**± 14**	
Study Intervention Administration															
Pembrolizumab administration		X			X		X	X	X	X					Pembrolizumab 200 mg IV Q3W.
Tumor Scans															
Tumor scans and response assessment (thorax, abdomen, pelvis)	X*	←								→	X				*Perform Q9W (63±7 days) from randomization through Week 54, then Q12W (84 ±7 days) subsequently (ie, after Weeks 9, 18, 27, 36, 45, 54, 66, 78, 90, etc.). -Schedule should be followed regardless of treatment delays. -If imaging was obtained within 4 weeks before DC, scan at DC is not mandatory.

Objectives	Endpoints
• Objective: To evaluate the safety and tolerability for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	• Adverse events (AEs) and study intervention discontinuation due to AEs
• Objective: To compare the mean change from baseline in the global health status/quality of life (QoL), cough, chest pain, dyspnea, and physical functioning for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	• Change from baseline for the following patient-reported outcomes (PROs) scales/items: global health status/QoL (EORTC QLQ-C30 items 29 and 30), cough (EORTC QLQ-LC13 item 31), chest pain (EORTC QLQ-LC13 item 40), dyspnea (EORTC QLQ-C30 item 8), and physical functioning (EORTC QLQ-C30 items 1-5)
• Objective: To compare the time to true deterioration (TTD) in global health status/QoL, cough, chest pain, dyspnea, and physical functioning for the combinations of pembrolizumab + lenvatinib versus pembrolizumab + matching placebo	• TTD, defined as the time from baseline to the first onset of a ≥10-point deterioration from baseline with confirmation by the subsequent visit of a ≥10-point deterioration from baseline in global health status/QoL (EORTC QLQ-C30 items 29 and 30), cough (EORTC QLQ-LC13 item 31), chest pain (EORTC QLQ-LC13 item 40), dyspnea (EORTC QLQ-C30 item 8), and physical functioning (EORTC QLQ-C30 items 1-5) • TTD in the composite endpoint (combination of cough [QLQ-LC13 item 31], chest pain [QLQ-LC13 item 40], or dyspnea [QLQ-C30 item 8]) defined as the time to first onset of a ≥10-point deterioration from baseline in any one of 3 scale items with confirmation by the subsequent visit of a ≥10-point deterioration from baseline in the same scale as the first onset

4.2.2 Rationale for the Use of Comparator/Placebo

Based on the results of KEYNOTE-024, pembrolizumab monotherapy has become the SOC for treating Stage IV NSCLC in patients with PD-L1 TPS ≥50% with no *EGFR* or *ALK* genomic tumor aberrations. The study resulted in a significant PFS benefit for pembrolizumab over the SOC at the time (platinum-doublet therapy), with an HR of 0.50 (95% CI, 0.37-0.68; $p<0.001$; median PFS, 10.3 months [95% CI, 6.7 to not reached] with pembrolizumab and 6.0 months [95% CI, 4.2-6.2 months] with platinum-doublet), as well as an OS benefit with an HR of 0.60 (95% CI 0.41-0.89; $p=0.005$; median OS was not reached in either treatment arm) [Reck, M., et al 2016]. The results from KEYNOTE-024 established pembrolizumab as 1L therapy for patients whose tumors have a TPS ≥50% and in whom *EGFR*- or *ALK*-directed therapies are not indicated, and the regimen has received regulatory approval for this use by the FDA and EMA. Additionally, based on results from KEYNOTE-042, pembrolizumab monotherapy is considered one of the treatment options in treating Stage IV NSCLC in patients with PD-L1 TPS ≥1% with no *EGFR* or *ALK* genomic tumor aberrations. The study resulted in a significant OS benefit for pembrolizumab over the SOC at the time (platinum-doublet therapy), with an HR of 0.81 (95% CI, 0.71-0.93; $p=0.0018$). Median OS was 16.7 months (95% CI, 13.9-19.7 months) with pembrolizumab and 12.1 months (95% CI, 11.3-13.3 months) with platinum-doublet therapy [Lopes, G., et al 2018].

The use of a lenvatinib matching placebo in combination with pembrolizumab will ensure the objectivity of the local investigators' treatment decision and AE causality assessments, while still providing participants the SOC treatment.

4.3 Justification for Dose

4.3.1 Maximum Dose/Exposure for This Study

NOTE: As of Amendment 007-06, lenvatinib, matching placebo, and Second-course treatment have been removed from the study. This section has been updated accordingly.

The maximum dose/exposure of pembrolizumab allowed in this study is 200 mg Q3W up to 35 cycles (approximately 2 years) for the Initial Treatment Phase.

4.3.2 Rationale for Dose Interval and Study Design

4.3.2.1 Lenvatinib Dosing

NOTE: As of Amendment 007-06, lenvatinib and matching placebo have been removed from the study. This section is no longer applicable, but has been left unchanged for reference.

The dosing regimen of lenvatinib was selected based on the results of the Phase 1b portion of Phase 1b/2 Study 111/KEYNOTE-146, of which the primary endpoint was to determine the MTD and RP2D for lenvatinib in combination with pembrolizumab 200 mg Q3W. Thirteen participants (lenvatinib 24 mg/day + pembrolizumab 200 mg IV Q3W: n=3; lenvatinib 20 mg/day + pembrolizumab 200 mg: n=10) were enrolled into the Phase 1b portion of the

- Result in individual participant exposures within a range that was well tolerated and safe in participants with melanoma and NSCLC

Clinical data have shown meaningful improvement in benefit-risk, including OS, at 200 mg Q3W across multiple indications. Additionally, a fixed-dose regimen simplified the dosing regimen to be more convenient for physicians and to reduce the potential for dosing errors. A fixed-dosing scheme also reduced complexity in the logistical chain at treatment facilities and reduced wastage.

4.4 Beginning and End of Study Definition

The overall study begins when the first participant (or their legally acceptable representative) provides documented informed consent. The overall study ends when the last participant completes the last study-related contact, withdraws consent, or is lost to follow-up (ie, the participant is unable to be contacted by the investigator).

4.4.1 Clinical Criteria for Early Study Termination

The clinical study may be terminated early if the extent (incidence and/or severity) of emerging effects/clinical endpoints is such that the risk/benefit ratio to the study population as a whole is unacceptable. In addition, further recruitment in the study or at (a) particular study site(s) may be stopped due to insufficient compliance with the protocol, Good Clinical Practice (GCP), and/or other applicable regulatory requirements, procedure-related problems or an unacceptably high number of discontinuations or withdrawals due to administrative reasons.

5 STUDY POPULATION

Male and female participants with Stage IV NSCLC, who express PD-L1 (TPS ≥1%), have received no systemic anticancer therapy for their Stage IV NSCLC, and are at least 18 years of age will be enrolled in this study.

Prospective approval of protocol deviations to recruitment and enrollment criteria, also known as protocol waivers or exemptions, is not permitted.

5.1 Inclusion Criteria

To be eligible for inclusion in this study, the participant must:

Type of Participant and Disease Characteristics

1. Have a histologically or cytologically confirmed diagnosis of NSCLC.

   Note: Mixed tumors will be categorized by the predominant cell type; if small-cell elements are present, the participant is ineligible.

2. Have Stage IV NSCLC (American Joint Committee on Cancer [AJCC], version 8).

- Bisphosphonates or denosumab
- Antihypertensive therapy (including additional antihypertensive treatment as appropriate if BP increases once the participant is enrolled)

6.5.2 Prohibited Concomitant Medications

Participants are prohibited from receiving the following therapies during the screening and treatment phases of the study:

- Concurrent anticancer therapies, such as chemotherapy, targeted therapies, antitumor interventions (surgical resection, surgical debulking of tumor, etc.), or cancer immunotherapy not specified in this protocol
  - Note: Topical anticancer agents to treat skin lesions (eg, in situ melanoma or squamous cell carcinoma) are allowed, excluding skin metastasis of melanoma.
- Other concurrent investigational drugs
- Live or attenuated vaccines within 30 days and while participating in the study. Note: killed vaccines are allowed.
- Systemic glucocorticoids for any purpose other than those listed in Section 6.5.4.1
- Radiation therapy for disease control
  - Note: Palliative radiotherapy is permitted for nontarget lesions if considered medically necessary by the treating physician and upon discussion with the Sponsor.

For participants who, in an assessment by the investigator, require the use of any of the aforementioned treatments for clinical management, continuation of the study intervention and further participation in the study must be discussed and agreed upon with the Sponsor.

If participants receive additional anticancer therapies, this will be judged to represent evidence of PD, and study intervention will be discontinued. These participants should complete all end-of-treatment assessments and continue to be followed for survival during the follow-up period.

6.5.3 Drug Interactions

NOTE: As of Amendment 007-06, lenvatinib and matching placebo have been removed from the study. This section is no longer applicable, but has been left unchanged for reference.

There are no DDI-related concomitant medication prohibitions or restrictions.

Lenvatinib is not expected to clinically meaningfully alter exposure to CYP3A4/ P-glycoprotein (Pgp) substrates based on results from a lenvatinib drug-drug interaction

Table 3 Dose Modification and Toxicity Management Guidelines for Immune-related Adverse Events Associated with Pembrolizumab Monotherapy, Coformulations or IO Combinations

General instructions:

1. Severe and life-threatening irAEs should be treated with IV corticosteroids followed by oral steroids. Other immunosuppressive treatment should begin if the irAEs are not controlled by corticosteroids.
2. Pembrolizumab monotherapy, coformulations or IO combinations must be permanently discontinued if the irAE does not resolve or the corticosteroid dose is not ≤10 mg/day within 12 weeks of the last treatment.
3. The corticosteroid taper should begin when the irAE is ≤ Grade 1 and continue at least 4 weeks.
4. If pembrolizumab monotherapy, coformulations or IO combinations have been withheld, treatment may resume after the irAE decreased to ≤ Grade 1 after corticosteroid taper.

irAEs	Toxicity Grade (CTCAEv4.0)	Action With Pembrolizumab Monotherapy, Coformulations or IO Combinations	Corticosteroid and/or Other therapies	Monitoring and Follow-up
Pneumonitis	Grade 2	Withhold	Administer corticosteroids (initial dose of 1-2 mg/kg prednisone or equivalent) followed by taper	Monitor participants for signs and symptoms of pneumonitis Evaluate participants with suspected pneumonitis with radiographic imaging and initiate corticosteroid treatment Add prophylactic antibiotics for opportunistic infections
	Recurrent Grade 2 or Grade 3 or 4	Permanently discontinue		

irAEs	Toxicity Grade (CTCAEv4.0)	Action With Pembrolizumab Monotherapy, Coformulations or IO Combinations	Corticosteroid and/or Other therapies	Monitoring and Follow-up
Hypophysitis	Grade 2	Withhold	Administer corticosteroids and initiate hormonal replacements as clinically indicated	Monitor for signs and symptoms of hypophysitis (including hypopituitarism and adrenal insufficiency)
	Grade 3 or 4	Withhold or permanently discontinue [a]		
Hyperthyroidism	Grade 2	Continue	Treat with nonselective beta-blockers (eg, propranolol) or thionamides as appropriate	Monitor for signs and symptoms of thyroid disorders
	Grade 3 or 4	Withhold or Permanently discontinue [a]		
Hypothyroidism	Grade 2-4	Continue	Initiate thyroid replacement hormones (eg, levothyroxine or liothyronine) per standard of care	Monitor for signs and symptoms of thyroid disorders
Nephritis and renal dysfunction	Grade 2	Withhold	Administer corticosteroids (prednisone 1-2 mg/kg or equivalent) followed by taper	Monitor changes of renal function
	Grade 3 or 4	Permanently discontinue		
Myocarditis	Grade 1	Withhold	Based on severity of AE administer corticosteroids	Ensure adequate evaluation to confirm etiology and/or exclude other causes
	Grade 2, 3 or 4	Permanently discontinue		

If the investigator recommends continuation of study intervention beyond disease progression, the participant or his/her legally acceptable representative will be asked to sign consent.

Specifics about the study and the study population are to be included in the study informed consent form.

Informed consent will adhere to IRB/IEC requirements, applicable laws and regulations, and Sponsor requirements.

8.1.2 Inclusion/Exclusion Criteria

All inclusion and exclusion criteria will be reviewed by the investigator who is a qualified physician to ensure that the participant qualifies for the study.

8.1.3 Participant Identification Card

All participants will be given a participant identification card identifying them as participants in a research study. The card will contain study site contact information (including direct telephone numbers) to be used in the event of an emergency. The investigator or qualified designee will provide the participant with a participant identification card immediately after the participant provides documented informed consent. At the time of intervention allocation/randomization, site personnel will add the treatment/randomization number to the participant identification card.

The participant identification card also contains contact information for the emergency unblinding call center so that a healthcare provider can obtain information about study intervention in emergency situations where the investigator is not available.

8.1.4 Medical History

8.1.4.1 General Medical History

A medical history will be obtained by the investigator or qualified designee.

Medical history will include all active conditions, drug allergies, significant medical procedures, and any condition diagnosed within the previous 10 years that are considered to be clinically important by the investigator. Any cancer, other than the cancer under study, will be recorded as medical history, even if diagnosed greater than 10 years before enrollment. Details regarding the cancer under study will be recorded separately and not listed as medical history. The medical history will also include an assessment of smoking history.

8.1.4.2 Oncologic Disease Details

The investigator or qualified designee will obtain historic and current details of the participant's cancer under study. This information will include, but is not limited to, date of

8.2.1.2 Tumor Scans During the Study

The first on-study scan assessment should be performed at 9 weeks (63 days ±7 days) from the date of randomization. Subsequent tumor scans should be performed every 9 weeks (63 days ±7 days) or more frequently if clinically indicated. After 54 weeks, participants who remain on treatment will have scans performed every 12 weeks (84 days ±7 days). Scan timing should follow calendar days and should not be adjusted for delays in cycle starts. Scans should continue to be performed until PD is identified by the investigator (unless the investigator elects to continue treatment and follow iRECIST), the start of new anticancer treatment, withdrawal of consent, death, or any other reason for discontinuation of pembrolizumab listed in Section 7.1 is met, whichever occurs first. All supplemental scans must be submitted to the central imaging vendor.

OR should be confirmed by a repeat scan assessment. Tumor scans to confirm PR or CR should be performed at least 4 weeks after the first indication of a response is observed. Participants will then return to regular scheduled scans, starting with the next scheduled time point. Participants who receive additional scans for confirmation do not need to undergo the next scheduled tumor scan if it is less than 4 weeks later; tumor scans may resume at the subsequent scheduled time point. Note: Response does not need to be verified in real time by the central imaging vendor.

Per iRECIST (Section 8.2.1.6), PD should be confirmed by the site 4 to 8 weeks after site-assessed first radiologic evidence of PD in clinically stable participants. Participants who have unconfirmed PD may continue on treatment at the discretion of the investigator until PD is confirmed by the site, provided they have met the conditions detailed in Section 8.2.1.6. Participants who receive confirmatory scans do not need to undergo the next scheduled tumor scan if it is less than 4 weeks later; tumor scans may resume at the subsequent scheduled time point, if clinically stable. Participants who have confirmed PD by iRECIST, as assessed by the site, will discontinue study intervention. Exceptions are detailed in Section 8.2.1.6.

8.2.1.3 End of Treatment and Follow-up Tumor Scans

For participants who discontinue study intervention, tumor scans should be performed at the time of treatment discontinuation (±4-week window). If previous scans were obtained within 4 weeks before the date of discontinuation, then scans at treatment discontinuation is not mandatory. For participants who discontinue study intervention because of documented PD, this is the final required tumor scan.

After amendment 007-06, imaging obtained on study will be submitted to the imaging vendor but these will not be assessed; similarly, verification of PD is no longer needed before treatment discontinuation.

8.2.1.4 Second-course Treatment Tumor Scans

NOTE: As of Amendment 007-06, the Second-course Treatment Phase has been removed from this study. This section is no longer applicable.

Table 6 Reporting Time Periods and Time Frames for Adverse Events and Other Reportable Safety Events

Type of Event	**Reporting Time Period:** **Consent to Randomization/ Allocation**	**Reporting Time Period:** **Randomization/ Allocation through Protocol-specified Follow-up Period**	**Reporting Time Period:** **After the Protocol-specified Follow-up Period**	**Time Frame to Report Event and Follow-up Information to Sponsor:**
Nonserious Adverse Event (NSAE)	Report if: - due to protocol-specified intervention - causes exclusion - participant is receiving placebo run-in or other run-in treatment	Report all	Not required	Per data entry guidelines
Serious Adverse Event (SAE) including Cancer and Overdose	Report if: - due to protocol-specified intervention - causes exclusion - participant is receiving placebo run-in or other run-in treatment	Report all SAEs, cancer, overdose associated with pembrolizumab and overdose associated with lenvatinib with an AE	Report if: - drug/vaccine related. (Follow ongoing to outcome)	Within 24 hours of learning of event
Pregnancy/ Lactation Exposure	Report if: - participant has been exposed to any protocol-specified intervention (eg, procedure, washout or run-in treatment including placebo run-in) Exception: A positive pregnancy test at the time of initial screening is not a reportable event.	Report all	Previously reported – Follow to completion/termination; report outcome	Within 24 hours of learning of event
Event of Clinical Interest (require regulatory reporting)	Report if: - due to intervention - causes exclusion	Report - potential drug-induced liver injury (DILI) - require regulatory reporting	Not required	Within 24 hours of learning of event
Event of Clinical Interest (do not require regulatory reporting)	Report if: - due to intervention - causes exclusion	Report - non-DILI ECIs and those not requiring regulatory reporting	Not required	Within 5 calendar days of learning of event

8.10 Medical Resource Utilization and Health Economics

All-cause hospitalizations and emergency room visits must be reported in the eCRF, from the time of treatment randomization through 90 days after cessation of study intervention, or 30 days after cessation of study intervention, if the participant initiates new anticancer therapy, whichever is earlier.

8.11 Visit Requirements

NOTE: As of Amendment 007-06, the Second-course Treatment Phase has been removed from the study. The following subsections have been updated accordingly.

Visit requirements are outlined in the SoA (Section 1.3). Specific procedure-related details are provided in Section 8.

Unscheduled visits are permitted at any time during the course of the study.

8.11.1 Screening

Documented informed consent must be provided before performing any protocol-specific procedure. Results of a test performed before the participant signs the ICF as part of routine clinical management are acceptable in lieu of a screening test if performed within the specified time frame. Screening procedures are to be completed within 28 days before the first dose of study intervention except for the following:

- Laboratory tests are to be performed per the SoA (Section 1.3).
- Evaluation of ECOG is to be performed at screening, within 7 days before Cycle 1 but before randomization.
- For WOCBP, a urine or serum pregnancy test will be performed within 24 hours before the first dose of study intervention. If urine pregnancy results cannot be confirmed as negative, a serum pregnancy test will be required (performed by the local study site laboratory).

Participants may be rescreened after initially failing to meet the inclusion/exclusion criteria. Results from assessments during the initial screening period are acceptable in lieu of a repeat screening test if performed within the specified time frame and the corresponding inclusion/exclusion criteria are met. Participants who are rescreened will retain their original screening number.

8.11.2 Initial Treatment Phase

Visit requirements are outlined in the SoA (Section 1.3). Specific procedure-related details are provided in Section 8.1.

Analysis Populations	Efficacy: Intention-to-Treat (ITT) Safety: All Participants as Treated (APaT)
Primary Endpoints	• PFS per RECIST 1.1 based on BICR • OS
Secondary Endpoints	• OR per RECIST 1.1 based on BICR • Safety and tolerability • Change from baseline in Global health status/QoL, cough, chest pain, dyspnea and physical functioning scores • Time to True Deterioration (TTD) in global health status/QoL, cough, chest pain, dyspnea, and physical functioning • TTD in the composite endpoint of cough, chest pain, or dyspnea items
Statistical Methods for Key Efficacy Analyses	**As of Amendment 007-06, the prespecified interim and final analyses of the study will not be performed. There will be no further analyses of efficacy endpoints.** The primary hypotheses will be evaluated by comparing pembrolizumab + lenvatinib to pembrolizumab + placebo in PFS and OS using a stratified log-rank test and in ORR using the stratified Miettinen and Nurminen method [Miettinen, O. 1985]. The HR will be estimated using a stratified Cox regression model. Event rates over time will be estimated within each treatment group using the Kaplan-Meier method. The difference in ORR will be estimated using the stratified Miettinen and Nurminen method with strata weighting by sample size.
Statistical Methods for Key Safety Analyses	The analysis of safety results will follow a tiered approach. The tiers differ with respect to the analyses that will be performed. There are no events of interest that warrant elevation to Tier 1 events in this study. Tier 2 parameters will be assessed via point estimates with 95% CIs provided for between-group comparisons; only point estimates by treatment group are provided for Tier 3 safety parameters. The 95% CIs for the between-treatment differences in percentages will be provided using the Miettinen and Nurminen method.
Interim Analyses	**As of Amendment 007-06, no interim analyses of the study will be performed.**
Multiplicity	**As of Amendment 007-06, no interim or final analyses of the study will be performed.**
Sample Size and Power	**As of Amendment 007-06, no interim or final analyses of the study will be performed.** The planned sample size is approximately 620 participants. For PFS, based on 416 events, the study has 86.5% power to detect a HR of 0.7 (pembrolizumab + lenvatinib vs pembrolizumab + placebo) at α=0.55% (1-sided). For OS, based on 388 events, the study has 90% power to detect a HR of 0.71 (pembrolizumab + lenvatinib vs pembrolizumab + placebo) at α=1.95% (1-sided).
China Extension Study	**As of Amendment 007-06, enrollment in the Extension Study in China has been stopped.** China participants randomized during the global study phase will be included in all global study analyses (efficacy and safety). China participants randomized during the China extension phase will be excluded from all global study analyses. China participants randomized during global and extension phases will both be included in any China-specific analyses.

Table 7 Analysis Methods for Key Efficacy Endpoints

Endpoint/Variable	Statistical Method	Analysis Population	Missing Data Approach
Primary Analyses:			
PFS (RECIST 1.1) by BICR	Testing: stratified log-rank test Estimation: Stratified Cox model with Efron's tie-handling method	ITT	Censored according to rules in Table 8
OS	Testing: stratified log-rank test Estimation: Stratified Cox model with Efron's tie-handling method	ITT	Censored at last known alive date
Secondary Analyses:			
ORR (RECIST 1.1) by BICR	Testing and estimation: stratified Miettinen and Nurminen method	ITT	Participants with missing data are considered nonresponders
Abbreviations: BICR = blinded independent central review; ITT = intention-to-treat; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors.			

The strategy to address multiplicity issues with regard to multiple endpoints and interim analyses is described in Section 9.7, Interim Analyses, and in Section 9.8, Multiplicity.

9.6.1.1 Progression-free Survival

The nonparametric Kaplan-Meier method will be used to estimate the PFS curve in each treatment group. The hypotheses of treatment difference in PFS will be tested by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to estimate the magnitude of the treatment difference (ie, HR) between the treatment arms. The HR and its 95% CI from the stratified Cox model with Efron's method of tie handling and with a single treatment covariate will be reported.

Since PD is assessed periodically, PD can occur any time in the time interval between the last assessment where PD was not documented and the assessment when PD is documented. For the primary analysis, for the participants who have PD, the true date of PD will be approximated by the date of the first assessment at which PD is objectively documented per RECIST 1.1 by the BICR vendor, regardless of discontinuation of study drug. Additional analyses will be performed for PFS per iRECIST.

To evaluate the robustness of the PFS endpoint per RECIST 1.1 via BICR by the imaging vendor, 1 primary and 2 sensitivity analyses with a different set of censoring rules will be performed. For the primary analysis, if the events (PD or death) are after more than 1 missed disease assessment, the data are censored at the last disease assessment before missing visits. Also, data after new anticancer therapy are censored at the last disease assessment before the initiation of new anticancer therapy. The first sensitivity analysis follows ITT principles (ie,

date of last known contact. The analysis using the restricted mean survival time method may be conducted for OS to account for the possible nonproportional hazards effect.

9.6.1.3 Objective Response Rate

The stratified Miettinen and Nurminen method will be used for comparison of ORR between the treatment groups. The difference in ORR and its 95% CI from the stratified Miettinen and Nurminen method with strata weighting by sample size will be provided.

9.6.1.4 Duration of Response

For participants who demonstrate CR or PR, DOR is defined as the time from first documented evidence of CR or PR until PD or death due to any cause, whichever occurs first.

The nonparametric Kaplan-Meier method will be used to summarize the DOR. The median and range of DOR will be provided.

Censoring rules for DOR are summarized in Table 9.

Table 9 Censoring Rules for Duration of Response

Situation	Date of Progression or Censoring	Outcome
No PD or death, no new anticancer therapy initiated	Last adequate disease assessment	Censor (nonevent)
No PD or death, new anticancer therapy initiated	Last adequate disease assessment before new anticancer therapy initiated	Censor (nonevent)
Death or PD immediately after ≥2 consecutive missed disease assessments or after new anticancer therapy, if any	Earlier date of last adequate disease assessment before ≥2 missed adequate disease assessments and new anticancer therapy, if any	Censor (nonevent)
Death or PD after ≤1 missed disease assessments and before new anticancer therapy, if any	PD or death	End of response (event)

Abbreviation: PD = progressive disease.
Note: A missed disease assessment includes any assessment that is not obtained or is considered inadequate for evaluation of response.

9.6.2 Statistical Methods for Safety Analyses

Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, laboratory tests, vital signs, and ECG measurements.

The analysis of safety results will follow a tiered approach (Table 10). The tiers differ with respect to the analyses that will be performed. AEs (specific terms as well as system organ

Table 10 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoint	95% CI for Treatment Comparison	Descriptive Statistics
Tier 2	Any AE (≥10% of participants in one of the treatment groups)	X	X
	Any Grade 3 to 5 AE (≥5% of participants in one of the treatment groups)	X	X
	Any serious AE (≥5% of participants in one of the treatment groups)	X	X
Tier 3	Any AE		X
	Change from baseline results (laboratory test toxicity grade)		X
Abbreviations: AE = adverse event; CI = confidence interval; X = results will be provided.			

9.6.3 Statistical Methods for Patient-reported Outcome Analyses

Mean change from baseline

The time point for the mean change from baseline will be determined based on blinded data review prior to the database lock for any PRO analysis and documented in the sSAP.

To assess the treatment effects on the PRO score change from baseline in the PRO endpoints defined in Section 9.4.3, a constrained longitudinal data analysis (cLDA) model proposed by Liang and Zeger [Liang, Kung-Yee and Zeger, Scott L. 2000] will be applied, with the PRO score as the response variable, and treatment, time, the treatment by time interaction, and stratification factors used for randomization (See Section 6.3.2) as covariates. The treatment difference in terms of least square (LS) mean change from baseline will be estimated from this model together with 95% CI. Model-based LS mean with 95% CI will be provided by treatment group for PRO scores at baseline and postbaseline time point.

Time to Deterioration (TTD)

For the TTD endpoints defined in Section 9.4.3, the Kaplan-Meier method will be used to estimate the TTD curve for each treatment group. The estimate of median time to deterioration and its 95% CI will be obtained from the Kaplan-Meier estimates. The treatment difference in TTD will be assessed by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling and with a single treatment covariate will be used to assess the magnitude of the treatment difference (ie, HR). The HR and its 95% CI will be reported. The same stratification factors used for the stratified PFS and OS analyses will be used as the stratification factors in both the stratified log-rank test and the stratified Cox model.

OS follows an exponential distribution, with a median of 16.7 months for the control group.

Enrollment period is approximately 21.7 months.

Annual drop-out rate is 20% and 2% for PFS and OS, respectively.

Follow-up period is approximately 8.8 and 18.5 months for PFS and OS, respectively, after the last participant is randomized.

The sample size and power calculations were performed in R (package "gsDesign") and EAST 6.4.

Extension Study in China

As of amendment 007-06, enrollment in the Extension Study in China has been stopped.

In order to evaluate the consistency of efficacy and safety in the subpopulation in China compared with the global population, following completion of global study enrollment, participants in China will continue to be randomized in a 1:1 ratio into the pembrolizumab + lenvatinib arm and pembrolizumab + placebo arm until the planned sample size of approximately 120 participants in China is reached. Participants in China randomized after completion of enrollment in the global study will not be included in the analysis of the global study.

9.10 Subgroup Analyses

To determine whether the treatment effect is consistent across various subgroups, the between-group treatment effect for OS, PFS, and OR (with a nominal 95% CI) will be estimated and plotted by treatment group within each category of the following subgroup variables:

- Geographic region (East Asia, non-East Asia)
- ECOG performance status (0, 1)
- Predominant tumor histology (squamous, nonsquamous)
- TPS (1%-49%, ≥50%)
- Age category (<65, ≥65 years)
- Sex (female, male)
- Race (white, nonwhite)
- Smoking status (never, former/current smoker)

and accidental trauma (eg, sprained ankle) that may interfere with or prevent everyday life functions but do not constitute a substantial disruption.

e. Is a congenital anomaly/birth defect

- In offspring of participant taking the product regardless of time to diagnosis.

f. Other important medical events

- Medical or scientific judgment should be exercised in deciding whether SAE reporting is appropriate in other situations such as important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the participant or may require medical or surgical intervention to prevent 1 of the other outcomes listed in the above definition. These events should usually be considered serious.

- Examples of such events include invasive or malignant cancers, intensive treatment in an emergency room or at home for allergic bronchospasm, blood dyscrasias or convulsions that do not result in hospitalization, or development of drug dependency or drug abuse.

10.3.3 Additional Events Reported in the Same Manner as SAE

Additional events that require reporting in the same manner as SAE

In addition to the above criteria, AEs meeting either of the below criteria, although not serious per ICH definition, are reportable to the Sponsor in the same time frame as SAEs to meet certain local requirements. Therefore, these events are considered serious by the Sponsor for collection purposes.

- Is a new cancer (that is not a condition of the study)

- Is associated with an overdose of pembrolizumab

10.3.4 Recording AE and SAE

AE and SAE recording

- When an AE/SAE occurs, it is the responsibility of the investigator to review all documentation (eg, hospital progress notes, laboratory, and diagnostics reports) related to the event.

- The investigator will record all relevant AE/SAE information on the AE CRFs/worksheets at each examination.

- It is not acceptable for the investigator to send photocopies of the participant's medical records to the Sponsor in lieu of completion of the AE CRF page.

3. **The following components are to be used to assess the relationship between the Sponsor's product and the AE; the greater the correlation with the components and their respective elements (in number and/or intensity), the more likely the Sponsor's product caused the AE:**

   - **Exposure:** Is there evidence that the participant was actually exposed to the Sponsor's product such as: reliable history, acceptable compliance assessment (pill count, diary, etc.), expected pharmacologic effect, or measurement of drug/metabolite in bodily specimen?

   - **Time Course:** Did the AE follow in a reasonable temporal sequence from administration of the Sponsor's product? Is the time of onset of the AE compatible with a drug-induced effect (applies to studies with investigational medicinal product)?

   - **Likely Cause:** Is the AE not reasonably explained by another etiology such as underlying disease, other drug(s)/vaccine(s), or other host or environmental factors.

   - **Dechallenge:** Was the Sponsor's product discontinued or dose/exposure/frequency reduced?

   - If yes, did the AE resolve or improve?

   - If yes, this is a positive dechallenge.

   - If no, this is a negative dechallenge.

   - (Note: This criterion is not applicable if: (1) the AE resulted in death or permanent disability; (2) the AE resolved/improved despite continuation of the Sponsor's product; (3) the study is a single-dose drug study; or (4) Sponsor's product(s) is/are only used 1 time.)

   - **Rechallenge:** Was the participant re-exposed to the Sponsor's product in this study?

   - If yes, did the AE recur or worsen?

   - If yes, this is a positive rechallenge.

   - If no, this is a negative rechallenge.

   (Note: This criterion is not applicable if: (1) the initial AE resulted in death or permanent disability, or (2) the study is a single-dose drug study; or (3) Sponsor's product(s) is/are used only 1 time.)

NOTE: IF A RECHALLENGE IS PLANNED FOR AN AE THAT WAS SERIOUS AND MAY HAVE BEEN CAUSED BY THE SPONSOR'S PRODUCT, OR IF RE-EXPOSURE

10.4 Appendix 4: Device Events, Adverse Device Events, and Medical Device Incidents: Definitions, Collection, and Documentation

This section is not applicable.

A urine pregnancy test will be obtained per the SoA (Section 1.3). Additional urine/serum testing may be performed if clinically warranted and/or as defined by local regulations. If a urine pregnancy test cannot be confirmed as negative, a serum pregnancy test is required.

Abbreviation	Expanded Term
IUD	intrauterine device
iUPD	unconfirmed PD
IUS	intrauterine hormone-releasing system
IV	intravenous
KL-6	Krebs von den Lungen-6
LEAP	lenvatinib and pembrolizumab
LVEF	left ventricular ejection fraction
mAb	monoclonal antibody
MRI	magnetic resonance imaging
MTD	maximum tolerated dose
MUGA	multigated acquisition scan
NCI	National Cancer Institute
NSCLC	non–small cell lung cancer
OR	objective response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PD-1	programmed cell death protein
PDGFR	platelet-derived growth factor receptor
PD-L1	programmed cell death ligand 1
PD-L2	programmed cell death ligand 2
PFS	progression-free survival
Pgp	P-glycoprotein
PK	pharmacokinetic
PopPK	population pharmacokinetics
PR	partial response
PRES	posterior reversible encephalopathy syndrome
PRO	patient-reported outcomes
Q3W	every 3 weeks
QD	once daily
QLQ-LC13	Quality of Life Questionnaire and Lung Cancer Module 13
QoL	quality of life

- Improvement for the dyspnea, cough, and chest pain scores is defined by a decrease of 10 points or more in score from baseline at any time during the study and confirmed by a decrease of 10 points or more in score at the next consecutive visit.

- Stability is defined as follows, when the criteria for improvement are not met:
  - an improvement confirmed by a less than 10 points change from baseline at the next consecutive visit
  - less than 10 points change in score confirmed by a less than 10 points change in score at the next consecutive visit
  - less than 10 points change in score and confirmed by an improvement at the next consecutive visit
- Stability + improvement is defined as the sum of proportions with stability and improvement as specified above.
- Deterioration is defined as a 10 points or greater worsening from baseline at any time during the study when the criteria for improvement or stability are not met.

3.5. Analysis Populations

3.5.1. Efficacy Analysis Populations

The analyses of the primary efficacy endpoints are based on the ITT population. All randomized participants will be included in this population. Participants will be analyzed in the treatment group to which they were randomized. Based on actual enrollment, all participants who were randomized on or prior to January 28, 2021 will be included in the ITT population of the global study. All participants who failed screening in addition to all randomized participants on or prior to January 28, 2021will be included in the screening population. Details on the approach to handling missing data are provided in Section 3.6.

3.5.2. Safety Analysis Populations

Safety analyses will be conducted in the APaT population, which consists of all randomized participants who received at least 1 dose of study intervention. Participants will be included in the treatment group corresponding to the study intervention they actually received for the analysis of safety data using the APaT population. This will be the treatment group to which they are randomized, except for participants who take incorrect study intervention for the entire treatment period; such participants will be included in the treatment group corresponding to the study intervention actually received. Any participant who receives the incorrect study intervention for 1 cycle, but receives the randomized treatment for all other cycles, will be analyzed according to the randomized treatment group, and a narrative will be provided for any events that occur during the cycle for which the participant is incorrectly dosed.

Cox model. Participants without documented death at the time of analysis will be censored at the date of last known contact.

In case the proportional hazards assumption is not valid, RMST method may be conducted for OS to account for the possible non-proportional hazards effect as a sensitivity analysis.

3.6.1.3. Objective Response Rate

The stratified Miettinen and Nurminen method will be used for comparison of ORR between the treatment groups. The difference in ORR and its 95% CI from the stratified Miettinen and Nurminen method with strata weighting by sample size will be provided. The same stratification factors used for randomization (Section 6.3.2 in the protocol) will be used as stratification factors in the analysis.

3.6.1.4. Duration of Response

For subjects who demonstrate CR or PR, DOR is defined as the time from first documented evidence of CR or PR until PD or death due to any cause, whichever occurs first.

The nonparametric Kaplan-Meier method will be used to summarize the DOR. The median and range of DOR will be provided.

Censoring rules for DOR are summarized in Table 3.

Table 3 Censoring Rules for Duration of Response

Situation	Date of Progression or Censoring	Outcome
No PD or death, no new anticancer therapy initiated	Last adequate disease assessment	Censor (nonevent)
No PD or death, new anticancer therapy initiated	Last adequate disease assessment before new anticancer therapy initiated	Censor (nonevent)
Death or PD immediately after ≥2 consecutive missed disease assessments or after new anticancer therapy, if any	Earlier date of last adequate disease assessment before ≥2 missed adequate disease assessments and new anticancer therapy, if any	Censor (nonevent)
Death or PD after ≤1 missed disease assessments and before new anticancer therapy, if any	PD or death	End of response (event)
Abbreviation: PD = progressive disease. Note: A missed disease assessment includes any assessment that is not obtained or is considered inadequate for evaluation of response.		

3.6.2. Statistical Methods for Safety Analyses

Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, laboratory tests, vital signs, and ECG measurements.

Table 4 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoint	p-Value	95% CI for Treatment Comparison	Descriptive Statistics
Tier 2	Grade 3-5 AE (incidence ≥5% of participants in one of the treatment groups)		X	X
	Serious AE (incidence ≥5% of participants in one of the treatment groups)		X	X
	AEs (incidence ≥10% of participants in one of the treatment groups)		X	X
Tier 3	Any AE			X
	Any Grade 3-5 AE			X
	Any Serious AE			X
	Any Drug-Related AE			X
	Any Serious and Drug-Related AE			X
	Any Grade 3-5 and Drug-Related AE			X
	Discontinuation due to AE			X
	Death			X
	Specific AEs, SOCs (incidence <10% of participants in all of the treatment groups)			X
	Change from Baseline Results (lab toxicity shift)			X
Abbreviations: AE = adverse event; CI = confidence interval; SOC = system organ class.				

Exposure-Adjusted Approach

To properly account for the potential difference in follow-up time between the study arms, AE incidence adjusted for treatment exposure analyses may be performed as appropriate.

Time to Grade 3-5 AE

In addition to the tiered approach, exploratory analysis may be performed on the time to the first Grade 3-5 AE. The time to the first Grade 3-5 AE is defined as the time from the first day of study drug to the first event of a Grade 3-5 AE. Summary statistics will be provided.

3.6.3. Statistical Methods for Patient-reported Outcome Analyses

3.6.3.1. Scoring Algorithm

EORTC QLQ-C30:

For each scale or item, a linear transformation will be applied to standardize the score as between 0 and 100, according to the corresponding scoring standard. For global health status/quality of life and all functional scales, a higher value indicates a better level of function; for symptom scales and items, a higher value indicates increased severity of symptoms.

imagine'. The recall period is current health today (the day of completion). For the EQ-5D VAS scale, A ≥7-point change from baseline in VAS is considered to be a MID (Packard et al., 2007).

3.6.3.2. PRO Compliance Summary

Completion and compliance of EORTC QLQ-C30, EORTC-QLQ-LC13, and EQ-5D by visit and by treatment will be described. Numbers and percentages of complete and missing data at each visit will be summarized. An instrument is considered complete if at least one valid score is available according to the missing item rules outlined in the scoring manual for each respective instrument.

Completion rate of treated participants (CR-T) at a specific time point is defined as the number of treated participants who complete at least one item over the number of treated participants in the PRO analysis population.

$$\text{CR-T} = \frac{Number\ of\ treated\ participants\ who\ complete\ at\ least\ one\ item}{Number\ of\ treated\ participants\ in\ the\ PRO\ analysis\ population}$$

The completion rate is expected to shrink in the later visit during study period due to the participants who discontinued early. Therefore, another measurement, compliance rate of eligible participants (CR-E) will also be employed as the support for completion rate. CR-E is defined as the number of treated participants who complete at least one item over number of eligible participants who are expected to complete the PRO assessment, not including the participants missing by design such as death, discontinuation, translation not available.

$$\text{CR-E} = \frac{Number\ of\ treated\ participants\ who\ complete\ at\ least\ one\ item}{Number\ of\ eligible\ participants\ who\ are\ expected\ to\ complete}$$

The reasons of non-completion and non-compliance will be provided in supplementary table:

- Completed as scheduled
- Not completed as scheduled
- Off-study: not scheduled to be completed.

In addition, reasons for non-completion as scheduled of these measures will be collected using "miss_mode" forms filled by site personnel and will be summarized in table format. The schedule (study visits and estimated study times) and mapping of study visit to analysis visit for PRO data collection is provided in Table 5. If there are multiple PRO collections within any of the stated time windows, the assessment completed closest to the target collection day will be used in the analyses.

Table 6 Censoring Rules for Time-to-Deterioration

Scenario	Outcome
Deterioration documented	Event observed at time of assessment (first deterioration)
Ongoing, death or discontinued from study without deterioration	Right censored at time of last assessment
No baseline assessments	Right censored at treatment start date

3.6.3.5. Overall Improvement and Overall Improvement/Stability

Overall improvement rate will be analyzed, which is defined as the proportion of participants who have achieved an improvement as defined in Section 3.4.3 PRO Endpoints. The point estimate of overall proportions of participants who have achieved an improvement, stability, and deterioration will be provided by treatment group together with 95% CI using exact binomial method by Clopper and Pearson (1934). Stratified Miettinen and Nurminen's method will be used for comparison of the overall improvement rate between the treatment groups. The difference in overall improvement rate and its 95% CI, along with nominal two-sided p-values, from the stratified Miettinen and Nurminen's method with strata weighting by sample size will be provided. The stratification factors used for randomization (see Section 6.3.2 in protocol) will be applied to the analysis.

The same method will be used to analyze overall improvement/stability rate, which is defined as the proportion of participants who have achieved improvement/stability as defined in Section 3.4.3 PRO Endpoints.

3.6.3.6. Analysis Strategy for Key PRO Endpoints

A summary of the analysis strategy for the key PRO endpoints is provided in Table 7.

Table 8 Summary of Interim and Final Analyses Strategy

Analyses	Key Endpoints	Timing	Estimated Time After First Participant Randomized	Primary Purpose of Analysis
IA	PFS OS	Both ~416 PFS events have been observed and ~8.8 months after last participant randomized	~30.5 months	• Final PFS analysis • Interim OS analysis
Final analysis	OS	Both ~388 deaths have occurred and ~18.5 months after last participant randomized	~40.2 months	• Final OS analysis
Note that for the IA and the FA, if the events accrue slower than expected such that the targeted number of events cannot be reached in the anticipated timeframe, the Sponsor may conduct the analysis with an additional 2 months and 7 months of follow-up for the IA and FA respectively, or the specified number of events is observed, whichever occurs first. Abbreviations: IA = interim analysis; OS = overall survival; PFS = progression-free survival.				

Non-binding futility analyses will be conducted in July 2021 (data cutoff in May 2021) to coincide with an anticipated safety DMC meeting and at the IA as per DMC recommendation. Further details of any evaluation are specified in the Appendix 5.1.

3.7.2. Safety Interim Analyses

The DMC will conduct regular safety monitoring. The timing of the safety monitoring will be specified in the DMC charter.

3.8. Multiplicity

The study uses the graphical method of Maurer and Bretz [Maurer, W., et al 2011] to control multiplicity for multiple hypotheses as well as interim analysis. According to this approach, study hypotheses may be tested more than once, and when a particular null hypothesis is rejected, the α allocated to that hypothesis can be reallocated to other hypothesis tests. Figure *1* shows the initial 1-sided α allocation for each hypothesis in the ellipse representing the hypothesis. The weights for reallocation from each hypothesis to the others are represented in the boxes on the lines connecting the hypotheses.

Table 10 Efficacy Boundaries and Properties for Overall Survival Analyses

Analysis	**Value**	**α=0.0195**	**α=0.025**
IA: 78%[a] N: 620 Events: 302 Month:30.5 [f]	Z	2.4029	2.2869
	p (1-sided)[b]	0.0081	0.0111
	~HR at bound[c]	0.7584	0.7686
	P(Cross) if HR=1[d]	0.0081	0.0111
	P(Cross) if HR=0.71[e]	0.7173	0.7548
Final N: 620 Events: 388 Month: 40.2 [f]	Z	2.1181	2.0193
	p (1-sided)[b]	0.0171	0.0217
	~HR at bound[c]	0.8064	0.8145
	P(Cross) if HR=1[d]	0.0195	0.0250
	P(Cross) if HR=0.71[e]	0.9000	0.9165

Abbreviations: HR = hazard ratio; IA = interim analysis.
The number of events and timings are estimated.

[a] Percentage of total planned events at each interim analysis.
[b] The nominal α for group sequential testing.
[c] The approximate HR required to reach an efficacy bound
[d] The probability of crossing a bound under the null hypothesis
[e] The probability of crossing a bound under the alternative hypothesis
[f] The approximate number of months since first participant randomized

The bounds provided in Table *10* are based on the assumption that the expected number of events at the IA and FA for OS are 302 and 388, respectively. At the time of an analysis, the observed number of events may differ substantially from the expected. To avoid overspending at an interim analysis and leave reasonable alpha for the final analysis, the minimum alpha spending strategy will be adopted. At an IA, the information fraction used in Lan-DeMets spending function to determine the alpha spending at the IA will be based on the minimum of the expected information fraction and the actual information fraction at each analysis. Specifically,

- In the scenario that the events accrue slower than expected and the observed number of events is less than the expected number of events at a given analysis, the information fraction will be calculated as the observed number of events at the interim analysis over the target number of events at FA.

- In the scenario that the events accrue faster than expected and the observed number of events exceeds the expected number of events at a given analysis, then the information fraction will be calculated as the expected number of events at the interim analysis over the target number of events at FA.

reduction from the global primary efficacy analysis assuming the same hazard ratio used in the sample size and power calculation for the global study. The primary analysis for PFS will be conducted in the Chinese subpopulation when approximately 75 PFS events have been collected.

4.6.1.2. Overall Survival

Analysis of OS for extension is the same to that for the global study if applicable.

The nonparametric Kaplan-Meier method will be used to estimate the survival curves. The treatment difference in OS will be assessed by the stratified log-rank test (based on the stratification factors defined in Section 6.3.2 in the protocol). A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (ie, the HR). The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported. The stratification factors used for randomization (Section 6.3.2 in the protocol) will be applied to both the stratified log-rank test and the stratified Cox model. The same stratification factors used in the global study will be used. For the Chinese subgroup analysis, the stratified method will only be used if applicable. The factor of Geography (East Asia vs. non-East Asia) will not be included in the stratified analysis for the Chinese subgroup analysis. Participants without documented death at the time of analysis will be censored at the date of last known contact. An analysis using the restricted mean survival time method may be conducted for OS to account for the possible nonproportional hazards effect.

Consistency in OS will be evaluated similarly as that in PFS. The primary analysis for OS will be conducted in the Chinese subpopulation when approximately 75 OS events have been collected.

4.6.1.3. Objective Response Rate

Analysis of ORR or DOR for extension is the same to that for the global study if applicable.

The stratified Miettinen and Nurminen method will be used for comparison of ORR between the treatment groups. The difference in ORR and its 95% CI from the stratified Miettinen and Nurminen method with strata weighting by sample size will be provided. The same stratification factors used for randomization (Section 6.3.2 of the study protocol) will be used as stratification factors in the analysis. The same stratification factors used in the global study will be used. For the Chinese subpopulation analysis, the stratified method will only be used if applicable. The factor of Geography (East Asia vs. non-East Asia) will not be included in the stratified analysis for Chinese subgroup analysis.

4.6.1.4. Duration of Response

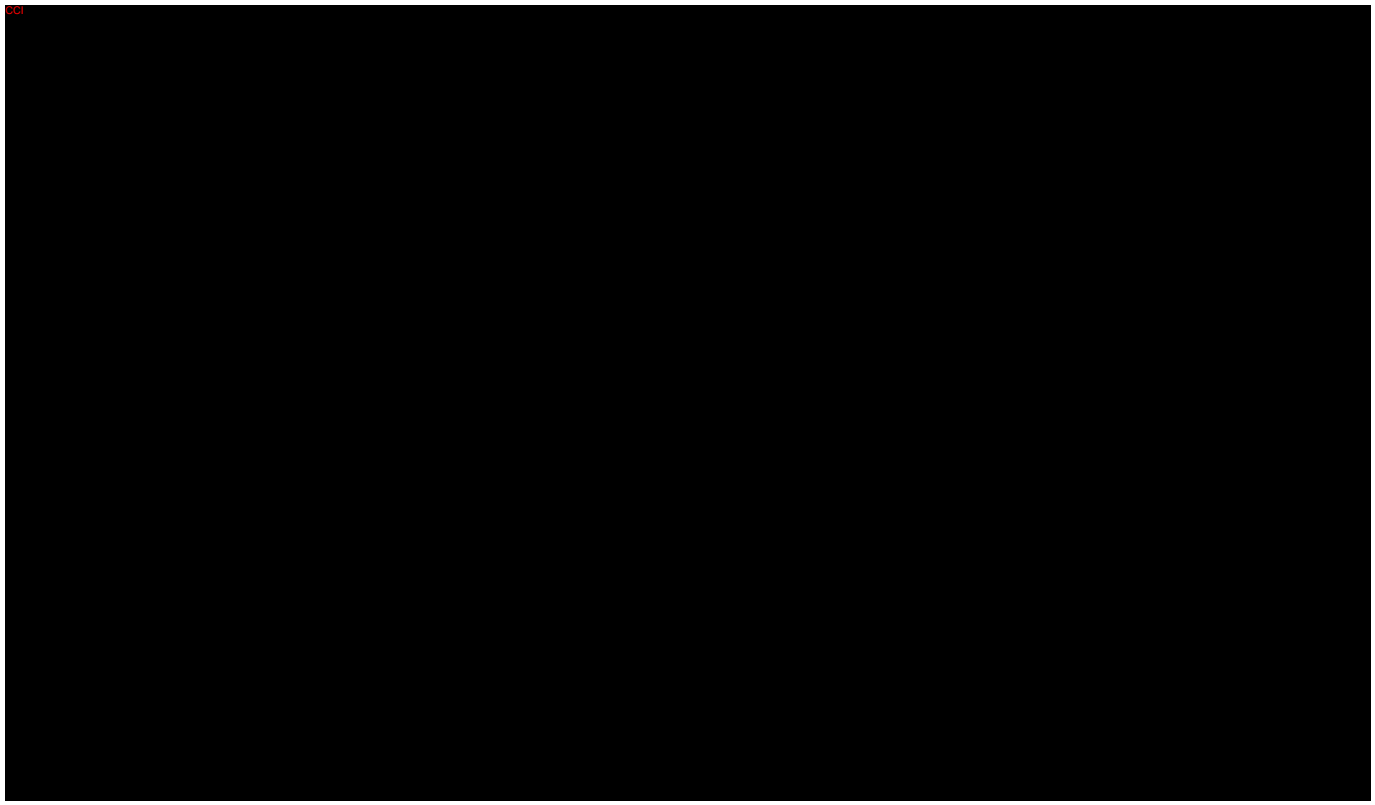
For subjects who demonstrate CR or PR, DOR is defined as the time from first documented evidence of CR or PR until PD or death due to any cause, whichever occurs first.

The nonparametric Kaplan-Meier method will be used to summarize the DOR. The median and range of DOR will be provided.

5. APPENDICES:

5.1. Rules for Non-Binding Futility Analyses

Non-binding futility analyses will be conducted to coincide with the safety DMC meeting in July 2021 (data cutoff in May 2021) and at the IA as per DMC recommendation. For each futility analysis, the DMC committee will evaluate the totality of the data and discuss with the EOC. The futility boundary for OS is based upon the Hwang-Shih-DeCani (HSD) beta-spending function with parameter of -15. The futility criteria are non-binding and only intended to provide guidance.



Abbreviation	Expanded Term
PK	pharmacokinetic
PR	partial response
PRO	patient-reported outcomes
QLQ-LC13	Quality of Life Questionnaire and Lung Cancer Module 13
QoL	quality of life
RECIST 1.1	Response Evaluation Criteria in Solid Tumors 1.1
SAE	serious adverse event
SAP	statistical analysis plan
sSAP	supplemental statistical analysis plan
TPS	tumor proportion score
TTD	time to true deterioration