Reporting and Analysis Plan

Study ID: 209229

Official Title of Study: A Randomized, Double-blind, Adaptive, Phase II/III Study of GSK3359609 or Placebo in Combination with Pembrolizumab for First-Line Treatment of PD-L1 Positive Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma

NCT ID: NCT04128696

Date of Document: 12-JUL-2021

Division	:	Worldwide Development
Information Type	:	Statistical Analysis Plan (SAP)

Title Page

Protocol Title: A Randomized, Double-blind, Adaptive, Phase II/III Study of GSK3359609 or Placebo in Combination with Pembrolizumab for First-Line Treatment of PD-L1 Positive Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma

Protocol Number: 209229

Compound Number: GSK3359609

Short Title: A Phase II/III study of GSK3359609 in combination with pembrolizumab compared with pembrolizumab plus placebo in participants with PD-L1 positive recurrent or metastatic HNSCC (INDUCE-3)

Sponsor Name: GlaxoSmithKline Research & Development Limited

Legal Registered Address: 980 Great West Road
Brentford
Middlesex, TW8 9GS
UK

Regulatory Agency Identifier Number(s)

Registry	ID
Clinicaltrials.gov	NCT04128696
EudraCT	2019-002263-99
IND	143221

Effective Date: Refer to document date

Copyright 2021 the GlaxoSmithKline group of companies. All rights reserved.
Unauthorised copying or use of this information is prohibited.

SAP AUTHOR(S):

Author	Date
PPD Principal Statistician, Oncology Clinical Statistics	02-JUL-2021

SAP TEAM REVIEW CONFIRMATIONS (E-MAIL CONFIRMATION):

Approver	Date of E-mail Confirmation
PPD Physician Project Leader, Oncology Clinical Development	09-JUL-2021
PPD Clinical Investigation Leader, Oncology Clinical Development	08-JUL-2021
PPD Safety Development Leader, Pharmacovigilance Safety & Medical Governance	09-JUL-2021
PPD Lead Programmer, Oncology Clinical Programming	09-JUL-2021
PPD Health Outcomes PPD Oncology Value Evidence and Outcomes	08-JUL-2021
PPD Senior Medical Writer, Medical Writing Science	09-JUL-2021
PPD PPD Clinical Pharmacology Modelling & Simulation	12-JUL-2021

Clinical Statistics Line Approval (eSignature):

Approver
PPD
PPD Early IO Clinical Statistics Head

TABLE OF CONTENTS

PAGE

Version history

This Statistical Analysis Plan (SAP) for study 209229 is based on the protocol amendment 3 (GlaxoSmithKline Document Number TMF-13842230) dated 29-JUN-2021.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	12-NOV-2019	Not Applicable	Original version
2	26-JAN-2021	General updates	Updated to align with protocol amendment 2 and correction of minor typographical errors
2	26-JAN-2021	Added exploratory analysis of healthcare resource utilization added.	Updated to align with protocol amendment 2.
2	26-JAN-2021	Added max-combo analysis method.	To further detect and evaluate the treatment effect when the proportional hazards assumption is violated.
2	26-JAN-2021	Updated the date of treatment start to the date of randomization in the definition of best overall response per RECIST and iRECIST.	Corrected for clarification purpose.
2	26-JAN-2021	Added additional analyses due to the COVID-19 pandemic.	To examine the impact of the COVID-19 pandemic which began after the finalization of protocol amendment 1.
2	26-JAN-2021	Added the implementation details of the blinded independent central review (BICR) based on the outcome of the	To clarify the implementation strategy of the BICR.

SAP Version	Approval Date	Change	Rationale
		adaptive decision and the BICR auditing.	
2	26-JAN-2021	Updated infusion to cycle when summarizing the exposure data.	To align the terminology.
2	26-JAN-2021	Removed the expected timing of analyses.	It is sufficient to include the number of expected events as the study is event driven.
2	26-JAN-2021	Removed reference to overall summary score derivation in EORTC QLQ-C30.	Not applicable when all domains are not administered.
2	26-JAN-2021	Clarified that participants receiving a second course of treatment will be included in the analysis of PFS2.	Updated to align with protocol amendment 2.
2	26-JAN-2021	Added COVID-19 Analysis Set.	Additional subset of safety set for COVID-19 related analyses.
2	26-JAN-2021	Updated Enrolled Analysis Set definition.	For clarification of how to handle randomization and dosing errors.
2	26-JAN-2021	Cisplatin-resistant subgroup removed.	Not a relevant subgroup of interest. Given the population, the time to recurrence subgroup analysis is sufficient.
2	26-JAN-2021	Added clarification of requirements for second course treatment safety analyses.	Clarification required following protocol amendment 2.
2	26-JAN-2021	Clarifications added to the PFS primary and supplementary analysis scenarios.	Logic clarification required.

SAP Version	Approval Date	Change	Rationale
3	19-FEB-2021	Added an analysis set definition for response analysis at adaptive decision making	Added to clarify who will be included in the response analysis for interim analysis 1, which is aligned with response assessment schedule, allowing for a 1 week visit window.
3	19-FEB-2021	Clarified that the point estimate refers to the percentage of participants with an event in the tiered approach to safety analyses.	To clarify what point estimate should be derived.
3	19-FEB-2021	Added BOR and iBOR summaries. Added 95% exact confidence intervals for within-treatment estimates of ORR/DCR, and similarly for iORR/iDCR.	To aide data interpretation and for alignment with GSK standards.
3	19-FEB-2021	Restructured the ORR/DCR definition section and clarified that a one week visit window should be considered when defining durable SD to be considered in the derivation of DCR. Removed example SD derivation. Similar updates made for iSD and iDCR.	To align with visit schedule and clarify the difference between the derivation of SD as BOR vs durable SD included in the derivation of DCR (similarly for iSD and iDCR). Removed example to avoid misinterpretation.
3	19-FEB-2021	Clarified the weighting strategy for the stratified Miettinen and Nurminen's Method.	Numerous weighting strategies are available. Text expanded (already stating the use of sample size weighting) for clarification.
3	19-FEB-2021	General updates	Correction of minor typographical and formatting errors

SAP Version	Approval Date	Change	Rationale
4	Refer to document date	Intent-to-Treat analysis set changed to Modified Intent-to-Treat	To address change in study design following IA, where decision was made to discontinue ICOS/placebo. Participants randomized or first dosed after this date will be excluded from efficacy analyses.
4	Refer to document date	Removal of supplementary, sensitivity, BICR and subgroup analyses for the purposes of an abbreviated CSR. Some exploratory analyses may be performed and reported outside of the CSR. Minimally required safety to be included. For safety analyses, the between-treatment difference will not be analyzed.	Early study termination.
4	Refer to document date	Removed reference to reporting of retreated/second course treatment.	No patients received second course treatment prior to early termination, so analysis no longer required.
4	Refer to document date	Clarified process for determining the deterioration threshold for TTD in pain and physical function.	Due to early study termination, blinded data will be pooled across 209229 and 209227 studies and threshold analysis will be completed following database lock.
4	Refer to document date	General updates	Correction of minor typographical and formatting errors

SAP Version	Approval Date	Change	Rationale
4	Refer to document date	Targeted number of events will not be updated in the case of early study termination	Adds no value to revise following decision to terminate study

1. INTRODUCTION

The purpose of this SAP is to describe the planned analyses to be included in the Clinical Study Report for Study 209229. Details of the planned interim analysis, in addition to the final analyses, are provided. The main CSR will be an abbreviated report including all primary and secondary endpoints. The final CSR will report updated key safety analyses after LSLV.

Additional detail with regards to data handling conventions and the specification of data displays will be provided in the Output and Programming Specification (OPS) document.

1.1. Changes to the Protocol Defined Statistical Analysis Plan

Due to the early termination of the study, the SAP has been developed for the purposes of an abbreviated CSR for reporting of primary and secondary endpoints and for the final safety update.

Intent-to-Treat Analysis Set changed to a Modified Intent-to-Treat. Participants who were first dosed or randomized after the date of dear investigator letter (DIL) requesting immediate discontinuation of GSK3359609/placebo will be excluded. The date of the DIL is the 13th April.

For the abbreviated CSR, supplementary, sensitivity, BICR and subgroup analyses will not be performed. Some exploratory analyses may be performed and reported outside of the CSR. Minimally safety analyses will be performed. For safety analyses, the between-treatment difference will not be analyzed.

1.2. Objectives, Endpoints and Estimands

1.2.1. Objectives and Endpoints

Objectives	Endpoints
Primary Objectives	**Primary Endpoints**
• Compare the efficacy of GSK3359609 in combination with pembrolizumab to pembrolizumab plus placebo in the Programmed Death Ligand 1 (PD-L1) expression positive (CPS ≥1) population and in the PD-L1 expression high (CPS ≥20) population.	• OS in PD-L1 CPS ≥1 and CPS ≥20 populations defined as the time from the date of randomization to the date of death due to any cause. • PFS per RECIST v1.1 by investigator assessment in PD-L1 CPS ≥1 population, defined as the time from the date of randomization to the date of first documented disease progression or death due to any cause, whichever comes first
Secondary Objectives	**Secondary Endpoints**
• Further compare the efficacy of GSK3359609 in combination with pembrolizumab compared with pembrolizumab plus placebo	• PFS per iRECIST (iPFS) by investigator assessment in the PD-L1 CPS ≥1 population

Objectives	Endpoints
	• PFS per RECIST v1.1 and iPFS by investigator assessment in PD-L1 CPS ≥20 population • Milestone OS rate at 12 and 24 months in the PD-L1 CPS ≥1 and CPS ≥20 populations • ORR per RECIST v1.1 by investigator assessment in the PD-L1 CPS ≥1 and CPS≥20 populations • DCR per RECIST v1.1 by investigator assessment in the PD-L1 CPS ≥1 and CPS ≥20 populations • DoR per RECIST v1.1 by investigator assessment in the PD-L1 CPS ≥1 and CPS ≥20 populations
• Evaluate the safety and tolerability of GSK3359609 in combination with pembrolizumab compared with pembrolizumab plus placebo	• Frequency and severity of AEs, AESI, SAEs • Dose modifications (i.e., interruptions, discontinuations)
• Evaluate and compare disease related symptoms and impact on function and health-related quality of life (HRQoL) of GSK3359609/pembrolizumab versus pembrolizumab plus placebo	• The time to deterioration in pain measured by the EORTC QLQ-H&N35 pain domain in the PD-L1 CPS ≥1 and CPS ≥20 populations • The time to deterioration in physical function measured by the PROMIS PF 8c in the PD-L1 CPS ≥1 and CPS ≥20 populations
Exploratory Objectives	**Exploratory Endpoints**
• Compare the efficacy of GSK3359609 in combination with pembrolizumab to pembrolizumab plus placebo	• ORR, DoR, DCR per iRECIST • PFS2, defined as the time from the date of randomization to the date of second objective disease progression per RECIST v1.1, or death due to any cause, whichever first
• Evaluate and compare disease-related symptoms, overall bother of treatment side effects, and impact on function and HRQoL of GSK3359609/pembrolizumab versus pembrolizumab plus placebo	• Symptomatic AEs as measured by the FACT GP5 • Changes in other domains of quality of life as measured by the selected EORTC IL50/51 (subset of domains of the EORTC QLQ-C30 and EORTC QLQ-H&N35), BPI-I3 and EQ-5D-3L
• Evaluate healthcare resource utilization of participants in the GSK3359609 combination with pembrolizumab arm versus participants	• Non-protocol healthcare encounters, such as provider visits, emergency room visits, hospitalizations, medications, tests, or procedures

Objectives	Endpoints
in the placebo combination with pembrolizumab arm	
• Evaluate GSK3359609 PK properties	• Summary of GSK3359609 concentrations and Cmax, Cmin, AUC (0- τ) as data permit
• Determine immunogenicity of GSK3359609	• Anti-drug antibody (ADA) incidence
• Explore relationship between biomarkers in tumor and blood, such as immune response biomarkers, target expression and efficacy endpoints	• Tumor and blood-based analysis of DNA, RNA, and protein analytes/profiles; OS, PFS, ORR, other efficacy parameters

• Genetics Research: Investigate the relationship between host genetic variations and response to therapy	• Germline genetic evaluations may be conducted for:
	  • Clinical response, including GSK3359609/ pembrolizumab or any concomitant medicines
	  • Disease susceptibility, severity, and progression and related conditions

Abbreviations: AE=adverse events; AESI=adverse events of special interest; Brief Pain Inventory-Item 3=BPI-I3; DCR=disease control rate; DNA=deoxyribonucleic acid; DoR=duration of response; EORTC QLQ-C30=European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire 30-item Core Module; EORTC H&N35=EORTC Head and Neck 35 Item Module; EQ-5D-3L=EuroQoL 5 Dimensions; FACT-GP5 = Functional Assessment of Cancer Therapy – General (Item GP5); HRQoL= health-related quality of life; iPFS = immune-based progression-free survival; iRECIST=immune-based Response Evaluation Criteria in Solid Tumors; ORR=overall response rate; OS=overall survival; PFS=progression-free survival; PROMIS-PF-8c= Patient-Reported Outcomes Measurement Information System-Physical Function-Short Form; RECIST= Response Evaluation Criteria in Solid Tumors; RNA=ribonucleic acid

1.2.2. Estimands

Primary and key secondary study objectives are presented in Table 2 with additional information, including prespecified estimands with related attributes.

Due to early study termination, only the primary estimands will be reported in the main CSR.

Table 2 Estimands

Objective (Hypothesis[1])	Estimand Category	Estimand			
		Variable/ Endpoint[2]	Population of interest (Analysis Set)	Intercurrent Event Strategy[3]	Population Level Summary Measure
Primary Objective 1: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab compared to pembrolizumab plus placebo in OS in PD-L1 CPS ≥1 and CPS ≥20 R/M HNSCC (H1, H2)	Primary	OS, defined as the interval of time from the date of randomization to the date of death due to any cause	CPS ≥1 R/M HNSCC, CPS ≥20 R/M HNSCC (mITT)	• New anti-cancer therapy: treatment policy • Treatment discontinuation: treatment policy	Hazard ratio for GSK3359609+pembrolizumab vs. placebo+pembrolizumab
Primary Objective 2: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab compared to pembrolizumab plus placebo in PFS in PD-L1 CPS ≥1 R/M HNSCC (H3)	Primary	PFS, defined as the time from the date of randomization to the date of the first objectively documented disease progression per RECIST v1.1 based on investigator assessment, or death due to any cause, whichever occurs first.	CPS ≥1 R/M HNSCC (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • ≥2 missed disease assessments: hypothetical • Death: composite	Hazard ratio for GSK3359609+pembrolizumab vs. placebo+pembrolizumab

Objective (Hypothesis[1])	Estimand Category	Estimand			
		Variable/ Endpoint[2]	Population of interest (Analysis Set)	Intercurrent Event Strategy[3]	Population Level Summary Measure
Key Secondary Objective 1: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab compared to pembrolizumab plus placebo in iPFS in PD-L1 CPS ≥1 R/M HNSCC (H4)	Primary	iPFS	CPS ≥1 R/M HNSCC (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • ≥2 missed disease assessments: hypothetical • Death: composite	Hazard ratio for GSK3359609+pembrolizumab vs. placebo+pembrolizumab
Key Secondary Objective 2: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab compared to pembrolizumab plus placebo in TTD in pain in PD-L1 CPS ≥1 and CPS ≥20 R/M HNSCC (H5, H6)	Primary	TTD in pain	CPS ≥1 R/M HNSCC, CPS ≥20 R/M HNSCC (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • Disease progression per RECIST v1.1 or iRECIST: treatment policy • Death: treatment policy • ≥2 missed corresponding PRO assessments if at least one missed assessment is due to the participant being too ill: hypothetical	Hazard ratio for GSK3359609+pembrolizumab vs. placebo+pembrolizumab
Key Secondary Objective 3: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab compared to pembrolizumab plus placebo in TTD in PF in PD-L1 CPS ≥1 and CPS ≥20 R/M HNSCC	Primary	TTD in PF	CPS ≥1 R/M HNSCC, CPS ≥20 R/M HNSCC (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • Disease progression per RECIST v1.1 or iRECIST: treatment policy • Death: treatment policy	Hazard ratio for GSK3359609+pembrolizumab vs. placebo+pembrolizumab

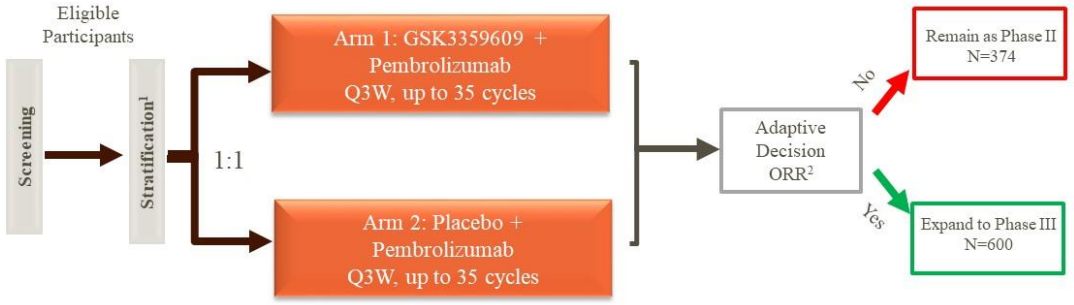
Objective (Hypothesis[1])	Estimand Category	Estimand			
		Variable/ Endpoint[2]	Population of interest (Analysis Set)	Intercurrent Event Strategy[3]	Population Level Summary Measure
(H7, H8)				• ≥2 missed corresponding PRO assessments if at least one missed assessment is due to the participant being too ill: hypothetical	

Abbreviations: iPFS = immune-based progression-free survival; iRECIST=immune-based Response Evaluation Criteria in Solid Tumors; mITT =Modified Intent-to-Treat; OS = overall survival; PF = physical function; PFS = progression free survival; PRO = Patient Reported Outcomes; RECIST= Response Evaluation Criteria in Solid Tumors; TTD = time to deterioration.

1. Refer to Section 2 for details in the hypotheses.
2. Refer to Section 5 for the definition of variable/endpoint.
3. Refer to Section 5.2.1 and Section 6.3 for details in the two or more missed disease assessments.

1.3. Study Design

Overview of Study Design and Key Features



Abbreviations: CPS=combined positive score; GSK609=GSK3359609; HPV=human papilloma virus; ORR=overall response rate; PD-L1=programmed death ligand-1; Q3W=every three weeks; vs=versus

1. Stratification Factors:
   a. PD-L1 Status (CPS ≥20 vs. 1≤ CPS <20)
   b. HPV status in oropharynx sites only (positive vs. negative/unknown) vs non-oropharynx)
2. ORR assessed by RECIST v1.1; refer to Section 9.5 of protocol for further details on adaptive decision analysis.

Design Features	• Design: This is a randomized, double-blind, multicentre, parallel, adaptive, phase II/III study comparing a combination of GSK3359609 (ICOS agonist) and pembrolizumab to pembrolizumab plus placebo in participants with PD-L1 CPS ≥1 recurrent/metastatic (R/M) head and neck squamous cell carcinoma (HNSCC) of the oropharynx, oral cavity, hypopharynx or larynx. • If at the interim adaptive analysis, the outcome meets the defined ORR positive criterion, the study will be expanded from a Phase II to a Phase III design with a total of 600 participants randomized. If the outcome does not meet the defined criteria, then the study will remain as Phase II with 374 participants randomized. • All participants will be stratified by two factors i) PD-L1 CPS status (CPS ≥20 vs 1≤ CPS <20); and ii) HPV status in oropharyngeal cancers (positive vs negative/unknown) vs non-oropharyngeal cancers then randomly assigned in a 1:1 ratio to the GSK3359609/pembrolizumab arm or pembrolizumab with placebo arm. • The study is comprised of three periods: screening, treatment, and follow-up. The total duration of study participation begins with the signing of the informed consent form (ICF) through the final protocol-defined follow-up assessment for survival. ○ For participants who meet all eligibility criteria and are randomized within the study, the maximum duration of treatment with pembrolizumab, GSK3359609 and placebo is expected to be approximately 2 years, up to 35 cycles. ○ The follow-up period begins when study treatment is permanently discontinued; participants will undergo follow-up assessments for safety, PFS on first subsequent anti-cancer therapy (PFS2) and survival as indicated in the Schedule of Activities (SoA) in the protocol.
Study intervention	• GSK3359609, placebo, and pembrolizumab are each administered as a 30-minute IV infusion once Q3W. • GSK3359609 or placebo will be administered first followed by pembrolizumab.
Study intervention Assignment	• Participants will be randomly assigned to either receive the combination of 24 mg GSK3359609 plus 200 mg pembrolizumab (GSK3359609/pembrolizumab) or placebo plus 200 mg pembrolizumab (placebo/pembrolizumab) in a 1:1 allocation ratio after stratification.
Interim Analysis	• The study is designed to have one interim analysis (IA) for adaptive decision making, one PFS/iPFS primary analysis (final analysis) and two OS analyses (one

Overview of Study Design and Key Features	
	IA and one final analysis) in PD-L1 CPS ≥1 participants. Two OS analyses will be performed in PD-L1 CPS ≥20 participants. ○ IA1: The interim analysis for adaptive decision making will be conducted using ORR/DCR per RECIST v1.1 based on investigator assessment when approximately the first 100 PD-L1 CPS ≥1 participants have a minimum follow-up of 6 months. ○ IA2: At the time of PFS/iPFS primary analysis, the OS interim analysis will be conducted in PD-L1 CPS ≥1 participants to allow for early stopping of the study due to efficacy or allow for non-binding futility analysis. The timing of PFS/iPFS analysis and the OS interim analysis is triggered by the pre-specified number of PFS events in PD-L1 CPS ≥1 population. ○ FA: The timing of the final OS analysis in the PD-L1 CPS ≥1 participants and CPS ≥20 participants will be triggered by the pre-specified number of OS events in the PD-L1 CPS ≥1 population. • The number of events for the PFS and OS analyses are specified in the protocol and is dependent on whether the study is expanded to Phase III or remains as Phase II. • The study will use an IDMC. The safety of the study intervention will be monitored.
Multiplicity	• The type I error rate is controlled at 2.5% with an initial assignment of 0.01% for PFS and 2.49% for OS. • Alpha will be re-allocated as detailed in the multiplicity testing strategy in Section 2.1.

2. STATISTICAL HYPOTHESES / SUCCESS CRITERIA

The following primary hypotheses will be tested:

Overall Survival (OS)

- Hypothesis (H1): GSK3359609 in combination with pembrolizumab prolongs OS compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥1 R/M HNSCC.

- Hypothesis (H2): GSK3359609 in combination with pembrolizumab prolongs OS compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥20 R/M HNSCC.

Progression-free Survival (PFS)

- Hypothesis (H3): GSK3359609 in combination with pembrolizumab prolongs PFS by investigator assessment compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥1 R/M HNSCC.

The following key secondary hypotheses will be tested:

Immune-based progression-free Survival (iPFS)

- Hypothesis (H4): GSK3359609 in combination with pembrolizumab prolongs iPFS by investigator assessment compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥1 R/M HNSCC.

Time to Deterioration (TTD) in Pain

- Hypothesis (H5): GSK3359609 in combination with pembrolizumab prolongs TTD in Pain (measured by EORTC IL51) compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥1 R/M HNSCC.

- Hypothesis (H6): GSK3359609 in combination with pembrolizumab prolongs TTD in Pain (measured by EORTC IL51) compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥20 R/M HNSCC.

TTD in Physical Functioning

- Hypothesis (H7): GSK3359609 in combination with pembrolizumab prolongs TTD in Physical Functioning (measured by PROMIS PF 8c) compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥1 R/M HNSCC.

- Hypothesis (H8): GSK3359609 in combination with pembrolizumab prolongs TTD in Physical Functioning (measured by PROMIS PF 8c) compared with pembrolizumab/placebo in participants with PD-L1 CPS ≥20 R/M HNSCC.

The study is considered to have met the study primary objective if GSK3359609 in combination with pembrolizumab is superior to pembrolizumab with placebo on either PFS or OS in PD-L1 CPS ≥1 participants.

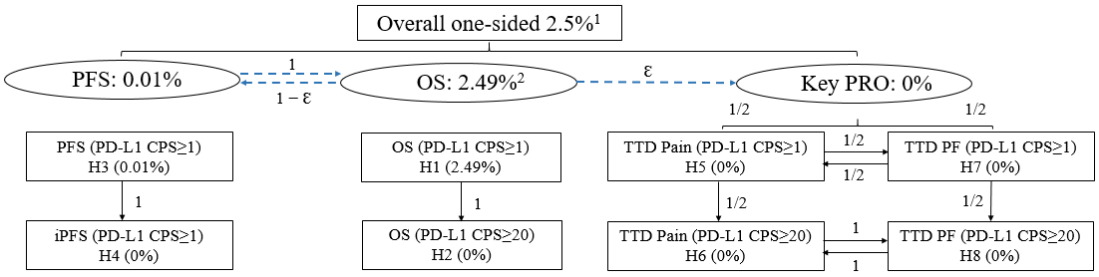
2.1. Multiple Comparisons and Multiplicity

The primary comparisons of interest are the comparisons between GSK3359609/pembrolizumab and placebo/pembrolizumab in PFS/iPFS in the PD-L1 CPS ≥1 participants, as well as in OS, TTD in Pain, and TTD in PF in the PD-L1 CPS ≥1 population and the PD-L1 CPS ≥20 population.

The study employs the graphical method [Maurer, 2013] to provide strong multiplicity control for multiple hypotheses as well as interim analyses.

The graphical approach is based on closed testing procedures using weighted group sequential Bonferroni tests for the intersection hypotheses. Under the mild monotonicity condition on the error spending functions which is met in this study, it allows the use of sequentially rejective graphical procedures in group sequential trials and controls the family wise error rate in the strong sense. By defining weighted directed graphs, it defines a weighting strategy between and within subfamilies which consist of a subset of null hypotheses each.

Figure 1 Multiplicity Testing Strategy for Comparisons Between GSK3359609 in Combination with Pembrolizumab and Pembrolizumab Administered with Placebo



Abbreviations: CPS = combined positive score; H = hypothesis; I = GSK3359609; iPFS = progression-free survival per iRECIST; OS = overall survival; P = pembrolizumab; PF=physical function; PFS = progression-free survival per RECIST v1.1; PD-L1 = programmed cell death receptor 1-ligand 1; TTD=time to deterioration

1. The alpha level assigned to a subfamily will be rolled over only if the hypotheses within the subfamily are all significant based on the weight for re-allocation presented on the dashed lines connecting subfamilies. Within each subfamily, the weights for reallocation from each hypothesis to the others are represented on the solid lines connecting hypotheses.
2. ε = 5/6 if both hypotheses in the OS subfamily (i.e. PDL1 CPS≥1 and PD-L1 CPS≥20 participants) are significant at the time of PFS/iPFS analyses; if the OS hypothesis is not significant in either population at the time of PFS/iPFS analyses, then ε = 1 at the final OS analysis.

The family-wise type I error for this study is strongly controlled at 2.5% (one-sided). Figure 1 shows the initial one-sided alpha-allocation for PFS/iPFS, OS, and key PRO endpoints. The multiplicity control strategy applies whether the study remains as a Phase II trial or expands to a Phase III trial.

A one-sided alpha was chosen as, in this placebo-controlled study, pembrolizumab is given in both treatment arms. It is not expected that the combination of GSK3359609 and pembrolizumab would lead to a detrimental effect due to the active pembrolizumab on both arms and the favourable toxicity profile to date from the addition of GSK3359609 to the SoC. Although the one-sided testing is proposed, as described in Section 9.4.1.1 of the protocol, the 95% two-sided confidence intervals of the treatment effects (hazard ratio for time-to-event endpoints) will be reported, which would provide conclusions on the negative results if there is any statistically significant detrimental effect. Section 6.2 presents the algorithm and the associated application of the graphical testing strategy illustrated in Figure 1. The R package gMCP [Rohmeyer, 2011] is available to provide functions and graphical user interfaces for graph based multiple test procedures.

The alpha re-allocation for PFS/iPFS, OS, and key PRO endpoints is further explained below.

Overall Survival

- An initial alpha level of 2.49% is allocated to the OS hypothesis in PD-L1 CPS ≥1 participants.

- When PFS and iPFS test are both significant in PD-L1 CPS ≥1 participants, the OS hypothesis in PD-L1 CPS≥1 participants will be tested at 2.5% (re-allocated alpha).
- OS in PD-L1 CP ≥20 participants is tested sequentially at the same overall alpha level if GSK3359609 in combination with pembrolizumab demonstrates superiority to pembrolizumab plus placebo in OS in PD-L1 CPS ≥1 participants. If OS is not significant in PD-L1 CPS ≥1 participants, the formal conclusion of statistical significance on OS in PD-L1 CPS ≥20 participants will not be drawn.

The following alpha- and beta-spending functions are used in OS hypothesis of PD-L1 CPS ≥1 participants and PD-L1 CPS ≥20 participants.

- For the OS hypothesis, the alpha-spending function based on Lan-DeMets O'Brien-Fleming approximation spending function [Lan, 1983] and the beta-spending function based on the Hwang-Shih-DeCani boundary method [Hwang, 1990] with a gamma parameter of -20 are constructed to implement group sequential boundaries that control the type I error rate as well as allow for non-binding futility analysis. Refer to Section 5.7.2 for details in the information fractions.

There is a total of 2 OS analyses in each of the PD-L1 CPS ≥1 and CPS ≥20 participants. Section 5.7.2 demonstrates the bounds and boundary properties for OS hypothesis testing. Efficacy boundaries and non-binding futility boundaries are based on initially assigned type I error rate before any alpha re-allocation and projected number of events at study milestones. The actual boundaries will be determined from the actual number of events at the time of the specified interim analysis using the alpha- and beta- spending functions. Actual futility bounds will be updated if overall beta is changed with respect to alpha roll-over.

Progression-free Survival/Immune-based progression-free Survival

- An initial alpha level of 0.01% is allocated to the PFS hypothesis in PD-L1 CPS ≥1 participants.
- When OS hypotheses in both populations (PD-L1 CPS≥1 participants and PD-L1 CPS ≥20 participants) are significant at the time of the OS interim analysis, the PFS hypothesis in PD-L1 CPS ≥1 participants may be tested at 0.425% (re-allocated alpha).
- If the PFS hypothesis in PD-L1 CPS ≥1 is significant, the hypothesis of iPFS in PD-L1 CPS ≥1 will be tested sequentially at the same alpha level.

Key Patient Reported Outcomes (TTD in Pain, TTD in PF)

- Only if GSK3359609 in combination with pembrolizumab demonstrates superiority to pembrolizumab plus placebo in OS in both populations (PD-L1 CPS ≥1 participants and PD-L1 CPS ≥20 participants), will key secondary PRO endpoints be tested. The alpha level from OS hypothesis will be propagated to key PRO hypotheses.

- If OS meets superiority in both populations at the time of the OS interim analysis, a total of 2.075% alpha level will be propagated to key PRO hypotheses.
- If OS meets superiority in both populations at the time of the final OS analysis but PFS or iPFS fails to demonstrate superiority in PD-L1 CPS ≥1 participants, a total of 2.49% alpha level will be propagated to key PRO hypotheses.
- If OS meets superiority in both populations at the time of the final OS analysis and PFS and iPFS both demonstrate superiority in PD-L1 CPS ≥1 participants, a total of 2.5% alpha level will be propagated to key PRO hypotheses.
- The alpha level propagated from OS will be equally split between TTD in Pain and TTD in PF in PD-L1 CPS ≥1 participants, with the possibility to further propagate the levels between each other.
- If one of two key PRO hypotheses in PD-L1 CPS ≥1 participants is rejected based on the re-allocated alpha level, the key PRO hypotheses in PD-L1 CPS ≥20 participants can be tested based on the (updated) weight, with the possibility to further propagate the levels between each other.
- Based on the re-allocated cumulative alpha level, the nominal significance level for each key PRO endpoint will be calculated based on the Lan-DeMets O'Brien-Fleming approximation alpha-spending function.

3. SAMPLE SIZE DETERMINATION

The 2-in-1 adaptive Phase II/III study design [Chen, 2018] allows expanding the Phase II study seamlessly into a Phase III confirmatory study; refer to Section 4 of the protocol for details of the study design.

The Phase II study will randomize approximately 374 participants with a 1:1 ratio between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm. The study is event-driven, and the sample size calculation is driven by overall survival events.

Assuming the prevalence rate of PD-L1 CPS ≥20 among the PD-L1 CPS ≥1 population is 53%, a sample size of 374 PD-L1 CPS ≥1 participants will provide approximately 198 PD-L1 CPS ≥20 participants.

A long-term survival benefit, observed as a long-lasting plateau towards the tail of the survival curve, and a delayed treatment effect, observed as a late separation in survival curves between the experimental and control arms; have been reported in randomized clinical trials among participants treated with immuno-oncology drugs.

Given the OS data reported in the KN-048 clinical trial, a long-term survival benefit in the pembrolizumab/placebo arm and potential non-proportional hazards with a delayed treatment effect are anticipated. Refer to protocol Section 9.2 for details, including power calculation and sample size assumptions.

Remain as a Phase II Study

The final OS analysis in the PD-L1 CPS ≥1 and CPS ≥20 participants for a Phase II study (i.e. without expansion) will be carried out after approximately 244 deaths in PD-L1 CPS ≥1 participants have occurred between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm, barring early stopping for futility or efficacy. It is expected that approximately 113 deaths in PD-L1 CPS ≥20 participants have occurred between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm.

With the above numbers of events and before any alpha re-allocation, the study provides a 79.8% power in the PD-L1 CPS ≥1 participants and a 60% power in the PD-L1 CPS ≥20 participants to demonstrate superiority of OS of GSK3359609 in combination with pembrolizumab relative to pembrolizumab plus placebo at the pre-specified initial alpha (one-sided) level of 2.49%.

The estimated numbers of PFS/iPFS events for a Phase II study at the final PFS evaluation are estimated to be 281 in the PD-L1 CPS ≥1 participants.

The estimated number of PFS/iPFS events in the PD-L1 CPS ≥1 participants provides a 67.6% power for detecting a HR of 0.61 in PFS/iPFS in the PD-L1 CPS ≥1 participants at the alpha level of 0.01% (one-sided).

Expand to a Phase III Study

Should expanding the Phase II study into a Phase III be decided, a total of 600 participants will be randomized in a 1:1 ratio between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm, including those participants already enrolled. Participants already used for the decision making in the ongoing study will be included in Phase III analyses. The Phase III study will have overall sample size of 600 participants. Assuming the aforementioned prevalence rate of PD-L1 CPS status, a sample size of 600 participants with PD-L1 CPS ≥1 status will provide 318 PD-L1 CPS ≥20 participants.

The final OS analysis in the PD-L1 CPS ≥1 and CPS≥20 participants for a Phase III study (i.e. with expansion) will be carried out after approximately 367 deaths in the PD-L1 CPS ≥1 participants have occurred between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm, barring early stopping for futility or efficacy. It is expected that approximately 171 deaths in PD-L1 CPS ≥20 participants have occurred between the GSK3359609 in combination with pembrolizumab arm and the pembrolizumab/placebo arm.

With the above numbers of OS events and before any alpha re-allocation, the study provides a 90.1% power in the PD-L1 CPS ≥1 participants, and a 73.5% power in the PD-L1 CPS ≥20 participants to demonstrate superiority of OS of GSK3359609 in combination with pembrolizumab relative to pembrolizumab plus placebo at the pre-specified initial alpha (one-sided) level of 2.49%.

The estimated numbers of PFS/iPFS events for a Phase III study (i.e. with expansion) at the final PFS evaluation are estimated to be 432 in the PD-L1 CPS ≥1 participants.

The estimated number of PFS/iPFS events in the PD-L1 CPS ≥1 participants provides a 92.5% power for detecting a HR of 0.61 in PFS/iPFS in the PD-L1 CPS ≥1 participants at the alpha level of 0.01% (one-sided).

3.1. Sample Size Sensitivity

In the case that there is no violation of the proportional hazards in OS between GSK3359609 in combination with pembrolizumab and the pembrolizumab/placebo arm throughout the entire course of the study; the powers in OS for PD-L1 CPS ≥1 and PD-L1 CPS ≥20 participants are summarized in protocol Section 9.2.3.

4. ANALYSIS SETS

Analysis Set	Definition / Criteria	Analyses Evaluated
Screened	• All participants who sign the ICF.	• Study population
Enrolled	• All participants who entered the study. • Participants who were randomized or dosed in error are included in the enrolled population. • Note that screening failures (who never passed screening even if rescreened) and participants screened but never enrolled into the study (Reserve, Not Used) are excluded from the Enrolled analysis set as they did not enter the study. • This population will be based on the study intervention the participant was randomized to.	• Study population
Modified Intent-To-Treat (mITT)	• All randomized participants whether or not randomized intervention was administered, excluding those who were first dosed or randomized after the date of dear investigator letter (DIL) requesting immediate discontinuation of GSK3359609 /placebo. The date of the DIL is the 13th April. • This analysis set will be based on the study intervention to which the participant was randomized and will be the primary analysis set for the analysis of efficacy data.	• Study population • Efficacy

Analysis Set	Definition / Criteria	Analyses Evaluated
	• Any participants who receives a study intervention randomization number will be considered to have been randomized.	
Safety	• All randomized participants who take at least 1 dose of study intervention. • Participants will be analyzed according to the actual study intervention received. • Participants will be assigned to the actual study intervention group of GSK3359609 + Pembrolizumab if the participant received any dose of GSK3359609.	• Safety • Immunogenicity
COVID-19	• All participants in the Safety set who had a confirmed, probable or suspected COVID-19 case diagnosis.	• Baseline Characteristics, COVID-19 Assessments, Treatment Compliance and Exposure, Medical History and Laboratory Data
IA1	• All randomized participants whether or not randomized intervention was administered, who have a minimum of 22 weeks follow-up from date of randomization to the time of data cut off, regardless of death or study withdrawal. • This analysis set will be based on the study intervention to which the participant was randomized to.	• IA1 efficacy

4.1. Protocol Deviations

Important protocol deviations (including deviations related to study inclusion/exclusion criteria, conduct of the trial, participant management or participant assessment) will be summarized for the mITT analysis set.

Protocol deviations will be tracked by the study team throughout the conduct of the study. These protocol deviations will be reviewed to identify those considered as important as follows:

- Data will be reviewed prior to freezing the database to ensure all important deviations are captured and categorized in the protocol deviations SDTM dataset.

- This dataset will be the basis for the summaries of important protocol deviations.

A separate listing of all inclusion/exclusion criteria deviations will also be provided. This listing will be based on data as recorded on the inclusion/exclusion page of the eCRF.

5. STATISTICAL ANALYSES

5.1. General Considerations

5.1.1. General Methodology

The modified Intent-to-Treat (mITT) analysis set will be used for all study population analyses, efficacy analyses and PRO analyses, unless otherwise specified. The safety analysis set will be used for all safety analyses, unless otherwise specified.

Stratified statistical analyses (the stratified logrank test, the stratified Cox model and the stratified max-combo test) will be based on the following stratification factors, PD-L1 expression (CPS ≥20 vs. 1≤ CPS <20), and HPV status (positive vs. negative). Participants with oropharynx HPV negative/unknown and non-oropharyngeal tumors will be combined as the HPV negative group in the stratified analyses. The analyses will be performed based on the data collected in Interactive Response Technology (IRT) at randomization will be used, even if it is subsequently discovered that these values were incorrect.

In the case of a substantial amount of wrong stratification assigned at the time of randomization, a sensitivity analysis may be performed based on the data collected in the CRF (or vendor data if collected outside of eCRF).

Confidence intervals will use 95% confidence levels unless otherwise specified.

Unless otherwise specified, continuous data will be summarized using descriptive statistics: n, mean, standard deviation (std), median, minimum and maximum. Categorical data will be summarized as the number and percentage of participants in each category.

Unless otherwise specified, subsequent anticancer therapy will include systemic anticancer therapy, follow-up radiotherapy that is not palliative, or on treatment or follow-up cancer-related surgery/procedures that are not palliative or diagnostic in nature.

5.1.2. Baseline Definition

For all endpoints unless otherwise specified the baseline value will be the latest pre-dose assessment with a non-missing value, including those from unscheduled visits. If time is not collected, Day 1 assessments are assumed to be taken prior to first dose and used as baseline. For participants who did not receive study treatment during the study, baseline will be defined as the latest, non-missing collected value.

Unless otherwise stated, if baseline data is missing no derivation will be performed and baseline will be set to missing.

5.1.3. Multicenter Studies

In this multicenter global study, enrollment will be presented by country and site.

Data from all participating centers will be integrated and no controlling for center-effect will be considered in the statistical analyses. It is anticipated that participant accrual will be spread thinly across centers and summaries of data by center would unlikely be informative and will not be provided.

5.2. Primary Endpoint(s) Analysis

5.2.1. Definition of Endpoint(s)

Overall Survival (OS)

Overall Survival (OS) is defined as the interval of time from the date of randomization to the date of death due to any cause.

Participants without documented death will be censored at last known alive date. The last date will be determined by the maximum collection/assessment date from among selected data domains within the clinical database; details will be provided in a separate Output and Programming Specification (OPS) document. When calculating overall survival, all deaths including those following subsequent anti-cancer therapy will be included.

Progression-Free-Survival per RECIST v1.1

Progression-free-survival (PFS) per RECIST v1.1 by investigator assessment is defined as the time from the date of randomization to the date of the first objectively documented disease progression per RECIST v1.1 based on investigator assessment, or death due to any cause, whichever occurs first.

The date of disease progression is defined as the date of radiological disease progression based on imaging data per RECIST v1.1. For cases where symptomatic progression is documented by the investigator, the derived response based on tumor assessment data will be utilized.

For participants who receive subsequent anticancer therapy, the following additional rules will apply:

- If a participant has only a baseline visit or does not have an adequate post-baseline radiological assessment on or prior to the date of initiation of anticancer therapy, PFS will be censored at the date of randomization.

- If anticancer therapy is started without documented disease progression or is started prior to documented disease progression, then PFS will be censored at the date of the last adequate radiological assessment on or prior to the initiation of anticancer therapy (i.e., if an assessment occurs on the same day as the start of new anticancer therapy, the assessment will be used as it is assumed that the assessment occurred prior to the administration of new anticancer therapy). The date of the last adequate radiological assessment will be used as the censoring date.

- If the start date of anticancer therapy is partial, the imputation rules described in the OPS will be applied.

Since PFS is interval censored, extended loss to follow-up prior to PD or death increases the uncertainty when the event occurs. As such, PFS will be analyzed censoring for extended time without an adequate assessment to account for missed response assessments prior to disease progression or death. Specifically, if there are two or more assessments which are missing followed by an assessment of PD or death, PFS will be censored at the last adequate assessment prior to PD or death.

A summary of the assignments for progression and censoring dates for the primary analysis of PFS per RECIST v1.1 is illustrated in Table 3.

Table 3 Censoring Rules for Primary Analysis of PFS per RECIST v1.1

Situation	Primary Analysis
No or incomplete baseline disease assessments and the participant has not died	Censored at the date of randomization
No adequate[1] post-baseline disease assessments (prior to anticancer therapy, if initiated) and the participant has not died[2]	Censored at the date of randomization
With adequate[1] post-baseline disease assessments, new anticancer treatment is not initiated, and no documented PD or death	Censored at the date of last adequate[1] radiological disease assessment
With adequate[1] post-baseline disease assessments before start of new anticancer therapy, and new anticancer treatment is initiated (prior to documented PD or death)[3]	Censored at the date of last adequate[1] radiological disease assessment on or prior to starting new anticancer treatment
PD or death documented after ≤1 missed disease assessment[4]	Progressed at the date of documented PD[5] or death, whichever occurs first
PD or death documented after ≥2 missed disease assessments[4,6]	Censored at the date of last adequate[1] radiological disease assessment prior to the ≥2 missed disease assessment[7]

Abbreviations: CR=complete response; PD=progressive disease; PFS=progression-free survival; PR=partial response; RECIST=response evaluation criteria in solid tumors; SD=stable disease

1. An adequate assessment is defined as an assessment where the investigator assessed response is CR, PR, or SD
2. If participant has documented PD but no other adequate assessments, see scenarios below.
3. If PD and new anti-cancer therapy occur on the same day, it is assumed that the progression was documented first (i.e., outcome is progression; the date is the date of the assessment for progression).
4. The case where PD or death documented after the initiation of new anticancer treatment is described above and is not included in this situation.
5. The earliest of (i) Date of radiological assessment showing new lesion (if progression is based on new lesion); or (ii) Date of radiological assessment showing unequivocal progression in non-target lesions, or (iii) Date of last radiological assessment of measured lesions (if progression is based on increase in sum of measured lesions).
6. Refer to Section 6.3 for details in extended time without an adequate assessment.
7. The date of randomization will be used if there are no adequate post-baseline assessments.

5.2.2. Main Analytical Approach

Due to early study termination, only the main analytical approaches using primary estimands will be completed for the main CSR.

Overall Survival (OS)

The non-parametric Kaplan-Meier method will be used to estimate the survival curves for OS. Kaplan-Meier plots of OS will be presented by study intervention. Kaplan-Meier estimates for the median overall survival and the first and third quartiles will be presented, along with 95% CIs. CIs for quartiles will be estimated using Brookmeyer-Crowley method (Brookmeyer, 1982). The treatment difference in survival will be assessed by the stratified log-rank test.

A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio). The hazard ratio (HR) and its corresponding 95% confidence interval from the stratified Cox

model with a single treatment covariate will be reported separately for the PD-L1 CPS ≥1 and CPS ≥20 group.

OS in PD-L1 CPS ≥20 participants will be tested sequentially at the 2.49% alpha level if GSK3359609 in combination with pembrolizumab demonstrates superiority to pembrolizumab with placebo in OS in PD-L1 CPS ≥1 participants. If OS is not significant in PD-L1 CPS ≥1 participants, the formal conclusion of statistical significance on OS in PD-L1 CPS ≥20 participants will not be drawn.

5.2.2.1. Statistical Methodology Specification

Endpoint / Variables
• OS
Model Specification
• OS will be estimated using Kaplan-Meier analysis for each study intervention (PROC LIFETEST). The median, 25th and 75th percentiles of OS will be estimated and corresponding 95% Confidence intervals will be estimated using the Brookmeyer-Crowley method (Brookmeyer, 1982). • Comparison of distributions of OS between study interventions will be based on the stratified log-rank test (PROC LIFETEST). • A stratified Cox proportional hazard model with Efron's method of tie handling and study intervention as the sole explanatory variable will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio) in OS between study interventions (PROC PHREG).
Model Checking & Diagnostics
• The proportional hazards assumption will be assessed prior to model fitting using the following methods: o Kaplan-Meier plot by study intervention: A non-parallel pattern is an indication of violation of the proportional hazard assumption. o Plot of log(-log(survival)) versus log(time) by study intervention: A non-parallel pattern is an indication of violation of the proportional hazard assumption. o Plot of Schoenfeld residuals versus time: A non-zero slope is an indication of a violation of the proportional hazard assumption. o Evaluation of time-dependency of treatment effect by adding an interaction term of treatment and time in the Cox model. If the interaction term is significant ($p< 0.05$), it is considered that the proportional hazards assumption is violated. If one or more of the procedures above demonstrates clear violation of the proportional hazards assumption in OS, it is considered the proportional hazards assumption does not hold. Hazard ratio and corresponding 95% CI estimated from the Cox model will still be reported.
Model Results Presentation
• Kaplan-Meier estimates for the median overall survival and the first and third quartiles will be presented, along with 95% CIs. • The p-value from the stratified log-rank test will be reported. • The hazard ratio and the corresponding 95% confidence interval from the Cox model will be reported.

Subgroup Analyses
• The stratified log-rank test and Cox proportional hazard model proportional hazard model analysis will be repeated in the subgroup analyses defined in Section 5.2.5 if data permit.

Progression-Free Survival per RECIST v1.1 (PFS)

A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio). The hazard ratio and its 95% confidence interval from the stratified Cox model with a single treatment covariate will be reported for PD-L1 CPS ≥1 and PD-L1 CPS ≥20 participants. Kaplan-Meier plots of PFS will be presented by study intervention. Summaries of number and percentage of participants experiencing a PFS event and the type of event (PFS per RECIST v1.1 or death) will be provided along with median PFS and the first and third quartiles and 95% CIs for each treatment.

5.2.2.2. Statistical Methodology Specification

Endpoint / Variables
• PFS per RECIST v1.1 based on investigator assessment
Model Specification
• PFS will be analyzed across study interventions using Kaplan-Meier analysis (PROC LIFETEST). The median, 25th and 75th percentiles of PFS will be estimated and corresponding 95% Confidence intervals will be estimated using the Brookmeyer-Crowley method (Brookmeyer,1982). • Comparison of distributions of PFS between study interventions will be based on the stratified log-rank test (PROC LIFETEST). • A stratified Cox proportional hazard model with Efron's method of tie handling and study intervention as the sole explanatory variable will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio) in PFS between the study interventions (PROC PHREG).
Model Checking & Diagnostics

• The proportional hazards assumption will be assessed prior to model fitting using the following methods:
  ○ Kaplan-Meier plot by study intervention: A non-parallel pattern is an indication of violation of the proportional hazard assumption.
  ○ Plot of log(time) against log(-log(survival)) by study intervention: A non-parallel pattern is an indication of violation of the proportional hazard assumption.
  ○ Plot of Schoenfeld residuals for treatment: A non-zero slope is an indication of a violation of the proportional hazard assumption.
  ○ Evaluation of time-dependency of treatment effect by adding an interaction term of treatment and time in the Cox model. If the interaction term is significant ($p < 0.05$), it is considered that the proportional hazards assumption is violated.
• If one or more of the procedures above demonstrates clear violation of the proportional hazards assumption in PFS, it is considered the proportional hazards assumption does not hold. Hazard ratio and corresponding 95% CI estimated from the Cox model will still be reported.

Model Results Presentation
• Kaplan-Meier estimates for the median PFS and the first and third quartiles will be presented, along with 95% CIs. • The p-value from the stratified log-rank test will be reported. • Hazard ratio and corresponding 95% confidence interval from the Cox model will be reported.

5.2.3. Sensitivity Analyses

Due to early study termination, sensitivity analyses will not be performed for the main CSR.

5.2.4. Supplementary Analyses

Due to early study termination, supplementary analyses will not be performed for the main CSR.

5.2.5. Subgroup analyses

Due to early study termination and minimal available data, subgroup analyses will not be performed for the main CSR.

5.3. Secondary Endpoint(s) Analysis

5.3.1. Key/Confirmatory Secondary Endpoint(s)

5.3.1.1. Definition of Endpoint

Progression-Free Survival per iRECIST (iPFS)

Progression-free survival per iRECIST (iPFS) is one of the key secondary endpoints of this study. It is defined as the interval of time from the date of randomization to the date of the first documented disease progression confirmed consecutively per iRECIST based on investigator assessment, or death due to any cause, whichever occurs first.

Unlike RECIST 1.1, iRECIST requires the confirmation of progression and uses the terms unconfirmed progressive disease (iUPD) and confirmed progressive disease (iCPD).

The progression event date (iPD date) to be used in the calculation of PFS per iRECIST should be the first date of documented iUPD provided that iCPD is confirmed at the next assessment.

If more than one assessment is recorded as iUPD then the final occurrence prior to iCPD will be used. The exception to the rule would be when consecutive iUPDs followed by iCPD, the first occurrence of iUPDs in a row prior to iCPD will be used. For example, if a participant has iUPD at time points 1 and 2 and iCPD at time point 3, then iPD date would be the date of time point 1. If iUPD occurs, but is disregarded because of later iSD, iPR, or iCR, that iUPD date should not be used as the progression event date.

If progression is not confirmed and there is no subsequent iSD, iPR, or iCR, then the iUPD date will be used as iPD date in the following scenarios:

- Participant discontinues study intervention because the participant was judged not to be clinically stable
- Participant does not undergo further response assessments due to any reason (i.e., participant refusal, protocol non-compliance, or participant death)
- Next timepoint response of iUPD, and iCPD never occurs

For the purpose of confirmation of progression per iRECIST, assessments that are not done or are not evaluable will be disregarded and the next evaluable assessment be considered the 'next assessment'. For example, an iUPD followed by an assessment that was not done or not evaluable, and then another unconfirmed progressive disease qualifying for the criteria of iCPD, would be indicative of iCPD.

Determination of dates of iPFS events and dates for censoring in the iPFS analysis are summarized in Table 4.

Table 4 Censoring Rules for Analysis of iPFS

Situation	Primary Analysis
No or incomplete baseline disease assessments and the participant has not died	Censored at the date of randomization
No post-baseline disease assessments and the participant has not died	Censored at the date of randomization
With post-baseline disease assessments, new anticancer treatment is not initiated, and no iPD date or death	Censored at the date of last adequate[2] radiological disease assessment
With post-baseline disease assessments, new anticancer treatment is initiated (prior to iPD date or death)[1]	Censored at the date of last adequate[2] radiological disease assessment on or prior to starting new anticancer treatment
iPD date or death after ≤1 missed disease assessment[4]	Progressed at iPD date[3] or death
iPD date or death documented after ≥2 missed disease assessments[4,5]	Censored at the date of last adequate radiological disease assessment prior to the ≥2 missed disease assessment

Abbreviations: iPD=immune-based progressive disease; iPFS=progression-free survival per iRECIST

1. If iPD and new anti-cancer therapy occur on the same day, it is assumed that the progression was documented first (i.e. outcome is progression; the date is the date of the assessment for progression).
2. An adequate assessment is defined as an assessment where the investigator assessed response is iCR, iPR, or iSD
3. The earliest date of radiological assessments of lesion in which progression criteria are met for the iPD.
4. The case where iPD date or death documented after the initiation of new anticancer treatment is described above and is not included in this situation.
5. Refer to Section 6.3 for details in extended time without an adequate assessment.

Time to Deterioration (TTD) in Pain and Physical Function (PF)

Time to deterioration (TTD) is defined as the time from randomization to the first definitive meaningful deterioration from baseline in the EORTC IL51 pain domain score and the TTD in Physical Function (PF) score. Deterioration is defined as an increase in the EORTC IL51 pain domain and a decrease in physical function as measured by the PROMIS PF 8c. Specifically, the deterioration has to be:

- Meaningful, i.e. greater than a clinically meaningful within-individual change in score, as defined below;

- Definitive, i.e. all subsequent assessment of the score are also showing a clinically meaningful deterioration compared with baseline, or no further score is available for the participants for any reason (including discontinuation, disease progression or death).

Participants who don't show meaningful deterioration will be censored at the time of the last available PRO assessment within the pain domain of EORTC IL51 for TTD in pain or within PROMIS PF 8c for TTD in PF. For TTD in pain or TTD in PF, the determination of dates of deterioration events and dates for censoring are summarized in Table 5.

Table 5 Censoring Rules for Analysis of TTD

Situation	Primary Analysis
No or incomplete baseline corresponding PRO assessments	Censored at the date of randomization
No post-baseline corresponding PRO assessments	Censored at the date of randomization
With post-baseline corresponding PRO assessments, new anticancer treatment is not initiated, and no deterioration	Censored at the date of last available corresponding PRO assessment
With post-baseline corresponding PRO assessments, and new anticancer treatment is initiated (prior to deterioration)[1]	Censored at the date of last available corresponding PRO assessment on or prior to starting new anticancer treatment
Deterioration after ≤1 missed corresponding PRO assessment[2]	Deteriorated at the date of deterioration
Deterioration after ≥2 missed corresponding PRO assessments if at least one missed assessment is due to the participant being too ill[2,3]	Censored at the date of last available corresponding PRO assessment prior to the ≥2 missed corresponding PRO assessment

1. If deterioration and new anti-cancer therapy occur on the same day, it is assumed that the deterioration was documented first (i.e. outcome is deterioration; the date is the date of deterioration).
2. The case where deterioration observed after the initiation of new anticancer treatment is described above and is not included in this situation.
3. In the case of deterioration after ≥2 missed corresponding PRO assessments and all missed assessments are due to reason other than that participant is too ill, it censors at the date of last corresponding PRO assessment on or prior to starting new anticancer treatment as the situation of with post-baseline corresponding PRO assessments, and new anticancer treatment is initiated (prior to deterioration), i.e. missing corresponding PRO assessments are ignored in this case. Refer to Section 6.3 for details in extended time without an adequate assessment.

As no threshold for meaningful within-individual change is established for the EORTC IL51 pain domain score or PROMIS PF 8c score, the value for use in the TTD analyses will be determined using pooled blinded data from study 209229 and 209227. The threshold may be finalised following study database lock and will be reported in the main CSR. The full procedure for determination of meaningful within-person change in EORTC IL51 pain domain score and PROMIS PF 8c score will be fully described in a standalone SAP. It will include using an anchor-based approach that utilizes the patient global impression of severity and change as an anchor, and possibly other clinical anchors (e.g. ECOG status). Supportive distribution-based methods may be applied as the sensitivity analysis.

5.3.1.2. Main analytical approach

Progression-free-survival per iRECIST

If the PFS hypothesis in PD-L1 CPS≥1 is significant, the hypothesis of iPFS in PD-L1 CPS≥1 will be tested sequentially at the same alpha level. The statistical analysis and reporting of iPFS follows the same method as PFS described in Section 5.2.2. Censoring rules for PFS per iRECIST are given in Table 4.

Time to Deterioration (TTD) in Pain and Physical Function (PF)

The time to deterioration (TTD) in Pain measured by the EORTC IL51 pain domain and the TTD in Physical Function (PF) measured by the PROMIS PF 8c will be analyzed separately for the PD-L1 CPS ≥ 1 and CPS ≥ 20 group using the non-parametric Kaplan-Meier method. Kaplan-Meier plots of TTD will be presented by study intervention. Kaplan-Meier estimates for the median TTD and the first and third quartiles will be presented, along with 95% CIs. CIs for quartiles will be estimated using the Brookmeyer-Crowley method (Brookmeyer, 1982). The TTD for both treatment groups will be compared by the stratified log-rank test.

A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio). The hazard ratio and its 95% confidence interval from the stratified Cox model with a single treatment covariate will be reported separately for the PD-L1 CPS ≥ 1 and CPS ≥ 20 group. Participants with no post-baseline assessments will be considered censored at day 1 and participants without definitive meaningful clinical deterioration for the PRO score will be treated as censored for that PRO score at the last visit.

Only if GSK3359609 in combination with pembrolizumab demonstrates superiority to pembrolizumab plus placebo in OS in both populations (PD-L1 CPS ≥1 participants and PD-L1 CPS ≥20 participants), will key secondary PRO endpoints be tested. The alpha level from OS hypothesis will be propagated to key PRO hypotheses as described in Section 2.1.

5.3.2. Supportive Secondary Endpoint(s)

PFS per RECIST v1.1 and iPFS in PD-L1 CPS ≥20

PFS per RECIST v1.1 and iPFS in PD-L1 CPS ≥20 participants will be analyzed and reported similarly to the primary analysis of PFS and iPFS in PD-L1 CPS ≥1 participants as described in Section 5.2.2 and Section 5.3.1.2 respectively.

OS rate at 12 and 24 months

The non-parametric Kaplan-Meier method will be used to estimate the survival curves. OS rate at 12 months and 24 months and the corresponding 95% CI will be estimated from the Kaplan-Meier analysis will be reported separately for the PD-L1 CPS ≥1 and CPS ≥20 group. The confidence intervals will be based on the Brookmeyer-Crowley method (Brookmeyer, 1982).

If no participants have follow-up duration exceeding the OS milestone, the corresponding summary will not be produced.

A supportive summary of the duration of follow-up will also be produced, presenting the minimum, maximum, median and 25^{th} and 75^{th} percentiles.

Objective Response Rate (ORR) and Disease Control Rate (DCR) per RECIST v1.1

Best Overall Response (BOR)

- The best overall response is the best response recorded from the date of randomization until disease progression or initiation of new anti-cancer therapy, whichever is earlier, as assessed by the investigator per RECIST v1.1. The order from best to worst of the available responses is CR, PR, stable disease (SD), progressive disease (PD) and not evaluable (NE).
- As a randomized double-blind study in which primary endpoints are OS and PFS, the confirmation of CR and PR is not required.
- A subject without any adequate post-baseline assessments will have a BOR of NE.
- To be assigned a status of SD as the best overall response, the minimum criteria for SD duration must be met. Since disease assessment starts at week 9 and a window of ± 7 days is allowed, 56 days since the date of randomization will be used as the minimum criteria for SD duration.

ORR per RECIST v1.1 is defined as the proportion of the participants who have a complete response (CR) or partial response (PR) as the best overall response per RECIST v1.1 based upon investigator assessment.

DCR per RECIST v1.1 is defined as the percentage of participants with a best overall response of CR or PR at any time plus the percentage of participants with SD durability meeting the minimum time of 15 weeks per RECIST v1.1 based upon investigator assessment. A status of SD≥15 weeks will be assigned if the follow-up disease assessment has met the SD criteria at least once after the date of randomization at a minimum of 14 weeks (98 days) considering a one-week visit window.

The number and percentage of participants achieving BOR of CR, PR, SD, PD, NE, as well as those meeting SD durability requirements (SD≥15 weeks) will be provided separately for the PD-L1 CPS ≥1 and CPS ≥20 participants.

The number and percentage of participants achieving ORR and DCR per RECIST v1.1 will be provided separately for the PD-L1 CPS ≥1 and CPS ≥20 participants. The 2-sided 95% exact (Clopper-Pearson) confidence limits for the binomial proportion will also be included.

Stratified Miettinen and Nurminen's method will be used for comparison of the ORR/DCR between two treatment groups. The difference in ORR/DCR and its 95% confidence interval from the stratified Miettinen and Nurminen's method with strata weighting by sample size [Chow, 2003] (i.e. where larger strata carry more weight as compared to smaller strata) with a single treatment covariate will be reported separately for the PD-L1 CPS ≥1 and CPS ≥20 participants. Participants with unknown or missing response will be treated as non-responders that is these participants will be included in the denominator when calculating the percentage of ORR and DCR.

Duration of Response (DoR) per RECIST v1.1

Duration of response (DoR) per RECIST v1.1 is defined as the time from first documented evidence of CR or PR until first documented disease progression per RECIST v1.1 based upon investigator assessment or death due to any cause, whichever occurs first, among participants who demonstrated CR or PR as the best overall response per RECIST v1.1. Censoring rule will follow those for the primary analysis of PFS per RECIST v1.1.

If the number of participants with a best overall response of CR or PR permits, the Kaplan-Meier method will be used to estimate the survival curves for DoR per RECIST v1.1. The Kaplan-Meier estimates for the median DoR and the first and third quartiles, along with 95% CIs estimated using the Brookmeyer-Crowley method (Brookmeyer, 1982) will be reported separately for the PD-L1 CPS ≥1 and CPS ≥20 group by study intervention.

Adverse Events/Serious Adverse Events

The safety analyses will be based on the safety analysis set, unless otherwise specified.

Adverse events analyses including the analysis of adverse events (AEs), Serious AEs (SAEs) and other significant AEs will be based on GSK Core Data Standards.

An overview summary of AEs, including but not limited to, counts and percentages of participants with any AE, AEs related to study intervention, Grade 3+ AEs, Grade 3+ AEs related to study intervention, AEs leading to permanent discontinuation of study intervention, study intervention related AEs leading to permanent discontinuation of study intervention, AEs leading to dose delays, SAEs, SAEs related to study intervention, fatal SAEs, and fatal SAEs related to study intervention and AESIs will be produced.

Adverse events will be coded using the standard Medical Dictionary for Regulatory Affairs (the latest MedDRA dictionary version at the time of reporting) and graded by the investigator according to the NCI-CTCAE v5.0 [NCI, 2017].

A summary of number and percentage of participants with any adverse events by maximum grade will be produced. AEs will be sorted by System Organ Class (SOC) and Preferred term (PT) in descending order. The summary will use the following algorithms for counting the participant:

- **Preferred term row:** Participants experiencing the same AE preferred term several times with different grades will only be counted once with the maximum grade.

- **Any event row:** Each participant with at least one adverse event will be counted only once at the maximum grade no matter how many events they have.

The frequency and percentage of AEs (all grades) will also be summarized and displayed in descending order by PT only.

A separate summary will be provided for study intervention-related AEs by PT and maximum grade. A study intervention-related AE is defined as an AE for which the

investigator classifies the possible relationship to study intervention as "Yes". A worst-case scenario approach will be taken to handle missing relatedness data, i.e. the summary table will include events with the relationship to study intervention as 'Yes' or missing.

All SAEs will be tabulated based on the number and percentage of participants who experienced the event by SOC and PT. A separate summary will also be provided for study intervention-related (fatal and non-fatal) SAEs. The summary tables will be displayed by PT.

A study intervention-related SAE is defined as an SAE for which the investigator classifies the relationship to study intervention as "Yes". A worst-case scenario approach will be taken to handle missing data, i.e. the summary table will include events with the relationship to study intervention as 'Yes' or missing.

A summary of fatal AEs by PT will also be produced.

A summary of non-serious AEs that occurred in strictly 5% of the participants or above will be provided (no rounding for the percentage will be used in terms of 5% threshold, e.g., event with 4.9% incidence rate should not be included in this table). The summary will be displayed by System Organ Class (SOC) and Preferred Term (PT) in descending order of total incidence.

All AEs will be listed. Additionally, a listing of subject IDs for each individual AE will be produced.

Adverse Events of Special Interest (AESI)

The summary of event characteristics will be provided for each AESI (refer to protocol for details) respectively, including number of participants with any event, number of events, number of participants with any event that is serious, number of participants with any event that is related to study intervention, number of occurrences (one, two, three or more), maximum grade, outcomes and the action taken for the event.

The percentage will be calculated in two ways, one with number of participants with event as the denominator and the other with total number of participants as the denominator.

The worst-case approach will be applied at participant level for the maximum grade, i.e. a participant will only be counted once as the worst-case from all the events experienced by the participant.

For action taken to an event, a participant will be counted once under each action. Summary statistics showing the time to onset and the duration of the first occurrence for AESI may also be presented as appropriate.

Dose Modifications

The summaries of dose modifications (dose interruptions, missed doses and dose delays) will be provided if the data warrant. All the dose interruptions, missed doses and dose delays will be listed separately if data warrant.

Dose interruptions will be summarized by number of interruptions and reasons for interruption.

5.4. Tertiary/Exploratory Endpoint(s) Analysis

Due to early study termination, with the exception of the below endpoints, exploratory endpoint analyses will not be performed.

5.4.1. Pharmacodynamic and Biomarker Analyses

The pharmacodynamic and biomarker analysis as well as germline genetic evaluation may be performed. If performed, these may be defined separately in an independent analysis plan and reported outside of the main CSR.

5.5. Other Safety Analyses

The safety analyses will be based on the Safety Analysis Set, unless otherwise specified.

A final CSR at the end of study will contain only key safety analyses and will be conducted after LSLV for the final CSR. Details around which safety outputs will be created for the final CSR will be provided in the OPS.

5.5.1. Extent of Exposure

Extent of exposure to GSK3359609 and Pembrolizumab will be summarized separately and overall, for the Safety Set.

The number of study intervention cycles administered will be summarized with mean, median, standard deviation, minimum, and maximum. The number and percentage of participants who received a given number of cycles (<4, 4-6, and >6) will be reported.

Dose intensity (dose delivered per 3-week period on treatment) will be summarized using mean, median, standard deviation, minimum, and maximum.

Missed doses will be summarized. Dose delays will be summarized by number of delays, reasons for the delays, and delay duration (days). The mean, standard deviation, median, minimum value, and maximum value will be computed for the duration of delay as well as the number and percentage of the delays ≤21, 22-42, and >42 days.

The duration of exposure to study intervention (days) will be calculated and summarized using mean, median, standard deviation, minimum, and maximum.

5.5.2. Deaths

All deaths will be summarized based on the number and percentage of participants. This summary will classify participants by time of death relative to the last dose of medication (>30 days or ≤30 days) and analyze the primary cause of death in the order listed in the CRF. The relationship to COVID-19 infection will also be summarised. A supportive listing will be generated to provide participant-specific details on participants who died.

5.5.3. Adverse Events Leading to Discontinuation of Study Treatment and/or Withdrawal from the Study and Other Significant Adverse Events

The following categories of AEs will be summarized separately in descending order of total incidence by PT only and separate supportive listings will be generated with participant level details for those participants:

- AEs leading to discontinuation of study treatment

A listing of all other significant adverse events will be produced.

5.5.4. Pregnancies

While pregnancy itself is not considered to be an AE or SAE, any pregnancy complication or elective termination of a pregnancy for medical reasons will be recorded as an AE or SAE as described in the protocol. If participants become pregnant while on the study, or the female partner of a male participant becomes pregnant during the study, the information will be included in the narratives and no separate table or listing will be produced.

5.5.5. Additional Safety Assessments

Laboratory evaluations including the analyses of chemistry, hematology, coagulation, cardiac function, thyroid function and routine urinalysis laboratory tests and other screening tests will be based on GSK Core Data Standards.

Unless otherwise specified, the denominator in percentage calculation at each scheduled visit will be based on the number of participants with non-missing value at each particular visit.

5.5.5.1. Laboratory Data

The assessment of laboratory toxicities will examine the laboratory tests listed in Appendix 2 in the protocol. Laboratory grades will be evaluated using CTCAE v5.0. However, some tests are not graded using CTCAE. For hematology, Red Blood Cell (RBC) is not gradable by CTCAE v5.0. For clinical chemistry, BUN and creatinine clearance are not gradable by CTCAE v5.0. For sodium, potassium, calcium, glucose, and magnesium there will be two bi-directional parameters (hyper and hypo) created and the tests will be graded by CTCAE v5.0 in both directions.

Summaries of worst-case grade increase from baseline grade will be provided for all the chemistry, hematology, coagulation and thyroid function lab tests that are gradable by CTCAE v5.0. These summaries will display the number and percentage of participants with a maximum post-baseline grade increasing from their baseline grade. Any increase in grade from baseline will be summarized along with any increase to a maximum grade of 3 and any increase to a maximum grade of 4. Missing baseline grade will be assumed as grade 0. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia separately.

Liver function laboratory tests will be included with chemistry lab tests. A listing of liver monitoring/stopping events will also be produced, where appropriate.

Hepatobiliary laboratory events including possible Hy's law cases will be provided in addition to what has been described above. Possible Hy's law cases are defined as any elevated alanine aminotransferase (ALT) ≥3×upper limit of normal (ULN) and total bilirubin ≥2×ULN, or ALT≥3×ULN and INR>1.5. Hy's law cases where ALT ≥3×ULN and total bilirubin ≥2×ULN, with alkaline phosphatase (ALP) <2×ULN at the time of bilirubin elevation will also be considered. Total bilirubin ≥2×ULN can be within 28 days following the ALT elevation and if direct bilirubin is available on the same day, it must be >35% of total bilirubin.

5.5.5.2. Vital Signs

Summaries of grade increase in systolic blood pressure (SBP) and diastolic blood pressure (DBP) will not be produced.

5.5.5.3. ECG

A 12-lead ECG will be performed at Screening to calculate the heart rate and measurements such as PR, QRS, and QT intervals. ECG after Screening will be performed as clinically indicated. Data will not be summarized.

5.5.5.4. Performance Status

ECOG performance status will be summarized at baseline only, for the mITT Analysis Set.

5.6. Other Analyses

5.6.1. Blinded Independent Central Review (BICR) Auditing Plan

Due to early study termination, BICR will not be performed.

5.7. Interim Analyses

The Independent Data Monitoring Committee (IDMC) will make recommendations for discontinuation or modification of the study based on ongoing reviews of safety data according to the IDMC Charter. In addition, the IDMC will also evaluate all interim efficacy data, including the adaptive decision, PFS/iPFS analysis and OS interim analysis, and make a recommendation based on observed results of the study.

In this double-blind study, all GSK and site personnel will be restricted from access to interim analysis results provided to the IDMC until the conclusion of the study, unless the IDMC recommends significant changes to study conduct that require a protocol amendment. In this case a select group from GSK will be unblinded to review the data to agree on future study conduct.

Due to early study termination as a result of the adaptive decision making, only the adaptive decision interim analysis will be performed, as planned.

5.7.1. Adaptive Decision Making

The analysis for adaptive decision will be conducted using ORR/DCR per RECIST v1.1 based on investigator assessment when approximately the first 100 PD-L1 CPS ≥1 participants have a minimum follow-up of 6 months.

The adaptive decision making will be guided by the analysis in the PD-L1 CPS ≥1 participants in the IA1 analysis set, i.e. those with a minimum follow-up of 22 weeks from date of randomization to the data cut off, regardless of death or study withdrawal. This is aligned with the one week visit window for the Week 21 assessment.

The adaptive decision criteria will be positive if there is at least 8% improvement of ORR in the GSK3359609 in combination with pembrolizumab arm comparing with the pembrolizumab/placebo arm in PD-L1 CPS ≥1 population. Confirmation of CR and PR is not required in the adaptive decision making.

1) If the ORR outcome per RECIST v1.1 is positive with ΔORR≥8% in PD-L1 CPS ≥1 population, the study continues to an originally planned Phase III sample size for a definitive Phase III evaluation.
2) If the ORR/DCR outcome per RECIST v1.1 is negative with ΔORR<0% and ΔDCR<0% in PD-L1 CPS ≥1 population, the study may stop for futility depending on the recommendation of IDMC based on the totality of the data; refer to the IDMC charter for details.
3) Otherwise, the study will continue as planned Phase II sample size for a definitive Phase II evaluation.

5.7.2. PFS/iPFS Analysis and Interim OS Analysis

The below planned analyses will not be performed due to early study termination.

The study is designed to have one PFS/iPFS analysis and two OS analyses in PD-L1 CPS ≥1 participants. Two OS analyses will be performed in PD-L1 CPS ≥20 participants. The safety of the treatment will also be assessed at the interim analysis.

The timing of PFS/iPFS analysis and the OS interim analysis is triggered by the pre-specified number of PFS events in the PD-L1 CPS ≥1 population.

At the time of PFS/iPFS analysis, the OS interim analysis will be conducted in PD-L1 CPS ≥1 participants to allow for early stopping of the study due to efficacy or allow for non-binding futility analysis. PD-L1 CPS ≥20 participants will be tested sequentially if OS is significant in PD-L1 CPS ≥1 participants.

The timing of the final OS analysis in the PD-L1 CPS ≥1 and CPS ≥20 participants will be triggered by the pre-specified number of OS events in the PD-L1 CPS ≥1 population

The nominal significance levels for the interim and final analyses of OS will be determined by the Lan-DeMets spending function based upon the O'Brien-Fleming boundary. The futility bounds of this study are non-binding and the bounds are considered guidance rather than strict bounds.

Table 6 summarizes the information fraction, sample size and decision guidance for the planned PFS/iPFS and OS analyses as a Phase II study. Table 7 summarizes the information fraction, sample size and decision guidance for the planned PFS/iPFS and OS analyses as a Phase III study.

Table 6 Summary of Sample Size and Decision Guidance at the Planned PFS and OS Analyses as a Phase II Study

Analysis	Key Endpt	Population	Expected Number of Events (Information Fraction)	Efficacy Boundary[1]		Non-Binding Futility Boundary[1]	
				p-value	Cumulative alpha	p-value	Cumulative beta
PFS FA, OS IA (H1-H4)	PFS/ iPFS	CPS≥1	~281 (100%)	0.0001	0.0001	NA	NA
	OS	CPS≥1	~205 (84%)	0.0144	0.0144	0.4483	0.0101
		CPS≥20	~95 (84.1%)	0. 0144	0.0144	0.5546	0.0184
OS FA (H1,H2)[2]	OS	CPS≥1	~244 (100%)	0.0208	0.0249	0.0208	0.2474
		CPS≥20	~113 (100%)	0.0208	0.0249	0.0208	0.4440

Abbreviations: CPS=combined positive score; Endpt=endpoint; FA=final analysis; H=hypothesis; HR=hazard ratio; IA=interim analysis; NA=not applicable

1. Efficacy boundaries and non-binding futility boundaries are based on initially assigned type I error rate (one-sided) before any alpha re-allocation and projected number of events at study mile stones. Actual efficacy boundaries will be based on actual numbers of events available at study milestones and actual futility bounds will be updated if overall beta is changed with respect to re-allocation.
2. The descendant OS hypothesis in PD-L1 CPS≥20 participants will be tested only if the parent OS hypothesis in PD-L1 CPS≥1 participants is significant.

Table 7 Summary of Sample Size and Decision Guidance at the Planned PFS and OS analyses as a Phase III Study

Analysis	Key Endpt	Population	Expected Number of Events (Information Fraction)	Efficacy Boundary[1]		Non-Binding Futility Boundary[1]	
				p-value	Cumulative alpha	p-value	Cumulative beta
PFS FA, OS IA (H1-H4)	PFS/ iPFS	CPS≥1	~432 (100%)	0.0001	0.0001	NA	NA
	OS	CPS≥1	~295 (80.4%)	0.0124	0.0124	0.4775	0.0020
		CPS≥20	~137 (80.1%)	0.0122	0.0122	0.5927	0.0050
OS FA (H1,H2)[2]	OS	CPS≥1	~367 (100%)	0.0213	0.0249	0.0213	0.0991
		CPS≥20	~171 (100%)	0.0213	0.0249	0.0213	0.2648

Abbreviations: CPS=combined positive score; Endpt=endpoint; FA=final analysis; H=hypothesis; HR=hazard ratio; IA=interim analysis; NA=not applicable

1. Efficacy boundaries and non-binding futility boundaries are based on initially assigned type I error rate (one-sided) before any alpha re-allocation and projected number of events at study mile stones. Actual

efficacy boundaries will be based on actual numbers of events available at study milestones and actual futility bounds will be updated if overall beta is changed with respect to re-allocation.

2. The descendant OS hypothesis in PD-L1 CPS ≥20 participants will be tested only if the parent OS hypotheses in PD-L1 CPS ≥1 participants is significant.

5.8. Additional Analyses Due to the COVID-19 Pandemic

5.8.1. Protocol Deviations

In addition to the overall summary of important protocol deviations, a separate listing of all protocol deviations related to COVID-19 will be produced for the Enrolled Analysis Set, where important protocol deviations will be flagged.

5.8.2. Additional Displays for Participants with a COVID-19 Infection

A participant is defined as having a suspected, probable or confirmed COVID-19 infection during the study if they answer “Confirmed”, “Probable” or “Suspected” to the case diagnosis question from the COVID-19 coronavirus infection assessment eCRF.

A listing of the numbers of participants with a suspected, probable or confirmed COVID-19 infection, and COVID-19 test results and additional symptoms will be produced, based on the COVID-19 Analysis Set.

5.8.3. Safety

5.8.3.1. Assessment of COVID-19 Adverse Events

A Standardized MedDRA Query (SMQ) will be used to identify all COVID-19 AEs.

The incidence of AEs and SAEs (Fatal and Non-Fatal) of COVID-19, COVID-19 AEs leading to study drug discontinuation, and COVID-19 AEs by severity, will be obtained from standard AE and SAE summaries.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: Abbreviations and Trademarks

6.1.1. List of Abbreviations

Abbreviation	Description
ADA	Anti-drug antibodies
AE	Adverse Event
AESI	Adverse Events of Special Interest
AIC	Akaike's Information Criteria
AUC(0-t)	Area under the plasma concentration-time curve from time 0 to the time of the last quantifiable concentration)
AUC(0-τ)	Area under the concentration-time curve over the dosing interval
BICR	Blinded Independent Central Review
BOR	Best Overall Response
BPI-I3	Brief Pain Inventory- Item 3
Cmax	Maximum observed concentration
Cmin	Minimum Observed Concentration
CI	Confidence Interval
CPD	Confirmed Progressive Disease
CPS	Combined Positive Score
CR	Complete Response
CSR	Clinical Study Report
CTCAE	Common Term Criteria for Adverse Events
DCR	Disease Control Rate
DNA	Deoxyribonucleic acid
DoR	Duration of Response
ECG	Electrocardiography
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Record Form
EORTC QLQ C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire – 30 item Core Module
EORTC H&N35	European Organization for Research and Treatment of Head and Neck 35 Item Module
EQ-5D-3L	EuroQoL 5 Dimensions 3 Levels
FACT-G	Functional Assessment of Cancer Therapy - General
GSK	GlaxoSmithKline
HNSCC	Head and Neck Squamous Cell Carcinoma/Cancer
HPV	Human Papilloma Virus
HRQoL	Health-related Quality of Life
IA	Interim Analysis
IA1	Interim Analysis 1
ICF	Informed Consent Form
ICOS	Inducible T Cell Co-Stimulatory Receptor
iCPD	immune-based Confirmed Progressive Disease

Abbreviation	Description
iCR	immune-based Complete Response
IDMC	Independent Data Monitoring Committee
iPFS	immune-based progression-free survival
iPR	immune-based Partial Response
iRECIST	immune-based Response Evaluation Criteria in Solid Tumors
iSD	immune-based Stable Disease
iUPD	immune-based Unconfirmed Progressive Disease
LVEF	Left Ventricular Ejection Fraction
mITT	Modified Intent-To-Treat
OPS	Output and Programming Specification
ORR	Overall Response Rate
OS	Overall Survival
PD	Progressive Disease
PD-L1	Programmed Death-Ligand 1
PF	Physical Function
PFS	Progression Free Survival
PK	Pharmacokinetic
PR	Partial Response
PRO	Patient Reported Outcomes
PROMIS	Patient-Reported Outcomes Measurement Information System
PT	Preferred Term
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic acid
Q3W	Every Three Weeks
SAP	Statistical Analysis Plan
SD	Stable Disease
SOC	System Order Class
TTD	Time to Deterioration

6.1.2. Trademarks

Trademarks of the GlaxoSmithKline group of companies
None

Trademarks not owned by the GlaxoSmithKline group of companies
NONMEM
SAS
WinNonlin

6.2. Appendix 2: Graphical Testing Strategy Between and Within Subfamilies

Let α be the total family-wise type I error, α = 2.5% (one-sided). In Figure 1, the elementary subfamily is represented by the node shown in ellipsoid with the local significance level $\alpha_S = \alpha W_S$ shown inside the node, where W_S representing the associated initial weight for the subfamily S, S = PFS, OS, or PRO. The transition weights g_{ij} for re-allocation between any two subfamilies i and j are represented on the dashed lines connecting subfamilies, where g_{ij} indicates the fraction of the significance level at the initial node (parent subfamily i) that is added to the significance level at the terminal node (descendent subfamily j) if all hypotheses in the subfamily i are rejected. The alpha level assigned to a descendent subfamily will be rolled over to the descendant subfamily based on the transition weight only if all hypotheses within the parent subfamily are significant.

Similarly, the elementary hypothesis is represented by the node shown in square with the initial local significance level $\alpha_S w_i$ for the hypothesis H_i ($i \in \{1, 2\}$ if S = OS; $i \in \{3, 4\}$ if S = PFS; $i \in \{5, 6, 7, 8\}$ if S = PRO) shown inside the node. The transition weights for re-allocation between any two hypotheses are represented on the solid lines connecting hypotheses within the subfamily S, S = OS, PFS, or PRO, where α_S denotes type I error rate for the corresponding subfamily.

Let H_i, $i \in I = \{1, \ldots, h\}$ denote the i^{th} hypothesis, $t = \{1, 2, \ldots, k\}$ denote the t^{th} planned interim analysis, and α denote the overall Type I error rate in I. The following algorithm shows the general sequentially rejective graphical testing procedures in group sequential trials based on consonant closed weighted Bonferroni tests using group sequential boundaries.

Algorithm:

Step 0: Set $t = 1$.

Step 1: At interim analysis t, compute unadjusted p-values $p_{i,t}$ and nominal significance levels $\alpha^*_{i,t}(\alpha w_i(I))$ for $i \in I$.

Step 2: Select $j \in I$ such that $p_{j,t} \leq \alpha^*_{j,t}(\alpha w_j(I))$ and reject H_j; go to step 3;

If no such j exists and $t < k$, the trial can be continued with $t \rightarrow t+1$; go to step 1 in this case, otherwise stop.

Step 3: Update the graph:

$$I \rightarrow I \setminus \{j\}$$

$$w_l(I) \rightarrow \begin{cases} w_l(I) + w_j(I)g_{jl}, & l \in I, \\ 0, & otherwise \end{cases}$$

$$g_{lk} \rightarrow \begin{cases} \frac{g_{lk} + g_{lj}g_{jk}}{1 - g_{lj}g_{jl}}, & l, k \in I,\ l \neq k,\ g_{lj}g_{jl} < 1 \\ 0, & otherwise \end{cases}$$

Step 4: If $|I| \geq 1$, go to Step 1; otherwise stop.

6.3. Appendix 3: Extended Time Without an Adequate Assessment

PFS/iPFS

Given the scheduled disease assessment (i.e. starts at week 9 and then every 6 weeks; if study treatment discontinued after Week 51 then assessments will be every 12 weeks thereafter until the start of subsequent systemic anticancer therapy), the definition of 2 missed disease assessments will change. The following rules will be used for identifying the duration of extended time without an adequate assessment for PFS.

If the time difference between the event (PD/death) and last adequate disease assessment prior to the new anticancer therapy is more than the window, PFS will be censored at the last adequate disease assessment prior to the event (PD/death) and the new anticancer therapy.

- If the event is after Week 15 + 7 days and on or prior to week 51 + 7 days, then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 98 days (12 weeks + 2-week windows) prior to the event;

- Else if the event is after Week 51 + 7 days and on or before Week 75 - 7 days then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 140 days (18 weeks + 2-week windows).

- Else if the event is after Week 75 - 7 days then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 182 days (24 weeks + 2-week windows).

Similar rules can be applied to the iPFS. For iPFS, if the time difference between iPD/death date and last adequate disease assessment prior to the new anticancer therapy is more than the window, iPFS will be censored at the last adequate disease assessment prior to iPD/death and the new anticancer therapy. The same window specified for PFS above will apply to iPFS.

TTD in Pain/PF

Given the scheduled PRO assessment for EORTC QLQ-H&N35 Pain domain and PROMIS PF 8c (i.e. every 3 weeks till Week 21 and every 6 weeks afterwards), the definition of 2 missed disease assessments will change. The following rules will be used for identifying the duration of extended time without an adequate assessment for TTD in Pain or TTD in PF.

If the time difference between the event (deterioration) and last adequate corresponding PRO assessment prior to the new anticancer therapy is more than the window, TTD will be censored at the last available assessment prior to the event (deterioration) and the new anticancer therapy.

- If the event is after Week 6 + 7 days and on or prior to week 21 + 7 days, then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 56 days (6 weeks + 2-week windows) prior to the event;

- If the event is after Week 21 + 7 days and on or prior to week 33 - 7 days, then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 77 days (9 weeks + 2-week windows) prior to the event;

- Else if the event is after Week 33 - 7 days then a subject will be identified as extended time without an adequate assessment if the subject did not have an adequate assessment during the time period of 98 days (12 weeks + 2-week windows).

6.4. Appendix 4: EORTC QLQ-C30 Scoring Information

Scores for each scale and single-item measure are averaged and transformed linearly to a score ranging from 0–100. (see below image for details). A high score for functional scales and for Global Health Status/QoL represent better functioning ability or HRQoL, whereas a high score for symptom scales and single items represents significant symptomatology [Basch, 2014].

Technical Summary

In practical terms, if items $I_1, I_2, \ldots I_n$ are included in a scale, the procedure is as follows:

Raw score
Calculate the raw score
RawScore = CCI [redacted]

Linear transformation
Apply the linear transformation to 0-100 to obtain the score *S*,

Functional scales:

Symptom scales / items:

Global health status / QoL:

CCI - This section contained Clinical Outcome Assessment data collection questionnaires or indices, which are protected by third party copyright laws and therefore have been excluded.

Range is the difference between the maximum possible value of *RS* and the minimum possible value. The QLQ-C30 has been designed so that all items in any scale take the same range of values. Therefore, the range of *RS* equals the range of the item values. Most items are scored 1 to 4, giving *range* = 3. The exceptions are the items contributing to the global health status / QoL, which are 7-point questions with *range* = 6, and the initial yes/no items on the earlier versions of the QLQ-C30 which have *range* = 1.

Handling of missing items

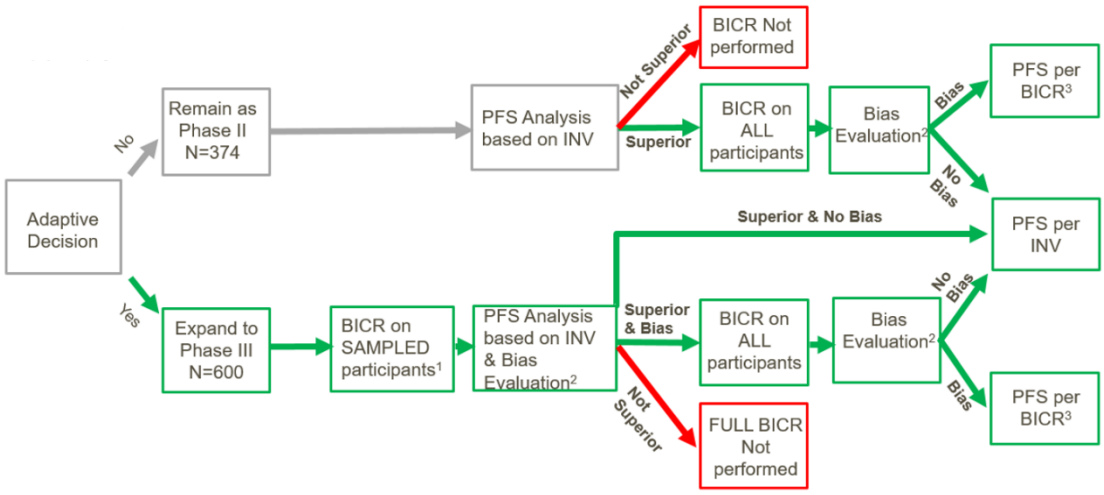
Single-item measures: if the item is missing, the score S will be set to missing.

Scales requiring multiple items: if at least half of the items from the scale are available, the score S will be calculated based on available items. If more than half of the items from the scale are missing, the score S will be set to missing (Fayers, 2001).

6.5. Appendix 5: BICR Implementation Plan

Depending on the results of the adaptive decision making and the PFS analysis, BICR if conducted will be performed in a random sampling of participants (BICR audit) or all participants (100% BICR) as shown in Figure 2. The decision of full study BICR will be guided by the analysis results of PFS per RECIST v1.1 in the CPS ≥1 population.

Figure 2 BICR Implementation Plan



Abbreviations: BICR = Blinded Independent Central Review; INV = investigator assessment; PFS = progression-free survival.

1. The sample BICR will be initiated prior to the databased lock of the PFS analysis when the study expands to phase III. The sampling schema will be generated by SDAC in batches where the first batch will be initiated close to the data base lock and the dynamic allocation will be used in the subsequent batch to ensure the balance across stratification factors.
2. SDAC will perform the analysis and the bias evaluation follows the method described in Section 5.6.1.
3. Based on the bias evaluation, the IDMC may recommend using PFS per BICR instead of investigator assessment and a corresponding protocol amendment may be warranted upon the IDMC recommendation.

If the study remains as a phase II study, BICR on all participants will be performed only when the primary analysis of PFS per RECIST v1.1 by investigator assessment demonstrates a statistically significant treatment effect.
In this case, BICR will be initiated after the database lock of the PFS analysis and bias will be evaluated for PFS based on the full study data.

If the study expands to phase III, a random sample-based BICR auditing approach will be performed.

If the BICR audit shows no evidence of investigator bias in favor of the GSK3359609 in combination with pembrolizumab arm and the PFS analysis results by BICR also

corroborate the results by investigator assessment, the treatment effect in PFS remains to be estimated based on investigator assessment.

If bias cannot be excluded based upon the audit but the primary analysis of PFS per RECIST v1.1 by investigator assessment demonstrates a statistically significant treatment effect, BICR will be conducted for all participants and bias will be further evaluated based on all participants.

If bias cannot be excluded based upon the audit and there is no evidence of a statistically significant treatment effect based on the primary analysis of PFS per RECIST v1.1 by investigator assessment, the third-party core imaging laboratory will not perform BICR for the remaining participants.

When full BICR is conducted and bias cannot be excluded based on the full BICR, the treatment effect in PFS will be estimated using the BICR results upon the recommendation of IDMC. Otherwise, the treatment effect in PFS remains to be based on investigator assessment and the analyses based on the BICR data are considered as supportive.

7. REFERENCES

Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014;106:1-11.

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982 Mar 1:29-41.

Chen C, Anderson K, Mehrotra DV, Rubin EH, Tse A. A 2-in-1 adaptive phase 2/3 design for expedited oncology drug development. Contemporary clinical trials. 2018 Jan 1;64:238-42.

Chow SC, Shao J, Wang H. Sample size calculations in clinical research. Boca Raton: Taylor & Francis. 2003:1–358.

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European Journal of Cancer. 2009 Jan 1;45(2):228-47.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. EORTC QLQ-C30 Scoring Manual, 3rd ed. European Organisation for Research and Treatment of Cancer, 2001.

GlaxoSmithKline Document Number TMF-13842230. Study 209229 Protocol (A Randomized, Double-blind, Adaptive, Phase II/III Study of GSK3359609 or Placebo in Combination with Pembrolizumab for First-Line Treatment of PD-L1 Positive Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma), 2021 Jun 29.

Hwang IK, Shih WJ and De Cani JS. Group sequential designs using a family of type 1 error probability spending functions. Statistics in medicine. 1990; 9(12): 1439-1445.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983; 70: 649-53.

Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research. 2013 Nov 1;5(4):311-20.

National Cancer Institute - Common Terminology Criteria for Adverse Events (NCI-CTCAE), version 5, DCTD, NCI, NIH, DHHS, November 27, 2017.

Rohmeyer K and Klinglmueller F. gMCP: A graphical approach to sequentially rejective multiple test procedures. 2011. R package version 0.6-5.

Seymour L, Bogaerts J, Perrone A, Ford R, Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, Lin NU, Litière S, Dancey J, Chen A, Hodi FS. iRECIST:

guidelines for response criteria for use in trials testing immunotherapeutics. The Lancet Oncology. 2017 Mar 1;18(3):e143-52.

Stone A, Macpherson E, Smith A, Jennison C. Model free audit methodology for bias evaluation of tumour progression in oncology. Pharmaceutical Statistics. 2015 Nov;14(6):455-63.