Statistical Analysis Plan

Study ID: 209227

Official Title of Study: A Randomized, Double-Blind, Adaptive, Phase II/III Study of GSK3359609 in Combination with Pembrolizumab and 5FU-Platinum Chemotherapy versus Placebo in Combination with Pembrolizumab and 5FU-Platinum Chemotherapy for First-Line Treatment of Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma

NCT ID: NCT04428333

Date of Document: 21-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 in Combination with Pembrolizumab and 5FU-Platinum Chemotherapy versus Placebo in Combination with Pembrolizumab and 5FU-Platinum Chemotherapy for First-Line Treatment of Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma

Protocol Number: 209227

Compound Number: GSK3359609
Short Title: A Phase II/III study of GSK3359609 in combination with pembrolizumab and 5FU-platinum chemotherapy compared with pembrolizumab in combination with 5FU-platinum chemotherapy in participants with recurrent or metastatic HNSCC

Acronym: INDUCE-4

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	NCT04428333
EudraCT	2019-003981-42
IND	143221

Effective Date: Refer to Document Date

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	12-JUL-2021
PPD Safety Development Leader, Pharmacovigilance Safety & Medical Governance	02-JUL-2021
PPD Lead Programmer, Oncology Clinical Programming	10-JUL-2021
PPD Health Outcomes PPD Oncology Value Evidence and Outcomes	12-JUL-2021
PPD Manager, Medical Writing Science-Oncology	09-JUL-2021

CLINICAL STATISTICS LINE APPROVAL (ESIGNATURE):

Approver
PPD PPD, Early IO Clinical Statistics Head

TABLE OF CONTENTS

TABLE OF CONTENTS 4

1. INTRODUCTION 10
   - 1.1. Changes to the Protocol Defined Statistical Analysis Plan 10
   - 1.2. Objectives, Endpoints and Estimands 10
     - 1.2.1. Objectives and Endpoints 10
     - 1.2.2. Estimands 12
   - 1.3. Study Design 17

2. STATISTICAL HYPOTHESES/ SUCCESS CRITERIA 19
   - 2.1. Statistical Hypotheses 19
   - 2.2. Multiple Comparisons and Multiplicity 20

3. ANALYSIS SETS 23
   - 3.1. Protocol Deviations 24

4. STATISTICAL ANALYSES 25
   - 4.1. General Considerations 25
     - 4.1.1. General Methodology 25
     - 4.1.2. Baseline Definition 25
     - 4.1.3. Multicenter Studies 26
   - 4.2. Primary Endpoint(s) Analysis 26
     - 4.2.1. Definition of Endpoint(s) 26
     - 4.2.2. Main Analytical Approach 28
     - 4.2.3. Sensitivity Analyses 31
     - 4.2.4. Supplementary Analyses 32
     - 4.2.5. Subgroup analyses 32
   - 4.3. Secondary Endpoint(s) Analysis 32
     - 4.3.1. Key/Confirmatory Secondary Endpoint(s) 32
     - 4.3.2. Supportive Secondary Endpoint(s) 34
   - 4.4. Tertiary/Exploratory Endpoint(s) Analysis 38
   - 4.5. Other Safety Analyses 38
     - 4.5.1. Extent of Exposure 38
     - 4.5.2. Deaths 38
     - 4.5.3. Adverse Events Leading to Discontinuation of Study Treatment and/or Withdrawal from the Study and Other Significant Adverse Events 39
     - 4.5.4. Pregnancies 39
     - 4.5.5. Additional Safety Assessments 39
   - 4.6. Other Analyses 40
     - 4.6.1. Blinded Independent Central Review (BICR) Auditing Plan 40
   - 4.7. Interim Analyses 40
   - 4.8. Additional Analyses Due to the COVID-19 Pandemic 43
     - 4.8.1. Protocol Deviations 43
     - 4.8.2. Additional Displays for Participants with a COVID-19 Infection 44
     - 4.8.3. Safety 44

5. SUPPORTING DOCUMENTATION 45
   - 5.1. Appendix 1: Abbreviations and Trademarks 45

5.1.1. List of Abbreviations 45
5.1.2. Trademarks 46
5.2. Appendix 2: Graphical Testing Strategy Between and Within Subfamilies 47
5.3. Appendix 3: Extended Time Without an Adequate Assessment 49
5.4. Appendix 4: EORTC QLQ-C30 Scoring Information 51
5.5. Appendix 5: BICR Implementation Plan 52

6. REFERENCES 54

VERSION HISTORY

This Statistical Analysis Plan (SAP) V2 for study 209227 is based on the protocol amendment 2 (GlaxoSmithKline Document Number TMF-13842020) dated 30-JUN-2021.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	20-July-2020	Not Applicable	Original version
2	22-Jan-2021	General updates	Updated to align with protocol amendment 1 and correction of minor typographical errors
2	22-Jan-2021	Medical resource use text modification.	Analysis details moved to standalone SAP for clarity.
2	22-Jan-2021	Patient reported outcomes (PRO) text modification.	Terminology clarification for EORTC IL50 and IL51 and analysis methodology clarification for PROMIS-PF-8c and BPI-I3 added. Investigation of missing PRO data added to aid analysis interpretation. Analysis details for PROs moved to a standalone SAP for clarity.
2	22-Jan-2021	Max-combo analysis method updated	Provides further details for the method where the proportional hazards assumption is not valid.
2	22-Jan-2021	Best overall response per RECIST and iRECIST changed to start from the date of randomization rather than from the date of treatment start.	Previous text corrected.
2	22-Jan-2021	COVID-19 additional analyses added	To examine the impact of the COVID-19 pandemic which began after the finalization of protocol version 1.

SAP Version	Approval Date	Change	Rationale
2	22-Jan-2021	Additional details added regarding the blinded independent central review both to the main text and to Appendix 5.	For clarity, particularly in terms of the sampling and auditing methodology.
2	22-Jan-2021	Updated infusion to cycle when summarizing the exposure data.	To align the terminology.
2	22-Jan-2021	Dose intensity calculation clarified.	Not previously addressed.
2	22-Jan-2021	EORTC QLQ-C30 Scoring Information updated	Removed reference to Overall summary score derivation. Not applicable as all domains are not administered.
2	22-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 1.
2	22-Jan-2021	Added COVID-19 Analysis Set.	Additional subset of safety set for COVID-19 related analyses.
2	22-Jan-2021	Updated Enrolled Analysis Set definition.	For clarification of how to handle randomization and dosing errors.
2	22-Jan-2021	Added clarification of requirements for second course treatment safety analyses.	Clarification required following protocol amendment 1.
2	22-Jan-2021	Added clarification for defining new anti-cancer therapy and clinical progression	Clarification required.

SAP Version	Approval Date	Change	Rationale
2	22-Jan-2021	Clarifications added to the PFS primary and supplementary analysis scenarios.	Logic clarification required.
3	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 GSK3359609/placebo. Participants randomized or first dosed after this date will be excluded from efficacy analyses.
3	Refer to document date	Removal of exploratory, supplementary, sensitivity and subgroup analyses for the purposes of an abbreviated CSR. Minimally required safety to be included.	Early study termination
3	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.
3	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.
3	Refer to document date	Added BOR summaries. Added 95% exact confidence intervals for within-treatment estimates of ORR/DCR.	To aide data interpretation and for alignment with GSK standards.
3	Refer to document date	Restructured the ORR/DCR definition section and clarified	To align with visit schedule and clarify the difference between the derivation of SD as BOR vs

SAP Version	Approval Date	Change	Rationale
		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.	durable SD included in the derivation of DCR. Removed example to avoid misinterpretation.
3	Refer to document date	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	Refer to document date	General updates	Correction of minor typographical and formatting errors
3	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 (CSR). 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, exploratory, supplementary, sensitivity, subgroup analyses will not be performed. Minimally required 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 + 5FU-platinum chemotherapy versus placebo in combination with pembrolizumab + 5FU-platinum chemotherapy in the total population (PD-L1 CPS all) and in the PD-L1 positive (CPS≥1) population	• OS in the total and the PD-L1 CPS≥1 populations defined as the time from the date of randomization until the date of death due to any cause • PFS per RECIST v1.1 by Investigator assessment in the total 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 + 5FU-platinum chemotherapy versus placebo in	• PFS per RECIST v1.1 by Investigator assessment in the PD-L1 CPS≥1 population

Objectives	Endpoints
combination with pembrolizumab + 5FU-platinum chemotherapy	• Milestone OS rate at 12, 24 and 36 months in the total and the PD-L1 CPS ≥1 populations • ORR per RECIST v1.1 by Investigator assessment in the total and the PD-L1 CPS ≥1 populations • DCR per RECIST v1.1 by Investigator assessment in the total and the PD-L1 CPS ≥1 populations • DoR per RECIST v1.1 by Investigator assessment in the total and PD-L1 CPS ≥1 populations
• Evaluate the safety and tolerability of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy compared with placebo in combination with pembrolizumab + 5FU-platinum chemotherapy	• Frequency and severity of AEs, AESI, SAEs • Dose modifications (i.e., interruptions, discontinuations)
• Compare disease and treatment related symptoms and impact on function and HRQoL of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy versus placebo in combination with pembrolizumab + 5FU-platinum chemotherapy	• The time to deterioration in Pain measured by the EORTC QLQ-H&N35 pain domain in the total and PD-L1 CPS ≥1 populations • The time to deterioration in Physical Function measured by the PROMIS-PF-8c in the total and PD-L1 CPS ≥1 populations
Exploratory Objectives	**Exploratory Endpoints**
• Compare the efficacy of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy to placebo in combination with pembrolizumab + 5FU-platinum chemotherapy	• ORR, DCR, DoR (sample-size permitting), PFS per iRECIST in the total population • 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 in the total population
• Evaluate and compare disease and treatment related symptoms and impact on function and HRQoL of GSK3359609 in combination with pembrolizumab and 5FU-platinum chemotherapy versus placebo in combination with	• 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

Objectives	Endpoints
pembrolizumab and 5FU-platinum chemotherapy	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 in the placebo combination with pembrolizumab arm	• Non-protocol healthcare encounters, such as provider visits, emergency room visits, hospitalizations, medications, tests, or procedures
• Evaluate GSK3359609 PK properties	• Summary of GSK3359609 concentrations and Cmax, Ctrough
• Determine immunogenicity of GSK3359609	• Anti-drug antibody (ADA) incidence
• Explore relationship between biomarkers in tumour 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 + 5FU-platinum chemotherapy compared to pembrolizumab+ 5FU-platinum therapy in OS in the total and PD-L1 CPS ≥1 populations in R/M HNSCC (H1, H2)	Primary	OS	• Total population (mITT) • PD-L1 CPS ≥1 population (mITT)	• New anti-cancer therapy: treatment policy • Treatment discontinuation: treatment policy	Hazard ratio for GSK3359609+pembrolizumab + 5FU-platinum chemotherapy vs. placebo+pembrolizumab + 5FU-platinum chemotherapy
Primary Objective 2: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy compared to placebo in combination with pembrolizumab + 5FU-platinum chemotherapy in PFS in total population R/M HNSCC (H3)	Primary	PFS	• Total population (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • ≥2 missed disease assessments: hypothetical • Death: composite	Hazard ratio for GSK3359609+pembrolizumab + 5FU-platinum chemotherapy vs. placebo+pembrolizumab + 5FU-platinum chemotherapy

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 + 5FU-platinum chemotherapy compared to placebo in combination with pembrolizumab + 5FU-platinum chemotherapy in PFS in PD-L1 CPS ≥1 R/M HNSCC (H4)	Primary	PFS	• PD-L1 CPS ≥1 population (mITT)	• New anti-cancer therapy: hypothetical • Treatment discontinuation: treatment policy • ≥2 missed disease assessments: hypothetical • Death: composite	Hazard ratio for GSK3359609+pembrolizumab + 5FU-platinum chemotherapy vs. placebo+pembrolizumab + 5FU-platinum chemotherapy
Key Secondary Objective 2: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy compared to placebo in combination with pembrolizumab + 5FU-platinum chemotherapy in TTD in pain in total population and PD-L1 CPS ≥1 R/M HNSCC (H5, H6)	Primary	TTD in pain	• Total population (mITT) • PD-L1 CPS ≥1 population (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	Hazard ratio for GSK3359609+pembrolizumab + 5FU-platinum chemotherapy vs. placebo+pembrolizumab + 5FU-platinum chemotherapy

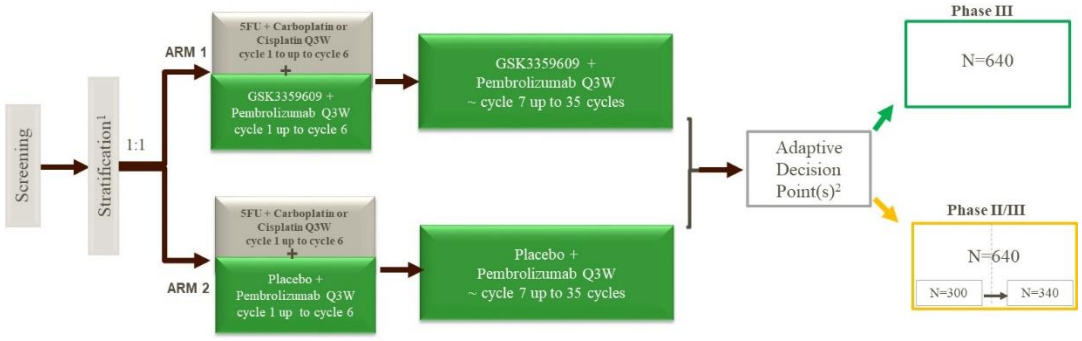
Objective (Hypothesis[1])	Estimand Category	Estimand			
		Variable/ Endpoint[2]	Population of interest (Analysis Set)	Intercurrent Event Strategy[3]	Population Level Summary Measure
				one missed assessment is due to the participant being too ill: hypothetical	
Key Secondary Objective 3: To demonstrate the superiority of GSK3359609 in combination with pembrolizumab + 5FU-platinum chemotherapy compared to placebo in combination with pembrolizumab plus placebo + 5FU-platinum chemotherapy in TTD in PF in total population and PD-L1 CPS ≥1 R/M HNSCC (H7, H8)	Primary	TTD in PF	• Total population (mITT) • PD-L1 CPS ≥1 population (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 + 5FU-platinum chemotherapy vs. placebo+pembrolizumab + 5FU-platinum chemotherapy

Abbreviations: 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 4 for the definition of variable/endpoint.
3. Refer to Section 4.2.1 and Section 5.3 for details in the two or more missed disease assessments.

1.3. Study Design

Overview of Study Design and Key Features



Abbreviations: 5FU=fluorouracil; AUC=area under the curve; CPS=combined positive score; HPV=human papilloma virus; m^2= meters squared; mg=milligram; PDL1=programmed death ligand-1; Q3W=every three weeks; vs=versus

1. **Stratification Factors:**
   - PD-L1 Status (CPS ≥20 vs. 1≤ CPS <20 vs CPS <1)
   - HPV status in oropharynx sites only (positive vs. negative/unknown) vs non-oropharynx)
2. Adaptive decision analyses based on external (Study 209229) and internal (Study 209227) analyses; refer to Section 4.1. of the protocol for details.

Design Features	• This is a randomized, double-blind, adaptive, phase II/III study comparing the combination of GSK3359609 (ICOS agonist) with pembrolizumab and 5FU-platinum chemotherapy to pembrolizumab and 5FU-platinum chemotherapy in participants with 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 of the 209229 study, the outcome is to expand 209229 to a phase III study; there will be no pause in enrollment for the 209227 study and the 209227 study will continue as a phase III and the sample size will be 640 participants. • If at the interim adaptive analysis of the 209229 study, the outcome is to remain 209229 as a phase II; the 209227 study will continue as a phase II/III seamless design and the adaptive decision will be made at the first OS analysis in the 209227 study. The first OS analysis will occur after 9 months of OS follow up in the first 300 participants; at which time recruitment will be paused. If the outcome is to expand to phase III an additional 340 participants will be recruited, otherwise the 209227 study will stop for futility; refer to Section 9.5 of the protocol for details. • If, at the interim adaptive analysis of the 209229 study, the outcome is to stop for futility, then the 209227 study may stop and no further accrual will occur. • All participants will be stratified by two factors: i) PD-L1 CPS status (CPS ≥20 vs. 1≤ CPS <20 vs CPS <1); and ii) HPV status in oropharyngeal cancers (p16 positive vs negative/unknown) vs non-oropharyngeal cancers, then randomly assigned in a 1:1 ratio to the two treatment arms • The study comprises three periods: screening, treatment, and follow-up. The total duration of study participation begins with the signing of the

Overview of Study Design and Key Features	
	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 duration of 5FU-platinum chemotherapy treatment will be 6 cycles (cycle=21 days). ○ 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	• 24 mg of GSK3359609 or placebo administered Q3W + 200 mg of pembrolizumab administered Q3W and 5FU-platinum chemotherapy (carboplatin AUC 5 or cisplatin) at 100 mg/m^2 are each administered as a 30-minute IV infusion once Q3W • GSK3359609 or placebo will be administered first followed by pembrolizumab; the 6 cycles of 5FU-platinum chemotherapy will be administered after pembrolizumab infusion is complete.
Study intervention Assignment	• Participants will be randomly assigned to either receive the combination of GSK3359609 24 mg + pembrolizumab 200 mg + 5FU-platinum chemotherapy **OR** placebo + pembrolizumab 200 mg + 5FU-platinum chemotherapy in a 1:1 ratio after stratification
Interim Analysis	• If the study continues to Phase III, there will be two interim analyses for OS, allowing for early stopping of the study due to efficacy or allow for non-binding futility analysis • For the Phase III study with no enrollment pause, the timing of the two interim analyses will be triggered by the pre-specified number of OS events in the total population. The final analysis of PFS will be aligned with the first OS interim analysis. • For the Phase II/III study with an enrollment pause, the first interim analyses of OS (IA1, adaptive decision making) will occur when the first 300 participants have a minimum follow-up of 9 months. The timing of the second interim analyses will be triggered by the pre-specified number of OS events in the total population. The final analysis of PFS will be aligned with the second OS interim analysis. • Results of the interim analyses will be reviewed by an independent data monitoring committee (IDMC). Further details of interim analyses will be provided in the IDMC Charter
Multiplicity	• The family-wise type I error rate is controlled at 2.5% (one-sided) with 0.01% allocated to the PFS hypotheses and 2.49% to the OS hypotheses • Alpha will be re-allocated as detailed in the multiplicity testing strategy in Section 2.2.

2. STATISTICAL HYPOTHESES/ SUCCESS CRITERIA

2.1. Statistical Hypotheses

The following primary hypotheses will be tested:

Overall Survival (OS)

- Hypotheses (H1): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs OS compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in all participants with R/M HNSCC (total population).

- Hypotheses (H2): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs OS compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in participants with PD-L1 CPS≥1 R/M HNSCC.

Progression-free Survival (PFS)

- Hypotheses (H3): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs PFS per RECIST v1.1 by investigator assessment compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in all participants with R/M HNSCC (total population).

The following key secondary hypotheses will be tested:

Progression-free Survival (PFS)

- Hypotheses (H4): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs PFS per RECIST v1.1 by investigator assessment compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in participants with PD-L1 CPS≥1 R/M HNSCC.

Time to Deterioration (TTD) in Pain

- Hypothesis (H5): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs TTD in Pain (measured by EORTC IL51) compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in all participants with R/M HNSCC (total population).

- Hypothesis (H6): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs TTD in Pain (measured by EORTC IL51) compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in participants with PD-L1 CPS≥1 R/M HNSCC.

TTD in Physical Functioning

- Hypothesis (H7): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs TTD in Physical Function (measured by PROMIS-PF-8c) compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in all participants with R/M HNSCC (total population).

- Hypothesis (H8): GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy prolongs TTD in Physical Function (measured by PROMIS-PF-8c) compared with placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in participants with PD-L1 CPS≥1 R/M HNSCC.

The study is considered to have met the study primary objective if GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy is superior to pembrolizumab + 5FU/platinum-based chemotherapy for either OS or PFS in the total population.

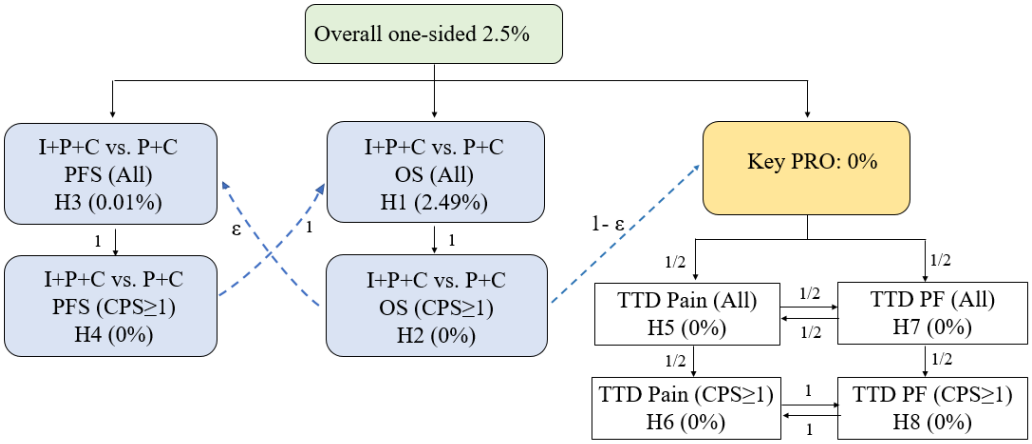
2.2. Multiple Comparisons and Multiplicity

The primary comparisons of interest are the comparisons between GSK3359609 plus pembrolizumab + 5FU/platinum-based chemotherapy and placebo plus pembrolizumab + 5FU/platinum-based chemotherapy in OS in the total population, OS in the PD-L1 CPS ≥1 population, and PFS in the total 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 Comparison Between GSK3359609 in Combination with Pembrolizumab + 5FU/Platinum-based Chemotherapy versus Placebo in Combination with Pembrolizumab + 5FU/Platinum-based Chemotherapy



Abbreviations: C=Chemotherapy; CPS=combined positive score; H = hypothesis; I=GSK3359609; OS=overall survival; P=pembrolizumab; PD-L1= programmed cell death receptor 1-ligand 1; PF=physical function; PFS=progression-free survival; PRO=patient reported outcome; TTD=time to deterioration vs=versus

1. The alpha inside of a box reflects the initial allocated alpha.
2. The alpha level assigned to a family will be rolled over only if the hypotheses within the family are all significant based on the splitting weights for re-allocation presented on the dashed lines connecting families. Within each family, the weights for reallocation from each hypothesis to the others are represented on the solid lines connecting hypotheses.
3. $\varepsilon = 49/249$ if H1 and H2 are rejected at the scheduled timing of PFS analysis; $\varepsilon = 0$ if H1 and H2 are rejected after the scheduled timing of PFS analysis.

The family-wise type I error for this study is controlled at 2.5% (one-sided). Figure 1 shows the initial one-sided alpha-allocation for OS, PFS, 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 plus chemotherapy is given in both treatment arms. It is not expected that the combination of GSK3359609 and pembrolizumab plus chemotherapy 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 5.2 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.

Overall Survival

- An initial alpha level of 2.49% is allocated to the OS hypotheses
- If PFS is significant in both the total population and the PD-L1 CPS ≥1 population, 0.01% alpha level will be reallocated to the OS hypothesis in the total population and be tested at 2.5% (re-allocated alpha).
- OS for the PD-L1 CPS ≥1 population is tested sequentially at the same overall alpha level as that of the OS in the total population if GSK3359609 in combination with pembrolizumab + 5FU/platinum chemotherapy demonstrates superiority to pembrolizumab + 5FU/platinum chemotherapy in OS in the total population. If OS is not significant in the total population, the formal conclusion of statistical significance of OS in the PD-L1 CPS≥1 population will not be drawn.

Refer to Section 3 for alpha- and beta-spending functions used in the OS hypothesis for each possible design.

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

- An initial alpha level of 0.01% is allocated to the PFS hypotheses
- If OS is significant in both the total population and PD-L1 CPS≥1 population at the time of the first OS interim analysis, 0.49% alpha level will be re-allocated to PFS hypothesis so that the PFS hypothesis in total population may be tested at 0.5%.
- If the PFS hypothesis in the total population is significant, the hypothesis of PFS in PD-L1 CPS≥1 population 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 +5FU/platinum chemotherapy demonstrates superiority to pembrolizumab + 5FU/platinum in OS in both populations (total population and PD-L1 CPS ≥1 population) 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 the total and PD-L1 CPS≥1 populations at the time of the first OS interim analysis, a total of 2.0% alpha level will be propagated to key PRO hypotheses.

- If OS meets superiority in both populations (total population and PD-L1 CPS ≥1 population) at the time of the second OS interim analysis or the final OS analysis but PFS fails to demonstrate superiority in either population (total population and PD-L1 CPS ≥1 population), a total of 2.49% alpha level will be propagated to key PRO hypotheses.

- If both OS and PFS demonstrates superiority in both the total population and the PD-L1 CPS≥1 population, 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, with the possibility to further propagate the alpha level between each other.

- If one of two key PRO hypotheses in the total population is rejected based on the re-allocated alpha level, the key PRO hypotheses in PD-L1 CPS≥1 population 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. 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 not needed (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

Analysis Set	Definition / Criteria	Analyses Evaluated
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 • Any participants who receives a study intervention randomization number will be considered to have been randomized	• Study population • Efficacy
Safety	• All randomized participants who take at least 1 dose of study intervention • This population will be the primary population for the analyses of safety data. • Participants will be assigned to the actual study intervention group of GSK3359609 plus pembrolizumab + 5FU-platinum chemotherapy 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.	• COVID-19 Assessments

3.1. 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.

4. STATISTICAL ANALYSES

4.1. General Considerations

4.1.1. 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 vs. CPS <1), 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.

4.1.2. 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.

4.1.3. 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.

4.2. Primary Endpoint(s) Analysis

4.2.1. 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.

Progression-Free-Survival (PFS) 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 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 also 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 post-baseline disease assessments, new anticancer treatment is not initiated, and no documented PD or death	Censored at the date of last adequate radiological disease assessment
With adequate post-baseline disease assessments before the 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 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 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. In case 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 5.3 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.

4.2.2. 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 (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 total and PD-L1 CPS ≥1 populations.

If PFS is significant in both the total and PD-L1 CPS≥1 populations for superiority of GSK3359609 in combination with pembrolizumab + 5FU/platinum-based chemotherapy compared to placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy, OS in the total population will be tested at the 2.5% alpha level. Otherwise, OS in the total population will be tested at the 2.49% alpha level. OS in the PD-L1 CPS≥1 population is tested sequentially at the same overall alpha level as that of OS in the total population. If OS is not significant in total population, the formal conclusion of statistical significance on OS in PD-L1 CPS≥1 population will not be drawn.

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 (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:
  ○ Kaplan-Meier plot by study intervention: A non-parallel pattern is an indication of violation of the proportional hazard assumption.
  ○ 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.
  ○ Plot of Schoenfeld residuals versus time: 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 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 4.2.5 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 the total and the PD-L1 CPS ≥1 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.

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 (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.

4.2.3. Sensitivity Analyses

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

4.2.4. Supplementary Analyses

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

4.2.5. Subgroup analyses

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

4.3. Secondary Endpoint(s) Analysis

4.3.1. Key/Confirmatory Secondary Endpoint(s)

4.3.1.1. Definition of Endpoint

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 4.

Table 4 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 5.3 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- individual 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.

4.3.1.2. Main analytical approach

PFS per RECIST v1.1 in PD-L1 CPS ≥1

PFS per RECIST v1.1 in PD-L1 CPS ≥1 participants will be analyzed and reported similarly to the primary analysis of PFS in PD-L1 CPS all participants as described in Section 4.2.2.

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 total and CPS ≥ 1 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 (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 total and CPS ≥ 1 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 + 5FU/platinum-based chemotherapy demonstrates superiority to placebo in combination with pembrolizumab + 5FU/platinum-based chemotherapy in OS in both populations (total and PD-L1 CPS ≥1 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 2.2.

4.3.2. Supportive Secondary Endpoint(s)

OS rate at 12, 24 and 36 months

The non- parametric Kaplan-Meier method will be used to estimate the survival curves. OS rate at 12 months, 24 months and 36 months, and the corresponding 95% CI will be estimated from the Kaplan-Meier analysis will be reported separately for the total and PD-L1 CPS ≥1 populations. The confidence intervals will be based on the Brookmeyer-Crowley method (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 25th and 75th 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 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 total and CPS ≥1 participants.

The number and percentage of participants achieving ORR and DCR per RECIST v1.1 will be provided separately for the PD-L1 total and CPS ≥1 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 total and CPS ≥1 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/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 (1982) will be reported separately for the PD-L1 total and CPS ≥1 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, 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.

4.4. Tertiary/Exploratory Endpoint(s) Analysis

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

4.5. 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.

4.5.1. Extent of Exposure

Extent of exposure to GSK3359609, Pembrolizumab, 5FU and platinum-based chemotherapy 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, accounting for delayed or missed doses) will be summarized using mean, median, standard deviation, minimum, and maximum.

Missed doses will be summarized by number of missed doses and reasons for missed dose, if available. 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.

4.5.2. 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.

4.5.3. 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.

4.5.4. 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.

4.5.5. 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.

4.5.5.1. 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.

A listing of all laboratory data and urinalysis data will be produced, separately.

Hepatobiliary laboratory events including possible Hy's law cases will be listed, 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.

4.5.5.2. Vital Signs

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

4.5.5.3. 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.

4.5.5.4. Performance Status

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

4.6. Other Analyses

4.6.1. Blinded Independent Central Review (BICR) Auditing Plan

Due to early study termination, no BICR will be performed as planned. Refer to Appendix 5: BICR Implementation Plan.

4.7. 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 adaptive decision analyses, PFS analysis and OS interim analysis, and make recommendations based on observed results of the study and the totality of the data available.

If the study continues as a phase III, there will be two interim analyses on OS, allowing for early stopping of the study due to efficacy or allow for non-binding futility analysis.

For the Phase III study with no enrollment pause, the timing of the two interim analyses will be triggered by the pre-specified number of OS events in the total population. The final analysis of PFS will be aligned with the first OS interim analysis.

For the Phase II/III study with an enrollment pause, the first interim analyses of OS (IA1, adaptive decision making) will occur when the first 300 participants have a minimum follow-up of 9 months. A futility criterion of p-value > 0.1 for log-rank test of OS is pre-specified for adaptive decision making. If the futility criterion is not met at the first interim analysis (IA1), the study will continue as a Phase III study and enrollment of an additional 340 participants to a total of 640 participants will occur. If the futility criterion is met in the OS analysis, the study will stop for futility. The timing of the second interim analyses will be triggered by the pre-specified number of OS events in the total population. The final analysis of PFS will be aligned with the second OS interim analysis.

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 5 and Table 6 summarize the required number of events, sample size and decision guidance for the planned PFS and OS analyses for the two Phase III study scenarios, respectively. 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 interim analyses conducted and 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 reallocation.

Results of the interim analyses will be reviewed by an independent data monitoring committee (IDMC). Further details of interim analyses are provided in the IDMC Charter. Any change to the planned event size will be described in the statistical analysis plan before any unblinding of the data except in the event of early study termination.

Due to early study termination, no interim analyses will be performed.

Table 5 Summary of Number of Events, Sample Size and Decision Guidance at the Planned Interim and Final Analyses for Phase III Study with no Pause in Enrollment

Analysis	Key Endpts	Expected Number of Events at the Planned Analysis (Information Fraction)	Efficacy Boundary[1]		Non-binding Futility Boundary[1]	
			p-value	Cumulative Alpha	p-value	Cumulative Beta
IA1: Final PFS Analysis (H3, H4), Interim OS Analysis 1 (H1, H2)	PFS	~470 (100%)	0.0001	0.0001	NA	NA
	OS	~310 (76%)	0.01	0.01	0.619	0.0007
IA2: Interim OS Analysis 2 (H1, H2)	OS	~375 (92%)	0.0016	0.019	0.143	0.017
FA: Final OS Analysis (H1, H2)	OS	~410 (100%)	0.019	0.0249	0.019	0.094

Abbreviations: Endpts=endpoints; FA: final analysis; H=hypothesis; IA: interim analysis; NA: not applicable; PFS=progression-free survival; OS=overall survival

1. Efficacy boundaries and non-binding futility boundaries are based on initially assigned type I error rate (one-sided) before any alpha roll-over and projected number of events at study milestones. Actual efficacy boundaries will be based on the actual number of interim analyses conducted and actual numbers of events available at study milestones. Actual futility bounds will be updated if overall beta is changed with respect to alpha roll-over.
2. The timing of interim analyses and final analysis for OS may be delayed in the presence of non-proportional hazards.

Table 6 Summary of Number of Events, Sample Size and Decision Guidance at the Planned Interim and Final Analyses for Phase II/III Study with Pause in Enrollment

Analysis	Key Endpts	Expected Number of Events at the Planned Analysis (Information Fraction)	Efficacy Boundary[1]		Non-binding Futility Boundary[1]	
			p-value	Cumulative Alpha	p-value	Cumulative Beta
IA1: Interim OS Analysis[1, 3] (H1, H2)	OS	~136 (32%)	0.00007	0.00007	0.1	0.263
IA2: Final PFS Analysis (H3, H4); Interim OS Analysis[2] (H1, H2)	PFS	~456 (100%)	0.0001	0.0001	NA	NA
	OS	~315 (74%)	0.009	0.009	NA	0.263
FA: Final OS Analysis (H1, H2)	OS	~425 (100%)	0.022	0.0249	0.022	0.288

Abbreviations: Endpts=endpoints; FA: final analysis; H=hypothesis; IA: interim analysis; NA: not applicable; PFS=progression-free survival; OS=overall survival

1. Efficacy boundaries and non-binding futility boundaries are based on initially assigned type I error rate before any alpha roll-over and projected number of events at study milestones. Actual efficacy boundaries will be based on the actual number of interim analyses conducted and actual numbers of events available at study milestones. Actual futility bounds will be updated if overall beta is changed with respect to alpha roll-over.
2. The timing of interim analyses and final analysis for OS may be delayed in the presence of non-proportional hazards.
3. Represents the OS adaptive decision analysis timepoint. If OS reaches the futility boundary then no further OS analyses will be performed and IA1 becomes the end of Phase II for OS.

4.8. Additional Analyses Due to the COVID-19 Pandemic

4.8.1. 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.

4.8.2. 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.

4.8.3. Safety

4.8.3.1. 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.

5. SUPPORTING DOCUMENTATION

5.1. Appendix 1: Abbreviations and Trademarks

5.1.1. List of Abbreviations

Abbreviation	Description
ADA	Anti-drug antibodies
AE	Adverse Event
AESI	Adverse Events of Special Interest
A&R	Analysis and Reporting
AUC	Area under the plasma concentration-time curve
BICR	Blinded Independent Central Review
BOR	Best Overall Response
BPI-I3	Brief Pain Inventory- Item 3
Cmax	Maximum observed concentration
CI	Confidence Interval
CPS	Combined Positive Score
CR	Complete Response
CSR	Clinical Study Report
CTCAE	Common Term Criteria for Adverse Events
Ctrough	Pre-dose (trough) concentration prior to the next dose for each cycle
DCR	Disease Control Rate
DNA	Deoxyribonucleic acid
DoR	Duration of Response
ECG	Electrocardiography
eCRF	Electronic Case Record Form
ECOG	Eastern Cooperative Oncology Group
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
ICF	Informed Consent Form
ICOS	Inducible T Cell Co-Stimulatory Receptor
IDMC	Independent Data Monitoring Committee
iPFS	immune-based progression-free survival
iRECIST	immune-based Response Evaluation Criteria in Solid Tumors
LVEF	Left Ventricular Ejection Fraction
mITT	Modified Intent-To-Treat

Abbreviation	Description
ORR	Overall Response Rate
OS	Overall Survival
OPS	Output and Programming Specification
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	Standard of Care
SOC	System Order Class
TTD	Time to Deterioration

5.1.2. Trademarks

Trademarks of the GlaxoSmithKline group of companies
None

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

5.2. 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:

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

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

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

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

5.3. Appendix 3: Extended Time Without an Adequate Assessment

PFS

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).

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).

5.4. 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: CCI [redacted]

Symptom scales / items: CCI [redacted]

Global health status / QoL: CCI [redacted]

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

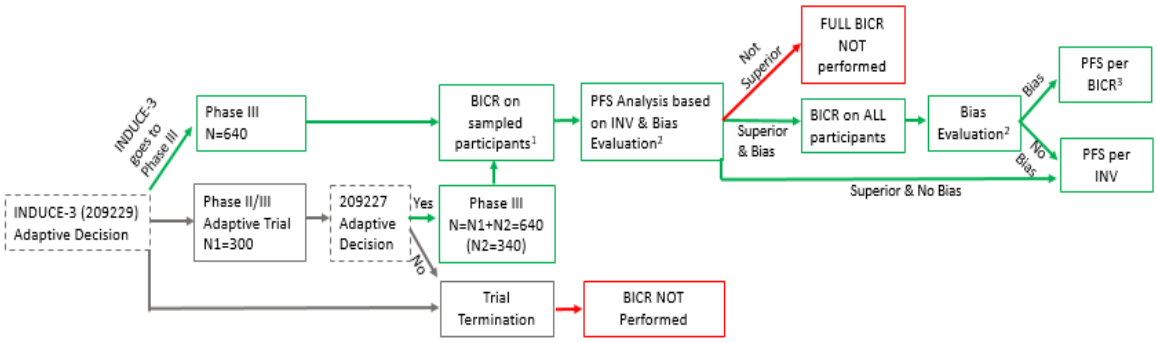
Single-item measures: if the item is missing, the score S will be set to missing.

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).

5.5. 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 (full 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 total 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.
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.

Phase III study with no enrollment pause

If the study continues to Phase III with no enrollment pause, 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 plus chemotherapy 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, full BICR (BICR for all participants) will be conducted and bias will be further evaluated based on all 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.

Phase II/III study with enrollment pause

If the study proceeds to Phase II/III adaptive design with enrollment pause, a random sample-based BICR auditing approach will be performed only when the futility criteria is not met at the interim analysis and the study continues to Phase III.

- If the BICR audit shows no evidence of investigator bias in favor of the GSK3359609 in combination with pembrolizumab plus chemotherapy 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, full BICR (BICR for all participants) will be conducted and bias will be further evaluated based on all 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.

For Phase II/III study with enrollment pause, if the study is decided to stop due to futility at the interim analysis, then BICR will not be conducted. Final analysis of PFS will be based upon the investigator assessment.

6. 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.

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

Fleming, TR, & Harrington, DP. A class of hypothesis tests for one and two sample censored survival data. Communications in Statistics-Theory and Methods. 1981; 10(8): 763-794.

GlaxoSmithKline Document Number TMF-13842020. Study 209227 Protocol (A Randomized, Double-Blind, Adaptive, Phase II/III Study of GSK3359609 in Combination with Pembrolizumab and 5FU-Platinum Chemotherapy versus Placebo in Combination with Pembrolizumab plus 5FU-Platinum Chemotherapy for First-Line Treatment of Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma). 2021 Jun 30.

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.

Signature Page for 209227 TMF-13852560 v1.0

Reason for signing: Approved	Name: PPD Role: Approver Date of signature: 21-Jul-2021 14:29:41 GMT+0000

Signature Page for TMF-13852560 v1.0