INTEGRATE IIb: Randomised Phase III Open Label Study of regorafenib + nivolumab vs standard chemotherapy in Refractory Advanced Gastro-Oesophageal Cancer (AGOC)

Protocol Number: AG0315OG/CTC0140 (version 9)

Statistical Analysis Plan

Prepared by: David Espinoza
NHMRC Clinical Trials Centre
University of Sydney

Version: v2

Version date: 26 May 2025

Approved by: Prof Ian Marschner_____ (Name)

Director Biostatistics_____ (Designation)

Ian Marschner Digitally signed by Ian Marschner Date: 2025.06.18 09:18:22 +10'00' _____ (Signature)

_____ (Date)

Contents

Version History

Version	*Date*	*Principal Changes*
1	30OCT2023	N/A
2	25MAY2025	Sec. 2.1 Add PFS supplementary analysis Sec. 3.1 Define randomisation date variable used for OS derivation, update list of eCRF fields for definition of date last known alive. Sec. 3.2 Define randomisation date variable used for PFS derivation. Sec. 3.3 Add listing of AE of special interest. Sec. 3.6 Add Time to Treatment Discontinuation definition. Sec. 4.0 Add details regarding data cut-off. Sec. 5.0 Define population for supplementary analysis excl. Italian participants. Sec. 7.0 Details re. ordering of STRAT factors added. Sec. 8.7 Add details regarding analysis for BOR, OTR, DCR, DOR. Sec. 9.0 Additional subgroups added to subgroup definition table. Sec 10 Update Sample size from 450 to 460. Sec 11.2 Add table Definitive Boundary for Rejection of the Null Hypothesis.

1 Introduction

This document defines the planned analyses for the INTEGRATE IIb trial. INTEGRATE IIb is an AGITG investigator-initiated study coordinated by the NHMRC Clinical Trials Centre. The trial is run with financial support from, but independently of, Bayer Healthcare Pharmaceuticals, Inc. Bayer reserve the right to use INTEGRATE IIb data to support regulatory labelling claims and have documented supplemental methodological strategies in a separate statistical analysis plan (SAP) specifically for that purpose. Instructions in that separate SAP will have primacy of other sections of this document with regards to analyses undertaken by Bayer to support regulatory labelling claims.

The aim of INTEGRATE IIb is to determine if the regorafenib and nivolumab combination (RegoNivo) improves overall survival compared with current standard chemotherapy options in refractory AGOC.

The study is an open-label phase III trial that randomises eligible patients to RegoNivo or standard chemotherapy options in a 2:1 ratio with stratification for geographic region (Asia vs. Rest of World (ROW)), prior VEGF inhibitors (Yes vs No), and prior immunotherapy (Yes vs No).

The target population is adults with histologically or cytologically confirmed AGOC, with evaluable metastatic or locally advanced disease, who have failed or were intolerant of a minimum of 2 lines of prior anti-cancer therapy which must have included a platinum agent and a fluoropyrimidine analogue.

Participants in the RegoNivo arm will self-administer 90mg (3x30mg) of regorafenib days 1-21 of each 28-day treatment cycle and receive intravenous nivolumab 240 mg day 1 of each 14 day cycle until disease progression or prohibitive adverse events as per protocol, or 2 years maximum treatment duration, whichever is applicable. Participants in the control arm will receive investigator choice of standard chemotherapy options (refer to protocol for details). Both treatment groups will also receive best supportive care.

2 Objectives

The primary objective is to determine the effect of RegoNivo compared to standard chemotherapy options on

1. Overall survival (OS) in the overall study population

The secondary objectives are to determine the effect of RegoNivo compared to standard chemotherapy on:

2. OS in the Asian subpopulation
3. Progression free survival (PFS)
4. Objective tumour response rate (OTRR)
5. Quality of life (QoL)
6. Safety (adverse events)
7. Time to treatment discontinuation (TTD)
8. Best overall response (BOR)
9. Disease Control Rate (DCR)
10. Duration of Response (DOR)

The tertiary/correlative objectives:

The analyses of the tertiary objectives specified in the protocol are considered out of scope for this SAP. A separate SAP will be prepared for the tertiary/correlative objectives. Biomarkers/PK data will not be joined to the trial database until the plan is finalised.

2.1 Efficacy Estimand Definition

The standard estimand definition for the efficacy objectives is based on the following specifications:

1. the treatment conditions of interest are randomisation to regorafenib + nivolumab or standard chemotherapy options. Participants will be analysed according to allocated treatment regardless of what treatment they received.
2. the population of interest is that defined by the protocol inclusion/exclusion criteria.
3. the endpoints are as per the definitions in Section 3.
4. a "treatment policy" approach will be used to account for intercurrent events.
5. the population-level summary measures used to compare treatments are as per the definitions in Section 8.

For the principal analysis of PFS, a "hypothetical strategy" will replace the "treatment policy" approach for the handling of intercurrent events, i.e., patients will be censored if they receive non-protocol anti-cancer treatment prior to meeting criteria for disease progression (See Section 3.2).

Supplementary analysis will be conducted for the analysis of PFS where patients will be censored if:

- they received non-protocol anti-cancer treatment prior to meeting criteria for disease progression (See Section 3.2), or
- death or progression is observed after 2 or more missed visits

3 Endpoint Derivation

3.1 Overall Survival (OS) (Primary Endpoint)

Overall survival is defined as the interval from the date of randomisation (RANDOMIZED_ATLOCAL on the RANDO eCRF) to date of death from any cause, or the date last known alive (for censored patients). The date last known alive will be based on the maximum date from the following eCRF fields detailed in the table below.

Field	Label	eCRF Form	Notes
VISDAT	Visit Date	DOV	VISITYN must be 'Yes'. Follow-up visit form explicitly asks about date last known alive. The date associated with follow-up visits will not be used.
DSLIVDAT	Date last known alive	FUPSTAT	
NPTSTDAT	Date this anti-cancer therapy started	FUPSTAT	
SVAEDAT	AE assessment date	AEYN	
AESTDAT	AE Start date	AE	Date cannot be used if date matches date of death.
AEENDAT	AE End date	AE	
SAESTDT	SAE Start date	AE	
SAEENDAT	SAE End date	AE	
HUADMDAT	Date of admission	HOSP	
HUDISDAT	Date of discharge	HOSP	
CMSTDAT	Start date	CONMED	
CMENDAT	End date	CONMED	
LBDAT	Collection date	LAB	Available wherever there have been TR collections

BRCLDAT	Date of procedure for tissue (date of surgery)	TISSUE	Remove any dates before randomisation date.
LBCOLDATTIM	Collection date	BLOODEXT	Cannot be used if the tumour marker value is ND.
QSC30CDAT	QoL completion dates	QLQC30	
QSC30STAMP	Paper QoL completion dates	QLQC30	
QSEQ5DDAT	QoL completion dates	EQ5D	
QSEQ5DSTAMP	Paper QoL completion dates	EQ5D	
QSSTO22CDAT	QoL completion dates	QLQSTO22	
QSSTO22STAMP	Paper QoL completion dates	QLQSTO22	
QSPDTADAT	QoL completion dates	PDTAF	
QSPDTASTAMP	Paper QoL completion dates	PDTAF	
OMTADAT	Date of tumour assessment	TADATE	
DSPROGDAT	Date of progression	PROG	
CELDDAT	Date last dose of study treatment	EOT	
DSCOMPDAT	Date of study completion	EOS	

3.2 Progression Free Survival (PFS)

Progression free survival (PFS) is defined as the interval from the date of randomisation (RANDOMIZED_ATLOCAL on the RANDO eCRF) to the date of first evidence of disease progression (DSPRGDAT on the PROG eCRF) or death, whichever occurs first. A PFS event is defined as the first occasion that either:

(i) RECIST v1.1 criteria for disease progression are met,

(ii) a participant is judged to have progressed by the responsible investigator (in the event no RECIST assessment is available), or

(iii) death occurs.

For participants judged to have progressed by the responsible investigator prior to, or in the absence of progression on RECIST v1.1 criteria, PFS is defined as the interval from the date of randomisation to:

- Date being recorded in the DSPRGDAT (“Date of progression”) field in the PROG eCRF from.

AND/OR

- If reason for ceasing study treatment (CEEOTRSN) is “Disease progression”, the date last dose of protocol treatment taken (CELDDAT) is used as the date of this event.

Participants who commence non-protocol anti-cancer therapy without prior evidence of progression will be censored at the prior RECIST assessment. Participants who are not observed to meet any of the above conditions will be censored at the date of the last post-baseline radiologic assessment (OMTADAT from the TADATE eCRF) or date of randomisation.

PFS according to iRECIST will be investigated as applicable (i.e., in the event there are instances of pseudoprogression in the REGONIVO arm).

3.3 Tumour Responses (BOR, OTR, DCR, DOR)

An objective tumour response (OTR) is defined according to RECIST v1.1 criteria as a best overall response (BOR) of complete response (CR) or partial response (PR). A subsequent assessment is used to confirm an initial CR/PR. Disease Control Rate (DCR) is defined as a BOR of complete or partial response or stable disease.

Duration of Response (DOR) is defined as the time (in days) from first response of PR or CR to radiological or clinical PD or death. A subsequent assessment is used to confirm an initial response of CR/PR, however if confirmed, the date of the first assessment of CR/PR is used. DOR is only defined for patients who achieve confirmed CR or PR. For participants who have not progressed or died at the time of analysis, DOR will be censored at the date of their last tumour evaluation. It is calculated as 'date of progression or death due to any cause – date of confirmed response'.

Response according to iRECIST will also be investigated as applicable (i.e., in the event there are instances of pseudoprogression in the REGONIVO arm).

3.4 Quality of Life (QoL)

The core EORTC QoL Questionnaire (QLQ-C30) will be used in conjunction with the disease specific module for gastric cancer (STO22). The EQ-5D will be used to obtain utility valuations on the health states experienced by participants. Additional items on cancer specific symptoms will be collected using the patient Disease and Treatment Assessment (PTDATA) form for English-speaking participants only. All instruments will be scored according to standard conventions (See Section 11.1).

3.5 Safety

The National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 5.0 will be used to classify and grade the intensity of adverse events after each treatment cycle.

Table below lists adverse events of special interest (AESI) and the CTCAE v5.0 terms which are relevant for this category.

No.	AESI	CTCAE v5.0
1	Gastrointestinal Haemorrhage	Anal haemorrhage, cecal haemorrhage, colonic haemorrhage, duodenal haemorrhage, oesophageal haemorrhage, gastric haemorrhage, ileal haemorrhage, jejunal haemorrhage, lower gastrointestinal haemorrhage, oral haemorrhage, rectal haemorrhage and upper gastrointestinal haemorrhage
2	Gastrointestinal Perforation	Gastrointestinal perforation, colonic perforation, duodenal perforation, oesophageal perforation, gastric perforation, ileal perforation, jejunal perforation, rectal perforation and small intestine perforation.
3	Intracranial haemorrhage	Intracranial haemorrhage
4	Cerebrovascular ischaemia	Cerebrovascular ischaemia

5	Rash	Rash, rash maculo-papular, papulopustular rash, rash pustular, rash acneiform and Stevens-Johnson syndrome
6	Hand-foot syndrome	Hand-foot syndrome
7	Fatigue	Fatigue
8	Blood bilirubin increased	Hyper-bilirubinaemia, blood bilirubin increased
9	ALT increased	ALT increased
10	AST increased	AST increased
11	ALP increased	ALP increased
12	GGT increased	GGT increased
13	Myelitis	Myelitis
14	Encephalitis	Encephalitis
15	Hyperglycaemia	Hyperglycaemia
16	Hypocortisolaemia	Adrenal insufficiency
17	Hypothyroidism	Hypothyroidism
18	Hypertension	Hypertension
19	Diarrhoea	Diarrhoea
20	Fever	Fever
21	Proteinuria	Proteinuria
22	Anorexia	Anorexia
23	Colitis/enterocolitis	Colitis/enterocolitis
24	Pneumonitis	Pneumonitis
25	Arthritis	Arthritis
26	Hyperthyroidism	Hyperthyroidism
27	Hepatic failure	Hepatic failure

3.6 Time to treatment discontinuation (TTD)

To align with other safety analysis, time to treatment discontinuation is defined for the safety population as the interval from the date of first treatment to the 'Date of last dose of study treatment' (CELDDAT on the EOT eCRF).

Participants who are still ongoing treatment, the censoring date will be the minimum of 'data cut-off date' or the 'last recorded treatment date', where:

- RegoNivo arm, the 'last recorded treatment date' is the last recorded date in the EXENDAT ('End date in this treatment period') field in the EOT eCRF for which a dose was delivered EXDOSE ('Dose per administration') > 0,
- Chemotherapy arm,

a) all participants except those on 'Trifluridine and Tipiracil', the 'last recorded treatment date' is the last recorded treatment date (EXCYCD1DAT, EXCYCD8DAT, EXCYCD15DAT, EXCYCD21DAT, EXORCYCSTDAT) for which a dose was delivered (as indicated by variables EXCYCD1ACDOS, EXCYCD8ACDOS, EXCYCD15ACDOS, EXCYCD21ACDOS)
b) for participants on 'Trifluridine and Tipiracil' the date of day 1 of any cycle is given by EXORCYCSTDAT. For all other days of that cycle, the date would be extrapolated from the day they receive an actual dose (variables EXORCYCD1DOS1, EXORCYCD1DOS2, …, EXORCYCD12DOS1, EXORCYCD12DOS2). The 'last recorded treatment date' is the last recorded treatment date extrapolated from EXORCYCSTDAT for which they receive a dose.

4 Timing of Analyses

Assuming the study is not terminated early, the final analysis is planned to be undertaken after the required number of deaths have occurred (i.e., N=380). The date that confirmation of 380 OS events has been reached will be declared as the data cut-off date(14JAN2025). (It is expected that upon completion of data cleaning, extra OS events prior to the cut-off date will be captured).

The study design incorporates formal interim analyses performed on OS once 2/3 of the required events are observed. The interim analysis allows for early rejection of the null hypothesis (of no difference between treatments) according to an error spending function with an O'Brien-Fleming boundary shape[2]. Indicative boundaries for these analyses are presented in the table below. The actual number of events observed at the time of the interim analyses would be used to construct the definitive rejection boundary (see appendix 11.2 for definitive boundaries for actual IA). The conditional power (CP) of the study will also be calculated for OS at the interim analyses[3]. The Independent Data Safety Monitoring Committee (IDSMC) may recommend altering aspects of the study (e.g., early termination on futility grounds) if the CP is unacceptably low. A CP <20% is suggested in the protocol as a guide for the IDSMC for defining 'an unacceptably low CP'.

Table 1: Indicative Boundary for Rejection of the Null Hypothesis

Proportion of Required Events	Events	Boundary Rejection of the Null Hypothesis*		2-Sided P-value Corresponding to Boundary	Power
		Z Score [logHR]	lower HR		
0.67	255	+/-2.50 [+/-0.333]	0.717	0.012	0.57158
1	380	+/-1.99 [+/-0.217]	0.805	0.046	0.90280

* Calculated using following SAS code: proc seqdesign plots=boundary(hscale=samplesize) pss(cref=1) stopprob(cref=1) boundaryscale= stdz /* mle p-value mle */; errorspendobrienfleming: design nstages=2 method=errfuncobf alpha=0.05 beta=0.0972 info=cum(0.67 1); samplesize model=twosamplesurv (nullhazard=0.116, hazardratio=0.7 acctime=24 foltime=12 weight= 2 1); run;

5 Analysis Sets

The intention-to-treat (ITT) population will comprise all randomised participants. For ITT analyses, participants will be analysed according to allocated treatment regardless of what treatment they received.

The safety population will comprise all randomised participants who received at least one administration of study medication. If a participant receives at least one dose of regorafenib and/or nivolumab treatment (irrespective of randomised allocation) in the period between randomisation and cessation of study treatment, they will be included in the REGONIVO treatment arm of the safety population. If a participant does not receive REGONIVO but does receive at least one dose of protocol defined standard chemotherapy, they will be included in the control arm of the safety population.

The primary analysis population used for the evaluation of the experimental treatment on efficacy parameters will be the ITT population. Safety analyses will be performed using the safety population.

To facilitate alignment with Bayer's planned analysis a supplementary analysis will be conducted with the ITT and Safety analysis sets with participants randomised in Italy excluded.

6 Type I Error (Alpha)

Unless otherwise specified (e.g. Section 8.5 specifies the use of a closed testing procedure for analyses of OS, and Section 9 specifies use of the Benjamini-Hochberg procedure), a two-sided α of 5% will be applied to interpret the results of hypothesis tests and to construct confidence intervals (which will accompany all treatment effect estimates). P-values from secondary analyses that are unadjusted for multiple comparisons will be interpreted conservatively.

7 Accounting for Stratification Factors

Randomisation was stratified by geographic region (Asia vs. ROW), prior VEGF inhibitors (Yes vs No), and prior immunotherapy (Yes vs No). Analyses will be based on corrected stratification factor information if this was wrongly recorded at randomisation. Sensitivity of conclusions arising from stratified analyses may be explored by performing unstratified analyses, as well as analyses that account for stratification factors as covariates.

For stratified analyses of time-to-event endpoints, if any of the eight strata defined by the three stratification factors has <10 events, then all three of the two stratification factor combinations will be assessed against the following rule: If more than one of these combinations meets the rule, that with the maximum of the minimum event counts across strata will be chosen. If there is a tie, the following order will be applied:

1. Stratification by prior VEGF inhibitors and prior immunotherapy,
2. Stratification by prior immunotherapy and geographic region,
3. Stratification by prior VEGF inhibitors and geographic region.

If none of the two factor combinations has at least 10 events, the single factor with the maximum of the minimum event counts across strata will be chosen. If there is a tie, the
the following order will be applied:

1. Stratification by prior immunotherapy,
2. Stratification by geographic region,
3. Stratification by prior VEGF inhibitors.

The same approach will be used with the Asian region cohort analysis and the two stratification factors of relevance (i.e., prior VEGF inhibitors and prior immunotherapy).

8 Analysis of Endpoints and Participant Characteristics

8.1 Participant Disposition

The number of participants in each Analysis Set will be presented along with reasons for any exclusions. The Kaplan-Meier method will be used to summarise follow-up time for OS by treatment allocation with deaths being treated as censored observations. A CONSORT flow diagram will be prepared.

8.2 Baseline Demographic and Clinical Characteristics

Descriptive statistics will be prepared to summarise baseline characteristics of the study participants by treatment allocation. Means, standard deviation, median, interquartile range, minimum and maximum will be presented for continuous variables, whilst counts and percentages will be presented for categorical variables.

8.3 Exposure to Study Treatment (TTD)

A summary of the cycles delivered and reasons for ceasing treatment will be presented. Kaplan-Meier estimates of median time to treatment discontinuation will be calculated.

8.4 Non-Protocol Anti-Cancer Treatments

A summary of non-protocol anti-cancer treatments recorded in the study database will be prepared by treatment allocation. Non-protocol anti-cancer therapy includes treatments received after the investigator indicated that assigned protocol therapy was ceased.

8.5 Overall Survival (OS) (primary endpoint)

The analysis on OS will involve testing the following two null hypotheses: (1) no treatment effect on OS in the whole trial cohort ($H_0^{(All)}$); and (2) no treatment effect on OS in the Asian region cohort ($H_0^{(Asian)}$). A sequential closed testing procedure will be used to constrain the overall type I error of these two tests to 5%. This will involve first testing $H_0^{(All)}$ at α=5%. If $H_0^{(All)}$ is rejected, then $H_0^{(Asian)}$ will be tested and rejected at α=5%. If $H_0^{(All)}$ is not rejected, then $H_0^{(Asian)}$ will not be rejected.

$H_0^{(All)}$ will be tested using a stratified log-rank test accounting for relevant stratification factors. A hazard ratio (HR) will be estimated from a stratified Cox proportional hazard regression (PH) model.

$H_0^{(Asian)}$ will be tested using a stratified log-rank test accounting for relevant stratification factors. A hazard ratio (HR) will be estimated from a stratified Cox proportional hazard regression (PH) model.

Kaplan-Meier curves will be prepared and used to estimate median OS with corresponding 95% confidence intervals.

8.6 Progression Free Survival (PFS)

The treatment groups will be compared on PFS using a stratified log-rank test accounting for stratification factors. Kaplan-Meier curves will be prepared and used to estimate median PFS with 95% confidence intervals. An estimate of the hazard ratio for the treatment effect will be obtained using a stratified PH model accounting for relevant stratification factors.

8.7 Tumour Response (BOR, OTR, DCR, DOR)

Objective tumour response rate will be calculated by summing the number of participants in respective arms that are assessed as having a complete or partial response (OTR) and dividing by the total number of participants in the corresponding arm of the analysis set. OTR rate and 95% confidence intervals will be prepared for each treatment group.

The treatment groups will be compared on OTR using a Cochran-Mantel-Haenszel test accounting for relevant stratification factors, and the corresponding odds ratio will be calculated. If the response rate is too low (expected cell count is less than 5 across strata), factors will be removed as necessary (in an analogous way to that described in Section 7 for time-to-event variables). If after removal of all factors the rule is not met, groups will be compared using Fisher's exact test.

DCR will be analysed in an analogous way to OTR, and DOR will be analysed in an analogous way to PFS. Frequency tables for BOR will be displayed.

8.8 Quality of Life (QoL)

The QoL analyses will comprise the primary, secondary, and tertiary set of approaches next described. These approaches were previously used for the INTEGRATE and INTEGRATE IIa trials.

8.8.1 Primary Approach: Deterioration-Free Survival

A deterioration-free survival (DFS) endpoint will be constructed as a marker of overall net clinical benefit of treatment. This is defined as the time from randomisation to the first of the following events: a 10-point deterioration in health status from baseline (without subsequent 10-point improvement compared with baseline), disease progression, death, or treatment discontinuation. Two DFS endpoints will be derived using different markers of health status deterioration based on the EORTC QLQ-C30: one using the Physical Function Scale (DFS_{PF}), and the other using the General Health Scale (DFS_{GHS}). The treatment groups will be compared on DFS_{PF} and DFS_{GHS} using a log-rank test accounting for stratification factors. Kaplan-Meier curves will be prepared and used to estimate median DFS with corresponding 95% confidence intervals. An estimate of the hazard ratio for the treatment effect will be obtained using a stratified PH model accounting for stratification factors.

8.8.2 Secondary Approach: Repeated Measures Model

A repeated measures model (RMM) will also be applied to the scales from EORTC QLQ-C30, EORTC STO22 and EQ-5D instruments in a secondary analysis out to 48 weeks. The RMM will be specified within the general linear-mixed model framework. The population-level summary measure used to compare treatments will be the predicted mean scores from the RMM. The RMM will include participant as a random effect (i.e., random intercept), and fixed effects for the relevant baseline score, treatment allocation, time point (as a factor), and a treatment allocation-by-time point interaction. All patients with a baseline measure and at least one post-baseline value will be included in analyses. Multiple imputation techniques may be investigated in a sensitivity analysis to account for patients who agreed to QoL assessment and returned a baseline form but did not supply any post-baseline form. Time point is constructed as a factor by assigning assessments to one of a series of contiguous visit windows of 4-week duration. An overall estimate will be obtained by removing the treatment allocation-by-time point interaction from the model. Whilst all available data will be used to fit the models, estimates of treatment differences at individual time points will be shown out to where a reasonable number of patients (e.g., 25) are still contributing information and estimates are reasonably precise. The applicability of a cumulative logit mixed effect model for accommodating highly skewed data transformed to an ordinal scale by quantile-split may be investigated.

8.8.3 Tertiary Approach: Frequency of Troublesome Symptoms

The proportion of participants experiencing troublesome symptoms, or troublesome impacts on general aspects of QoL, will be calculated across the post-baseline assessment period (based on the worse grade reported during the on-treatment period), and comparisons between treatment arms will be made using logistic regression adjusting for baseline in exploratory analyses. Symptoms will be defined as troublesome if they are rated with an intensity of >= 3 points relative to 0="no trouble at all" and 10="worst I can imagine". Reductions on general aspects of QoL (e.g., Physical well-being)

will be defined as troublesome if a 3-point decrement from the optimal score of 10 is reported (i.e. a rating of <=7 points relative to 0="worst possible" and 10="best possible").

8.9 Safety Data Analysis

Number of subjects with Adverse Events (AEs) and Serious Adverse Events (SAEs) will be tabulated by treatment received and CTCAE criteria including system organ class, term, and worst grade.

Counts and event rates of AESI will be reported by treatment allocation. Event rates of AESI will be presented via tornado plots.

9 Tests of Heterogeneity

Heterogeneity of treatment effecting the OS and PFS endpoints, with respect to a given covariate, will be tested by fitting a covariate-by-treatment interaction term in an unstratified Cox PH regression model along with the associated main effects terms. Tests of heterogeneity of treatment effect will be applied to the stratification factors used at randomisation. Additional subgroup analyses are listed in table 2. The p-values from these tests of heterogeneity will be appraised with respect to the multiple testing involved (using the Benjamini-Hochberg Procedure) with one family of tests for OS.

Table 2: Subgroup Definitions

Subgroups	Subgroup Definitions
Stratification factor: Region	Asia, v ROW
Stratification factor: Prior Immunotherapy	Yes, v No
Stratification factor: Prior VEGF inhibitors	Yes, v No
Primary Site	GOJ, v gastric
Age	≤64, v >64
ECOG Performance Status	0, v 1+
Metastatic disease sites	≤2, v >2
Lines of prior therapy	≤2, v >2
PD-L1 Combined Positive Score (CPS)	To be determined in a blinded analysis.
Planned chemotherapy (Control arm only)	Trifluridine v Paclitaxel/Docetaxel v Irinotecan

Where categories are sparsely populated or data quality is poor consideration will be given to collapse the levels of the subgroups or, if inappropriate the analysis may be omitted altogether.

10 Sample Size Determination

A sample of 460 participants randomised in a 2:1 ratio (RegoNivo: chemotherapy) and followed until 380 deaths occur provides 90% power to detect a hazard ratio (HR) for OS of 0.70 with a 2-sided α of 0.05 allowing for a margin for loss to follow-up. The design accommodates early stopping for benefit at an interim analysis performed at 2/3 of the required events (using the error spending approach of Lan-DeMets - see Table 1).

11 Appendices

11.1 Quality of life questionnaire scoring

11.1.1 EORTC QLQ

Instrument/Scale/Item	Scale Code	No. items	Range	Item number	High score
QLQ-C30					
Global health status/QoL	QL2	2	6	29, 30	+ve
Functional scales					
Physical	PF2	5	3	1 - 5	+ve
Role	RF2	2	3	6, 7	+ve
Emotional	EF	4	3	21 - 24	+ve
Cognitive	CF	2	3	20, 25	+ve
Social	SF	2	3	26, 27	+ve
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	-ve
Nausea and vomiting	NV	2	3	14, 15	-ve
Pain	PA	2	3	9, 19	-ve
Dyspnoea	DY	1	3	8	-ve
Insomnia	SL	1	3	11	-ve
Appetite loss	AP	1	3	13	-ve
Constipation	CO	1	3	16	-ve
Diarrhoea	DI	1	3	17	-ve
Financial difficulties	FI	1	3	28	-ve
QLQ- STO22					
Symptom scales					
Dysphagia	STODYS	3	3	31 – 33	-ve
Pain	STOPAIN	4	3	34 – 37	-ve
Reflux symptoms	STORFX	3	3	38 – 40	-ve
Eating restrictions	STOEAT	4	3	41 – 43,46	-ve
Anxiety	STOANX	3	3	47,48,50	-ve
Dry mouth	STODM	1	3	44	-ve
Taste	STOTA	1	3	45	-ve
Body image	STOBI	1	3	49	-ve
Hair loss	STOHL	2/1	3	51,52*	-ve

Definition:

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

Calculate the raw score: $$RS = \frac{I_1 + I_2 + \cdots + I_n}{n}$$

Apply the linear transformation to 0-100 to obtain the score S:

Functional Scales: $$S = \{1 - \frac{(RS - 1)}{range}\} * 100$$

Symptom scales/items: $$S = \{\frac{(RS - 1)}{range}\} * 100$$

Global health status/QoL: $$S = \{\frac{(RS - 1)}{range}\} * 100$$

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 the items in any scale take the same values. Therefore the range of RS equals the range of the items. Most items are scored 1 to 4 giving a *range* of 3. Exceptions are the items contributing to the global health status/QoL, which are 7-point questions with a *range* of 6.

Note a high score for a functional scale represents a *high/*healthy level of functioning, a high score for the global health status/QoL represents a *high QoL*, but a high score for a symptom/item scale represents a *high level of symptomatology/problems.*

Missing Items:

If at least half of the items from the scale have been answered, assume that the missing items have values equal to the average of those items which *are* present for that respondent.

Thus:

- Have at least half of the items from the scale been answered?
- If Yes, use all the items that were completed, and apply the standard equations for calculating the scale scores; ignore any items with missing values when making the calculations.
- If No, set scale score to missing.
- For single-item measures, set score to missing.

11.1.2 EQ5D-5L

The EQ5D-5L comprises the following items:

1. Mobility (5 response levels)
2. Personal Care (5 response levels)
3. Usual Activity (5 response levels)
4. Pain/Discomfort (5 response levels)
5. Anxiety/Depression (5 response levels)
6. a visual analogue scale assessing overall health (0-100)

Items 1 to 5 collectively define 5^5=3,125 health profiles. The UK utilities for these profiles will be applied. These weights are available from the EQ5D website (https://euroqol.org/).

11.1.3 PTDATA Questionnaire

The PTDATA questionnaire comprises questions on aspects of QoL and symptoms.
There are 7 items on aspects of QoL (Energy, Physical well-being, Appetite, Emotional well-being, Overall well-being, Mood, Mobility) rated on an 11-point scale ranging from 0 = “worst possible” to 10 = “best possible”. There are 47 questions on symptoms rated on a 11-point scale ranging from 0 = “no trouble at all” and 10 = “worst I can imagine”. The items are not collapsed to form scales.

11.2 Definitive Boundary for Rejection of the Null Hypothesis

Analysis	Number of Events	Boundary Rejection of the Null Hypothesis Z Score		2-Sided P-value Corresponding to Boundary
		Lower	Upper	
IA	271	- 2.41065	+ 2.41065	0.01592
Final	380	- 2.00297	+ 2.00297	0.04518