- To evaluate the efficacy of JS001 plus chemotherapy compared with placebo plus chemotherapy as measured by investigator and IRC-assessed PFS, ORR, DoR, and DCR according to irRECIST.

Exploratory:

- To assess predictive, prognostic, and pharmacodynamic exploratory biomarkers (including but not limited to PBMC, PD-L1, PD-1, tumor mutation burden and others) in archival and/or fresh tumor tissue and blood and their association with disease status, mechanisms of resistance, and/or response to Toripalimab Injection (JS001).
- To evaluate the utility of biopsy at the time of apparent disease progression to distinguish apparent increases in tumor volume related to the immunomodulatory activity of JS001 (i.e., pseudoprogression and tumor immune infiltration) from true disease progression.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a randomized, placebo-controlled, multi-center, double blinded, Phase III study to determine the efficacy and safety of Toripalimab Injection (JS001) in combination with gemcitabine/cisplatin compared with placebo in combination with gemcitabine/cisplatin as first-line treatment in patients with histological/cytological confirmation of recurrent or metastatic NPC.

Approximately 280 patients who fulfill all of the inclusion criteria and none of the exclusion criteria will be randomized in a 1:1 ratio to one of the two treatment arms according to the following stratification factors:

- ECOG performance status (0 versus 1)
- Disease stage (recurrent versus metastatic)

After stratification, patients will be randomly assigned by Interactive Web Response System (IWRS) to the combination of JS001 (Arm A) or placebo (Arm B) with gemcitabine and cisplatin given every 3 weeks (Q3W) in 3-week cycles. Day 1 (baseline) will be defined as the first day a patient receives study medication. Patients will receive JS001 (Arm A) or placebo (Arm B) on Day 1 of each 3-week cycle. In Arms A & B, patients will receive gemcitabine on Days 1 & 8 and cisplatin on Day 1 of each cycle. The chemotherapy will continue until progressive disease, excessive toxicity, noncompliance, withdrawal of consent, or a maximum of 6 cycles, whichever occurs first in the 'during chemotherapy' phase. During the 'post-chemotherapy' phase, patients randomized to Arm A or Arm B will continue treatment with JS001 or placebo as maintenance therapy Q3W until excessive toxicity or progressive disease, withdrawal of consent or Investigator's judgement or a maximum of 2 years. Patients may continue treatment with JS001 (Arm A) or placebo (Arm B) beyond radiographic progression by the Response Evaluation Criteria In Solid

Tumors (RECIST) version 1.1, provided they are experiencing clinical benefit, as assessed by the Investigator in the absence of unacceptable toxicity or symptomatic deterioration attributed to disease progression, as determined by the Investigator after an integrated assessment of radiographic data and clinical status. Patients will be permitted by sponsor medical monitor or designee and treating physician to continue JS001/placebo after RECIST v1.1 criteria for progressive disease are met if they meet all of the criteria according to the protocol.

Tumor evaluation scans will be performed at screening (as baseline) then every 6 weeks in the first 12 months then every 9 weeks thereafter until objective disease progression. The management of patients will be based solely upon the results of the tumor evaluation scans conducted by the Investigator.

The IRC, consisting of independent experts, will review all radiologic scans according to RECIST 1.1 and irRECIST.

The total length of the study, from screening of the first patient to the end of the study, is expected to be approximately 32 months.

At the time of end of study, sponsor will continuously provide investigational product to patients who still are under treatment in accordance with local regulations.

This is a double blinded and one interim efficacy analysis of PFS is planned and pre-defined stopping boundary is set for two-sided p-value. The iDMC will provide the recommendation as to whether to unblind the study or not according to the data of interim analysis and the iDMC charter if the stopping boundary were met. If the Sponsor accepted the recommendation and unblind the study, JS001 will be provided to the patients who were randomized to arm A and still on JS001 treatment until the treatment discontinuation criteria are met according to protocol, the placebo treatment will be terminated for the patients who were randomized to arm B. After the study is unblinded, the tumor evaluation, survival follow-up, safety information, and so on information collection should be performed as required by the protocol.

'Treatment Day' will be calculated relative to the first dose of study drug, i.e. Treatment Day = Assessment Date - The First Day a Patient Receives Study Medication + 1.

Continuous data will be summarized in terms of the mean, standard deviation (Sd), median, minimum, maximum and number of observations, unless otherwise stated. Continuous data that are expected to be skewed will be presented in terms of the maximum, upper quartile, median, lower quartile, minimum and number of observations. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The Sd will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of patients providing data at the relevant time point (n), frequency counts and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.

Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator. If sample sizes are small, the data displays will show the percentages, but any textual report will describe frequencies only.

Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.0001, in general, will be presented to four decimal places. P-values less than 0.0001 will be presented as "<0.0001".

Confidence interval (CI) will be presented to one more decimal place than the raw data.

All report outputs will be produced using SAS® version 9.3 or a later version in a secure and validated environment. All report outputs will be provided to the Sponsor in a single Microsoft Word document per each output.

Baseline for Efficacy and Safety

Baseline measurements are the last available data obtained prior to the first dose of study drug.

Scheduled and Unscheduled Assessments

Scheduled assessments taken outside of protocol allowable windows will be displayed according to the eCRF assessment recorded by the Investigator. The by-visit tables will only include the scheduled assessments, unless specifically stated otherwise. If more than one value is available for a given visit, the first valid observation will be used in summary tables, unless specifically stated otherwise, and all observations will be presented in listings.

Unscheduled assessments will be included in listings, but not in summary tables.

4.7.1.4 Examination of Subgroups

The primary efficacy endpoint of investigator-assessed PFS according to RECIST v1.1 will be analyzed including but not limited to following subgroups:

- Age: <=50 versus >50
- Sex: Male versus Female
- Baseline ECOG status per CRF: 0 versus 1
- Baseline ECOG status per IWRS: 0 versus 1
- Disease stage per CRF (recurrent versus metastatic)
- Disease stage per IWRS (recurrent versus metastatic)
- EBV copy number: <=500 versus >500; <=2000 versus >2000
- Baseline PD-1 expression status: PD-1 will be categorized to <=1% vs. > 1%, <=5% vs. > 5%, and <=10% vs. > 10%.

4.7.2 Primary Efficacy Endpoint – IRC-Assessed PFS

PFS is defined as the time from randomization to the time of first documented disease progression, as determined by IRC assessments per RECIST v1.1, or death from any cause, whichever occurs first. Patients who have not experienced disease progression or death at the time of analysis will be censored at the time of the last tumor assessment. Patients with no post-baseline tumor assessment and no death will be censored on the date of randomization. For patients who do not experience disease progression but have started any new anti-tumor therapy, PFS will be censored at the time of the last tumor assessment before the new therapy. For patients who have two or more consecutive missing tumor assessments, PFS will be censored at the time of the last available tumor assessment before the missing, or at the date of randomization if there was no post-baseline tumor assessment before the missing. In the COVID-19 pandemic period, for patients who have two or more consecutive missing tumor assessments due to COVID-19, in the circumstance if subsequent tumor assessments become available and there is no immediate disease progression, the subsequent tumor assessments will be used in the PFS analysis; if the subsequent immediate tumor assessment is disease progression, PFS will still be censored at the last tumor assessment before the missing, or at the date of randomization if there is no post-baseline tumor assessment before the missing.

PFS (month) will be calculated as:

(event date or censoring date – randomization date + 1) / 30.4375.

If the day of death is missing, it would be imputed with day 15.

The two-sided log-rank test, stratified by ECOG performance status (0 vs. 1), and disease stage (recurrent vs. metastatic), as recorded in the IWRS, will be used as the primary analysis to compare PFS between the two treatment arms.

The hazard ratio (HR) for disease progression or death will be estimated with the use of the stratified Cox proportional hazards model by including strata of ECOG performance status (0 vs. 1), and disease stage (recurrent vs. metastatic). Efron's method will be used

to handle ties. The 95% CI for the HR will be constructed. Results from an unstratified analysis will also be provided.

The Kaplan-Meier methodology will be applied to estimate the median PFS for each treatment arm, and Kaplan-Meier curves will be developed. The Brookmeyer Crowley methodology[4] by using log-log transformation to survival function, will be used to construct the 95% CI for the median PFS for each treatment arm.

Primary analysis will be performed using ITT population. Sensitivity analysis will be performed using PPS as well.

4.7.3 Secondary Efficacy Variables

4.7.3.1 Investigator-Assessed PFS

IRC-assessed PFS is defined as the time from randomization until the earliest occurrence of disease progression, as determined by the independent review committee, per RECIST v1.1, or death from any cause, whichever occurs first.

The methods outlined for the investigator-assessed PFS will be used for the IRC-assessed PFS analysis.

4.7.3.2 Overall Survival (OS)

OS is defined as the time from randomization to death from any cause. Data from patients who are alive at the time of the analysis will be censored as of the last date they are known to be alive. OS (month) is calculated as:

(death date or censoring date – date of randomization + 1) / 30.4375

If the day of death is missing, it would be imputed with day 15.

The methods outlined for PFS will be used for the OS analysis.

A by-patient listing of survival follow-up status will be provided.

4.7.3.3 Objective Response Rate (ORR)

An objective response is defined as either a confirmed complete response (CR) or a partial response (PR), as determined by IRC and the investigator using RECIST v1.1. Patients not meeting these criteria, including patients without any post baseline tumor assessment, will be considered non-responders.

Any tumor assessment, either by the IRC or the investigator, after the first PD or the new anti-cancer therapy will not be used to determine the best overall response evaluation.

ORR is defined as the proportion of patients who had an objective response. ORR and its 95% CI will be calculated for each treatment arm, and its 95% CI will be constructed using the Clopper-Pearson method (Clopper and Pearson, 1934).

Treatment difference in ORR will be analyzed using the Cochran-Mantel-Heanszel test, stratified by ECOG performance status (0 vs. 1), and disease stage (recurrent vs. metastatic). The 95% CI for the difference in ORRs between the two treatment arms will be determined using the Mantel-Heanszel method.

4.7.3.4 Duration of Response (DoR)

DoR will be assessed in patients who have an objective response CR/PR as determined by IRC and the investigator using RECIST v1.1.

DoR is defined as the time interval from the date of the first occurrence of a CR or PR (whichever status is recorded first) until the first date that progressive disease or death is documented, whichever occurs first.

Patients who have not progressed and who have not died at the time of analysis will be censored at the time of last tumor assessment date. If no tumor assessments are performed after the date of the first occurrence of a CR or PR, duration of response will be censored at the date of the first occurrence of a CR or PR. The DoR (month) will be calculated as:

(progression date or censoring date – first date of CR or PR + 1) / 30.4375

The methods outlined for PFS will be used for the DoR analysis.

4.7.3.5 Disease Control Rate (DCR)

DCR is defined as the rate of patients with CR or PR as best response or stable disease (SD) maintained for 6 weeks (>= 6 weeks) as determined by IRC and the investigator according to RECIST v1.1.

The methods outlined for ORR will be used for the DCR analysis.

4.7.3.6 PFS rate at 1 and 2 years

The PFS rate at 1 and 2 years after randomization will be estimated with the use of the Kaplan-Meier methodology for each treatment arm, along with 95% CI that are calculated with use of the standard error derived from Greenwood's formula, where the PFS is the IRC and investigator-assessed PFS according to RECIST v1.1. The 95% CI for the difference in the PFS rate between the two treatment arms will be estimated with use of the normal approximation method.

4.7.3.7 OS rate at 1 and 2 years

The methods outlined for the PFS rate will be used for the OS rate analysis.

suggest that the adverse event started prior to the first dose of study treatment or more than 60 days after the last dose of study treatment.

The incidence of some AEs, which should have been summarized under the same preferred term, could be underestimated because they were coded to different preferred terms due to different reported terms. Aside from the MedDRA dictionary, additional aggregation rule are applied to the following AEs: Leukopenia, Neutropenia, Thrombocytopenia, Lymphopenia, Hypokalemia, Fatigue, Rash, Musculoskeletal pain, Cough, Diarrhea, Pneumonia, Neuropathy peripheral, Upper respiratory tract infection, Abdominal pain, Arrhythmia, Oedema, Proteinuria, Hypertension, Blood bilirubin increased, Insomnia (please see Section 6.4 Appendix D).

Following summaries will be provided:

- An overview of all TEAEs: the number and percentage of patients with any TEAEs, any >= grade 3 TEAE, any treatment-emergent immune-related adverse event (irAE), any treatment emergent adverse events of special interest (AESI), any >= grade 3 treatment-emergent AESI, any SAE, any TEAEs with relationship to study drug, any TEAEs with relationship to gemcitabine, any TEAEs with relationship to cisplatin, any TEAEs leading to drug withdrawn, any TEAEs leading to drug interruption, any TEAEs leading to dose reduced, and any TEAEs leading to death, by treatment arm
- A summary of the number and percentage of patients reporting a TEAE and the number of TEAEs, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a TEAE, by treatment arm, maximum severity (grade 1-2, grade 3, grade 4, grade 5, >= grade 3), system organ class, and preferred term
- A summary of the number and percentage of patients reporting a treatment-emergent irAE and the number of irAEs, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a treatment-emergent irAE, by treatment arm, maximum severity (grade 1-2, grade 3, grade 4, grade 5, >= grade 3), system organ class, and preferred term
- A summary of the number and percentage of patients reporting a study drug related TEAE and the number of study drug related TEAEs, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a study drug related TEAE, by treatment arm, maximum severity (grade 1-2, grade 3, grade 4, grade 5, >= grade 3), system organ class, and preferred term
- A summary of the number and percentage of patients reporting a gemcitabine related TEAE and the number of gemcitabine related TEAEs, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a cisplatin related TEAE and the number of cisplatin related TEAEs, by treatment arm, system organ class, and preferred term

- Is life threatening (i.e., the adverse event, in the view of the investigator, places the patient at immediate risk of death).
- Requires or prolongs inpatient hospitalization (see Section 5.3.5.11 of the protocol).
- Results in persistent or significant disability/incapacity (i.e., the adverse event results in substantial disruption of the patient's ability to conduct normal life functions).
- Is a congenital anomaly/birth defect in a neonate/infant born to a mother exposed to study drug.
- Is a significant medical event in the investigator's judgment (e.g., may jeopardize the patient or may require medical/surgical intervention to prevent one of the outcomes listed above).

Other significant adverse events are those adverse events reported as leading to an intervention e.g. discontinuation of study treatment.

Following summaries will be provided:

- A summary of the number and percentage of deaths during the study treatment period, and the number and percentage of deaths during the follow-up period after treatment completion/discontinuation, by treatment arm
- A summary of the number and percentage of patients reporting a SAE and the number of SAEs, by treatment arm, system organ class and preferred term

4.8.3 Vital Signs and Weight

Vital signs will include pulse rate (beats/minute), respiration rate (breaths/min), systolic and diastolic blood pressure (mmHg) while the patient is in a seated position, and temperature (˚C). Vital sign will be measured and the relevant assessment will be performed at the period of pre-dose, JS001 infusion, and post-dose, respectively.

Both raw values and changes from baseline will be summarized using descriptive statistics at each scheduled visit by treatment arm. Baseline is defined as the last evaluation prior to the first dose of study drug.

Vital sign assessment will be summarized using frequencies and percentages.

Following summaries will be provided:

- A summary of the observed values and changes from baseline for each vital sign by each period (i.e. pre-dose, JS001 infusion, and post-dose) and by visit and treatment arm
- A summary of vital sign assessment by each period (i.e. pre-dose, JS001 infusion, and post-dose) and by visit and treatment arm.
- A summary of weight and change from baseline by visit and treatment arm.

If vital sign data is missing, no imputation will be made. During the study, the measurement of vital signs may be repeated at the discretion of the Investigator for safety reasons. If

- Coagulation (INR, PT, aPTT)
- Serum pregnancy test should only be conducted at screening period for women of childbearing potential, including women who have had a tubal ligation; urine pregnancy tests will be performed at each cycle during treatment. A serum pregnancy test must be performed if the urine pregnancy test is positive.

  Childbearing potential is defined as not having undergone surgical sterilization, hysterectomy, and/or bilateral oophorectomy or not being postmenopausal (≥12 months of amenorrhea).
- Urinalysis (specific gravity, pH, glucose, protein, ketones, and blood), dipstick permitted, and microscopic examination (RBCs, WBCs).
- Thyroid function testing (thyroid-stimulating hormone [TSH], free T3, free T4)
- HBV serology (HBsAg, HBsAb, HBcAb)
- HBV DNA should be obtained prior to randomization if patient has positive serology for HBcAb or HBsAg.
- HCV serology (HCV antibody)
- HCV RNA if HCV antibody is positive.
- HIV testing
  All patients will be tested for HIV prior to the inclusion into the study and HIV positive patients will be excluded from the clinical trial.

The continuous test results for each test parameter and change from baseline by visit will be summarized using the descriptive statistics. The categorical test results for each test parameter will be summarized using frequencies and percentages. A shift table of the categorical test results at each visit compared to the categorical test results at baseline will be provided.

The toxicity grade and abnormality worsened from the baseline will be summarized according to NCI-CTCAE version 5.0 (please see Section 6.2 Appendix B).

A by-patient listing of all laboratory data will be provided by treatment arm.

4.8.6 12-Lead Electrocardiogram (ECG) and Echocardiogram

Frequency and percentage of patients who had abnormal (clinically significant or not clinically significant) ECG and echocardiogram overall assessment will be summarized by treatment arm. When there are multiple ECGs taken at a visit, the worst ECG overall assessment will be used.

By-patient listings of 12 ECG and echocardiogram, including the LVEF result (%), will be provided.

4.8.7 ECOG Performance Status

Following summaries will be provided:

All results of analysis will be summarized using the ITT population .

4.9.3 Subgroup Analyses

To assess the consistency of the study results, the IRC and investigator-assessed PFS will be analyzed according to the subgroups defined in Section 4.7.1.4.

Summaries of PFS, including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of median PFS, will be produced separately for each level of the categorical variables for the comparisons between treatment arms.

Subgroup analyses will be performed using ITT population.

4.9.4 Immunogenicity Analysis

Anti-JS001 antibodies, immunoglobulins, and JS001 plasma level will be summarized with descriptive statistics by treatment arm using the Safety Population (Section 4.4.4). Ad hoc analyses may be pursued after data collection.

4.9.5 Biomarker Analysis

Biomarker data will be summarized with descriptive statistics by treatment arm using the Safety Population (Section4.4.4). Ad hoc analyses may be pursued after data collection.

4.9.6 Pharmacokinetic Analysis

Pharmacokinetic analyses will be limited to descriptive statistics conducted using the JS001 Treated Subjects. Individual and average concentration-collect time curve will be plotted.

4.10 Interim Analysis

An iDMC will be set up to evaluate safety data on an ongoing basis, as well as the efficacy data from the planned interim efficacy analyses. All summaries/analyses by treatment arm for the iDMC's review will be prepared by an independent party. Members of the iDMC will be external to the Sponsor and will follow a charter that outlines their roles and responsibilities.

4.10.1 IRC-Assessed PFS

One interim efficacy analysis of PFS was planned to be conducted when approximately 130 PFS events in the ITT population had been observed. This was expected to occur approximately 18 months after the first patient was randomized, although the exact timing of the interim analysis depended on the actual occurrence of PFS events.

The final analysis of PFS was planned to be conducted when approximately 200 PFS events in the ITT population have been observed. This was expected to occur approximately 25

Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

Table A2. Scoring the QLQ-H&N35

Scale name	Scale	Number of items	Item range	QLQ-H&N35 Item numbers
Symptom scales / items				
Pain	HNPA	4	3	1 – 4 (CRF 31 – 34)
Swallowing	HNSW	4	3	5 – 8 (CRF 35 – 38)
Senses problems	HNSE	2	3	13,14 (CRF 43 – 44)
Speech problems	HNSP	3	3	16,23,24 (CRF 46, 53, 54)
Trouble with social eating	HNSO	4	3	19 – 22 (CRF 49 – 52)
Trouble with social contact	HNSC	5	3	18,25 – 28 (CRF 48, 55 – 58)
Less sexuality	HNSX	2	3	29,30 (CRF 59, 60)
Teeth	HNTE	1	3	9 (CRF 39)
Opening mouth	HNOM	1	3	10 (CRF 40)
Dry mouth	HNDR	1	3	11 (CRF 41)
Sticky saliva	HNSS	1	3	12 (CRF 42)
Coughing	HNCO	1	3	15 (CRF 45)
Felt ill	HNFI	1	3	17 (CRF 47)
Pain killers	HNPK	1	1	31 (CRF 61)

Table 1. Schedule of Activities

	Screening [a]		All treatment cycles									Treatment Discontinuation [c]	Survival Follow-Up [d]
			During chemotherapy						Post chemotherapy		Unplanned Visit [b]		
Day (Window)	–28 to –1	–7 to –1	C1D1	C1D8	C2D1 (± 3)	C2D8	C3-C6D1 (± 3)	C3-C6D8	C7D1 (± 3)	C8D1 to PD or intolerable toxicity (± 3)	NA		
Informed consent	x[e]												
Demographic data	x												
Medical history and baseline conditions	x												
Patient-reported outcomes [f]			x		x		x		x	x			
Vital signs[g]		x	x	x	x	x	x	x	x	x	x	x	
Weight		x			x		x					x	
ECOG		x	x	x	x	x	x	x	x	x	x	x	
Height		x											
ECG[h]	x		As clinically indicated										
Echocardiogram[h]	x	As clinically indicated											
Complete physical examination [i]	x												
Limited physical examination [j]			x		x		x		x	x		x	
Hematology [k]		x		x	x	x	x	x	x	x		x	
Coagulation test (INR, PT, aPTT)		x	As clinically indicated										

Missing Data Conventions

In general, missing data will not be imputed for both efficacy analysis and safety analysis, unless specifically stated otherwise.

Detection of Outliers

Any outliers that are detected during the review of the data before database lock will be investigated. If necessary, queries will be issued to the Investigator, to either correct or confirm the outlier.

4.3 Study Subjects

4.3.1 Disposition of Patients

A clear accounting of the disposition of all patients who enter the study will be provided, from screening to study completion.

The following summaries will be provided:

- A summary of the number and percentage of patients who were randomized, the number and percentage of patients who were excluded prior to randomization by major reason and overall, and the number of percentage of patients who were re-screened, using the Enrolled Population.
- A summary of the number and percentage of patients who were treated (with at least one dose of study drug) in each phase of the study (during-chemotherapy phase and post-chemotherapy phase) and the number and percentage of patients who were discontinued from treatment in each phase of the study (during-chemotherapy phase and post-chemotherapy phase), and the number and percentage of major reason of patient discontinuation from treatment in each phase of the study (during-chemotherapy phase and post-chemotherapy phase), by treatment arm and overall, using the ITT population .
- A summary of the number and percentage of patients completed the study and discontinued from the study, and the number and percentage of major reason of patient discontinuation from the study, by treatment arm and overall, using the ITT population.
- A summary of follow-up time (months), which is defined as the time from randomization to a specified event (death, end of the study or the data cut-off for the analysis, whichever occurs first), by treatment arm and overall, using the ITT population. The study follow-up time (months) is calculated as

(event date - randomization date + 1) / 30.4375

A by-patient listing of disposition data will be provided, using the Enrolled Population.

4.7.3.8 Patient-Reported Outcomes (PRO)

The EORTC QLQ-C30 and QLQ-H&N35 data will be scored according to the EORTC scoring manual, the scoring can be seen in Section 6.1 6.1Appendix A.

In the event of incomplete data, if the scale has more than 50% of the constituent items completed, a pro-rated score will be computed that is consistent with the scoring manual and the validation papers of the measure. For subscales with less than 50% of the items completed, the subscale will be considered missing.

Summary statistics (mean, Sd, median, and range) of absolute scores and mean changes from the baseline will be calculated for all disease and/or treatment related symptom items and subscales of the EORTC QLQ-C30 and QLQ-H&N35 at each assessment time point for each treatment arm during the administration of the treatment and the survival follow up period.

The mean (and 95% CI) and median of the absolute scores and the changes from the baseline will be reported for interval and continuous variables.

Previously-published minimally-important differences will be used to identify meaningful change from the baseline within each treatment arm on the disease and/or treatment related symptoms scales (Osoba et al., 1998).

Questionnaire completion rate for EORTC-QLQ-C30 and QLQ-H&N35 is defined as the proportion of questionnaires actually received out of the expected number (i.e., number of patients during the administration of the treatment), which will be calculated and summarized for each assessment time point by treatment arm.

Following listings will be provided:

- A by-patient listing of EORTC-QLQ-C30
- A by-patient listing of QLQ-H&N35

4.8 Safety Evaluation

All safety summaries and analyses will be based upon the Safety Population unless otherwise indicated.

4.8.1 Adverse Events

AEs will be coded using the MedDRA version 23.0.

Treatment-emergent adverse events will be tabulated and are defined as those adverse events that either start or worsen in severity on or after the date/time of first dose of study treatment and on or before 60 days after the date/time of last dose of study treatment.

Where dates are missing or partially missing, adverse events will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to

- A summary of the commonly occurring TEAEs, i.e., reported in at least 5% of patients in any treatment arm, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a treatment emergent AESI and the number of AESI, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting a severe TEAE (AE with grade equal to or greater than 3) and the number of severe AEs, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting an TEAE leading to infusion reaction, by treatment arm, system organ class, and preferred term
- A summary of the number and percentage of patients reporting the TEAE leading to death, by treatment arm, system organ class, and preferred term

For the summarization of number of patients, a patient will be counted only once for each system organ class and each preferred term, even if the patient reported one or more event under each subcategory; for the summarization of number of events, the number of AEs will count all AEs of each patient under each system organ class and each preferred term.

Unless specified otherwise, the table of a summary will be ordered in terms of decreasing number of patients for MedDRA system organ class and then preferred term within system organ class in the overall group.

For each AE, severity will be graded based on the National Cancer Institute-Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 5.0, and the most extreme grade recorded in eCRF will be attributed and used in the by-severity summaries. If the severity is missing, it will not be imputed. Besides, if relationship to study drug or chemotherapy (gemcitabine, cisplatin) is missing, the relationship will be imputed as Yes.

Following listings will be provided:

- A by-patient listing of all AEs
- A by-patient listing of all immune-related AEs

By-patient listings will be presented by treatment arm and will include: center, patient identifier, age, sex, race, adverse event (system organ class, preferred term, and verbatim term), AE start date, AE stop date, severity, serious (yes, no), seriousness criteria, relationship (study drug, gemcitabine, cisplatin), causality, action taken, outcome, infusion reaction (yes, no) and timepoint of AE occurrence if yes (during infusion, with 24 hours after end of infusion).

4.8.2 Deaths, Serious Adverse Events, and Other Significant Adverse Events

An SAE is any AE that meets any of the following criteria:

- Is fatal (i.e., the adverse event actually causes or leads to death).

repeat measurements are taken at a particular time point, the last valid measurement will be used in the summaries.

By-patient listings of vital signs and weight will be provided.

4.8.4 Physical Examinations

A complete physical examination, according to local practice, should be performed at the screening visit, which includes

- head, eyes, ears, nose, and throat
- cardiovascular
- dermatological
- musculoskeletal
- respiratory
- gastrointestinal
- genitourinary
- neurological systems
- cranial nerve
- Lymph node (head and neck)
- Lymph node (other than head and neck)

At subsequent visits, only limited, symptom-directed physical examinations will be performed.

Following summaries will be provided:

- A summary of the number and percentage of patients for the complete physical examination at the screening visit will be provided by category (normal, abnormal and clinically significant, abnormal but not clinically significant), body system, and treatment arm.
- A summary of the number and percentage of patients for any abnormal findings in the subsequent visits will be provided by category (abnormal and clinically significant, abnormal but not clinically significant), visit, and treatment arm.

By-patient listings of physical examination will be provided.

4.8.5 Laboratory Evaluation

Samples for the following laboratory tests will be sent to the study site's local laboratory for analysis:

- Hematology (CBC, including RBC count, hemoglobin, hematocrit, WBC count with differential [neutrophils, eosinophils, lymphocytes, monocytes, basophils, and other cells], and platelet count)
- Serum chemistries (glucose, BUN or urea, creatinine, sodium, potassium, magnesium, chloride, bicarbonate, calcium, phosphorus, total bilirubin, ALT, AST, alkaline phosphatase, LDH, total protein, and albumin)

- A summary of the number and percentage of patients for each ECOG performance status score by visit, and treatment arm.
- A shift summary of baseline score by maximum post-baseline score by treatment arm

A by-patient listing of ECOG performance status at each visit will be provided.

4.8.8 Pregnancy Test

A by-patient list of pregnancy test results will be provided only for female patients.

4.9 Other Evaluation

4.9.1 Sensitivity Analyses for IRC-Assessed PFS

The following analyses will be performed as the sensitivity analyses for the primary analysis of investigator-assessed PFS:

- The unstratified analysis in which the log-rank test will be performed at two-sided alpha level of 0.05, the HR will be estimated using the proportional hazards model. Efron's method will be used to handle ties. The 95% CI for the HR will be constructed.
- The analysis using the methods outlined for the primary analysis in which the stratification factors are the actual variables recorded in CRF rather than the stratification factors recorded in the IWRS.
- The analysis using the methods outlined for the primary analysis in which patients did not experience disease progression or death but started new anti-tumor therapy whose event date is defined as the start date of new treatment.
- The analysis using the methods outlined for the primary analysis in which patients with disease progression or death being documented after missing two consecutive tumor assessments will be handled as event and will not be censored.
- The analysis using the methods outlined for the primary analysis. Regardless patients with disease progression or death being documented after missing two consecutive tumor assessments due to COVID-19, PFS will be censored at the last tumor assessment before the missing, or at the randomization date if there was no post-baseline tumor assessment before the missing.
- The analysis using the methods outlined for the primary analysis in which the analysis population is the PPS.

4.9.2 PFS, ORR, DoR, and DCR per Immune-Related RECIST

Analyses using irRECIST for PFS, ORR, DoR and DCR as determined by the IRC and investigator will be conducted. The applicable methods outlined for PFS and ORR will be used for these analyses.

The ORR and its 95% CI will be calculated, and its 95% CI will be constructed using the Clopper-Pearson method (Clopper and Pearson, 1934).

months after the first patient was randomized, while the exact timing of the final analysis will depend on the actual occurrence of PFS events.

To control the type I error for PFS analyses at a two-sided significance level of 0.05, the stopping boundaries for the PFS interim and final analyses were computed with use of the Lan-DeMets approximation to the O'Brien-Fleming boundary as shown in the following table:

Timing and Stopping Boundary of PFS Analyses

Type of Analysis	Timing Since FSI (month)	Planned Information Fraction (Event #)	Stopping Boundary (Two-Sided p-Value)
PFS interim analysis	18	65% (130)	0.011
PFS final analysis	25	100% (200)	0.047

FSI = First subject in; PFS = progression-free survival; HR = hazard ratio.

The planned interim analysis was performed with the data cutoff date of 30 May 2020, when 128 PFS events, as assessed by IRC per RECIST 1.1, had been observed in the ITT population. The PFS result generated at the interim analysis was highly statistically significant and was treated as definitive. The final PFS analysis will be conducted 18 months after the last patient was randomized, when it is estimated that approximately 150 IRC-assessed PFS events will be observed in the ITT population. The final PFS analysis will be performed for the descriptive purposes only.

4.10.2 OS

Two interim analyses of OS were planned at the interim and final analyses of PFS for the descriptive purpose only, where it was expected to observe approximately 49 and 74 deaths respectively. The final analysis of OS will be performed when approximately 130 deaths have been observed in the ITT population.

4.11 Determination of Sample Size

The sample size calculation is based on the primary endpoint PFS. Patients will be randomized in a 1:1 ratio. A total of 280 patients (140 per arm) are needed to observe 200 PFS events at approximately 25 months after the first patient is randomized in order to detect the PFS improvement of HR=0.67 with 80% power at a 2-sided significance level of 0.05; one interim analysis is planned when approximately 130 PFS events are observed.

The calculation of the sample size is based on the following assumptions:

- PFS is exponentially distributed
- The median PFS is 7 months for the standard chemotherapy
- The interim and final analyses of PFS use the Lan DeMets alpha spending function to approximate the O'Brien Fleming boundary
- The recruitment of 280 patients will take place over 14 months,

Nutritional supplements	HNNU	1	1	32 (CRF 62)
Feeding tube	HNFE	1	1	33 (CRF 63)
Weight loss	HNWL	1	1	34 (CRF 64)
Weight gain	HNWG	1	1	35 (CRF 65)