Statistical Analysis Plan

Protocol Title:	A Phase 2b, open-label, single-arm study of zanidatamab (ZW25) monotherapy in subjects with advanced or metastatic HER2-amplified biliary tract cancers
Protocol Number:	ZWI-ZW25-203
EudraCT Number:	2020-000459-11
Phase:	2b
Investigational Product:	Zanidatamab or ZW25
US IND Number:	[redacted]
Sponsor (Rest of World):	Zymeworks Inc. 114 East 4th Avenue, Suite 800 Vancouver, BC, Canada V5T 1G4

Sponsor (China and South Korea):	BeiGene, Ltd. c/o BeiGene USA, Inc. 2955 Campus Drive, Suite 200 San Mateo, CA 94403 USA	BeiGene (Beijing) Co., Ltd, Room 101, 201, 402, 502, Building 1, No. 30 Science Park Road, Changping District, 102206, Beijing, China

Version - Date:	Version 2.0 – 28-October-2022
Author:	[redacted]

Approval

Zymeworks:

BeiGene:

1 TABLE OF CONTENTS

1 TABLE OF CONTENTS 4
2 LIST OF FIGURES 5
3 LIST OF TABLES 5
4 LIST OF ABBREVIATIONS 6
5 INTRODUCTION 8
6 STUDY DESIGN 8
6.1 Description 8
6.2 Study Objectives and Endpoints 9
6.3 Treatment Assignment and Blinding 12
6.4 Statistical Hypotheses 12
6.5 Determination of Sample Size 12
6.6 Timing of Analyses and Interim Analyses 14
7 STATISTICAL CONSIDERATIONS 14
7.1 Analysis Sets 14
7.2 Subgroups 15
7.3 Handling of Missing Data 16
7.4 Multicenter Studies 18
7.5 Adjustments for Covariates 18
7.6 Multiple Comparisons, Multiplicity, and Multiple Testing 18
7.7 Data Conventions, Definitions, and Formulas 19
8 STATISTICAL METHODOLOGY 20
8.1 Trial Details 20
8.2 Analysis of Efficacy 24
8.3 Analysis of Safety 29
8.4 Quality of Life, Disease-Related Pain Analyses and Opioid Use 36
8.5 Immunogenicity Analyses 38
8.6 Pharmacokinetic Analyses 38
8.7 Biomarker Analyses 41
8.8 Exploratory Analyses 41
8.9 Subgroup Analysis of Subjects Enrolled from China sites 41
9 CHANGES IN THE PLANNED ANALYSIS 43
10 REFERENCES 44

2 LIST OF FIGURES

Figure 1: Study Design 9

3 LIST OF TABLES

Table 1: Clopper-Pearson 95% Confidence Intervals for N=75 Subjects 12

Table 2: Clopper-Pearson 95% Confidence Intervals for N=25 Subjects 13

Table 3: Probabilities of Detecting ≥1 Adverse Events 13

Table 4: Analysis Sets 14

Table 5: Determination of Confirmed Best Overall Response 25

Table 6: Censoring and Event Scheme for PFS 26

Table 7: Censoring and Event Scheme for OS 29

Table 9: Toxicity Terms for Laboratory Parameters Graded for both High and Low values 34

Table 10: PK Sampling time points for Zanidatamab 38

Table 11: Steady-State Extensive Pharmacokinetic Sampling Schedule on Cycle 4 or any Subsequent Cycle 39

Table 12: Analysis Sets for Efficacy Analysis for Subjects Enrolled from China sites 42

4 LIST OF ABBREVIATIONS

Abbreviation	Definition
λ_z	terminal elimination rate constant
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ANOVA	analysis of variance
AST	aspartate aminotransferase
$AUC_{0-\infty}$	area under the serum concentration-time curve from zero to infinity
AUC_{0-t}	area under the serum concentration-time curve from zero to the last measurable concentration
AUC_{tau}	area under the serum concentration-time curve from zero to the end of dosing interval
BOR	best overall response
BTC	biliary tract cancer
C_{ave}	average concentration
CC	cholangiocarcinoma
CDE	Center for Drug Evaluation (China)
CFR	Code of Federal Regulations
CI	confidence interval
C_{max}	maximum concentration
C_{min}	minimum concentration
COVID-19	Coronavirus Disease 2019
CR	complete response
CRF	case report form
C_{SS}	steady state concentration
CT	computed tomography
ctDNA	circulating tumor DNA
DCR	disease control rate
DOR	duration of response
ECC	extrahepatic cholangiocarcinoma
ECD	extracellular domain
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	end of treatment
EQ-5D	European Quality of Life 5-Dimensions questionnaire
FDA	Food and Drug Administration
HER2	human epidermal growth factor receptor 2
ICC	intrahepatic cholangiocarcinoma
ICH	International Council for Harmonisation
ICR	independent central review
IDMC	independent data monitoring committee
IHC	immunohistochemistry
IND	investigational new drug
ISH	in situ hybridization
LVEF	left ventricular ejection fraction

MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
MUGA	multigated acquisition scan
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NMPA	National Medical Products Administration
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
PRO	patient-reported outcome
PS	performance status
QxW	every ‘x’ weeks
QOL	quality of life
QTcF	Corrected QT interval using Fridericia formula
RECIST 1.1	Response Evaluation Criteria in Solid Tumors, version 1.1
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SMQ	Standardised MedDRA Query
SOC	system organ class
TEAE	treatment-emergent adverse event
$t_{1/2}$	half-life
t_{max}	time to maximum concentration
ULN	upper limit of normal
US	United States
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

5 INTRODUCTION

This Statistical Analysis Plan (SAP) describes the statistical methods, assumptions, and details of the planned statistical analyses to be performed for Study ZW25-203.

6 STUDY DESIGN

6.1 Description

This pivotal, multicenter, open-label, single-arm trial will evaluate the anti-tumor activity of zanidatamab monotherapy in subjects with HER2-amplified, inoperable and advanced or metastatic biliary tract cancer (BTC), including intrahepatic cholangiocarcinoma (ICC), extrahepatic cholangiocarcinoma (ECC), and gallbladder cancer (GBC). Subjects must have received at least 1 prior gemcitabine-containing systemic chemotherapy regimen for advanced disease, and experienced disease progression after or developed intolerance to the most recent prior therapy. New or archival tumor tissue is required from all subjects for HER2 amplification and HER2 protein expression testing at a central lab using in situ hybridization (ISH) and immunohistochemistry (IHC) assays. Subjects may be tested for HER2 status at the central lab any time after diagnosis of advanced or metastatic disease and before study enrollment. Subjects who elect to be pre-screened for HER2 status must sign a separate informed consent for collection, storage, and analysis of the tumor tissue. Once a prospective subject experiences disease progression or develops intolerance to his/her most recent prior therapy, the main consent for the remaining screening assessments must be signed.

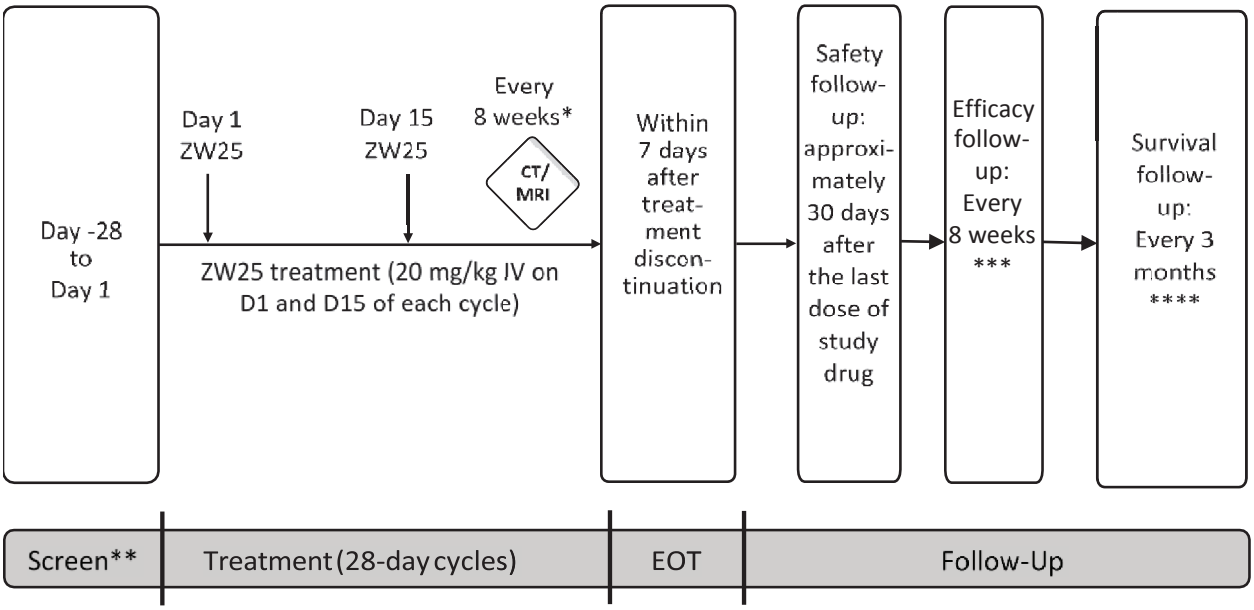
Two cohorts of subjects will be enrolled:

- Cohort 1, comprising subjects with HER2 amplification by ISH and HER2 overexpression by IHC, i.e., IHC 2+ or 3+
- Cohort 2, comprising subjects with HER2 amplification by ISH and HER2 IHC 0 or 1+

Enrolled subjects will receive zanidatamab, 20 mg/kg given intravenously (IV) every 2 weeks (Q2W) until 1 of the protocol-defined treatment discontinuation criteria are met. The primary endpoint of confirmed objective response rate (ORR) per the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) will be evaluated by independent central review (ICR). Disease response will be evaluated every 8 weeks (Q8W) with a ± 7-day window, using computed tomography (CT) or magnetic resonance imaging (MRI) scans.

Subjects who discontinue treatment with zanidatamab for any reason should enter the 30-day follow-up period, as well as survival follow-up. The study schema is presented in Figure 1.

Figure 1: Study Design



CT = computed tomography; EOT = end of treatment; IV = intravenous; MRI =magnetic resonance imaging; ZW25 = zanidatamab
* Timed from Cycle 1 Day 1.
** Subjects may be tested for HER2 status any time after diagnosis of advanced or metastatic disease and before study enrollment. Subjects who elect to be pre-screened for HER2 status must sign a separate informed consent for collection, storage, and analysis of the tumor tissue.
*** Every 8 weeks until disease progression or start of subsequent anticancer therapy.

**** Every 3 months until death, lost to follow-up, withdrawal of consent, study completion, or study termination by sponsor.

6.2 Study Objectives and Endpoints

Objectives	Endpoints
Primary: • To evaluate the anti-tumor activity of zanidatamab monotherapy in subjects with advanced or metastatic HER2-amplified biliary tract cancers (BTC)	• Confirmed objective response rate (ORR) by the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1), assessed by independent central review (ICR)

Objectives	Endpoints
Secondary:	
• To further evaluate the anti-tumor activity of zanidatamab monotherapy in subjects with advanced or metastatic HER2-amplified BTC	• Duration of response (DOR) by RECIST 1.1 assessed by ICR. • Proportion of subjects with a DOR ≥16 weeks by RECIST 1.1 assessed by ICR. • Disease control rate (DCR) by RECIST 1.1 assessed by ICR. • Progression-free survival (PFS) by RECIST 1.1 assessed by ICR. • ORR by RECIST 1.1 assessed by investigator. • DOR by RECIST 1.1 assessed by investigator. • Proportion of subjects with a DOR ≥16 weeks by RECIST 1.1 assessed by investigator. • DCR by RECIST 1.1 assessed by investigator. • PFS by RECIST 1.1 assessed by investigator. • Overall survival (OS)
• To evaluate the safety and tolerability of zanidatamab monotherapy in subjects with advanced or metastatic HER2-amplified BTC	• Frequency and severity of adverse events (AEs) • Frequency of serious adverse events (SAEs) and deaths • Frequency and severity of clinical laboratory abnormalities • Frequency of dose modifications of zanidatamab

Objectives	Endpoints
• To evaluate the pharmacokinetics (PK) of zanidatamab	• Serum concentrations of zanidatamab as a function of time post-dosing • PK parameters for single (first) dose and multiple doses • Population pharmacokinetics and exposure-response analyses. To be conducted according to the data analysis plan entitled "Population Pharmacokinetic and Exposure-Response Analysis to Support Regulatory Biologics License Application Submission of Study ZWI- ZW25- 203 in Biliary Tract Cancer"
• To evaluate the immunogenicity of zanidatamab	• Frequency, duration, and time of onset of anti-drug antibodies (ADA) and neutralizing antibodies, if applicable
CCI [redacted]	[redacted]

6.3 Treatment Assignment and Blinding

This is a single-arm study that will enroll subjects into 2 separate cohorts, based on their HER2 status. Prior to the primary analysis for the study, access to the ICR response assessment data will be limited only to a small team as defined in the Data Access Plan for data cleaning purposes to limit potential bias.

6.4 Statistical Hypotheses

This is a single-arm, open-label study. No statistical hypotheses will be tested.

6.5 Determination of Sample Size

This study will enroll approximately 100 subjects: approximately 75 subjects in Cohort 1 and approximately 25 subjects in Cohort 2. No formal sample size calculations were performed. Approximately 75 subjects will be enrolled in Cohort 1 to evaluate the primary endpoint of confirmed ORR for the primary analysis of efficacy. The 2-sided, exact Clopper Pearson binomial 95% confidence intervals (CIs) that could be observed with this sample size are listed in Table 1. A minimum of 23 subjects with confirmed objective response from a sample-size of 75 would result in the lower bound of a two-sided, exact Clopper Pearson binomial 95% CI exceeding 20 percentage points.

Table 1: Clopper-Pearson 95% Confidence Intervals for N=75 Subjects

Confirmed ORR		95% CI	
Percent	**Responses**	**Lower**	**Upper**
20.0%	15	11.7%	30.8%
25.3%	19	16.0%	36.7%
30.7%	23	20.5%	42.4%
36.0%	27	25.2%	47.9%
40.0%	30	28.9%	52.0%
45.3%	34	33.8%	57.3%
50.7%	38	38.9%	62.4%

Approximately 25 subjects will be enrolled in Cohort 2 to evaluate the primary endpoint of confirmed ORR as a secondary analysis of efficacy. The 2-sided, exact Clopper-Pearson binomial 95% CIs that could be observed with a sample size of 25 subjects are listed in Table 2.

Table 2: Clopper-Pearson 95% Confidence Intervals for N=25 Subjects

Confirmed ORR		95% CI	
Percent	**Responses**	**Lower**	**Upper**
20.0%	5	6.8%	40.7%
24.0%	6	9.4%	45.1%
28.0%	7	12.1%	49.4%
32.0%	8	15.0%	53.5%
36.0%	9	18.0%	57.5%
40.0%	10	21.1%	61.3%
44.0%	11	24.4%	65.1%
48.0%	12	27.8%	68.7%

With a sample size of 100 subjects (Cohorts 1 and 2 combined), the probabilities of detecting 1 or more treatment-emergent adverse events (TEAEs) for various frequencies of occurrence in this population are listed in Table 3.

Table 3: Probabilities of Detecting ≥1 Adverse Events

Frequency of Adverse Event	Binomial Probability
0.1%	10%
1%	63%
5%	99%

6.6 Timing of Analyses and Interim Analyses

No interim analyses of the efficacy will be performed for this study. The primary analysis of efficacy will be performed approximately 6 months after the last subject has been enrolled into Cohort 1.

Safety and study conduct will be monitored throughout the study by an independent data monitoring committee (IDMC). The committee is tasked with monitoring the safety of participants in this study through regular and/or ad hoc meetings. The first IDMC meeting will be held after the first 25 subjects have received at least 1 dose of study treatment; subsequent meetings will be held quarterly to review cumulative safety data to identify any potential new safety signals. Additional IDMC meetings will be held at least 28 days after the tenth subject enrolled in China has received 1 dose of zanidatamab to review safety. The IDMC will be responsible for making recommendations to the sponsor as to appropriate study direction; further details will be provided in the IDMC charter.

7 STATISTICAL CONSIDERATIONS

7.1 Analysis Sets

For the purposes of analysis, the subject analysis sets are defined in Table 4.

Table 4: Analysis Sets

Analysis Set	Description
Screened	All participants who sign the main study informed consent
Enrolled	All participants who sign the main study informed consent, were eligible for the study, and received approval to enroll at the end of Screening.
Safety	All safety evaluable subjects enrolled in Cohorts 1 and 2. A subject is considered safety evaluable if they received any amount of zanidatamab.
Efficacy	The following Efficacy Analysis Sets will be evaluated for this study: • Cohort 1 (Primary) - Includes all safety evaluable subjects enrolled in Cohort 1. The primary analysis of efficacy will be performed using this analysis set. • Cohort 2 - Includes all safety evaluable subjects enrolled in Cohort 2.
Response Evaluable	The following Response Evaluable Analysis Sets will be evaluated for this study: • Cohort 1 - Includes all response evaluable subjects enrolled in Cohort 1 • Cohort 2 - Includes all response evaluable subjects enrolled in Cohort 2 Response Evaluable includes all subjects in the safety analysis set with measurable disease at baseline and at least 1 evaluable post-baseline disease assessment (per RECIST 1.1) or who discontinued study treatment due to death or unequivocal clinical progression.
Pharmacokinetics	All subjects who received any amount of zanidatamab and have at least 1 post-baseline PK assessment.

Immunogenicity	All subjects who received any amount of zanidatamab and have both baseline ADA and at least 1 post-baseline ADA results available

7.2 Subgroups

The following subgroups may be evaluated for efficacy. Subgroups with 5 or fewer subjects may not be evaluated.

- Disease subtype
  - Gallbladder cancer (GBC)
  - Intrahepatic cholangiocarcinoma (ICC)
  - Extrahepatic cholangiocarcinoma (ECC)
- . Intolerance to the most recent prior therapy
  - Yes
  - No (progressed on the most recent regimen)
- Number of prior regimens for treatment of metastatic disease
  - Less than 2 regimens
  - 2 or more regimens
- HER2 expression
  Cohort 1
    - IHC3+
    - IHC2+
    - IHC Result Not available

  Cohort 2
    - IHC1+
    - IHC0
- Geographic region
  - North America (Canada, US)
  - Asia (China, South Korea)
  - Other (Chile, France, Italy, Spain, UK)
- Sex (Female, Male)
- Age
  - <65 years
  - ≥65 years
  - <75 years
  - ≥ 75 years
- Baseline ECOG
  - ECOG = 0
  - ECOG = 1

- Disease stage at baseline (study entry)
  - Stage IIb and Stage III
  - Stage IV
- Race
  - Asian
  - Non-Asian

7.3 Handling of Missing Data

7.3.1 Cohort Assignment

Subjects with missing or indeterminate IHC results will be assigned to Cohort 1. The number and proportion of subjects with missing IHC results will be presented.

7.3.2 Efficacy Data

Partial or Missing Dates

Partial or missing death dates, tumor assessment dates, dates of last contact, dates of onset of post-study treatment anti-cancer therapy, and dates used in computing secondary efficacy endpoints will be imputed as follows:

- Missing day only: For partial dates with only the day of the month missing, the day will be imputed as the 15th day of the month, provided that a preceding date of interest does not occur on/after the 15th day of the month. If the preceding date does occur on/after the 15th day in the same month, then the missing day will be imputed as half the distance between the preceding date and the end of the month.
- Missing month and day: For partial dates with the month and day missing, the month and day will be imputed as July 1st, provided that a preceding date of interest does not occur in the same year. If the preceding date occurs in the same year, then the month and day will be imputed as half the distance from the preceding date to the end of the known year.
- Missing month, day, and year: Where appropriate, specific rules for handling completely missing dates are described in the applicable endpoint and analysis section.

In situations where the above rules result in an illogical time (e.g., negative study day, negative cycle day, etc.) the date will be imputed as EOT date + 1 day.

Missing Post-Baseline Assessments

For RECIST based endpoints, subjects without any post-baseline tumor assessments will be handled as follows:

- For the analysis of PFS, subjects will be handled as described in Table 6.
- Subjects with no evaluable post-baseline response assessments will be considered as Not Evaluable (NE) for best overall response.

7.3.3 Quality of Life Data

There will be no imputation of missing data for either the EQ-5D or BPI.

7.3.4 Safety Data

7.3.4.1 Adverse Event Start Dates

For AEs where the date of onset is during or after administration of the first dose of study treatment, missing or partial start dates will be imputed as the earliest possible date that is on or after the date of the first dose of study treatment (zanidatamab). For events indicated on the CRF as occurring prior to first dose of study treatment, missing or partial dates will not be imputed.

- Missing day only: If the start date and date of first dose share the same month and year, the missing start day will be imputed as the day of first dose. If the start date month is after the month of first dose, day will be imputed as the 1st (i.e., 01-MMM-YY).
- Missing month and day: If the start date and date of first dose share the same year, day and month will be imputed as the day and month of first dose. If the start date year is after the year of first dose, the month and day will be imputed as January 1st (i.e., 01-JAN-YY).
- Missing month, day, and year: Missing start dates will be imputed as the date of first dose.

7.3.4.2 Adverse Event End Dates

Missing or partially missing end dates for AE which are known to have resolved will be imputed as follows:

- Missing day only: The day will be imputed as last day of the month (e.g., 31-MMM-YY or 30-MMM-YY). The AE end date is then imputed as the minimum of {last day of the month} and {death date}. If the month and year are the same as the data cutoff date month and year, then the data cutoff date will be used as the imputed AE end date.

- Missing month and day: The missing month and missing day will be imputed as December 31. The AE end date is then imputed as the minimum of {Dec 31 of the year}, {death date}, {End of Study date} and {data cutoff date}.
- Missing month, day, and year: The AE end date will be imputed as the minimum of {death date}, {End of Study date} and {data cutoff date}.

7.3.5 Pharmacokinetic Data

Missing values for PK measurements will be listed as missing and excluded from the calculation of summary statistics. If a PK result is available but the collection time is missing or incomplete, then the nominal collection time indicated in the case report form (CRF) will be used. Sample concentrations that are below the lower limit of quantification (LLOQ) will be shown as less than reportable (<LLOQ) in the individual subject listings and imputed as follows for the purposes of calculating summary statistics and producing graphical summaries:

- If the sample was collected prior to the 1st dose of zanidatamab, then the value will be imputed as zero.
- If the sample was collected after the 1st dose of zanidatamab, then the value will be imputed as the LLOQ.

Exclusion of PK data will be based on documented errors in dosing, sample labeling, sample handling, or bioanalysis. Excluded data will not be used in calculations for any tabulated or graphical summaries. Subjects that were not dosed or were not dosed in accordance with their assigned dose level, will be excluded from any tabulated or graphical summaries. Biologically implausible data may also be excluded in the summaries; however, all individual data will be listed and any exclusion, along with the corresponding rationale, will be noted.

7.4 Multicenter Studies

The primary set of analyses will be conducted using pooled data from all sites. There are multiple sites in this study, however it is not anticipated that any center will accrue enough subjects to warrant an analysis by center.

7.5 Adjustments for Covariates

No adjustments for covariates will be performed.

7.6 Multiple Comparisons, Multiplicity, and Multiple Testing

No adjustments for multiple comparisons will be performed.

7.7 Data Conventions, Definitions, and Formulas

The following data conventions will be used for the tables, listings, and figures.

- **Study drug:** Zanidatamab or ZW25
- **Baseline value of a study assessment:** The last non-missing observation prior to the first dose of zanidatamab, unless otherwise specified. If both a planned assessment and repeat assessment (e.g., retest) meet the above criteria and were collected on the same date and time, the repeat assessment will be used as baseline.
- **Study day of a study event or assessment:** The difference between the date of the event or assessment and the date of first dose plus 1 for events or assessments occurring on or after the day of first dose. For example, if an event occurs on 08JUN2014 and the date of first dose was 06JUN2014, the study day of the event is 3. For events or assessments occurring before the day of first dose, it will be the difference between the date of the event or assessment and the date of first dose. For example, if an event occurs on 05JUN2014 and the date of first dose was 06JUN2014, the study day of the event is -1, which is consistent with the Study Data Tabulation Model (Version 3.2) from Clinical Data Interchange Standards Consortium (CDISC).
- **Pre-treatment Period:** Period of time prior to first dose of zanidatamab
- **Study Treatment Period:** Period of time that begins on the date of the first dose of zanidatamab through 30 days after the date of the final dose of zanidatamab.
- **Post-treatment Follow up Period:** Period of time that begins 1 day after EOT and ends at End of Study (EOS).
- **Disease Assessment dates:** In the event, radiographic assessment for a subject occurs over multiple days for a visit, the following convention will be used to determine the date of the RECIST disease response assessment:
  - For a response of Progressive Disease (PD), the date of the earliest radiographic scan will be used.
  - For responses other than PD, the date of the latest radiographic scan will be used.
- **Duration of zanidatamab Treatment:**

  Duration of zanidatamab treatment is defined as the difference in days between date of first dose and the earliest of the following dates:
  - Date of last dose + 14 days
  - Date of death
  - Date of subsequent anticancer therapy

  Duration (days) = (Earliest date above) – (date of first dose) + 1.

- **Total Cumulative Dose:**

  Total cumulative dose (mg) = sum of all doses (mg) received during the entire treatment period duration.

- **Actual Dose Intensity (ADI):**

  ADI is the actual dose (mg/kg) per week that the subject received over the entire treatment period. For calculating ADI, the treatment period is defined as time from the first dose of study drug to date of last dose + 14 days, regardless of death status during this period.

- **Intended Dose Intensity (IDI):**

  IDI is the intended dose of drug (mg/kg) per week

- **Relative Dose Intensity (RDI):**

  RDI is the percent of the intended dose intensity over the entire treatment period:

  $$RDI = \frac{ADI \times 100}{IDI}$$

8 STATISTICAL METHODOLOGY

8.1 Trial Details

The following information will be summarized for the safety analysis set unless otherwise specified.

8.1.1 Enrollment

The following study enrollment information will be summarized, for all screened subjects.

- Number of subjects treated and the percentage Number of screen failures and the percentage Number and percentage of subjects enrolled and treated in each region and country.
- Screen failures
  - Demographic variables: age, gender, race, country, and ethnicity.
  - Number and percentage of subjects who failed screening by reason.
  - Number and percentage by region and country
- A listing of subjects who failed screening will also be produced, with reasons for screen failure and available demographic information.

In addition, the following summaries will be provided for treated subjects.

- Number and percentage of subjects enrolled and treated in each cohort.
- .

8.1.2 Subject Disposition

The following disposition information will be summarized.

- Number and percentage of subjects who discontinued zanidatamab and the associated reasons for discontinuation.
- Number and percentage of subjects who discontinued study and the associated reasons for discontinuation from the study
- Number and percentage of subjects actively in post-treatment (survival) follow-up.
- Number of subjects in each analysis set
- Duration of study follow-up, defined as the time from first dose to the data cut off date.

8.1.3 Protocol Deviations

Protocol deviations (as defined in the ZWI-ZW25-203 Monitoring Plan) will be identified by site monitors, the medical monitor and by checks of the clinical database. Protocol deviations will be classified into prespecified sub-types and categorized as either important deviations or non-important deviations. Important deviations are deviations that are considered significant. Important deviations will be summarized for each cohort. Per the study's COVID-19 deviation guideline, all deviations due to COVID-19 are documented on source documents and tracked in EDC. The frequencies of missed tumor scans and missed infusions due to COVID-19 will provided.

8.1.4 Baseline Characteristics and Disease History

Baseline characteristics will be summarized using counts and percentages for categorical variables and summary statistics (mean, quartiles, standard deviation, and range) for continuous variables. Categorical variables listed in this section that are also identified as efficacy subgroups of interest (Section 7.2) will be summarized using the categories defined in that section. Characteristics to be summarized include the following:

- Demographic variables: age, sex, race and ethnicity (Hispanic, Non-Hispanic)
- Physical characteristics: height (cm), weight (kg)
- ECOG performance status

- Region or Country
- Disease history:
  - Primary disease (disease subtype)
    - Gallbladder cancer (GBC)
    - Intrahepatic cholangiocarcinoma (ICC)
    - Extrahepatic cholangiocarcinoma (ECC)
  - Stage at initial diagnosis (Stage I, II, …IV and including substages e.g., IIb)
  - Duration (months) from initial diagnosis to metastatic/locally advanced
  - History of brain metastases (Yes, No)
  - Number of years since brain metastases was diagnosed
  - Tumor classification/characteristics
    - HER2 IHC tumor expression score (0, 1+, 2+, 3+, No result)
  - **Investigator** assessed tumor characteristics per RECIST 1.1.
    - Sum of diameters for target lesions (mm)
  - **ICR** assessed tumor characteristics per RECIST 1.1.
    - Measurable disease status (Yes, No)
    - Sum of diameters for target lesions (mm)
  - Prior radiotherapy
    - Any (Yes, No)
    - Total number of treatments
  - Prior cancer surgery
    - Any (Yes, No)
    - Total number of surgeries
  - Prior systemic anticancer therapy
    - Any (Yes, No)
    - Total number of regimens
    - Received Gemcitabine+Cisplatin
  - Prior systemic anticancer therapy for metastatic/locally advanced disease
    - Any (Yes, No)

        - Number of regimens received
        - Number of regimens received (<2, 2+)
        - Received Gemcitabine
        - Prior exposure to:
            - PD1 / PDL1 inhibitor
            - Gemcitabine monotherapy
            - Gemcitabine and Oxaliplatin in combination
            - Gemcitabine and Cisplatin in combination
            - Gemcitabine and Fluoropyrimidine in combination
            - Gemcitabine and other combination therapies
            - Fluoropyrimidine based therapies (without Gemcitabine)
    - Baseline hepatic impairment (Per the National Cancer Institute Organ Dysfunction Working Group)
        - None
        - Mild
        - Moderate
        - Severe
    - Baseline renal impairment (Per the Cockcroft-Gault formula for estimating creatinine clearance and FDA guidance titled: Pharmacokinetics in Patients with Impaired Renal Function - Study Design, Data Analysis, and Impact on Dosing and Labeling [September 2020])
        - Mild to moderate
        - Normal

8.1.5 Prior Systemic Anticancer Therapy, Radiotherapy and Surgeries

Prior systemic anticancer therapies will be coded using the World Health Organization Drug Dictionary (WHO-DD) Version 2009Q3 or higher and listed by ATC level 4 code. Prior surgery and radiotherapies will also be listed.

8.1.6 Concomitant Medications

Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) and listed and summarized by preferred term using counts and percentages. Multiple occurrences of the same medication within a subject will be summarized only once.

8.1.7 Concomitant Procedures

Concomitant procedures will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 24.0 or higher and listed and summarized by procedure and indication using counts and percentages. Multiple occurrences of the same procedure within a subject will be summarized only once.

8.2 Analysis of Efficacy

The primary efficacy analysis will be based on the Cohort 1 Efficacy Analysis Set. In addition, a pooled analysis combining Cohorts 1 and 2 may be presented. Additional analyses will be conducted on the Response Evaluable Analysis sets as described in Table 4.

8.2.1 Best Overall Response

Best overall response (BOR) is defined as the best response observed per RECIST 1.1 from start of the study treatment until any of the following occurring for a subject:

- Receipt of non-protocol specified anticancer therapy
- Documented disease progression
- Death from any cause
- Subject withdrawal of consent

Confirmed BOR requires confirmation of CR or PR by a subsequent tumor assessment that is performed a minimum of 4 weeks from when the initial response is observed. See Table 5 for more details.

BOR and confirmed BOR will be summarized using counts and percentages for both the ICR and investigator assessments. Confirmed best overall response will be determined as illustrated in Table 5. Subjects who die and have no post baseline radiographic scans will be categorized as not evaluable (NE) for BOR and confirmed BOR (Section 7.3.2).

Concordance and discordance of ICR and investigator assessment of confirmed BOR will be summarized using counts and percentages.

Table 5: Determination of Confirmed Best Overall Response

First Time Point Response	Second Time Point Response	Confirmed Response (Best Response) *
PD	No further evaluation	PD
NE	PD	PD
CR	PD	SD or PD (1)
PR	PD	SD or PD (1)
SD	PD	SD or PD (1)
CR	CR	CR
CR	NE **	SD or NE (2)
PR	CR	PR
PR	PR	PR
PR	SD (3) **	SD
PR	NE **	SD or NE (2)
SD	CR	SD
SD	PR	SD
SD	SD	SD
SD	NE	SD or NE (2)
NE	CR	SD
NE	PR	SD
NE	SD	SD
NE	NE	NE

*** A Best Response of SD can only be made after the subject is on-study for a minimum of eight (8) weeks (56 days) ± fourteen (14) days. If the subject is not on-study for this amount of time, any tumor assessment indicating stable disease before eight (8) weeks (56 days) will have a Best Response of NE unless PD is identified.**

**** Subsequent documentation of CR may provide confirmation of a previously identified CR for subjects where the second integrated response was NE. Subsequent documentation of PR may provide confirmation of a previously identified PR for subjects where the second integrated response was NE or SD. If the third TPR confirms the CR (or PR) then the Confirmed Response will be CR (or PR). For this study, only one (1) intervening NE is allowed between CRs/PRs. For example:**
CR NE CR = CR; PR NE PR = PR. Additionally, one (1) SD is allowed between PRs (e.g., PR SD PR = PR). For this study, one (1) intervening NEs are allowed between CRs/PRs but one (1) intervening SD is allowed only between PRs. For example, CR NE CR = CR; PR NE PR = PR; CR SD CR ≠ CR; PR SD PR = PR.
Note: in the following scenario, PR SD NE PR, the second PR is not a confirmed PR.
(1) Best response will be SD if the first TPR is after eight (8) weeks (56 days) ± **fourteen (14) days** on-study. Otherwise, the best response will be PD.
(2) Best response will be SD if the first TPR is after eight (8) weeks (56 days) ± **fourteen (14) days** on-study. Otherwise, the best response will be NE.
(3) TPR is SD if the increase from the first to the second assessment does not qualify for PD.

Source Independent review charter, Version 1.0; Bioclinica: 03Aug2020

8.2.2 Objective Response and Confirmed Objective Response

Objective response is defined as achieving a BOR of CR or PR per RECIST 1.1. The proportion of subjects with an objective response and the corresponding two-sided, exact Clopper Pearson binomial 95% CI will be calculated for both the ICR and investigator assessments of disease response.

Confirmed objective response is defined as achieving a confirmed BOR of CR or PR per RECIST 1.1. The primary endpoint for this study is confirmed ORR based on the ICR response assessments. A secondary endpoint of confirmed ORR will also be computed based on the investigator response assessments. For each endpoint, the proportion of subjects with a

confirmed objective response and the corresponding two-sided, exact Clopper Pearson binomial 95% CI using the Clopper-Pearson method (Clopper 1934) will be calculated.

The time of first confirmed objective response is defined as the time from the first dose of study treatment to the earliest date a subject had a confirmed objective response (CR or PR). This will be summarized for subjects who had a confirmed CR or PR as follows:

- Descriptive statistics (n, mean, median, standard deviation, quartiles, range) of the time in weeks
- Cumulative percentage responding by Week 9, Week 17, Week 25, Week 33, Week 41, and after Week 41

Concordance and discordance of confirmed objective response per ICR and per investigator assessment will be summarized using counts and percentages.

8.2.3 Progression-Free Survival (PFS)

PFS is defined as the time from the first dose of study treatment to the date of documented disease progression (per RECIST 1.1) or death from any cause, whichever occurs first. Subjects who are alive and have not progressed at the time of the analysis will be censored at the time of their last tumor assessment that was a CR, PR, SD, or non-CR/non-PD. Details of the censoring scheme for this endpoint are described in Table 6.

Kaplan-Meier plots and estimates of the quartiles and their corresponding two-sided 95% CI will be computed using the Brookmeyer and Crowley method (Brookmeyer and Crowley 1982) with log-log transformation (Collett 1994). The proportion of subjects with progression free survival at defined timepoints e.g., at 3, 6, 9 and 12 months will also be provided. Two-sided 95% CIs for these landmark PFS estimates will be based on the Greenwood estimator (Greenwood 1926).

Table 6: Censoring and Event Scheme for PFS

Scenario	Progression/Censor Date	Outcome
No baseline or no post-baseline response assessments and no death	Date of first dose	Censored
No disease progression (PD)	Date of last CR, PR, SD, or non-CR/non-PD	Censored
New anti-cancer treatment started before first PD	Date of last CR, PR, SD, or non-CR/non-PD on or prior to date of new anti-cancer treatment	Censored
Progressive disease (PD)	Date of PD, if response assessment prior to PD was either CR, PR, SD, or non-CR/non-PD	Progressed
	Date of last CR, PR, SD, or non-CR/non-PD, if PD occurred after 2 or more consecutive missed and/or NE overall response assessments*	Censored
Death	Death date, if last response assessment prior to death was CR, PR, SD, or non-CR/non-PD	Progressed

	and death occurred ≤18 weeks from the last response assessment	
	Death date, if there are no baseline or no post-baseline response assessments and death occurred ≤18 weeks from the first dose.	Progressed
	Date of last CR, PR, SD, or non-CR/non-PD, if death occurred after 2 or more consecutive missed and/or NE overall response assessments*	Censored
	Date of first dose, if there are no baseline or no post-baseline response assessments and death occurred >18 weeks from the first dose	Censored

Note: * Two consecutive post-baseline tumor assessments refers to the next two protocol scheduled tumor assessments. Time is measured from the last adequate response assessment date. Subject will be considered to have missed two consecutive scheduled visits if no scans, scheduled or unscheduled, have occurred within the protocol-mandated disease assessment schedule; this will be within 2*(8 + 1) = 18 weeks. Note: The scanning schedule is every 8 weeks.

Partial or missing dates of death dates, dates of last contact, and tumor assessment dates will be imputed as described in Section 7.3.2. Separate PFS analyses will be presented for the ICR and investigator tumor assessments.

8.2.3.1 PFS Sensitivity Analyses

The following sensitivity analyses may be performed to assess the robustness of the estimates of PFS per ICR using the same statistical methods described in Section 8.2.3 for the analysis of PFS.

- An analysis where clinical progression is treated as an event in addition to radiographic progression and death. Censoring rules per Table 6 and as stated in the below table will be followed..

Clinical progression	Clinical progression date, if last response assessment prior to clinical progression is CR, PR, SD, or non-CR/non-PD and clinical progression occurred ⩽18 weeks from the last response assessment	Progressed
	Clinical progression date if there are no post-baseline, response assessments and clinical progression occurred ⩽18 weeks from the first dose.	Progressed
	Date of last CR, PR, SD, or non-CR/non-PD, if clinical progression occurred after 2 or more consecutive missed and/or NE overall response assessments	Censored
	Date of first dose, if there are no baseline or no post-baseline response assessments and	Censored

	clinical progression occurred >18 weeks from the first dose

- An analysis in which subjects who initiate a new therapy prior to experiencing disease progression are considered to have had an event (PD) at the time of new therapy.
- An analysis where subjects who die or progress after 2 or more consecutive missed or non-evaluable tumor assessments are considered to have had an event (PD) on the date of the first missed visit.

8.2.4 Disease Control Rate (DCR)

Disease control is defined as a BOR of SD or confirmed CR or PR per RECIST 1.1. The proportion of subjects who achieve disease control and the corresponding two-sided, exact Clopper Pearson binomial 95% CI will be estimated for both the ICR and investigator assessments.

8.2.5 Clinical Benefit Rate

Clinical benefit is defined as achieving SD for ≥ 24 weeks or a confirmed, BOR of CR or PR per RECIST 1.1. The proportion of subjects with clinical benefit and the corresponding two-sided, exact Clopper Pearson binomial 95% confidence interval will be calculated for both the ICR and investigator assessments.

8.2.6 Duration of Response

DOR is defined as the time from the first confirmed objective response (CR or PR) to documented PD per RECIST 1.1 or death from any cause. Subjects who are alive and have not progressed at the time of the analysis will be censored at the time of their last tumor assessment that was a CR, PR, SD, or non-CR/non-PD. The same censoring rules outlined in Table 6 for PFS will be applied to DOR with date of first dose replaced by date of first response. DOR will be calculated based on the ICR response assessments and separately for the investigator response assessments. Only subjects who achieve a confirmed objective response will be included in the analysis of each endpoint.

DOR will be analyzed using Kaplan-Meier methodology and a Kaplan-Meier plot of DOR will also be provided. Kaplan-Meier estimates of the quartiles (median, 25th and 75th) will be computed. Two-sided 95% CI for the quartiles using the Brookmeyer and Crowley method (Brookmeyer and Crowley 1982) with log-log transformation (Collett 1994) will be presented. Tumor assessment dates will be imputed as described in Section 7.3. In the event there are too

few confirmed responses to generate Kaplan-Meier estimates, DOR may be summarized descriptively (i.e., mean, median, standard deviation, minimum, and maximum).

The proportion of subjects with a DOR ≥ 16 weeks and the corresponding two-sided, exact Clopper Pearson binomial 95% CI will also be calculated. In addition, the Kaplan-Meier probability and corresponding 2-sided 95% CI at Week 16 will be calculated.

8.2.7 Overall Survival (OS)

Overall survival (OS) is defined as the time from the first dose of study treatment until the date of death from any cause or date last known alive for those who did not die.

Specifically,

OS = Date of death or date last known alive – Date of first dose + 1

Subjects alive at the time of analysis will be considered as censored on the date the subject was last known to be alive (i.e., date of last contact). Subjects lacking data beyond the day of first dose will have their survival time censored on the date of first dose (i.e., OS duration of 1 day). Partial or missing dates of death dates and last known alive date will be imputed as described in Section 7.3.2. If resulting imputed death date is on or before last known alive date, then death of death will be set to last known alive date + 1 day. Details of the censoring scheme are described below in Table 7.

Table 7: Censoring and Event Scheme for OS

Scenario	Progression/Censor Date	Outcome
Not known to have died by data cutoff date	Date last known alive	Censored
Death	Death date	Death

Completely imputed dates (e.g., the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last known alive date.

Kaplan Meier estimates of the survival curve and quartiles will be calculated. The 2-sided 95% CIs for the quartiles will be calculated using the Brookmeyer and Crowley method (Brookmeyer and Crowley 1982) with log-log transformation (Collett 1994). Additionally, the probability of survival at timepoints from 3 months to the end of the follow-up period will be reported.

8.3 Analysis of Safety

All analyses of safety will be based on the safety analysis set which includes all safety evaluable Cohort 1 and 2 subjects.

8.3.1 Extent of Exposure

Zanidatamab exposure will be summarized using counts and percentages for categorical variables and summary statistics (mean, median, standard deviation, and range) for continuous variables. The following variables will be summarized.

- Duration of exposure (see Section 7.7 for definition)
- Number of treatment cycles initiated,
- Total number of infusions received,
- Relative dose intensity percent (see Section 7.7 for definition)
- Total cumulative dose received (see Section 7.7 for definition)
- Frequency of missed infusions and associated reasons
- Frequency of dose reductions and associated reasons
- Frequency of interrupted infusions and associated reasons
- Frequency of delayed infusions and reasons

8.3.2 Adverse Events

Definitions

An adverse event (AE) is defined as any untoward medical occurrence in a clinical study subject administered a medicinal product which does not necessarily have a causal relationship with this treatment. All AEs will be coded to standard "preferred terms" and system organ classifications (SOC) using MedDRA. Severity will be graded by study investigators using NCI-CTCAE v5.0.

TreatmentEmergent Adverse Events (TEAEs) are defined as AEs with an onset during or after receipt of the first dose of zanidatamab and up to and including 30 days after the last dose.

Summaries

Adverse events will be summarized by preferred term (and SOC, or severity grade where appropriate). Multiple occurrences of the same AE within a subject will be summarized only once at the most severe grade level for the time frame under consideration. For summaries by severity, only the worst grade for an AE will be counted for a particular subject. AEs occurring prior to the first dose of zanidatamab or more than 30 days after the last dose of zanidatamab will be excluded from summaries but included in data listings.

The following summaries will be produced:

- Overall Summary of Adverse Events – Subject Incidence

- Overall Summary of Adverse Events – Event Frequency
- Incidence of Treatment Emergent Adverse Events
- Incidence of Grade 3 and Higher Treatment Emergent Adverse Events
- Incidence of Zanidatamab Related Treatment Emergent Adverse Events
- Incidence of Zanidatamab Related Grade 3 and Higher Treatment Emergent Adverse Events
- Incidence of Treatment Emergent Adverse Events Leading to Discontinuation of zanidatamab
- Incidence of Zanidatamab Related Treatment Emergent Adverse Events Leading to Discontinuation of Zanidatamab
- Incidence of Treatment Emergent Adverse Events Leading to Dose Reduction
- Incidence of Zanidatamab Related Treatment Emergent Adverse Events Leading to Dose Reduction
- Incidence of TEAEs Leading to Dose Delay of Zanidatamab
- Incidence of TEAEs Leading to Dose Interruption of Zanidatamab
- Incidence of Grade 5 TEAE

In addition to summary tables, the following listings will be produced.

- All TEAEs
- All non TEAEs
- AEs Resulting in Death
- AEs Leading to Discontinuation of Zanidatamab
- AEs Leading to Dose Reduction of Zanidatamab
- AEs Leading to Dose Delay of Zanidatamab
- AEs Leading to Infusion Interruption

8.3.3 Serious Adverse Events and Deaths

Serious adverse events (SAEs) will be summarized by preferred term and SOC using counts and percentages. Multiple occurrences of the same SAE within a subject will be counted only once for the timeframe under consideration. The following summaries of SAEs will be produced.

- Incidence by decreasing frequency of preferred term
- Incidence by SOC and decreasing frequency of SOC and preferred term
- Incidence of zanidatamab related SAEs by decreasing frequency of preferred term

In addition to summary tables, listings of SAEs will be produced.

The following summaries of death will be produced.

- A table of death showing cause of death that occurred ≤ 30 days after last dose versus that occurred > 30 days after last doseA listing of all deaths

8.3.4 Adverse Events of Special Interest and Select Adverse Events

AESIs to be summarized include the following:

a. **Infusion Related Reactions (IRR):** Defined using the MedDRA preferred term of, "Infusion Related Reaction" or identified by the investigator as an infusion related reaction.

b. **Non-infectious Pulmonary toxicities:** Defined as events in the Broad Interstitial Lung Disease SMQ.

c. **Potential Cardiac events:** Defined as grade 2 or higher adverse events in the Broad Cardiac Failure SMQ or ECHO/MUGA scan results that indicate a post baseline decrease in LVEF of ≥ 10 percentage points from pretreatment baseline and a value < 50%.

d. **Confirmed cardiac events:** Defined as the subset of potential cardiac events (bullet c. above), that have been clinically reviewed by Zymeworks and determined to be consistent with cardiac events of absolute decrease in LVEF of ≥ 10 percentage points from pretreatment baseline and a value < 50%, and/or grade ≥ 2 heart failure.

Select AEs to be summarized include the following:

- **Treatment-emergent diarrhea:** Defined using the MedDRA preferred term of "Diarrhoea"
- **Treatment-emergent embryo-fetal toxicity:** Defined using the modified Pregnancy and neonatal topics SMQ which excludes *the sub SMQ Lactation related topics including neonatal exposure through breast milk (SMQ)*.

Adverse events of special interest (AESIs) may be listed and summarized by MedDRA preferred term using counts and percentages. Multiple occurrences of the same AESI within a subject will be counted only once for the time frame under consideration. For summaries by severity, only the worst grade for an AESI will be counted for a particular subject. If 5 or fewer subjects have a particular AESI/Select AE, a data listing will be provided in lieu of a summary table.

The time to first onset and duration will be calculated and summarized for each AESI and Select AE. For a particular AESI/Select AE, the time to first onset for a subject is defined as either:

- Date of 1st occurrence – date of 1st infusion (C1D1) + 1 day

Or

- Date and time of 1st occurrence – date and time of 1st infusion + 1 min

The formula used above will depend on the unit of time used in the data collection for the event.

Time to first onset will be summarized using summary statistics (n, mean, median, standard deviation, range).

Time to resolution of AESI/Select AEs will be summarized for events that have an AE end date. Estimates of median, minimum and maximum time to resolution will be provided using descriptive statistics. Time to resolution is defined as the time from the start of the AE to the end date of the AE. The number and percentage of events without resolution will also be provided.

a.

8.3.4.1 Characteristics of Infusion Related Reactions and Pulmonary Toxicities

Symptoms associated with AESIs will be summarized using counts and percentages. Multiple occurrences of the same symptom within a subject will be summarized only once for the time frame under consideration.

Additional summaries to characterize IRRs will include summary statistics of the time from the most recent zanidatamab exposure to the first subsequent IRR (for all infusions during which IRRs occur); percent of subjects with a history of any risk factor for IRRs including atopy, hypersensitivity, anaphylaxis, and/or allergy; the percent of subjects who underwent a tryptase test (and the proportion who had elevated tryptase levels); a frequency distribution of IRR onset by treatment cycle, and a frequency distribution of treatments (if any) provided for the IRRs.

Additional summaries to characterize non-infectious pulmonary toxicities will include the proportion of subjects with symptoms; a frequency distribution of symptoms; the proportion of subjects with a history of risk factors for non-infectious pulmonary toxicity, and a frequency distribution of risk factors.

8.3.4.2 Characteristics Confirmed Cardiac Events

Additional summaries to characterize confirmed cardiac events are data driven and may include summaries of the proportion of subjects with symptoms; the proportion of subjects with limitations to physical activity (most severe if a subject has multiple limitation events); and the proportion of subjects with any elevated brain natriuretic peptide (BNP) or troponins and elevation status.

8.3.5 Clinical Laboratory Results

Per protocol, samples for hematology, serum chemistry, and urinalysis will be collected at the following timepoints: Screening visit, day 1 of each cycle, EOT visit and at the 30-day post last dose visit. Samples for coagulation will be collected at the screening visit, and at day 1 of cycle 1. All laboratory results will be converted into Système International (SI) units for analysis and graded using the laboratory reference ranges from individual laboratories and the criteria from NCI CTCAE (Common Terminology Criteria for Adverse Events, Version 5) by the Sponsor. If one or more normal ranges for a laboratory are missing, then a standard normal range from the Merck Manual will be used to perform toxicity grading.

8.3.5.1 Laboratory Results

Listings of select laboratory results will be produced for all subjects by nominal visit and where appropriate, may include the change from baseline, laboratory toxicity grade, and reference range. Each abnormal value will be flagged to show whether it is a value below or above the reference range. Laboratory abnormalities will be summarized in a table using counts and percentages.

8.3.5.2 Laboratory Toxicities

Laboratory toxicities by grade will be summarized using counts and percentages. For each graded laboratory test, only the worst (i.e., highest) toxicity grade will be counted for subjects with multiple toxicities within a time period (including scheduled and unscheduled assessments). For each graded laboratory test, shift tables will be produced to summarize the shift in toxicity grade from baseline to the worst treatment-emergent grade. In addition to summary tables, listings of laboratory test results may be provided for subjects with any Grade 3 or higher toxicities.

Table 8: Toxicity Terms for Laboratory Parameters Graded for both High and Low values

Analyte	**Term (high values)**	**Term (low values)**
Calcium	Calcium high	Calcium low
Blood glucose	Blood glucose high	Blood glucose low
Potassium	Potassium high	Potassium low
Sodium	Sodium high	Sodium low
Hemoglobin	Hemoglobin high	Hemoglobin low
Lymphocyte Count	Lymphocyte count high	Lymphocyte count low

Magnesium	Magnesium high	Magnesium low
Phosphorous	Phosphorous high	Phosphorous low
WBC	White blood cell count high	White blood cell low

8.3.6 Cardiac Function

8.3.6.1 Left Ventricular Ejection Fraction

The following left ventricular ejection fraction (LVEF) endpoints will be summarized using counts and percentages.

- LVEF result at baseline (<50%, ≥ 50%)
- Minimum LVEF result post-baseline (<50%, ≥ 50%)
- Minimum post-baseline LVEF result with a decrease of ≥10 percentage points or more from the baseline result
- Post-baseline LVEF result that is < 50% with a decrease of ≥10% or more from the baseline result

In addition, a shift table based on the <50% and ≥50% cut-offs will be used to summarize changes of the baseline LVEF values against the lowest treatment-emergent values.

8.3.6.2 ECG

ECG interpretations (e.g., normal, or abnormal) will be summarized using counts and percentages. For abnormal ECG, the frequency of clinically significant findings will also be summarized using counts and percentages. The following QTcF interval endpoints will be summarized using counts and percentages:

- QTcF interval increase from baseline (<30 milliseconds, 30–60 milliseconds, and >60 milliseconds) during the treatment period.
- QTcF interval prolongation <450 msec, 450-480 msec, >480 msec during the treatment period

For both endpoints, the worst value for each subject during the treatment period (first dose date through the end of the safety reporting period, i.e. 30 days after the last dose of study treatment) will be summarized.

8.3.7 Vital Signs

Vital signs (blood pressure, heart rate, respiratory rate, and temperature) will be summarized using descriptive statistics (i.e., mean, median, standard deviation, and range).

8.3.8 ECOG Performance Status

ECOG performance status over time will be listed by subject. The baseline status, best status post-baseline, worst status post-baseline, and status at end of treatment will be summarized using counts and percentages.

8.3.9 Post-Study Treatment Systemic Anti-Cancer Therapies

Post-study treatment systemic anti-cancer therapies (i.e., subsequent treatment) will be coded using the World Health Organization Drug Dictionary (WHO-DD) summarized by preferred term using counts and percentages. Multiple occurrences of the same therapy within a subject will be summarized only once.

8.4 Quality of Life, Disease-Related Pain Analyses and Opioid Use

Analysis of Quality-of-life endpoints will be based on the safety analysis set for cohort 1 and cohort 2 separately.

CCI

CCI

8.5 Immunogenicity Analyses

All immunogenicity analyses will be based on the immunogenicity analysis set described in Section 7.1.

Subjects developing an anti-drug antibody response will be listed. Subjects 'positive' in Tier 1 and Tier 2 will be listed along with their Tier 3 titers. If available, domain specificity data and anti-ZW25 neutralizing antibodies will be listed and summarized as well.

The occurrence of anti-drug antibodies (ADAs) will be summarized using counts and percentages by nominal time point and by cohort. Titers for confirmed positive ADA will be summarized using descriptive statistics (e.g., n, geometric mean, geometric coefficient of variance). When determining post-baseline incidence, patients are considered to be ADA positive if they are initially ADA negative at baseline and test positive for ADAs following study drug exposure ("treatment-induced ADA response"), or if they are ADA positive at baseline and the titer of one or more post-baseline samples is at least 4-fold greater than the titer of the baseline sample ("treatment-enhanced ADA response"). Patients are considered to be ADA negative if they are ADA negative and all post-baseline samples are negative, or if they are ADA positive at baseline but do not have any post-baseline samples with a titer that is at least 4-fold greater than the titer of the baseline sample ("treatment unaffected"). Duration of antibody response is also monitored and identified as either "transient" (ADA positive result detected at only one post-baseline sampling timepoint [excluding last timepoint] or at 2 or more timepoints during treatment or safety follow-up where the first and last ADA positive samples are separated by a period of <16 weeks, irrespective of any negative samples in between) or "persistent" (ADA positive result detected at the last post-baseline sampling timepoint or at 2 or more timepoints during treatment or safety follow-up where the first and last ADA positive samples are separated by a period ≥16 weeks, irrespective of any negative samples in between).

Additional analyses to evaluate the association between development of ADAs and efficacy/adverse events may also be performed based on treatment-emergent ADA results.

8.6 Pharmacokinetic Analyses

All pharmacokinetic (PK) analyses will be performed on the Pharmacokinetics analysis set defined in Section 7.1. Pharmacokinetic sampling will be performed according to two different schedules. Table 9 presents the PK sampling timepoints for both extensive and sparse sampling.

Table 9: PK Sampling time points for Zanidatamab

Cycle number / Cycle length	Day	Time Point	Time window	Extensive	Sparse
1 / 28 days	1	Pre-dose		X	X

		End of ZW25 infusion	+15 min	X	X
		2 hours post-dose	± 15 min	X	
		4 hours post-dose	± 15 min	X	
		8 hours post-dose	± 3 h	X	
	2	24 hours post-dose	±4 h	X	
	5	96 hours post-dose	±4 h	X	
	15	Pre-dose	– 4 h	X	X
		End of ZW25 infusion	+15 min	X	X
Cycle 2 / 28 days	1 and 15	Pre-dose	– 4 h	X	X
		End of ZW25 infusion	+15 min	X	X
Additional Even-Numbered Cycles [a]	1	Pre-dose	– 4 h	X	X
		End of ZW25 infusion	+15 min	X	X
EOT	-	-		X [b]	X [b]

EOT = end of treatment; h = hour; min = minute; PK = pharmacokinetics; ZW25=Zanidatamab.
Extensive sampling will be performed for the first 6 enrolled subjects only.

a Day 1 of Cycles 4, 6, 8, 10, and 12.
b Only required if subject has completed less than 6 months of treatment.

Table 10: Steady-State Extensive Pharmacokinetic Sampling Schedule on Cycle 4 or any Subsequent Cycle

Day	Time Point	Time window
1	Pre-dose	
	End of zanidatamab infusion	+15 min
	2 hours post-dose	± 15 min
	4 hours post-dose	± 30 min
	8 hours post-dose	± 3 hours
3	48 hours post-dose	± 1 days
5	96 hours post-dose	± 1days

15	336 hours post-dose	± 3days

Extensive sampling at steady state: Performed for the first 10 suitable subjects in China who reach Cycle 4 and the first 20 suitable subjects in the rest of world who reach Cycle 4

8.6.1 Serum Concentrations

Peak and trough serum concentrations of zanidatamab will be summarized using descriptive statistics (e.g., n, mean, coefficient of variance) by nominal collection timepoint. Tabulations and graphs such as individual plots and mean plots may be presented.

8.6.2 Pharmacokinetic Parameters

Serum concentrations of zanidatamab will be measured as a function of time post-dosing.

For the extensive PK analysis set (dose 1 and steady-state), serum concentration-time data of each patient will be tabulated and graphically presented on linear and semilogarithmic scales. Pharmacokinetic parameters will be determined using standard noncompartmental method with actual collection time.

For each subject, the parameters listed below will be reported when there is sufficient data available to support the accurate estimation of the parameter based on actual sample collection times.

The following PK parameters will be derived for single (first) dose using extensive PK sampling data after dose 1:

- AUC_{last} area under the serum concentration-time curve from zero to the last measurable concentration for dose 1
- $AUC_{tau,1}$ area under the serum concentration-time curve tau (the time from the start to the end of the dosing interval) for dose 1
- AUC_{INF} area under the serum concentration-time curve from zero to infinity for dose 1
- $C_{max,1}$ maximum observed serum concentration for dose 1
- $C_{trough,1}$ trough serum concentration for dose 1
- $t_{½}$ terminal half-life for dose 1
- CL serum clearance for dose 1
- V_z volume of distribution in the terminal phase for dose 1

The following PK parameters will be reported for extensive subjects with multiple dosing at steady-state using PK sampling data:

- $AUC_{tau,ss}$ area under the serum concentration-time curve tau (the time from the start to the end of the dosing interval) at steady-state
- $C_{max,ss}$ maximum observed serum concentration at steady-state
- $C_{trough.ss}$ trough serum concentration at steady-state
- $C_{ave,ss}$ average concentration over the dosing interval, calculated as AUC_{tau}/tau. Where tau is the dosing interval for Dose 1, at steady-state
- $t_{1/2,ss}$ terminal half-life at steady-state
- V_{ss} volume of distribution at steady-state
- Accumulation index for AUC_{tau} and C_{max} steady-state PK parameter value divided by the PK parameter value from dose 1
- Peak-to-trough ratio: C_{max}/C_{trough}

The following parameters will be derived and summarized for all subjects with either extensive or sparse sampling data:

- $R_{Ctrough}$ accumulation index, calculated as C_{trough} (last dose)/C_{trough} (first dose)
- $C_{trough,ss}$ trough concentration at steady state

PK parameters will be summarized using summary statistics (e.g., n, geometric mean, geometric coefficient of variance) by nominal timepoint.

CCI [redacted]

[redacted]

[redacted]

[redacted]

8.9 Subgroup Analysis of Subjects Enrolled from China sites

All the above analyses, including efficacy, PK, immunogenicity and safety will be conducted for subjects enrolled from China sites, unless otherwise specified. These analyses will be used for submission to the CDE.

For efficacy analysis, there are 4 analysis sets that will be used for efficacy analysis for subjects enrolled from China sites. The primary efficacy analysis set includes subjects enrolled in Cohort

1 who received any amount of Zanidatamab. The other 3 efficacy analysis sets are supportive secondary analyses, with definition is provided in Table 11. A pooled analysis combining Cohorts 1 and 2 may be performed in addition, if a clinically meaningful treatment effect is observed for Cohort 2.

For safety evaluation, subjects enrolled from China sites who received any amount of Zanidatamab will be used for safety analysis.

In addition, trial conduct information will also be summarized for subjects enrolled from China sites. Trial conduct information includes screen failures and enrollment, subject disposition, protocol deviations, baseline demographic and disease history, prior systemic cancer therapy, radiotherapy and surgery, concomitant medications, and procedures.

At the timing of analysis for the overall subject population, analyses for subjects enrolled from China sites will be performed concurrently with same cutoff date. The subgroup analysis of subjects enrolled from China sites may not be included in the clinical study report.

Table 11: Analysis Sets for Efficacy Analysis for Subjects Enrolled from China sites

Analysis Set	**Description**
Efficacy Analysis Set China Subgroup **[E1]**	Includes subjects enrolled from China sites who received any amount of zanidatamab. Cohort 1: includes subjects enrolled in Cohort 1 from E1, this is the primary efficacy analysis set.
Modified Efficacy Analysis Set China Subgroup **[E2]**	Includes subjects enrolled from China sites who received any amount of zanidatamab and excludes intolerant subjects who are not adequately treated per NMPA CDE's requirement. Cohort 1: includes subjects enrolled in Cohort 1 from E2, this is the secondary efficacy analysis set.
Response Evaluable Analysis Set China Subgroup **[RE1]**	Includes subjects enrolled from China sites who are response evaluable. A subject is considered response evaluable if they have measurable disease at baseline and at least 1 evaluable, post-baseline disease assessment (per RECIST 1.1) or discontinued study treatment due to death or unequivocal clinical progression and received any amount of zanidatamab. Cohort 1: includes subjects enrolled in Cohort 1 from RE1, this is the secondary efficacy analysis set.
Modified Response Evaluable Analysis Set China Subgroup **[RE2]**	Includes subjects enrolled from China sites who are response evaluable and excludes intolerant subjects who are not adequately treated per NMPA CDE's requirement. Cohort 1: includes subjects enrolled in Cohort 1 from RE2, this is the secondary efficacy analysis set.

9 CHANGES IN THE PLANNED ANALYSIS

The following changes and/or clarifications to the analysis described in the study protocol are described.

Change	Reason for change
Definition of treatment-emergent adverse event (TEAE). Protocol definition: "TEAEs are defined as events with an onset during or after receipt of the first dose of zanidatamab and up to and including 30 days after the last dose but prior to the start of a new anti-cancer therapy." SAP definition: "TEAEs are defined as AEs with an onset during or after receipt of the first dose of zanidatamab and up to and including 30 days after the last dose."	The new definition in the SAP is more conservative than the definition in the study protocol.
Table 4: Analysis Set Pharmacokinetics definition was modified to consider subjects who did not have a baseline sample Immunogenicity analysis set definition was added	The new definition is more robust
Section 8.6: Pharmacokinetics Analysis Definitions and analysis were added/clarified	Added clarification
Definition of PFS and DOR were updated to remove clinical progression as an event.	To align with standard definition of PFS
Definition of duration of response	Clarified compared to protocol language that response (PR, CR) must be confirmed.
Definition of disease control rate	Clarified compared to protocol language that PR and CR must be confirmed.

10 REFERENCES

1. Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1-Update and clarification: From the RECIST committee. Eur J Cancer. 2016; 62:132–137. doi: 10.1016/j.ejca.2016.03.081
2. Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26: 404-413.
3. Independent Review Charter. A Phase 2b, open-label, single-arm study of ZW25 monotherapy in subjects with advanced or metastatic HER2-amplified biliary tract cancers. Version 1.0; Bioclinica: 03August2020.
4. Brookmeyer, R. and Crowley, J. (1982) A confidence interval for the median survival time. Biometrics, 38, 29-41. doi:10.2307/2530286.
5. Collett D (1994). Interval-censored survival data. Modelling survival data in medical research. Boca Raton, Fla., Chapman & Hall/CRC: 237-251.
6. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228-47. doi: 10.1016/j.ejca.2008.10.026.
7. Major Greenwood, Jr. (1926). The Natural Duration of Cancer. Reports of Public Health and Related Subjects, Vol. 33, HMSO, London.
8. Guidance for Industry. Pharmacokinetics in Patients with Impaired Renal Function – Study Design, Data Analysis, and Impact on Dosing. September 2020, Clinical Pharmacology, Revision 2.
9. Cockcroft, D.W. and M.H. Gault. Prediction of creatinine clearance from serum creatinine. Nephron. 1976. 16(1):31-41.
10. Mohamed Elmeliegy et-al. Discordance Between Child-Pugh and National Cancer Institute Classifications for Hepatic Dysfunction: Implications on Dosing Recommendations for Oncology Compounds. The Journal of Clinical Pharmacology 2021, 61(1) 105–115
11. Cleeland C.S. The Brief Pain Inventory, User Guide, 2009.