Statistical Analysis Plan

Phase I/II trial of humanized 3F8 bispecific antibody (Hu3F8 BsAb) in patients with relapsed/refractory neuroblastoma, osteosarcoma, and other GD2(+) solid tumors

Statistical Analysis Plan: Final
Statistical Analysis Plan: 1.0
Statistical Analysis Plan Date: 06Oct2021

Investigational Product: Hu3F8 BsAb (Nivatrotamab)

Protocol Reference: 18-034
[redacted] Study ID: 000000214175
Sponsor: Y-mAbs Therapeutics Inc.
NCT No.: 03860207

Author: [redacted], Principal Statistician

Table of Contents

List of Tables

List of Figures

Reviewers

The following reviews of the Statistical Analysis Plan (SAP) were conducted:

Name and Title	Role	Version Last Reviewed	Company / Organization
[redacted], Principal Statistician	Peer Review Statistician	Internal Draft 0.1	[redacted]
[redacted], Manager Statistical Programmer	Statistical Programmer	Internal Draft 0.1	[redacted]
[redacted], Project Manager II, Oncology	Project Leader	Internal Draft 0.1	[redacted]
[redacted], Director Clinical Pharmacokinetics	Pharmacokinetics Expert	Internal Draft 0.1	[redacted]
[redacted], Oncology Medical Director	Medical Monitor	Internal Draft 0.1	[redacted]
[redacted], Associate Director Biostatistics	Expert Statistician	Sponsor Draft 0.1	[redacted]
[redacted], Senior Project Statistician	Senior Project Statistician	Sponsor Draft 0.1	Y-mAbs Therapeutics Inc
[redacted], Associate Director, Corporate Drug Safety	Safety Manager	Sponsor Draft 0.1	Y-mAbs Therapeutics Inc
[redacted], Senior Project Statistician	Senior Project Statistician	Sponsor Draft 0.1	Y-mAbs Therapeutics Inc

Glossary of Abbreviations

Abbreviation	Term
AE	Adverse Event
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	Below the Limit of Quantification
BM	Bone Marrow
BMI	Body Mass Index
BP	Blood Pressure
BSA	Body Surface Area
BUN	Blood Urea Nitrogen
BW	Body Weight
CBC	Complete Blood Count
CI	Confidence Interval
CR	Complete Response/ Remission
CRC	Clinical Research Coordinator
CRDB	Clinical Research Database
CRF	Case Report Form
CRM	Continual Reassessment Method
CRP	C-reactive protein
CRS	Cytokine Release Syndrome
COVID-19	Coronavirus Disease of 2019
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CTMS	Clinical Trial Management System
CV	Coefficient of Variation
DAS	DLT Evaluable Analysis Set
DBP	Diastolic BP
DCR	Disease Control Rate
DI	Dose Intensity
DLT	Dose Limiting Toxicity
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic Acid
EC	Ethics Committee
ECG	Electrocardiogram
ECHO	Echocardiogram
EOD	Extent of Disease Evaluation
FAS	Full Analysis Set
FDA	Food and Drug Administration
FDG-PET	Fluorine-18-Fluorodeoxyglucose Positron Emission Tomography
FLACC	Face, Legs, Arms, Cry, Consolability
GD2	Disialoganglioside
GD3	Ganglioside Precursor Disialohematoside
GM2	Disialotetrahexosylganglioside
HAHA	Human Anti-Human Antibody
HR	Heart Rate
Hu3F8-BsAb	humanized anti-GD2 x anti-CD3 bispecific antibody
INRC	International Neuroblastoma Response Criteria

Abbreviation	Term
IRB	Institutional Review Board
IV	Intravenous
LDH	Lactate dehydrogenase
LLN	Lower Limit of Normal
MCH	Mean Corpuscular Hemoglobin
MCV	Mean Corpuscular Volume
MD	Minimal Disease
MedDRA	Medical Dictionary for Regulatory Activities
MIBG	Metaiodobenzyl-guanidine/ MIBG scintigraphy
MR	Mixed Response
MRD	Minimal Residual Disease
MRI	Magnetic Resonance Imaging
MSK	Memorial Sloan Kettering
MTD	Maximum Tolerated Dose
NB	Neuroblastoma
NC	Not Calculated
NCI-CTCAE	National Cancer Institute-Common Terminology Criteria for Adverse Events
NE	Not Evaluable
NI	Not Involved
ORR	Overall Response Rates
OS	Overall Survival
PDI	Planned Dose Intensity
PET	Positron Emission Tomography
PFS	Progression Free Survival
PI	Principal Investigator
PK	Pharmacokinetic
PKAP	PK Analysis Plan
PPS	Per-Protocol Set
PR	Partial Response/ Remission
PT	Preferred Term
QTcB	Bazett corrected QT interval
QTcF	Fridericia corrected QT interval
RBC	Red Blood Cell Count
RDI	Relative Dose Intensity
RDW	Red cell Distribution Width
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic Acid
RP2D	Recommended Phase II Dose
SAE	Serious AE
SAP	Statistical Analysis Plan
SBP	Systolic BP
SD	Standard Deviation
SI	International System of Units
SOC	System Organ Class
SpO2	Pulse Oxymeter
TFLs	Tables, Figures and Listings
TEAE	Treatment Emergent AE
TESAE	Treatment-Emergent Serious AE
ULN	Upper Limit of Normal

Abbreviation	Term
WBC	White Blood Cell
WHO DD	World Health Organization Drug Dictionary

Statistical Analysis Plan Amendment

Not Applicable

1. Source Documents

The SAP was written based on the following documentation:

Document	Date	Version
Protocol Amendment 8	22Mar2021	Version A(8)
eCRF	10Feb2021	Final Version
DMC Charter	08Feb2021	Final Version
DMC Shells	19Mar2021	Draft Version 0.3

2. Protocol Details

2.1. Overall Trial Design

This is a phase I/II trial which will assess the toxicity and pharmacokinetics (PK) of the humanized anti- disialoganglioside (GD2) x anti-CD3 bispecific antibody (Hu3F8-BsAb) in phase I and the anti-tumor activity of hu3F8-BsAb in phase II.

In phase I, dosages of hu3F8-BsAb will follow a dose-finding design in order to determine the Maximum Tolerated Dose (MTD) in cohorts of patients with relapsed/refractory neuroblastoma (NB), osteosarcoma, or other GD2(+) solid tumors. The Recommended Phase II Dose (RP2D) will be established. Hu3F8-BsAb PK will be also evaluated.

In phase II, patients with relapsed/refractory NB (Group 1) and osteosarcoma (Group 2) will be treated with hu3F8-BsAb at the RP2D and anti-tumor activity will be assessed.

2.1.1. Phase I

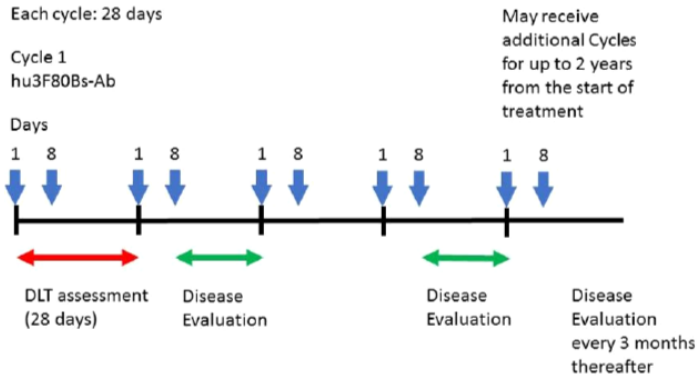
2.1.1.1. Phase I Design

Initially the phase I trial will follow a dose-escalation schema with 1 patient assigned per dose until the first ≥ Grade 2 toxicity (except for Grade 2 sinus bradycardia, sinus tachycardia, urticaria, fever, nausea, vomiting, diarrhea, urticaria, headache, paresthesia, electrolyte disturbances, pain) is observed, then the subsequent dose allocation will follow the Continual Reassessment Method (CRM)[1,2]. Eligible patients will include relapsed or refractory NB, osteosarcoma, and GD2 (+) solid tumors. In each cycle, hu3F8-BsAb is given on Days 1 and 8. PK is assessed by blood draws. Patients are monitored for Dose Limiting Toxicity (DLT) until Day 28. Patients with DLT will come off treatment. Human Anti-Human Antibody (HAHA) titer will be measured after each cycle. Extent of disease evaluation will be performed after the second dose of cycle 2 and cycle 4, then every 3 months thereafter. A total of 4 cycles will be given.

If patients experience no life-threatening toxicity clearly attributable to hu3F8-BsAb, do not develop progression of disease and continue to meet all eligibility criteria, they have the option of receiving additional cycles up to 2 years from the start of hu3F8-BsAb treatment. Patients may continue therapy beyond 4 cycles even if they achieve remission (see Figure 1).

The duration of each subsequent cycle is 28 days with hu3F8-BsAb being given on Days 1 and 8. MTD is defined as the dose level that leads to DLT in approximately 15% of patients (posterior probability of DLT that comes closest to the toxicity level of 0.15). The RP2D will be decided based on the safety data in the phase I and can be MTD or lower. If MTD is not reached, RP2D may be determined on the basis of PK studies. If no DLT is observed, the RP2D will be determined as a minimum dose that gives the highest hu3F8-BsAb serum levels based on the PK data. For example, the RP2D of hu3F8 monoclonal antibody was decided as 9 mg/kg/cycle because the serum hu3F8 levels reached a plateau beyond 8.4 mg/kg/cycle with no DLT (see Protocol Section 3.2.3). After cycle 2, intra-patient dose escalation will be permitted i.e. patients can receive the highest dose that has been evaluated and proven safe by DLT assessment.

Figure 1 Trial Design



2.1.1.2. Phase I Dose-Escalation Schedule

The phase I dose-escalation schedule schema is summarized in Table 1 below. Dose Level 1 is the starting dose, and the escalation and de-escalation schedule are described in Section 2.1.1.3. DLTs are defined in Section 2.1.1.5.

Hu3F8-BsAb is given intravenous (IV) over ~1-3 hours on Days 1 and 8 for each cycle. In cycle 1, blood is drawn for PK studies as described in Section 6.8.1.

Table 1 Dose-Escalation Schedule

Level	Dose on day 1	Dose on day 8	Dosage per cycle
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

2.1.1.3. Dose-Escalation Plan

This is a Phase I trial to determine the MTD of hu3F8-BsAb that leads to DLT in approximately 15% of patients. Hu3F8-BsAb is dosed based on Body Weight (BW) rather than Body Surface Area (BSA). Unlike chemotherapy for which clearance ultimately depends upon cardiac output which is dependent upon BSA, antibody distribution and clearance depend on total blood volume which is determined not by BSA, but by BW. Dose escalation will start at dose level 1, 0.0045 mcg/kg/dose (0.009 mcg/kg/cycle). Dose-escalation will be performed in two stages.

In the first stage, the trial will accrue 1 patient per dose until the first ≥ Grade 2 toxicity (except for Grade 2 sinus bradycardia, sinus tachycardia, fever, nausea, vomiting, diarrhea, urticaria, headache, paresthesia, electrolyte disturbances, pain) is encountered. The subsequent dose allocation will follow the CRM. CRM assumes a simple model for the probability of a DLT as a function of dose and uses the occurrence of toxicities in the patients enrolled in the trial to sequentially determine which dose to administer to a new patient. Patients will be assigned at each dose one at a time. New patients will be allocated a dose as suggested by the CRM algorithm based on the toxicities of previously accrued patients. If DLT data are not available for the previously accrued 1 or 2 patients, the CRM will proceed with the dose as calculated with available toxicities/data, however if there are 3 enrolled patients with incomplete /outstanding DLT data, accrual will be paused. DLT data from at most 2 patients can be pending but no more than 2 patients are allowed to have incomplete DLT data. To protect patient safety, a dose escalation of more than one dose level is not permitted.

12 dose levels will be examined (See Table 1). The initial estimates of DLT probabilities are: **0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.10, 0.15, 0.25, 0.35, 0.45, 0.50** for doses 1-12, respectively. Thus, the a priori belief is that dose level 8 is the MTD. It is assumed that the dose-toxicity follows a hyperbolic tangent model P(DLT=yes at dose d)=((tanh d+1)/2)^a, where a is the unknown parameter that will have to be estimated in order to determine which dose is the MTD and d is “an adjusted dose level”. A value of a=1.0 indicate that our prior beliefs were correct; while a value of a less than (greater than) 1.0 indicates that the combinations are more (less) toxic than believed.

To reflect the uncertainty in the prior probability estimates, it is assumed that it follows an exponential distribution (prior distribution) with mean 1.0. Dose escalation will be guided by the model after the occurrence of the first DLT. At that time, the above initial estimates of DLT might be revisited before model initiation to reflect the current knowledge on the safety profile and dose levels based on the observed data.

It is anticipated that approximately 30 evaluable patients will be enrolled and treated in Phase I. All patients who belongs to the DLT Evaluable Analysis Set (see section 4.5 for the definition) are considered evaluable for toxicity. Additional patients will be enrolled to replace any patients who are enrolled, but do not receive treatment. In the unlikely event that patient stops the drug due to HAHA before completing the first cycle, this patient will be considered non evaluable and will be replaced. If life-threatening toxicity occurs in any patient, further accrual at that dosage will be stopped pending review by the Principal Investigator (PI) or Co-PI. Toxicity at each dosage level is evaluated during the 28 days

following the first dose of hu3F8-BsAb. No more than two patients will receive their first dose of hu3F8-BsAb within a 24 hour period.

If no DLT is observed up to dose level 12 during the initial patient per dose escalation (e.g. no MTD), the RP2D will be determined as a minimum dose that gives the highest hu3F8-BsAb serum levels based on the PK data.

2.1.1.4. Dose Allocation in Clinical Research Database (CRDB) Section for the Protocol

Patient Registration and the CRM module are run independently. Once new patients are registered, the CRM module can be run to generate the recommended dose level for the newly registered patient.

During the first stage, one patient is accrued at a time starting at the first dose level and followed for DLT before the next patient is accrued. Patients must be followed for 28 days to determine if they had a DLT. If no DLT is seen, then the next patient is accrued at the next higher dose level.

This process will continue until a DLT is seen. Once a DLT is seen, the trial team should notify the CRDB Helpline and the trial statistician in order to have the protocol switched to the model-based stage. The dose level of subsequent patients will be assigned by the CRM module.

The CRM module uses information specified in the Biostatistics section of the protocol and the toxicity data of previously treated patients to determine the next dose level to which the patient will be assigned. Only eligible and evaluable patients are used in determining the next dose level.

Patients accrued during the model-based stage must also be followed for 28 days to determine if they had a DLT. Since the protocol allows 2 patients to have incomplete toxicity data, then the CRM algorithm can still be applied to generate a dose assignment even if toxicity/DLT data are missing for up to 2 patients who were last registered to the protocol. If toxicity data is missing on more than 2 patients or other required data is missing, no new dose assignment will be performed.

2.1.1.5. Dose Limiting Toxicities

Toxicity will be monitored using Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 or later. The DLT period corresponds to days 1 through 28 in cycle 1. Allowance will be made for the expected toxicities of hu3F8 from which hu3F8-BsAb was derived (See Protocol Section 11.2). As stated above, in the first stage (1 patient per dose-level) dose-escalation will be performed until ≥ Grade 2 related toxicity (except for Grade 2 sinus bradycardia, sinus tachycardia, urticaria, fever, nausea, vomiting, diarrhea, urticaria, headache, paresthesia, electrolyte disturbances, pain) is encountered, then the subsequent dose allocation will follow the CRM.

During the CRM phase, any ≥Grade 3 toxicity attributable to hu3F8-BsAb will be considered dose-limiting with the following **exceptions**:

- Grade 3 Cytokine Release Syndrome (CRS) improving to ≤ Grade 1 CRS within ≤ 72 hours after starting systemic steroids (See Protocol Section 11.3 for a grading system and algorithm)
- Grade 3 pain improving to ≤ Grade 2 within ≤ 72 hours of its onset. Grade 3 pain associated
- with myalgia is an expected clinical toxicity associated with CRS.
- Grade 3 neutropenia or anemia
- Grade 3 thrombocytopenia without clinically significant bleeding.
- Grade 4 lymphopenia
- Grade 3 elevations of Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT) if they return to < Grade 3 by day 28
- Catheter related infections
- DLT of hypertension is defined as persistent hypertension higher than 99th percentile for > 48 hours.

The following Grade 3 toxicities are related to infusion of monoclonal antibodies and will not be considered DLT if they resolve to < Grade 3 within 24 hours:

- Allergic reactions controlled with supportive care measures
- Vasovagal reaction
- Sinus bradycardia
- Sinus tachycardia
- Urticaria (improved to at least Grade 2)
- Fever

The following Grade 3 toxicities will not be considered a DLT if they resolve within 48 hours:

- Nausea
- Vomiting
- Diarrhea
- Peripheral neuropathy (paresthesia)
- Electrolyte disturbances

In all other cases, DLT is defined as Grade 3 or greater toxicities occurring during cycle 1 with the exception of toxicities clearly related to disease activity or co-interventions.

2.1.1.6. Criteria for Continuing on Therapy beyond Cycle 1

If patients experience DLT as described above, they will discontinue treatment. Patients will need to fulfill all eligibility criteria before each cycle. However, patients can continue treatment without delay if they have ≤ grade 4 lymphopenia. Cycles 2-4 may be delayed due

to toxicities other than DLT for a maximum of four weeks i.e. maximum duration between dose 1 of the prior cycle and dose 1 of the next cycle can be ≤ 56 days. If HAHA titer of >1300 U/ml develops, cycles beyond cycle 1 may be delayed for up to 120 days. However, if the duration between cycles is > 28 days, patients will be required to undergo extent of disease evaluation to ensure that there is no progressive disease.

2.1.2. Phase II

After RP2D is determined, a phase II trial will be carried out to assess the anti-tumor activity of hu3F8-BsAb. Two groups of patients will be studied: Group 1 will consist of patients with relapsed or refractory NB and Group 2 of relapsed or refractory measurable osteosarcoma. Hu3F8-BsAb is given on Days 1 and 8 of each cycle. The duration of each cycle is 28 days. However, cycles may be delayed for up to 56 days if >grade 2 toxicity is encountered.

HAHA titer will be measured after each cycle. If HAHA develops, cycles are deferred for up to 120 days until HAHA titer <1300 U/ml. Extent of Disease Evaluation (EOD) will be performed after the second dose of cycle 2 and cycle 4, then every 3 months thereafter.

A total of 4 cycles will be given, and patients with no life-threatening toxicity clearly attributable to hu3F8-BsAb, no progressive disease, continuing to meet all eligibility criteria and HAHA titer is <1300U/ml have the option of receiving additional cycles up to 2 years from the start of the Hu3F8-BsAb treatment. Patients may continue therapy beyond 4 cycles even if they achieve remission.

After cycle 4, duration between cycles can range from 28-120 days. Patients will undergo physical examination, Complete Blood Count (CBC) and complete chemistry and immune correlates will be tested as described in Protocol Section 10.

For Group 1, the overall response rate will be evaluated and is defined as the proportion of patients achieving Complete Remission (CR) or Partial Response (PR) based on revised International Neuroblastoma Response Criteria (INRC)[3] as the best of the responses at the end of cycle 2 and cycle 4. (See Section 2.4.2.1) excluding potential response assessments after initiation of any new prohibited anticancer treatment).

For Group 2, 4-month Progression Free Survival (PFS) will be assessed and is defined as the proportion of patients who have not developed progressive disease nor died at the end of 4 months from the first dose of Hu3F8-BsAb. In addition, number of patients who have CR or PR as the best of the responses at the end of cycle 2 and cycle 4 will be determined. (see Section 2.5.2.2)

Hu3F8-BsAb is given IV over ~1-3 hours on Days 1 and 8 for each cycle.

2.1.3. Evaluation During Treatment/Intervention

The following evaluations will be performed during the treatment period.

Table 2 Phase I - Evaluation During Treatment/Intervention

Tests	Pre-treatment	During Treatment	During Follow-up
Complete history and physical	✓	Cycle 1: Daily from Day 1 through Day 4, on Day 8, then weekly thereafter until Day 28 Subsequent Cycles: Before each dose of hu3F8-BsAb and within 72 hours of treatment	-
Pharmacokinetic studies (research blood)	✓	See Protocol Section 9.1 Cycle 1: See PK Section 6.8 Subsequent Cycles: Before each dose of hu3F8-BsAb, ~5 minutes after hu3F8-BsAb infusion	
Complete Blood Count	✓	Cycle 1: Approximately Days 2, 5, 8 (prior to the second dose of hu3F8- BsAb), and approximately Days 15 and 28. Subsequent Cycles: Before each dose of hu3F8-BsAb	-
Liver and Renal function tests (ALT, AST, alkaline phosphatase, albumin, bilirubin, BUN and creatinine) and serum electrolytes (sodium, potassium, chloride, bicarbonate, magnesium, and phosphate)	✓	Cycle 1: Before each dose of hu3F8- BsAb, within 72 hours after each dose of hu3F8-BsAb and approximately Day 15 Subsequent Cycles: Before each dose of hu3F8-BsAb	-
C3 & CH50	✓	Cycle 1 Only: Day 2	-
Lactate dehydrogenase (LDH)	✓	Cycle 1: Day 8 and approximately Day 28 Subsequent Cycles: If CRS is suspected based on clinical symptoms	
Ferritin and C-reactive protein (CRP)	✓	Cycle 1: Day 8 and approximately Day 28 Subsequent Cycles: If CRS is suspected based on clinical symptoms	-
Pulse Oxymeter (SpO2) and Blood pressure measurement	Blood pressure (BP) measurement (standing, sitting, lying down)	SpO2: During hu3F8-BsAb infusion BP: No later than 30 minutes after the end of hu3F8-BsAb infusion Cycle 1: Blood pressure measurement (standing, sitting, lying down) pre infusion, once daily throughout hospital stay and at discharge. For outpatient treatment: pre infusion and at discharge	-
Blood for HAHA	✓	Before each cycle	Approximately q3 months for 24 months from first hu3F8-BsAb dose and then q6 months for up to 5 years from first dose or until patient is off trial, whichever is earlier
Urinalysis	✓	Cycle 1: Prior to each dose of hu3F8-BsAb and approximately one week after the second dose	-
Pregnancy test, if applicable	✓	Within 2 weeks before each cycle	-
Electrocardiogram (ECG)	✓	Cycle 1: Post-infusion of the first dose and prior to the second dose on Day 8	-

Echocardiogram (ECHO)	✓	-	-
Cytokine studies (research blood)	✓	Cycle 1: prior to and ~ 6-8 hrs after the first and second hu3F8-BsAb dose, and on Days 2, 3, 4, 9, 10 and 11 Subsequent Cycles: If CRS is suspected based on clinical symptoms	-
Lymphocyte subsets, T cell activation/exhaustion markers (research blood)	✓	Cycle 1: on Days 2, 8 (prior to hu3F8-BsAb), 9, 10, and 11, and approximately on Days 15 and 28 Subsequent Cycles: Before each dose of hu3F8-BsAb	-
Bone Marrow (BM) studies*	✓	Any time after the second dose in cycle 2 and cycle 4 and approximately every 3 months thereafter	-
Computed Tomography (CT) / Magnetic Resonance Imaging (MRI), Metaiodobenzyl-guanidine (MIBG) scan* with/without Positron Emission Tomography (PET), bone scan	✓	Any time after the second dose in cycle 2 and cycle 4 and approximately every 3 months thereafter	-
Pain Assessment	✓	On days of hu3F8-BsAb treatment: prior to hu3F8-BsAb administration, during acute pain episode, and ~1 hour after the end of infusion or prior to discharge from the PDH	-
Stool Microbiome	✓ , where feasible	-	-

* Only for NB patients

Table 3 Phase II - Evaluation During Treatment/Intervention

Tests	Pre-treatment	During Treatment	During Follow-up
Complete history and physical	✓	Before each dose of hu3F8-BsAb and within 72 hours of treatment	-
Complete Blood Count	✓	Before each dose of hu3F8-BsAb	-
Liver and Renal function tests (ALT, AST, alkaline phosphatase, albumin, bilirubin, BUN and creatinine) and serum electrolytes (sodium, potassium, chloride, bicarbonate, magnesium, and phosphate)	✓	Before each dose of hu3F8-BsAb	-
LDH	✓	If CRS is suspected based on clinical symptoms	
Ferritin and CRP	✓	If CRS is suspected based on clinical symptoms	-
SpO2 and Blood pressure measurement	Blood pressure measurement (standing, sitting, lying down)	SpO2: During hu3F8-BsAb infusion BP: No later than 30 minutes after the end of hu3F8-BsAb infusion Cycle 1: Blood pressure measurement (standing, sitting, lying down) pre infusion, once daily throughout hospital stay and at discharge. For outpatient treatment: pre infusion and at discharge	-
Blood for HAHA	✓ (If applicable)	Before each cycle	Approximately q3 months for 24 months and then q6 months for up to 5 years from first hu3F8 dose or until patient is off trial, whichever is earlier
Urinalysis	✓	Cycle 1: Prior to each dose of hu3F8-BsAb and approximately one week after the second dose	-
Pregnancy test, if applicable	✓	Within 2 weeks before each cycle	-
ECG	✓	-	-
ECHO	✓	-	-
Cytokine studies (research)	✓	If CRS is suspected based on clinical symptoms	-
Lymphocyte subsets, T cell activation/exhaustion markers (research)	✓	Before each dose of hu3F8-BsAb	-
BM studies*	✓	After the second dose of cycle 2 and cycle 4 and approximately every 3 months thereafter	-
CT/MRI, MIBG scan* with/without PET, bone scan	✓	After the second dose of cycle 2 and cycle 4 and approximately every 3 months thereafter	-
Stool Microbiome	✓ , where feasible	-	-

* Only for NB patients

2.2. Trial Objectives

2.2.1. Primary Objective(s)

⇨ Phase I

The primary objective of phase I is to establish the safety of hu3F8-BsAb and determine the MTD and the RP2D.

⇨ Phase II

The primary objectives of phase II are:

- In Group 1: Assess the Overall Response Rates (ORR) of hu3F8-BsAb in patients with relapsed or refractory NB
- In Group 2: Assess the PFS in patients with relapsed or refractory osteosarcoma at 4 months from the first hu3F8-Bsab treatment.

2.2.2. Secondary Objective(s)

⇨ Phase I

The secondary objectives are:

- To study the PK of hu3F8-BsAb.
- To evaluate HAHA.
- To assess the anti-tumor activity and overall response rates of hu3F8-BsAb in patients with relapsed or refractory NB, osteosarcoma and other GD2(+) tumors.

⇨ Phase II

The secondary objectives are:

- To evaluate the duration of complete remission in groups 1 and 2.
- In Group 2: Assess the overall response rates of hu3F8-BsAb in patients with relapsed or refractory osteosarcoma.
- To continue to assess the safety/toxicity of hu3F8-BsAb.
- To evaluate HAHA.

2.2.3. Exploratory Objective(s)

⇨ Phase I

The exploratory objectives are:

- To study immunological effects of hu3F8-BsAb
- To study the effects of hu3F8-BsAb on BM Ribonucleic Acid (RNA)

- To study the effects of stool microbiome of the activity of hu3F8-BsAb.

⇨ Phase II

The exploratory objectives are:

- To assess Overall Survival (OS) and PFS after treatment with hu3F8-BsAb.
- To study immunological effects of hu3F8-BsAb
- To study the effects of hu3F8-BsAb on BM Deoxyribonucleic Acid (DNA) and RNA (Neuroblastoma patients only).
- To study the effects of stool microbiome of the activity of hu3F8-BsAb.

2.3. Sample Size and Power

2.3.1. Phase I

Phase I will enroll up to 30 evaluable patients. 10-15 patients/year are expected to be enrolled, and it is expected that the trial will be open to accrual for 36 months. All patients who are enrolled in the trial and receive at least one dose of hu3F8-BsAb are considered evaluable for toxicity. Additional patients will be enrolled to replace any patients who are enrolled, but do not receive any treatment. An amendment in the dose levels was done after the inclusion of 5 patients to dose levels 1 to 4; however, the modifications only impacted dose level 5 and higher, therefore the 5 first patients are still part of the CRM and will be analyzed as planned. The amendment does not modify the operating characteristics of the CRM.

2.3.2. Phase II

2.3.2.1. For Group 1 (relapsed or refractory NB)

The null hypothesis is that the hu3F8-BsAb treatment has ORR < 25% (assumption based on results of previous clinical trials / see protocol section 14.2.1.1). The alternative hypothesis is that the ORR after the hu3F8-BsAb treatment is ≥ 25%. The power is calculated assuming ORR=50%.

30 patients will be recruited. As pre-specified in the protocol a 95% one-sided confidence interval (CI) (corresponding to a two-sided 90% CI) for ORR will be computed using Wilson's method. If the confidence interval does not include the null ORR rate of 25% this treatment will be considered as successful. The sample size is based on confidence interval half-width 0.15 and desirable ORR rate of 50%. Based on simulation if the true rate is 50% the CI will not include 25% with probability 90%. In addition to 90% CI, 95% CI will also be presented.

Patients from the Full Analysis Set (FAS) (see section 4.3 for the definition) will be included into the assessment even if they are removed from the trial before completing cycle 4 due to any reasons. 10-15 patients/year are expected to be enrolled, therefore accrual of this group will be completed in 2-3 years.

2.3.2.2. For Group 2 (relapsed or refractory osteosarcoma)

The null hypothesis is that the hu3F8-BsAb treatment has 4-month PFS <12% (assumption based on results of seven clinical trials / see protocol section 14.2.1.2). The alternative hypothesis is that the 4-month PFS after hu3F8-BsAb treatment is ≥12%, whereas the power is calculated assuming that the 4-month PFS after hu3F8-BsAb treatment is 40%. 18 patients will be accrued.

According to the protocol, a 95% one-sided confidence (corresponding to a two-sided 90% CI) interval for 4 months PFS will be computed using Wilson's method. If the confidence interval does not include the null rate of 12%, this treatment will be considered as successful. The sample size is based on confidence interval half-width 0.190 and desirable 4 months PFS rate of 40%. Based on simulation if the true rate is 40% the CI will not include 12% with probability 90%.

5-10 patients/year are expected to be enrolled, therefore accrual of this group will be completed in 2-4 years.

Patients from the FAS (see section 4.3 for the definition) will be included. In order to take censoring prior to 4 months into account the CIs will be calculated using Kaplan-Meier methods rather than Wilsons's method. This is a change to the protocol and has been described in section 7 below. In addition to 90% CI, 95% CI will also be presented.

2.4. Primary Variable(s)

2.4.1. Phase I

The primary endpoint will be the occurrence of DLT during cycle 1. Based on this, the MTD and the RP2D will be determined.

2.4.2. Phase II

2.4.2.1. ORR For Group 1 (relapsed or refractory NB)

For relapsed or refractory NB patients, the primary endpoint will be the ORR defined as the proportion of patients achieving CR or PR based on INRC as the best response at the end of cycle 4.

For this analysis, Best Overall Response (BOR) will be derived as the best response at the end of cycle 4, excluding any potential response assessments after initiation of any new prohibited anticancer treatment.

Overall response will be defined by the investigator in the Response Summary eCRF form by combining response of the individual components (i.e. soft tissue, bone, and bone marrow disease):

- Complete Response/Remission (CR): All components meet criteria for CR.

- Partial Response/Remission (PR): PR in at least one component and all other components are either CR, Minimal Disease (MD) (bone marrow), PR (soft tissue or bone), or Not Involved (NI)*; no component with progression disease.
- Mixed Response (MR): PR or CR in at least one component but at least one other component with Stable Disease; no component with PD.
- Stable Disease: Stable Disease in one component with no better than Stable Disease or NI* in any other component; no component with progressive disease.
- Progressive Disease: Any component with Progressive Disease.
  *NI – Site not involved at trial entry and remains uninvolved.

Mixed Response corresponds to Minor Response in the *Revisions to the International Neuroblastoma Response Criteria: A Consensus Statement From the National Cancer Institute*[3]. Thus it will be considered as Minor Response for the analyses.

Primary (soft tissue) Tumor Response:

- CR: <10mm residual soft tissue at primary site AND Complete resolution of MIBG or Fluorine-18-Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) uptake (for MIBG-nonavid tumors) at primary site.
- PR: ≥30% decrease in longest diameter of primary site AND MIBG or FDG-PET uptake at primary site stable, improved, or resolved.
- Progressive Disease: >20% increase in longest diameter taking as reference the smallest sum on trial (this includes the baseline sum if that is the smallest on trial) AND Minimum absolute increase of 5mm in longest dimension.
- Stable Disease: Neither sufficient shrinkage for PR or sufficient increase for PD at the primary site.

Tumor Response at Metastatic Soft Tissue and Bone Sites:

- CR: Resolution of all sites of disease, defined as: Nonprimary target and nontarget lesions measure <10mm AND Lymph nodes identified as target lesions decrease to a short axis <10mm AND MIBG uptake or FDG-PET uptake (for MIBG-nonavid tumors) of nonprimary lesions resolves completely.
- PR: ≥30% decrease in sum of diameters of nonprimary target lesions compared with baseline AND all of the following:
  Nontarget lesions may be stable or smaller in size AND No new lesions AND ≥50% reduction in MIBG absolute bone score (relative MIBG bone score ≥ 0.1 to ≤ 0.5) or ≥50% reduction in number of FDG-PET-avid bone lesion
- Progressive Disease: Any of the following:
  Any new soft tissue lesion detected by CT/MRI that is also MIBG avid or FDG-PET avid
  Any new soft tissue lesion seen on anatomic imaging that is biopsied and confirmed to be neuroblastoma or ganglioneuroblastoma
  Any new bone site that is MIBG avid

A new bone site that is FDG-PET avid (for MIBG-nonavig tumors) AND has CT/MRI findings consistent with tumor OR has been conferment histologically to be neuroblastoma or ganglioneuroblastoma
>20% increase in longest diameter taking as reference the smallest sum on trial (this includes the baseline sum if that is the smallest on trial) AND minimum absolute increase of 5mm in sum diameters of target soft tissue lesions
Relative MIBG score ≥1.2

- Stable Disease: Neither sufficient shrinkage for PR or sufficient increase for PD of non primary lesions

Bone Marrow Metastatic Response:

- CR: Bone marrow with no tumor infiltration on reassessment, independent of baseline tumor involvement.
- Progressive Disease: Any of the following:
  Bone marrow without tumor infiltration that becomes >5% tumor infiltration on reassessment OR Bone marrow with rumor infiltration that increases by > two-fold and has >20% tumor infiltration on reassessment.
- MD: Any of the following:
  Bone marrow with ≤5% tumor infiltration and remains >0 to ≤5% tumor infiltration on reassessment OR Bone marrow with no tumor infiltration that has ≤5% tumor infiltration on reassessment OR Bone marrow with >20% tumor infiltration that has >0 to ≤5% tumor infiltration on reassessment
- Stable Disease: Bone marrow with tumor infiltration that remains positive with >5% tumor infiltration on reassessment but does not meet CR, MD, or Progressive Disease criteria.

2.4.2.2. 4-month PFS for Group 2 (relapsed or refractory osteosarcoma)

For relapsed or refractory osteosarcoma patients, the primary endpoint is the 4-month PFS. PFS is defined at the time from first dose of treatment to the date of first progression or death from any cause, whichever occurs first.

The assessment of progression will be done based on response criteria assessed from tumor assessments. Patients without progression or death at time of analysis will be censored at the date of their last evaluable response assessment based on Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1)[4] assessment.

PFS will be censored on the date of the last adequately evaluable tumor assessment:

- for patients who do not have an event (progression or death),
- for patients who start new anti-cancer treatment prior to an event,
- for patients with an event after a time period without assessments of response corresponding to at least two consecutive missing response assessments.

Patients who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of the first treatment unless death occurred on or before the time of the second planned tumor assessment (Cycle 4 scan), in which case the death will be considered an event.

PFS (in months) will be computed as [date of event or censoring – first dose of treatment + 1]/30.4375.

PFS will be calculated with one decimal.

2.5. Secondary Variable(s)

2.5.1. Phase I

2.5.1.1. PK of hu3F8-BsAb

See section 6.8 for more details

2.5.1.2. HAHA

Blood (~5ml per draw) drawn before each cycle for HAHA testing will be measured using ELISA. Antiidiotype antibodies (Ab3, Ab3') will be also tested with the same blood samples. This blood sample may also be tested for HAMA. HAMA positivity is allowed and will not affect patients' eligibility for treatment.

2.5.1.3. ORR

For NB, anti-tumor activity will be measured by revised INRC.
For osteosarcoma and other GD2(+) tumors, the response and progression will be evaluated in this trial using the RECIST Committee, version 1.1.
The proportion of NB, osteosarcoma, and other GD2(+) tumor patients responding to therapy will be determined for each disease type.
ORR will be defined as the proportion of patients achieving CR or PR as BOR at the end of cycle 4.

In addition, a similar analysis will be performed corresponding to end of trial but before any new prohibited anticancer therapy intakes. Potential response assessments after initiation of any new prohibited anticancer treatment will not be included in the assessment of BOR.

Thus, the BOR will be the best response recorded from the initiation to the following timepoints:

- At the end of cycle 4
- At the end of the trial.

For patients with osteosarcoma and other GD2(+) tumors, each patient will be assigned one of the following categories: CR, PR, Stable Disease, Progressive Disease, Not Evaluable (NE).
Patients' response will also be classified as "unknown" if insufficient data were collected to allow evaluation per these criteria.

According to RECIST v1.1, the following table provides a summary of the overall response status calculation at each time point for patients with or without measurable disease at baseline:

Table 4 Integration of target, non-target and new lesions into response assessment

Target Lesions	Non-Target Lesions	New Lesions	Overall Response
CR	CR	No	CR
CR	Non-CR/Non-Progressive Disease	No	PR
CR	Not all evaluated	No	PR
PR	Non-Progressive Disease / not all evaluated	No	PR
Stable Disease	Non-Progressive Disease / not all evaluated	No	Stable Disease
Not all evaluated	Non-Progressive Disease	No	NE
Progressive Disease	Any	Any	Progressive Disease
Any	Progressive Disease	Any	Progressive Disease
Any	Any	Yes	Progressive Disease
Note: Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression (or evidence of unequivocal disease progression) at that time should be reported as "*symptomatic deterioration"*. This is a reason for stopping therapy, but is NOT objective Progressive Disease. Every effort should be made to document the objective progression even after discontinuation of treatment.			

The overall response will be determined for all protocol specified assessments, ie scheduled assessments (with or without missing information) and also for all unscheduled assessments with non-missing information.

If a scheduled assessment is completely missing (target, non-target, new lesion), the overall response will be NE.

When Stable Disease is believed to be the BOR, it needs to be assessed a minimum of 6 weeks (+/- 7 days) after start of treatment. Otherwise, the BOR will be NE, unless any Progressive Disease was further documented, in which case BOR will be Progressive Disease.

If a patient progressed before this first assessment, he will be considered as 'early progressive'. If this event occurs, the overall response of the patient will be presumed as progression. This corresponds to patient who withdraws without any scan and whose reason for ending treatment or ending study is progressive disease (according to investigator opinion). In case of patient with a reason different to progressive disease, he will not be considered as progressive disease but will be at least considered as non-responder.

2.5.1.4. PFS

PFS is defined, like primary endpoint of Group 2 of phase II, at the time from first dose of treatment to the date of first progression or death from any cause, whichever occurs first.

Same censoring rules as those used for Group 2 of phase II will be applied considering INRC assessments instead of RECIST v1.1 assessments for NB patients.

PFS (in months) will be computed as [date of event or censoring – First dose of treatment + 1]/30.4375.

PFS will be calculated with one decimal.

2.5.1.5. OS

OS will be defined as the time from the first dose of treatment to the time of death or last follow up known to be alive for patient without death date.

For patients who have not died, overall survival will be censored at the date of last contact (date the patient was last known to be alive).

OS (in months) = [date of event or censoring – First dose of treatment + 1]/30.4375.

OS will be calculated with one decimal.

2.5.2. Phase II

2.5.2.1. Duration of Complete Remission

Duration of CR is defined only for patients who achieve CR after hu3F8-BsAb and will be calculated from the time of remission to the date of progression or death, whichever occurs first.

Same censoring rules as for PFS will be applied for CR duration analysis (See section 2.4.2.2).

CR duration (in months) will be computed as [date of progression or death or censoring – CR date + 1]/30.4375.

CR duration will be calculated with one decimal.

2.5.2.2. ORR for Group 2

ORR will be defined as the proportion of patients achieving CR or PR based on RECIST v1.1 as the best response at the end of cycle 4 (excluding potential response assessments after initiation of any new prohibited anticancer treatment).

2.5.2.3. HAHA

HAHA will be measured as described for phase I (section 2.5.1.2).

2.6. Exploratory Efficacy Variable(s)

2.6.1. Phase II

2.6.1.1. Overall Survival (OS)

OS will be defined as the time from the first dose of treatment to the time of death or last follow up known to be alive for patient without death date.

For patients who have not died, overall survival will be censored at the date of last contact (date the patient was last known to be alive).

OS (in months) = [date of event or censoring – First dose of treatment + 1]/30.4375.

OS will be calculated with one decimal.

2.6.1.2. PFS

As defined in section 2.4.2.2, PFS will be calculated as:

PFS (in months) = [date of event or censoring – first dose of treatment + 1]/30.4375.

PFS will be calculated with one decimal.

2.6.2. Phases I and II

Exploratory endpoints below are common to both Phase I and Phase II. They will be assessed separately for Phase I and Phase II and by disease group.

2.6.2.1. Immunological effects of hu3F8-BsAb

Lymphocyte subsets (including CD3, CD4, CD8, Foxp3, CD127, CD45RA, CCR7) and T cell activation/exhaustion markers (including CD25, PD-1) will be tested before and after the hu3F8-BsAb treatment. Each draw will collect about 16 mL of serum. Analysis will be performed in the Memorial Sloan Kettering (MSK) Immune Monitoring Core Facility. CBC should be checked with blood collected on the same day. Cytokine levels (including IFN-γ, IL-6, IL-8, IL-10) will be measured before and after the hu3F8-BsAb treatment during cycle 1 of phase I, and analyzed in the MSK Immune Monitoring Core Facility. During subsequent

cycles of phase I or in phase II, these cytokine levels will be measured if CRS is clinically suspected. Each draw for cytokine levels will collect about 4mL of serum.

Lymphocytes from heparinized blood will be collected for lymphocyte subset markers (including CD3, CD4, CD8, Foxp3, CD127, CD45RA, CCR7) and T cell activation/exhaustion markers (including CD25 and PD-1).

- Phase I Cycle 1: On Days 2, 8 (prior to hu3F8-BsAb), 9, 10 and 11, and approximately Days 15 and 28.
- For subsequent cycles in phase I, and phase II: Before each dose of hu3F8-BsAb.

2.6.2.2. Tumor DNA, RNA for NB patients only

For NB patients, tumor RNA will be used for transcriptome and microRNA profiling. Tumor RNA in BM from blood will be used for Minimal Residual Disease (MRD) testing, to be completed in the Cheung Lab at MSK. The samples will not have any associated patient-specific identifying information, but will be linked to clinical information such as patient age, gender, tumor stage, tumor recurrence status, and tumor site. Frozen tumors will be tested for tissue antigens and gangliosides (including but not limited to Disialotetrahexosylganglioside (GM2), GD2, Ganglioside Precursor Disialohematoside (GD3)).

2.6.2.3. Stool Microbiome

Based on emerging data on the impact of stool microbiome on function of immunotherapeutic agents, a stool sample will be collected on all patients prior to first dose of hu3F8-BsAb, where feasible.

2.6.2.4. Disease Control Rate (DCR) at Cycle 4

DCR is defined as :

- patients whose BOR at cycle 4 is CR, PR, MR or Stable Disease using INRC
- patients whose BOR at cycle 4 is CR, PR or Stable Disease, using RECIST v1.1.

2.7. Safety Variable(s)

The following safety endpoints will be assessed:

- **Adverse Events (AE)/ Serious Adverse Events (SAE)**, coded using the Medical Dictionary for Regulatory Activities (MedDRA) 23.1 and graded as per investigator judgement according to the NCI-CTCAE v4.0
- **Clinical laboratory** assessments as performed by local laboratories and entered in the eCRF by the site.
- **Vital Signs**
- **Physical Examination**

- **Pain Assessment for phase I if evaluable**

Pain scores will be assessed using the Face, Legs, Arms, Cry, Consolability (FLACC) scale [5],[6] and/or the numeric scale[7], as appropriate (see Appendix 2 for pain scales). Pain scores will be evaluated at the following time points on days with hu3F8-BsAb treatment:

- (a) once prior to commencement of any drug administration,
- (b) at least once during the acute pain episode when rescue pain medication doses are required (for hu3F8 this is towards the end of the 30 minute infusion or soon after completion of antibody infusion)
- (c) ~1 hour after the end of infusion or prior to discharge from the Pediatric Day Hospital.

Pain scores will be assessed for all treatment cycles. Opioid requirements, rather than pain score, will be used to establish safety for the Phase I primary objective (Section 2.2.1).
Pain scores will be used to assist clinical judgment (e.g. decision of pain medication use, observation of pain relief) and for documentation.

3. Estimand(s)

Not Applicable since no estimand was defined in the protocol.

4. Analysis Populations

In accordance with ICH E3 and E9[8], the following analysis sets will be used for the analyses.

4.1. All Included Analysis Set

The all included analysis set will be defined as all patients who are not screening failures, and hence included in the trial. Screening failures who had been treated by mistake will also be in the all included analysis set

4.2. All Screened Set

The All Screened Set will include every patient who has signed the informed consent form. The All Screened Set will be used for summaries of disposition and the associated listing.

4.3. Full Analysis Set

The Full Analysis Set (FAS) will consist of all patients treated with at least one dose of treatment.
The FAS will be used for summaries and analysis of demographics and baseline characteristics, treatments and medications, secondary efficacy endpoints and will be the primary analysis set for the analysis of the primary efficacy endpoints (phase II analyses).

4.4. Safety Analysis Set

All patients treated with at least one dose of the treatment will be included in the Safety Analysis Set.

4.5. DLT Evaluable Analysis Set (DAS)

The DLT evaluable analysis set (DAS) is a subset of the SAF. This analysis population will include patients in the dose escalation phase who receive both doses of hu3F8-BsAb in cycle 1 and had sufficient safety evaluations, or who had a DLT during the DLT evaluation period.

4.6. Pharmacokinetics Analysis Set

The PK Analysis Set will include all patients who received at least 1 dose of treatment (Hu3F8-BsAb) and have at least 1 valid PK concentration.

4.7. Special Subpopulations

Not Applicable

5. Data Handling

5.1. Time Points and Visit Windows

5.1.1. General Definitions

All assessment days will be related to the first day of first dose of treatment.

Day 1 is defined as first dose of treatment. Relative days after Day 1 are calculated as (assessment date – Day 1 date) + 1. Relative days prior to Day 1 are calculated as (assessment date – Day 1 date). The day prior to Day 1 is Day -1. Day 0 is not defined.

The date of the first dose of treatment for each patient will be taken from the Drug Administration eCRF page. If the date in this eCRF page is missing, alternatively the date of corresponding visit date in the Visit Date eCRF page will be used.

The date of the last dose of treatment for each patient will be taken from the End of Treatment eCRF page. If the date in this eCRF page is missing, alternatively the last date of Drug Administration eCRF page will be used.

5.1.2. Screening / Baseline Period

For all patients, the screening / baseline period is defined as the period from informed consent to the first dose of treatment. For some variables, data from more than one assessment within the screening / baseline period can be collected prior to the first dose of treatment.

The baseline value for a variable is therefore defined as the last non-missing value collected before the first dose of treatment in the screening / baseline period.

5.1.3. Treatment Period

Data collected at Day 1 will be assigned to the Treatment Period unless the time (HH:MM) of data collection and time (HH:MM) of first dose of treatment are both recorded and the data collection time is before the time of first dose of treatment. In this case, the assessment will be assigned to the screening / baseline period. If the time (HH:MM) of data collection is not recorded but the protocol and / or eCRF includes an instruction to the effect that all Day 1 assessments are to be performed prior to the first dose of treatment, the data collected at Day 1 will be assigned to the screening / baseline period. However, adverse events and medications starting on Day 1, will be assigned to the Treatment Period.

The Treatment Period is defined as the period from the date / time of the first dose of treatment up to and including the date / time of the last dose of treatment. For a definition of the treatment period for AEs, see section 6.7.3.

5.1.4. Visit Windows

All data will be analyzed using nominal trial visit as defined in Table 2 and Table 3 (see section 2.1.3) and eCRF. No visit windows will be applied for summary and analysis except for scan assessments for which the following visit windows will be considered:

- Cycle 2: 61 days +/- 30 days
- Cycle 4: 122 days +/- 30 days

For scans after cycle 4, supposed to occur every 3 months, the visit windows will be calculated using the date of the previous scan. For example:

- For the first scan after cycle 4, the windows will be: Cycle 4 scan date + 91 days (3 months) +/- 30 days;
- For the second scan after cycle 4, the windows will be: Date of first scan after cycle 4 + 91 days +/- 30 days; and so on for the next scans…

5.2. Handling of Dropouts, Missing Data, and Outliers

5.2.1. Handling of Missing Efficacy Data

Patients with missing data for ORR analysis will be treated as non-responders.
Additionally, for PFS analyses, patients without or missing evaluable tumor scan assessment will be censored as described in section 2.4.2.2.

5.2.2. Handling of Missing Safety Data

In general, missing clinical laboratory data, vital signs, and ECG data will not be imputed. Adverse event imputations for missing severity or relationship are given in Section 6.7.3. Unknown or partial medication and AE date imputations are given below and to be used only for the assessment of prior / concomitant status for medications and treatment-emergent status for AEs.

5.2.3. Handling of Partial and Missing Dates for Date of Birth, AE, Prior / Concomitant Medications

Partial and Missing Dates of Birth

Where the Date of Birth is missing the following convention will be used:

Where the day is missing and month and year are available the day will be completed as the 15th. For example, Date of Birth specified as --JAN1980 will be completed as 15JAN1980.

If the day and month are missing and the year is available, the day and month will be completed as 02JUL (the 183rd day of the year) if this is not the same year as the screening date. For example, Date of Birth specified as -----1980 will be completed as 02JUL1980.If this is the same year as screening date and screening date is after 02JULyyyy, same rule as above will be applied.

If this is the same year as screening date and screening date is before 02JULyyyy, it will be completed as the first day of the year (01JANyyy).

Partial and Missing Start Dates of Medical History

Where the Start Date of Medical History is missing, same rules as for dates of birth will be used.

If Start Date is completely missing, no imputation will be performed.

Missing or Partial Adverse Event and Prior / Concomitant Medication Start Dates

Missing and / or incomplete dates for medications and AEs are imputed in a manner resulting in the earliest onset or the longest duration during the Treatment Period, whilst ensuring that the start date does not occur after the stop date. The stop date will not be imputed if the medication or AE is "Ongoing". Technically, this will be done as follows:

For a missing / incomplete start date the earliest date of the following will be imputed:

- The later date of: the earliest possible start date, and the date of first dose of treatment.
- The latest possible start date.
- The latest possible stop date.

For a missing / incomplete stop date the later date of the following will be imputed:

- The earlier date of the latest possible stop date and the date of last dose of treatment.
- The earliest possible stop date.
- The earliest possible start date.

Here, the earliest possible date is defined as:

- The date itself if available.
- The date of the first day of the month, if month and year are available but the day is missing.
- The date of the first day of the year, if year is available but day and month are missing.
- the day of informed consent, if the date is completely missing.

The latest possible date is defined as:

- The date itself if available.

- The date of the last day of the month, if month and year are available but the day is missing.
- The date of the last day of the year, if year is available but day and month are missing.
- the date of last known date on the trial for the patient plus one year, if the date is completely missing.

5.2.4. Handling of Serum Concentrations that are Below the Lower Limit of Quantification

Serum concentrations that are below the lower limit of quantification (BLQ) will be handled as follows for descriptive statistics:

- Values that are BLQ will be set to LLOQ/2 for the calculation of summary statistics.
- Arithmetic mean or median values that are BLQ will be presented as LLOQ/2.

6. Statistical Methods

6.1. General Principles

All data processing, summarization and analyses will be performed using [redacted] SAS Environment / Version 9.4 (or later) of the SAS® statistical software package, except for the DLT analysis to estimate the MTD which will be performed using the CRM module in R software.

The default summary statistics for quantitative variables will be the number of observations (n), mean, standard deviation (SD), median, minimum (min) and maximum (max), for those patients with data.

All summary statistics will be rounded (using the SAS® function ROUND) and presented to one more decimal place than the raw value, except for the minimum and maximum values that will be presented with the same decimal precision as the raw value.

For qualitative variables, the number (n) and percentage (%) of patients with non-missing data per category will be the default summary presentation, and where appropriate and present, the number of missing values as a "Missing" category. When "Missing" is a category, number of patients in the analysis population will be used as denominator for percentages calculation, otherwise, the number of patients with non-missing data will be used as denominator (unless stated otherwise in the TFLs mock shell(s)).

All statistical comparisons will be made using two sided tests at the α=0.05 significance level unless specifically stated otherwise.

All laboratory test results will be received from the laboratories, and the results will be provided in both standard internal (SI) and conventional units. For the TFLs, the results will be summarized or presented in International System of Units (SI) units. Refer to Appendix of the TLFs mock shells for the SI unit corresponding to each laboratory test.

Refer to Appendix of the TLF shells for the precision level in which each laboratory test is reported by the laboratories.

Specifications for table, figures and data listing formats can be found in the TFL shells specifications for this trial. Please refer to "2. General Format Guidelines" section within TFL shells for more details on presentation of results.

The following treatment labels will be used for tables and listings in the different phases:

Table 5 Treatment Labels for Phase I

Ordering	Full treatment group (a minima for listings)	Short format labels (for use when space is limited) for tables
1	Hu3F8-BsAb 0.009 mcg/kg/cycle	0.009 mcg/kg
2	Hu3F8-BsAb 0.09 mcg/kg/cycle	0.09 mcg/kg
3	Hu3F8-BsAb 0.9 mcg/kg/cycle	0.9 mcg/kg
4	Hu3F8-BsAb 2.6 mcg/kg/cycle	2.6 mcg/kg
5	Hu3F8-BsAb 4.8 mcg/kg/cycle	4.8 mcg/kg
6	Hu3F8-BsAb 9.3 mcg/kg/cycle	9.3 mcg/kg
7	Hu3F8-BsAb 21.3 mcg/kg/cycle	21.3 mcg/kg
8	Hu3F8-BsAb 41.3 mcg/kg/cycle	41.3 mcg/kg
9	Hu3F8-BsAb 81.3 mcg/kg/cycle	81.3 mcg/kg
10	Hu3F8-BsAb 161.3 mcg/kg/cycle	161.3 mcg/kg
11	Hu3F8-BsAb 301.3 mcg/kg/cycle	301.3 mcg/kg
12	Hu3F8-BsAb 501.3 mcg/kg/cycle	501.3 mcg/kg

If dose is changed after cycle 2 in phase I, tumor response will be evaluated based on the combination of the doses. Thus, some additional columns may be added.

Table 6 Treatment Labels for Phase II

Ordering	Full treatment group (a minima for listings)	Short format labels (for use when space is limited) for tables
13	Hu3F8-BsAb RP2D Group 1 (Neuroblastoma)	Hu3F8-BsAb RP2D NB
14	Hu3F8-BsAb RP2D Group 2 (Osteosarcoma)	Hu3F8-BsAb RP2D OS
15	Hu3F8-BsAb RP2D Overall	Hu3F8-BsAb RP2D Overall

6.2. Patient Disposition and Data Sets Analyzed

Patient disposition will be summarized by treatment group and overall, where appropriate, for the All Screened Set. The following information will be reported:

- Number of patients for the following categories:
  - Screened,
- Number and percentage of patients for the following categories:
  - Included patients (100%),
  - Safety Analysis Set,
  - DLT Analysis Set,
  - Full Analysis Set,
  - PK Analysis Set
  - Discontinued the treatment,
    - Reasons for Treatment Discontinuation

- Ongoing in the trial,
- Discontinued the treatment during Cycle 1,
    - Reasons for Treatment Discontinuation during Cycle 1,
- Discontinued the treatment during Cycle 2,
    - Reasons for Treatment Discontinuation during Cycle 2,
- Discontinued the treatment during Cycle 3,
    - Reasons for Treatment Discontinuation during Cycle 3,
- Discontinued the treatment during Cycle 4,
    - Reasons for Treatment Discontinuation during Cycle 4,
- Discontinued the Trial,
    - Reasons for trial discontinuation.

- Number and percentage of patients at each country/ site.

A patient will be considered as having discontinued the treatment if they have a non missing reason for ending treatment in the End of Treatment form at time of reporting. To discontinue treatment during a cycle is interpreted to mean that the last imp received by the subject is one of the two doses in the cycle.

A patient will be considered as having discontinued the trial if they have a reason for coming off trial in the Off Study form at time of reporting. Otherwise, the patient will be considered as ongoing in the trial.

If applicable, treatment discontinuation due to Coronavirus Disease of 2019 (COVID-19) will also be described as well as missed visits/ missed assessments due to Covid-19.

A listing of all included patients with their treatment and trial completion status, including the respective reasons for treatment and trial discontinuation will be presented for the All Included Analysis Set.

A listing of all screen failed patients with their reasons for screen failure will be presented for the All Screened Set. A separate listing of patients who failed at least one inclusion / exclusion criteria including a text description of the criterion failed will be presented for All Included Analysis Set.

A listing of all patients excluded from at least one analysis set will be presented for the All Included Analysis Set, together with the reason for exclusion, if applicable.

6.3. Protocol Deviations

Deviations from the protocol, as defined in the protocol and / or protocol deviation plan, will be documented by the trial monitors and project management throughout the trial period.

Criteria which require clinical or medical monitoring interpretation will be reviewed prior to database lock. All important protocol deviations occurring during the trial will be reviewed and approved by Y-mAbs prior to database lock.

All important protocol deviations will be summarized for the Safety Analysis Set by treatment group and overall as described below:

- The number of unique patients with at least one important protocol deviation as well as the number of patients in each important protocol deviation category will be presented by default descriptive summary statistics for categorical variables.

A listing of all patients with one or more important / non important protocol deviations will be presented for the All Screened Set. If applicable, a listing of the patients affected by COVID-19 and the type of COVID-19 disruption will be provided for the All Screened Set.

Number and percentage of patients with at least one deviation related to COVID-19 will also be described by treatment group and type of deviations.

6.4. Demographic and Other Baseline Characteristics

6.4.1. Demographic Characteristics

Demographic characteristics will be summarized for the Safety Analysis Set by treatment group and overall and as described below. All missing data will be presented as part of a missing category, if appropriate. Standard descriptive statistics will be presented for the continuous variables of:

- Age (years)
  [calculated as (screening visit date – date of birth) / 365.25 and reported to 1 decimal place. See Section 5.2.3 for missing date of birth imputation.]
- Height (cm) at baseline
- Weight (kg) at baseline
- Body mass index (kg/m^2) at baseline
  [calculated as (body weight / $height^2$) where weight is in kg and height is in m] and presented to one decimal precision
- Systolic Blood Pressure (SBP) [mmHg] at baseline
- Diastolic Blood Pressure (DBP) [mmHg] at baseline
- Pulse Rate (beats per minute [bpm]) at baseline

Total counts and percentages of patients will be presented for the categorical variables of:

- Age group (years):
  - 0-<7
  - 7-<12
  - 12-<18
  - ≥ 18
- Sex
- Race
- Ethnicity
- Female of Childbearing potential.

Demographic characteristics will be listed for the SAS.

6.4.2. Baseline Characteristic

Baseline characteristics will be summarized for the Safety Analysis Set by treatment group and overall as described below. All missing data will be presented as part of a missing category, if appropriate.

Standard descriptive statistics will be presented for the continuous variables of:

- Time since initial Diagnosis (in months) calculated as (Screening date – Diagnosis Date +1)/30.4375

Total counts and percentages of patients will be presented for the categorical variables of:

- Disease Type (Neuroblastoma / High Grade Osteosarcoma / Other GD2-expressing solid tumor)
- Disease Histology (Neuroblastoma / Ganglioneuroblastoma)
- INSS Stage (1 / 2A / 2B / 3 / 4 / 4S)
- MYCN Amplification (Yes / No / Unknown)

No formal tests of statistical significance will be performed on the demographic and baseline characteristics data.

Baseline characteristics will be listed for the Safety Analysis Set.

6.4.3. Medical History

Medical history is defined as any condition, with the exception of the trial indication, that the patient may have had prior to enrollment in the trial, including any chronic conditions

diagnosed prior to entry in the trial. See section 5.2.3 for imputation of missing or partial dates for medical history.

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) [23.1] and will be presented by System Organ Class (SOC) and Preferred Term (PT) and total. The SOCs and PTs are to be sorted by Internationally Agreed order SOCs and descending PTs in the total column.

Medical history records will be summarized for the Safety Analysis Set by treatment group and overall as follows:

- The number and percentage of patients with at least one medical history record will be presented.
- The number and percentage of patients with at least one medical history record within each primary SOC and PT will be presented. The summary will be sorted using the internationally agreed order for SOC and using descending order of overall numerical counts for PT. Where terms tie, these will be sorted alphabetically.

Medical history records will be listed by-patient and within-patient by medical history start date for the Safety Analysis Set.

6.4.4. Prior and Concomitant Medications

All medications will be coded using the WHO Drug Global Dictionary, Global B3 Sep2020 Anatomical Therapeutic Chemical (ATC) Classification codes.

Prior medications and concomitant medications are defined as follows:

- Prior medications are those taken prior to screening with a stop date and time prior to the start of the Treatment Period.
- Concomitant medications are those with a start date and time on or after the start of the Treatment Period, or those with a start date and time before the start of the Treatment Period and either a stop date and time on or after the start of the Treatment Period, or are ongoing at the end of the trial.
- Concomitant medications will be further divided according to whether they started before (both prior and concomitant) or after (concomitant-only) the start of the Treatment Period.

See Section 5.2.3 for imputation of missing or partial dates for medication.

Prior and concomitant medications will be summarized separately for the Safety Analysis Set by treatment group and overall as follows:

- The number and percentage of patients with at least one prior / concomitant medication will be presented.

- The number and percentage of patients with at least one prior / concomitant medication within each Anatomical Group (ATC Level 1), Therapeutic Subgroup (ATC Level 2), and preferred term will be presented. The summary will be sorted using numerical counts by descending order of Anatomical Group, then descending order of Therapeutic Subgroup, then descending order of preferred term in the total column. Where groups or terms tie these will be sorted alphabetically.

Concomitant Medications to treat COVID-19 will also be summarized and listed by ATC code level 4, Preferred Drug Name and presented by treatment arm using the Safety Analysis Set.

Prior medications and concomitant medications will be listed separately for the Safety Analysis Set. In the listings the relative start and stop day of prior / concomitant medication use will be calculated relative to the first dose date and time of treatment and will be presented for those patients who received at least one dose of treatment. If the concomitant medication is "Ongoing" it will be indicated as such in the listing and the relative stop day will not be calculated. Only original dates will be presented in the listing even though the relative day may be based on an imputed date.

6.5. Measurements of Treatment Compliance

Treatment compliance is defined as the number of infusions that were actually received relative to the number of planned infusions as per the protocol for the duration of actual treatment exposure.

In general, the percentage overall compliance, assessed by infusion count, will be calculated as follows:

$$Compliance\ (\%) = \frac{Number\ of\ infusions\ actually\ received}{Number\ of\ planned\ infusions} \times 100$$

Infusions are planned on days 1 and 8 of each cycle, each cycle being of 28 days' duration therefore the number of planned infusions on treatment will be calculated as follows:

- If Day of last visit of the patient is between 1 and 7 during Cycle N, then, number of planned infusions is 2 x N – 1
- If Day of last visit of the patient is between 8 and 28 during Cycle N, then, number of planned infusions is 2 x N

Percent Compliance will be summarized for the Safety Analysis Set by treatment group.

Treatment compliance will be listed together with exposure for the Safety Analysis Set. Missing data will not be imputed and only original data for those fields (for example, date fields) will be presented in the listing together with derived variables such as the calculated compliance (%) and exposure duration.

6.6. Efficacy

The primary objectives of phase II are:

- In Group 1: Assess the ORR of hu3F8-BsAb in patients with relapsed or refractory NB
- In Group 2: Assess the PFS in patients with relapsed or refractory osteosarcoma at 4 months from the first hu3F8-Bsab treatment.

All information collected about scan assessments and bone marrow assessments will be presented in a data listing in patients from the FAS.

6.6.1. Primary Efficacy Analysis for Phase II

6.6.1.1. ORR for Group 1

The primary efficacy variable of relapsed or refractory NB patients of phase II is defined as ORR at the end of cycle 4 excluding potential response assessments after initiation of any new prohibited anticancer treatment.

The null hypothesis is that the hu3F8-BsAb treatment has ORR < 25%. The alternative hypothesis is that the ORR after the hu3F8-BsAb treatment is ≥ 25%.

90% and 95% two-sided confidence intervals for ORR will be computed using Wilson's method. If the two-sided 90% confidence interval does not include the null ORR rate of 25% the treatment will be considered as successful.

Patients in the FAS will be included into the assessment even if they are removed from the trial before completing cycle 4 due to any reasons.

The number and percentage of responder patients (Yes/ No) with details by best response categories (CR, PR and MR versus Stable Disease, Progressive Disease, Not Evaluable, Missing) will be described in the FAS and the confidence interval using the Wilson's method will be provided.

In addition, overall responses at each cycle will excluding potential response assessments after initiation of any new prohibited anticancer treatment be provided with the 90% and 95% confidence intervals using the Wilson's method.

Patients whose CRF response is "Very Good Partial Response" will be considered as "Partial Response" for the analyses.

ORR values will be also presented in a data listing as well as overall responses at each cycle.

6.6.1.2. 4-month PFS for Group 2

The primary efficacy variable of relapsed or refractory osteosarcoma patients of phase II is defined as 4-month PFS.

The null hypothesis is that the hu3F8-BsAb treatment has 4-month PFS <12%. The alternative

hypothesis is that the 4-month PFS after hu3F8-BsAb treatment is ≥ 12%.

The Kaplan-Meier estimate of the median time will be provided with 25th percentile and 75th percentile as well as PFS rates at 4 months with corresponding 90% and 95% Confidence Intervals.

In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk.

Frequency (number and percentage) of patients with an event (progression or death) and censoring reasons will be presented. Patients in the FAS will be included into the assessment even if they are removed from the trial before completing cycle 4 due to any reasons. See section 2.4.2.2 for details of censoring reasons.

4-month PFS values will be also presented in a data listing.

6.6.2. Sensitivity for the Primary Analyses

For ORR analysis in group 1 of phase II, the ORR at the end of the trial excluding potential response assessments after initiation of any new prohibited anticancer medication will be also studied and the two-sided confidences intervals (90% and 95%) using the Wilson's method will be provided.

6.6.2.1. Analyses Addressing Underlying Assumptions of Models

Not applicable

6.6.2.2. Analyses Addressing the Effect of Stratification Variable(s)

Not applicable

6.6.2.3. Analyses Addressing the Effect of Intercurrent Events

Not applicable

6.6.2.4. Analyses Addressing the Effect of Missing Data

Not applicable

6.6.3. Secondary Efficacy Analysis

6.6.3.1. ORR for Phase I

ORR for patients from phase I will be estimated for each disease type (excluding potential response assessments after initiation of any new prohibited anticancer treatment) and provided

with 90% and 95% two-sided Confidence Intervals using Wilson's method, by treatment group in the FAS.

For NB, anti-tumor activity will be measured by revised INRC, as described in section 2.4.2.1.

For osteosarcoma and other GD2(+) tumors, the response and progression will be evaluated in this trial using the RECIST Committee, version 1.1. as described in Section 2.5.1.3. The proportion of NB, osteosarcoma, and other GD2(+) tumor patients responding to therapy will be determined for each disease type.

In addition, overall response at each cycle excluding potential response assessments after initiation of any new prohibited anticancer treatment will be provided with the two-sided 90% and 95% confidence intervals using the Wilson's method.

Swimmer plot will also be produced (split by disease type if needed) with the following characteristics:

- Responses at each cycle over time for X axis starting from the first treatment date
- Dose levels and patient numbers for the Y axis.

Patients whose CRF response is "Very Good Partial Response" will be considered as "Partial Response" for the analyses.

ORR values will be also presented in a data listing as well as overall responses at each cycle.

6.6.3.2. PFS for Phase I

PFS analyses will be performed in patients from the FAS.

The number of overall censored patients, number of patients with events and Kaplan-Meier estimates median time will be provided with 25th percentile and 75th percentile and 95% Confidence Interval.

In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk.

Frequency (number and percentage) of patients with an event (progression or death) and censoring reasons will be presented by treatment group. See section 2.4.2.2 for details of censoring reasons.

PFS values will be also presented in a data listing.

6.6.3.3. OS for Phase I

OS analyses will be performed in patients from the FAS.

The number of overall censored patients, number of patients with events and Kaplan-Meier estimates median time will be provided with 25th percentile and 75th percentile and 95% Confidence Interval.

In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk.

In addition, OS rates at different time points (every 6 months) will be estimated with corresponding 2-sided 95% CIs.

Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment group. Censoring reasons are as follows:

- Alive
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include patients that the investigator states were lost to follow-up prior to the analysis cut-off.

OS values will be also presented in a data listing.

6.6.3.4. Duration of Complete Remission for Phase II

Duration of CR is defined only for patients who achieve CR after hu3F8-BsAb and will be calculated from the time of remission by cancer type group in patients from the FAS.

The number of overall censored patients, number of patients with events and Kaplan-Meier estimates median time will be provided with 25th percentile and 75th percentile and 95% Confidence Interval. In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk. Reasons for censoring will also be provided.

Duration of CR will also be presented in a data listing.

6.6.3.5. ORR for Group 2 of Phase II

ORR for patients from group 2 will be analyzed as ORR for patients from group 1 (See section 6.6.1.1) in the FAS.

The number and percentage of responder patients (Yes/ No) with details by best response categories (CR, PR versus Stable Disease, Progressive Disease, Not Evaluable, Missing) will be described in the FAS and the two-sided 90% and 95% confidence intervals using the Wilson's method will be provided.

In addition, overall responses at each cycle excluding potential response assessments after initiation of any new prohibited anticancer treatment will be provided with the 90% and 95% confidence intervals using the Wilson's method.

Patients whose CRF response is "Very Good Partial Response" will be considered as "Partial Response" for the analyses.

ORR values will be also presented in a data listing as well as overall responses at each cycle.

6.6.3.6. HAHA for Phases I and II

HAHA will be measured as previously described. The proportion of patients developing HAHA response (Positive / Negative) after hu3F8-BsAb therapy, as well as the continuous titer will be summarized at each timepoint for patients from the Safety Analysis Set.

In addition, HAHA data will be listed based on the Safety Analysis Set.

6.6.3.6.1. Additional analyses HAHA for Phase I

HAHA titer value over time will be presented graphically with time of dosing for patients in the Safety Analysis Set.

In addition, the correlation between HAHA response and ORR after cycle 4 (for neuroblastoma patients) will be analyzed using a logistic regression with inclusion of log(therapeutic dose) as a covariate and whether patients have HAHA at any point up until the last pre-dose HAHA assessment before cycle 4. If patients have switched dose after cycle 2, the average therapeutic dose will be used.

Similar analysis will be performed with DCR after cycle 4 (for osteosarcoma patients).

The association between HAHA and OS and PFS up to 4 months will be analysed by cancer type group using Cox-Regression analyses with log(therapeutic dose) and HAHA (Yes/ No) as time-dependent covariates. These analyses will be based on patients from the FAS.

This information will be summarized separately by cancer type in patients.

6.6.4. Sensitivity Analyses for the Secondary Efficacy Analysis

For ORR analysis of phase I and in group 2 of phase II, the ORR at the end of trial but before new prohibited anticancer therapy will be also studied and the two-sided confidences intervals (90% and 95%) using the Wilson's method will be provided.

6.6.5. Subgroup Analysis

Not applicable

6.6.6. Exploratory Analysis

6.6.6.1. OS for Phase II

OS is defined as the time from first dose of treatment until death from any cause. All deaths will be included, whether they occur during the trial or following treatment discontinuation.

Data will be presented by cancer type (NB versus Osteosarcoma) based on the FAS.

The number of overall censored patients, number of patients with events and Kaplan-Meier estimates median time will be provided with 25^{th} percentile, 75^{th} percentile and 95%

Confidence Interval. In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk.

In addition, OS rates at different time points (every 6 months) will be estimated with corresponding 2-sided 95% CIs.

Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment group. Censoring reasons are as follows:

- Alive
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include patients that the investigator states were lost to follow-up prior to the analysis cut-off.

OS values will be also presented in a data listing.

6.6.6.2. PFS for Group 1 of Phase II

PFS will be defined as the time from first dose of treatment to the earliest documented date of disease progression, or death due to any cause, whichever occurs first.

The number of censored patients, number of patients with events and Kaplan-Meier estimates median time will be provided with 25th percentile and 75th percentile and 95% Confidence Interval. In addition, the results will be presented graphically in Kaplan-Meier plots including count for numbers of patients at risk.

Frequency (number and percentage) of patients with an event (progression or death) and censoring reasons will be presented by treatment group. See section 2.4.2.2 for details of censoring reasons.

PFS values will be also presented in a data listing.

6.6.6.3. Immunological effects of hu3F8-BsAb for Phases I and II

Biologic correlate studies include (1) host immunity, (2) host T cell cytotoxic capacity, and (3) T cell homing into the tumor.

The parameters measured for each patient under “host immunity” include:

- the induction of HAHA,
- the induction of Ab3,
- the induction of Ab3’.

Under “T cell cytotoxic capacity”, the following measurements will be made:

- (1) cytokine profiles including IFN-γ, IL-6, IL-8, IL-10,

- (2) changes in lymphocyte populations and their differentiation including CD3, CD4, CD8, Foxp3, CD127, CD45RA, CCR7, and
- (3) activation and exhaustion markers on T cells including CD25 and PD-1.

"T cell homing" will be tested by immunohistochemistry of tumor samples if available.

For phase I, spaghetti plots of available data will be performed based on patients from the Safety Analysis Set.

For phase II, available data will be described by dose group (Neuroblastoma vs Osteosarcoma) based on patients from the Safety Analysis Set.

Available data will be also presented in a data listing.

6.6.6.4. Tumor DNA, RNA for NB patients only for Phases I and II

For NB patients only: tumor DNA and RNA matched to normal RNA, will be archived to profile tumor gene aberrations and gene expressions as well as to test correlation with MRD.

This information will become relevant when combined with genome and expression data from current and future studies of hu3F8-BsAb.

Given the heterogeneity of the diagnosis and stage in this trial, this objective is exploratory. No formal statistical comparisons will be made.

Tumor DNA and RNA will be described in patients from the Safety Analysis Set.

Relevant, available data will be presented in summaries/listing if deemed applicable.

6.6.6.5. Stool Microbiome

Sequencing of stool microbiota will be performed and correlated with clinical and immunological response in patients from the Safety Analysis Set.

Relevant, available data will be presented in summaries/listing if deemed applicable.

6.6.6.6. DCR

DCR at the end of cycle 4 for patients from phase I will be estimated for each disease type and provided with 90% and 95% Confidence Intervals using Wilson's method, by treatment group in the FAS.

For patients from phase II, DCR analyses will be performed by group (NB versus Osteosarcoma) and two-sided 90% and 95% Confidence Intervals using Wilson's method will be provided.

DCR values will be also presented in a data listing.

6.7. Safety

6.7.1. Extent of Exposure

Duration of exposure will be defined in months as:

- If last dose is a priming dose, then:

Exposure (months) = ([date of last dose – date of first dose] + 7)/30.4375

- If last dose is a therapeutic dose, then:

Exposure (months) = ([date of last dose – date of first dose] + 21)/30.4375

Duration of exposure in days will be summarized for the Safety Analysis Set by treatment group and overall. Descriptive statistics will be presented for duration of exposure.

The duration of exposure will also be computed in weeks (dividing by 7 the above computations), and summarized by treatment group, in patients from the Safety Analysis Set for following variables:

- Duration as continuous (in weeks) using descriptive statistics
- Categorized by time intervals (< 4 weeks, 4-<8 weeks, …) for which frequency counts and percentages of patients will be provided.
- Descriptive statistics will also be summarized for the total number of cycles received per patient, the number of doses received at Day 1, the number of doses received at Day 8.

The total dose of treatment in mcg/kg taken over the trial duration will be calculated as the total sum of all infusions received during the trial.

The total dose of treatment in mcg/kg taken over the trial duration will be summarized alongside dosing changes for the Safety Analysis Set by treatment group and overall as:

- Descriptive statistics will be presented for total dose of treatment (sum of actual dose received).

A listing of overall treatment exposure data, including the first and last dates of treatment will be presented together with compliance for the Safety Analysis Set. Further, trial treatment infusion data will be listed for the Safety Analysis Set.

Treatment interruptions due to Covid-19 will also be described if applicable.

6.7.2. DLT for Phase I

A DLT is defined as one or more events occurring during cycle 1. For more details, see Section 2.1.1.5.

A summary table of DLTs by treatment group occurring during cycle 1 will be provided, presenting the number and percentage of patients with DLTs, the number of DTLs per patient, and the type of DLTs in patients of phase 1 from the safety DLT Evaluable Analysis Set.

A by-patient listing of DLTs that occur during cycle 1 is planned by treatment group and for all patients from the DLT Evaluable Analysis Set during phase I of this trial. Patients will be grouped by the treatment group to which they are originally assigned, including those who may receive subsequent treatment at a lower or higher dose level.

Patients with and without DLTs from the DLT Evaluable Analysis will be plotted in a graph with dose level on the vertical axis and patient number (patient id sorted in the order in which they received the first dose) on the horizontal axis.

MTD will initially be estimated as the next suggested dose level from the CRM model described above based on all data in the DLT analysis set. This corresponds to the dose level with a posterior probability of DLT that comes closest to the toxicity level of 0.15.

Final MTD is determined by the DMC and the safety group, as other safety data than the DLTs may be relevant

6.7.3. Adverse Events

An AE is any untoward medical occurrence in a patient administered a pharmaceutical product which does not necessarily have a causal relationship with the treatment.

An AE can, therefore, be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a product, whether or not considered related to the product.

An AE includes:

- A clinically significant worsening of a concomitant illness.
- A laboratory abnormality which is clinically significant, i.e. an abnormality that suggests a disease and/or organ toxicity and is of a severity that requires active management. Active management includes interventional treatment or further investigations, for example change of medicine dose or more frequent follow-up due to the abnormality.

Each AE is to be classified by the Investigator as either serious or non-serious. This classification of the seriousness of the AE determines the reporting procedures to be followed. An AE that meets one or more of the following criteria/outcomes is classified as serious:

- Is fatal or life-threatening
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability/incapacity
- Is a congenital anomaly/birth defect
- Medically important event

All AEs recorded on the eCRF will be coded using the MedDRA dictionary [23.1].

The overall observation period for AEs will be divided into 3 parts:

- Pre-treatment period: From the day of the patient's informed consent to the day before the first dose of Hu3F8-BsAb
- On-treatment period: From the day of the first dose of Hu3F8-BsAb to 30 days after the last dose of Hu3F8-BsAb
- Post-treatment period: From 31 days after the last dose of Hu3F8-BsAb and thereafter

If dates are incomplete in a way that clear assignment to pre-, on-, post-treatment period cannot be made, the respective data will be assigned to the on-treatment period.

AE will be classified as Treatment–Emergent AEs (TEAEs) as follows:

- TEAEs are either events with start date and time after initiation of treatment, or events with start date and time prior to the start of the on-treatment period whose severity worsens on or after initiation of treatment.

  Some AE of the on-treatment period will not be TEAE if they occur on the day of the first treatment but before the time of the first treatment. If start time is incomplete in a way that clear assignment as TEAE/Non-TEAE cannot be made, the AE will be considered as TEAE.

  In addition, if AE starts (date/time) before first dose and grade is greater than 1 and stop (date date/time) is after first dose or missing, then this AE will be considered as a TEAE.
- Treatment-Emergent Serious AEs (TESAEs) will be defined as TEAEs regarded by the investigator as Serious = "Yes".
- The relationship between a TEAE and Hu3F8-BsAb is assessed as definite, probable, possible, or not related. A Hu3F8-BsAb-related TEAE will be defined as a TEAE considered by the investigator as definitely, possibly, or probably related to Hu3F8-BsAb or with unknown / missing relationship to Hu3F8-BsAb.
- Assessment of AE severity / intensity will be based on the National Cancer Institute-Common Terminology Criteria for AEs (NCI-CTCAE[9], version 4.0) apart from CRS events that will be graded using Lee grading system.. Severe TEAEs are defined as TEAEs assessed as being "Severe (Grade 3) / Grade 4 / Grade 5" in intensity.
-
- TEAEs leading to discontinuation of treatment are defined as TEAEs where "Action Taken with Study Treatment" is indicated as "Treatment discontinued".

Adverse events will be summarized by default descriptive summary statistics for categorical variables for the Safety Analysis Set by treatment group (phase I) and cancer type (phase 2) and overall as follows:

- An overview of TEAEs with onset during the pre-treatment or the treatment period including the number and percentage of patients with at least one of each mentioned TEAE type as well as the number of events:
    - Any TEAE
        - Leading to discontinuation of trial treatment
        - Leading to death
        - Grade 1 severity (mild)
        - Grade 2 severity (moderate)
        - Grade 3 severity (severe)
        - Grade 4 severity (life-threatening or disabling)
        - Grade 5 severity (death related to TEAE)
        - Missing severity grade
    - Any treatment related TEAE
        - Leading to discontinuation of trial treatment
        - Leading to death
    - Any serious TEAE
        - Leading to discontinuation of trial treatment
        - Leading to death
    - Any serious treatment related TEAE
        - Leading to discontinuation of trial treatment
        - Leading to death

- For TEAs with onset during the pre-treatment or the treatment period the number and percentage of patients reporting each TEAE and the count of number of events will be summarized by SOC and PT for the following types of TEAEs:
    - TEAEs
    - TEAEs Leading to Discontinuation of Trial Treatment
    - TEAEs Leading to Death
    - TEAE Related to COVID-19 if applicable
    - Grade 3, 4 or 5 TEAE
    - Infusion Related TEAE (defined as all related TEAEs occurring on the day of infusion after initiation of infusion if times are not missing; otherwise, if time is missing, it will correspond to all related TEAEs occurring on the day of infusion)
    - Grade 3, 4 or 5 Infusion Related TEAE
    - Trial Treatment Related TEAEs
    - Grade 3, 4 or 5 Related TEAEs
    - Treatment Related TEAEs Leading to Discontinuation of Trial Treatment
    - Treatment related TEAEs Leading to Death
    - Serious TEAEs
    - Serious TEAEs of grade 3 or higher
    - Serious TEAEs Leading to Discontinuation of Trial Treatment
    - Serious TEAEs Leading to Death
    - Related Serious TEAEs of grade 3 or higher

  - Treatment Related Serious TEAEs
  - Treatment Related Serious TEAEs Leading to Discontinuation of Trial Treatment
  - Treatment Related Serious TEAEs Leading to Death
- The number and percentage of patients who died will be summarized by the primary reason of death

In the above summaries, patients with more than one TEAE within a particular SOC are counted only once for that SOC. Similarly, patients with more than one TEAE within a particular PT are counted only once for that PT.

Summaries by SOCs and PTs will be sorted by SOCs by their Internationally Agreed Order (MedDRA 23.1) and PTs within SOC by descending order of total incidence. Where counts preferred terms tie, PTs will be sorted alphabetically.

No statistical comparisons of AEs between treatment groups will be performed.

All AE data will be listed separately and Pre-treatment period, the On-Treatment period and the Post-treatment period. Treatment-emergence status will be flagged. The listings will present the relative start and stop day of the AE calculated relative to the first dose of treatment and will be presented for those patients who received at least one dose of treatment. If the AE is "Ongoing" it will be indicated as such in the listings and the relative stop day will not be calculated. Only original dates will be presented in the listings even though the relative day may be based on an imputed date.

In addition, the following listings will be presented combined for all three periods (a column with onset period (pre-treatment, on-treatment, post-treatment) will be displayed):

- Listing of Deaths
- Listing of Serious TEAEs
- Listing of AEs Leading to Interruption of Trial Treatment
- Listing of AEs Leading to Discontinuation of Trial Treatment
- Listing of AE related to COVID-19, if applicable
- Listing of non-TEAEs

6.7.4. Laboratory Evaluations

Data for the following hematology, serum chemistry, and urinalysis parameters recorded in the eCRF are to be measured at the scheduled visits / time points indicated in Table 2 and Table 3 (see section 2.1.3).

Table 7 Laboratory Tests

Hematology Test (SI unit)	Serum Chemistry Test (SI unit)	Blood Urinalysis

• Red Blood Cell Count (RBC) (10^{12}/L)
• Hemoglobin (g/L)
• Hematocrit (%)
• White Blood Cell Count (WBC) (10^9/L)
• Differential WBC (10^9/L and %)
  ○ Neutrophils
  ○ Lymphocytes
  ○ Eosinophils
  ○ Basophils
  ○ Monocytes
• Platelet count (10^9/L)
• Mean Corpuscular Volume (MCV)
• Mean Corpuscular Hemoglobin (MCH)
• Mean Corpuscular Hemoglobin Concentration (MCHC)
• Red cell Distribution Width (RDW)
	• ALT (U/L)
	• AST (U/L)
	• Total Bilirubin (mmol/L)
	• Total Protein (g/L)
	• Phosphate (mmol/L)
	• Alkaline phosphatase (U/L)
	• Albumin (g/L)
	• Creatinine (mmol/L)
	• Blood Urea Nitrogen (BUN) (mmol/L)
	• Potassium (mmol/L)
	• Sodium (mmol/L)
	• Magnesium (mmol/L)
	• Chloride (mmol/L)
	• Carbon Dioxide (Bicarbonate) (mmol/L)
	• Ferritin (nmol/L)
	• C-reactive Proteins (mg/L)
	• Lactate Dehydrogenase (ukat/L)
		• Color
		• Specific Gravity
		• pH
		• Protein/ Albumin
		• Glucose
		• Ketones
		• Bilirubin
		• Urobilinogen
		• Leucocyte Esterase
		• Nitrate
		• RBC
		• WBC
		• Epithelial Cells
		• Small Round Cells
		• Casts
		• Pathological Casts
		• Sperm
		• Bacteria
		• Yeast

In accordance with the baseline value definition in Section 5.1.2, the absolute and percentage change from baseline will be derived as follows:

Absolute change (unit) = (post-baseline value – baseline value)

Percentage change from baseline = [(post-baseline value– baseline value) / baseline value] × 100

All laboratory data will be reported in SI units. All quantitative laboratory test values at each assessed timepoint will be compared with the relevant reference range in SI units and categorized as:

- Low: Below the lower limit of the reference range.
- Normal: Within the reference range (upper and lower limits included).
- High: Above the upper limit of the reference range.

For analysis purposes, values preceded by a "<" (i.e. those below the limits of quantification) will be considered equal to the half of the Lower Limit Of The Quantification (LLOQ/2). Those preceded by a ">" sign (i.e. those above the limits of quantification) will be considered equal to the upper limit of quantification.

Laboratory values will be assigned toxicity grades, when available*, using criteria based on the NCI CTCAE version 4.03[9] .

Parameters whose grade can be derived are the followings:

- Hematology
    - Hemoglobin (Low)
    - Hemoglobin (High)
    - WBC (Low)
    - WBC (High)
    - Neutrophils (absolute) (Low)
    - Lymphocytes (absolute) (Low)
    - Lymphocytes (absolute) (High)
    - Platelet (Low)

- Serum chemistry
    - Sodium (Low)
    - Sodium (High)
    - Potassium (Low)
    - Potassium (High)
    - Magnesium (Low)
    - Magnesium (High)
    - Creatinine (High)
    - Total Bilirubin (High)
    - ALT (High)
    - AST (High)
    - ALP (High)
    - Albumin (Low)

Laboratory data will be summarized by default descriptive summary statistics for continuous and categorical variables for the Safety Analysis Set by treatment group and overall as follows:

- Observed values and absolute change from baseline at each assessed timepoint for each standard continuous laboratory parameter;
- Number and percentage of patients with categorized shift (low, normal and high) values relative to the reference range at baseline compared to each post-baseline value for hematology and serum chemistry (for phase II only);
- Number and percentage of patients with worst NCI-CTCAE toxicity values;
- Number and percentage of patients with categorized shift NCI-CTCAE toxicity values at baseline compared to worst post-baseline timepoint.

Listings of all clinical laboratory data including derived percentage change from baseline will be provided for the Safety Analysis Set. Within each listing, laboratory values outside the

normal ranges will be flagged as either high or low and NCI-CTCAE grading will be provided for applicable laboratory assessments.

In addition, the following figures will be displayed:

- Longitudinal plot of laboratory parameters in patient with at least one abnormal value (Three different plot symbols will indicate whether the values were Low, Normal or High)
- Baseline and post baseline plot of laboratory parameters per cycle and per week

6.7.5. Vital Signs

The analyses described below will be conducted for the following vital signs assessments respectively:

- SBP (mmHg);
- DBP (mmHg);
- Heart Rate (HR) (bpm);
- Pulse Oximeter at rest (%);
- Pulse Oximeter at exertion (%);

In accordance with the baseline value definition in Section 5.1.2, the absolute and percentage change from baseline will be derived as follows:

Absolute change (unit) = (post-baseline value – baseline value)

Percentage change from baseline = [(post-baseline value – baseline value) / baseline value] × 100

The following will be summarized by treatment group (phase I) and cancer type (phase II) and overall for the Safety Analysis Set:

- Observed values and absolute change from baseline at each assessed timepoint for each standard vital sign parameter using default summary statistics for continuous variables

A listing of all vital signs data including derived percentage change from baseline will be provided for the Safety Analysis Set.

In addition, the following figures will be displayed:

- Longitudinal plot of vital signs parameters in patient with at least one abnormal value
- Baseline and post baseline plot of vital signs parameters per cycle and per week
- Scatter plot of vital signs parameters per cycle and per week

6.7.6. Physical Examination

Abnormalities identified from physical examination are recorded in the eCRF as Medical History or AEs as appropriate and will be listed and summarized as such [See Sections 6.4.3 (Medical History) and 6.7.3 (Adverse Events)].

For each physical examination body system, the number and percentage of patients with abnormalities at baseline and at each assessed timepoint will be summarized by treatment group and overall for the Safety Analysis Set.

Physical examination findings (normal / abnormal) and details of abnormalities will be listed for each patient at each assessed timepoint.

6.7.7. Other Safety Variables

6.7.7.1. Pain Assessment for Phase I if evaluable

Pain assessment data analyses are detailed below. They will be performed if data are evaluable.

In accordance with the baseline value definition in Section 5.1.2, the absolute and percentage change from baseline will be derived as follows:

Absolute change = (post-baseline value – baseline value)

Percentage change from baseline = [(post-baseline value – baseline value) / baseline value] × 100

⇨ FLACC Scale

Number and percentages of patients by treatment group and overall in patients from the Safety Analysis Set will be provided for the following categories:

- Face (0 / 1 / 2)
- Legs (0 / 1 / 2)
- Activity (0 / 1 / 2)
- Cry (0 / 1 / 2)
- Consolability (0 / 1 / 2)

In addition, Assessment of Behavioural Score will be derived by summing all individual scores of the FLACC scale (Face + Legs + Activity + Cry + Consolability) and will result in a total score between 0 and 10.

In addition, from this score, the following categories will be derived:

- 0 = Relaxed and comfortable
- 1-3 = Mild discomfort
- 4-6 = Moderate pain

- 7-10 = Severe discomfort/pain

Descriptive statistics of this score will be provided in patients from the Safety Analysis Set:

- Observed values and absolute change from baseline at each assessed timepoint using default summary statistics for continuous variables
- A categorical summary will be provided using counts and percentages for baseline and post-baseline values

⇨ Numeric Pain Rating Scale

Descriptive statistics of Numeric Pain Rating Scale value will be performed in patients from the Safety Analysis Set:

- Observed values and absolute change from baseline at each assessed timepoint using default summary statistics for continuous variables.

⇨ Faces Pain Rating Scale

A categorical summary of Face Pain Rating Scale will be provided using counts and percentages for baseline and post-baseline values in patients from the Safety Analysis Set.

The categories are the followings:

- 0 = No Hurt
- 2 = Hurts Little Bit
- 4 = Hurts Little More
- 6 = Hurts Even More
- 8 = Hurts Whole Lot
- 10 = Hurts Worst

6.7.7.2. Orthostatic Hypotension

All information related to orthostatic hypotension will be described in patients from the Safety Analysis Set:

- Signs of orthostatic hypotension
- SBP (mmHg)
- DBP (mmHg)
- Heart Rate

All data collected in the corresponding eCRF form will also be presented in a data listing.

6.7.7.3. Pregnancy Test

Listing of pregnancy test data will be provided at each visit in patients from the Safety Analysis Set.

6.7.8. Interim Analysis and Data Monitoring

6.7.8.1. Data Monitoring Committee (DMC) Analyses

A DMC is established to assure patient safety and will function independently of all other individuals associated with the conduct of the trial, including site investigators participating in the trial. The DMC will consist of a minimum of 2 physicians whose expertise covers relevant specialties and, where applicable, a statistician.

During the conduct of Phase I the DMC will evaluate available safety information and recommend trial continuation or termination. If continuation is recommended, the DMC will recommend to the sponsor whether the dose should be escalated, de-escalated, or held at the same level based on the CRM.

At the completion of Phase I, the DMC will review relevant data and provide a recommendation regarding the MTD and RP2D.

During the conduct of Part 2, DMC meetings will be held on a regular basis to review cumulative safety data and evaluate whether the trial should be modified, stopped, or continue unchanged. The Sponsor Safety Committee will evaluate the recommendations from the DMC after each DMC meeting. Any significant finding/recommendation from the DMC and endorsed by the Sponsor Safety Committee will be communicated to the regulatory authorities and Institutional Review Board (IRB)/Ethics Committee (EC) as appropriate, and to the sites.

Responsibilities, procedures, content of the DMC packages, and workflow of the DMC are specified in the DMC Charter and DMC shells.

6.7.8.2. Phase I Analysis

As described above, at the end of Phase I, all efficacy and safety analyses corresponding to phase I as well as PK analyses will be performed in order to let DMC members provide recommendation regarding the MTD and RP2D to be selected for Phase II.

6.8. Pharmacokinetic Assessments

A listing of PK blood sample collection times and serum concentrations will be presented for Hu3F8-BsAb separately for all patients for the Safety Analysis Set.

Pharmacokinetic concentrations will be summarized for the Pharmacokinetics Analysis Set for each timepoint by treatment group using protocol scheduled times and appropriate summary statistics.

See Section 5.2.4 for the handling of serum concentrations that are BLQ.

6.8.1. Pharmacokinetic Analysis

The following PK parameters will be determined where possible from the concentrations of Hu3F8-BsAb in Phase I Cycle 1, assuming single dose parameters ('split-dose' on Days 1 and 8) using noncompartmental methods in validated software program Phoenix WinNonlin (Certara, Version 8.1 or higher):

Table 8 PK Parameters

Parameter	Definition
$AUC_{0\text{-}tlast}$	area under the concentration-time curve from time 0 to the time of the last quantifiable concentration (t_{last})[a]
$AUC_{0\text{-}8d}$	area under the concentration-time curve over the time interval 0 to 8 days postdose
$AUC_{0\text{-}28d}$	area under the concentration-time curve over the time interval 0 to 28 days postdose
$AUC_{0\text{-}\infty}$	area under the concentration-time curve from time 0 extrapolated to infinity[c]
$\%AUC_{extrap}$	percentage of area under the concentration-time curve due to extrapolation from the last quantifiable concentration to infinity
C_{max}	maximum observed concentration over 0 to 28 days
t_{max}	time of the maximum observed concentration
$t_{1/2}$	apparent terminal elimination half-life over 0 to 28 days
MRT	Mean residence time
CL	total clearance
V_{ss}	volume of distribution at steady state
$DAUC_{0\text{-}tlast}$	$AUC_{0\text{-}tlast}$ normalized by dose [c]
$DAUC_{0\text{-}28d}$	$AUC_{0\text{-}28d}$ normalized by dose [c]
$DAUC_{0\text{-}\infty}$	$AUC_{0\text{-}\infty}$ normalized by dose [b,c]
DC_{max}	C_{max} normalized by dose [c]

[a] The AUC will be calculated using the linear trapezoidal rule for increasing concentrations and the logarithmic rule for decreasing concentrations (linear up/log down rule).
[b] Based on the last observed quantifiable concentration after 2nd dose in cycle
[c] Calculated by dividing the parameter by mcg/kg/cycle

Additional PK parameters may be determined where appropriate.

Pharmacokinetic analysis will be carried out where possible using actual dose administered (mcg/kg/cycle) and actual postdose blood sampling times after start of infusion. If an actual time is missing nominal time may be used.

The parameters C_{max}, and t_{max} will be obtained directly from the concentration-time profiles. If C_{max} occurs at more than 1 timepoint, t_{max} will be assigned to the first occurrence of C_{max}.

6.8.1.1. Criteria for the Calculation of Apparent Terminal Elimination Rate Constant and Half-Life

The start of the terminal elimination phase for each patient will be defined by visual inspection and generally will be the first point at which there is no systematic deviation from the log-linear decline in concentrations.

The apparent terminal elimination rate constant (λ_z) will only be calculated when a reliable estimate can be obtained using at least 3 data points, preferably not including C_{max}, and the adjusted coefficient for determination of exponential fit (R^2-adj) of the regression line is ≥0.7. Parameters requiring λ_z for their calculation (eg, $AUC_{0-\infty}$, $t_{1/2}$, MRT, CL and V_{ss}) will only be calculated if the R^2-adj value of the regression line is ≥0.7.

The following regression-related diagnostic PK parameters will be determined, when possible:

Table 9 Regression-Related Diagnostic PK Parameters

Parameter	Units	Definition
λ_z	1/h	apparent terminal elimination rate constant
λ_z Upper	h	end of exponential fit
λ_z Lower	h	start of exponential fit
λ_z N	NA	number of data points included in the log-linear regression
λ_z Span Ratio	NA	time period over which λ_z was determined as a ratio of $t_{1/2}$
R^2-adj	NA	adjusted coefficient for determination of exponential fit

Where possible, the span of time used in the determination of λ_z (ie, the difference between λ_z Upper and λ_z Lower) should be ≥2 half-lives. If the λ_z Span Ratio is <2, the robustness of the $t_{1/2}$ values will be discussed in the Clinical Study Report (CSR).

6.8.1.2. Criteria for the Calculation and Reporting of Area Under the Concentration-time Curve

The minimum requirement for the calculation of AUC will be the inclusion of at least 3 consecutive concentrations above the lower limit of quantification. If there are only 3 consecutive concentrations, at least 1 should follow C_{max}.

6.8.1.3. Criteria for Handling Concentration Below the Limit of Quantification or Missing Concentration for PK Analysis

Serum concentrations below the limit of quantification (BLQ) will be assigned a value of 0 before the first measurable concentration and thereafter BLQ concentrations will be treated as missing. The following rules apply to the specific situations defined below:

- If an entire concentration-time profile is BLQ, it will be excluded from PK analysis.
- Where 2 or more consecutive concentrations are BLQ at the end of a profile, the profile will be deemed to have terminated and any further quantifiable concentrations will be set to missing for the calculation of the PK parameters, unless they are considered to be a true characteristic of the profile of the drug.
- If a predose serum concentration is missing, it will be set to 0 by default within Phoenix WinNonlin.

6.8.1.4. Treatment of Outliers in PK Analysis

If a value is considered to be anomalous due to being inconsistent with the expected PK profile, it may be appropriate to exclude the value from the PK analysis. However, the exclusion of any data must have strong justification and will be documented in the CSR.

Any quantifiable predose concentration value in the first cycle will be considered anomalous and set to missing for the PK analysis. This will be set to 0 by default in Phoenix WinNonlin.

6.8.2. Presentation of PK Data

If the actual time of sample collection deviates from the nominal time by more than ±10%, the serum concentration will be flagged and may excluded from the summary statistics.

Individual concentrations deemed to be anomalous will be flagged in the listings and excluded from the summary statistics.

For PK parameters the following rule will apply:

Geometric mean and coefficient of variation will not be calculated for t_{max}.

Summary tables, arithmetic mean (+ standard deviation [SD]) figures, overlaying individual figures, and individual figures by treatment and time postdose will be provided for serum PK concentrations in the intensive PK sampling period of Phase 1. All figures will be produced on both linear-linear and linear-logarithmic scales. The +SD bars will only be displayed on the linear-linear scale. Summary tables by treatment will be provided for all PK parameters, with the exception of diagnostic regression-related PK parameters.

For sparse PK sampling in Phase 1 and 2 (prior to and ~ 5 minutes after hu3F8-BsAb infusion), serum concentrations will be listed and summarized by phase, cycle and timepoint.

7. Changes in the Conduct of the Trial or Planned Analysis

The changes indicated below are being made before any data analysis and correspond to changes to protocol version 8. None of these changes are documented as formal protocol amendments:

- MTD is not exactly defined as a maximum dose level at which ≤ 1 of 6 patients experiencing DLT but it is defined as the dose level that leads to DLT in approximately 15% of patients (posterior probability of DLT that comes closest to the toxicity level of 0.15). This has been clarified.
- A DLT Evaluable Analysis Set (defined as all patients in the dose escalation phase who receive both doses of hu3F8-BsAb in cycle 1 and had sufficient safety evaluations, or who had a DLT during the DLT evaluation period) will be considered to evaluate the toxicity instead of "all enrolled patients who received their assigned dose"
- For PFS and OS definition, first dose of treatment will be considered instead of enrollment date as defined by Tinazzi[10]
- Death is considered for PFS definition to be consistent with Tinazzi[11]
- CR duration will be calculated from the time of remission (which is consistent with the definition in the statistical section of the protocol) instead of from first day of treatment (mentioned in the efficacy section of the protocol)
- 4-months PFS for group 2 of phase II will not be analysed using Wilson's method to provide CIs. Instead, PFS rates at 4 months using Kaplan Meier method will be displayed in order to take censoring into account. It corresponded to the exploratory analyses of the protocol. It will be now considered as the primary analysis of phase II for group 2 patients.
- ORR analyses at end of treatment excluding potential response assessments after initiation of any new prohibited anticancer treatment will be performed as sensitivity analyses
- The correlation between HAHA response and overall response rate (for neuroblastoma) or disease control rate (for osteosarcoma) will not be tested using chi-square test. Instead, logistic regression analysis with inclusion of log(therapeutic dose) as a covariate will be applied.
- Analyses about immunological effects will not be adjusted with Hochberg as they are exploratory analyses.
- Disease Control Rates will be analysed as an exploratory efficacy endpoint for phases I and II.
- Accumulation index for AUC (RA,AUC), C_{max} (RA, C_{max}) will not be calculated since the PK parameter calculation methods are based on a 'split' single dose approach.

8. Appendices

8.1. Appendix 1: Document History

Document Version, Status, Date	Summary / Reason for Changes
Version 0.1, Draft, 11Jun2021	Not applicable; the first version
Version 0.2, Draft, 06Jul2021	Sponsor comments following Draft 0.1 review
Version 1.0, Final, 06Oct2021	Sponsor comments following Draft 0.2 review

8.2. Appendix 2: Pain Scales

Table 10 FLACC Scale

CATEGORIES	SCORING		
	0	**1**	**2**
FACE	No particular expression or smile	Occasional grimace or frown, withdrawn, disinterested	Frequent to constant quivering chin, clenched jaw
LEGS	Normal position or relaxed	Uneasy, restless, tense	Kicking, or legs drawn up
ACTIVITY	Lying quietly, normal position moves easily	Squirming, shifting back and forth, tense	Arched, rigid or jerking
CRY	No cry, (awake or asleep)	Moans or whimpers; occasional complaint	Crying steadily, screams or sobs, frequent complaints.
CONSOLABILITY	Content, relaxed.	Reassured by occasional touching hugging or being talked to, distractible.	Difficulty to console or comfort

- Manworren RC, Hynan LS. *Clinical validation of FLACC: preverbal patient pain scale*. Pediatr Nurs 2003;29:140-6.
- Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. *The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment*. Paediatr Anaesth 2006;16:258-65[11].
- Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. *The FLACC: a behavioral scale for scoring postoperative pain in young children*. Pediatr Nurs 1997;23:293-7.

Figure 2 Numeric Pain Rating Scale

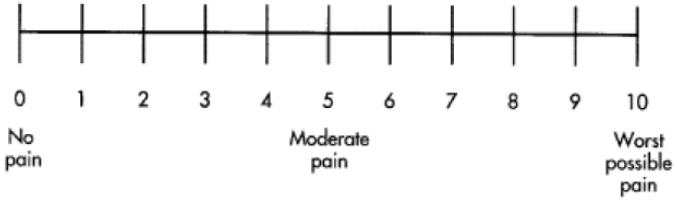


Figure 3 Faces Pain Rating Scale

- Wong DL, Baker CM. *Pain in children: comparison of assessment scales*. Pediatr Nurs 1988;14:9-17.

9. References

[1] Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. *A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in Phase I dose-finding studies*. Clin Trials 2008;5:465-77.

[2] O'Quigley J, Pepe M, Fisher L. *Continual reassessment method: a practical design for phase 1 clinical trials in cancer*. Biometrics 1990;46:33-48.

[3] Park JR, Bagatell R, Cohn SL, et al. *Revisions to the International Neuroblastoma Response Criteria: A Consensus Statement From the National Cancer Institute*. Clinical Trials Planning Meeting. J Clin Oncol 2017:JCO2016720177.

[4] E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, J. Verweij. *New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)*

[5] Manworren RC, Hynan LS. *Clinical validation of FLACC: preverbal patient pain scale*. Pediatr Nurs 2003;29:140-6.

[6] Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. *The FLACC: a behavioral scale for scoring postoperative pain in young children.* Pediatr Nurs 1997;23:293-7.

[7] Wong DL, Baker CM. *Pain in children: comparison of assessment scales*. Pediatr Nurs 1988;14:9-17.

[8] ICH. *Structure and Content of Clinical Study Reports*, Guideline E3, 1995. Available at https://database.ich.org/sites/default/files/E3_Guideline.pdf

[9] *Common Terminology Criteria for Adverse Events (CTCAE) v4.03.*

[10] Tinazzi. *Efficacy endpoints in Oncology*, Phuse 2013

[11] Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. *The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment*. Paediatr Anaesth 2006;16:258-65

Statistical Analysis Plan (SAP)

Type of Approval (select one) : ☒ **SAP** ☐ **Initiation of Programming**

Sponsor Name:	Y-mAbs Therapeutics Inc		
Sponsor Protocol/CIP ID:	18-034	[redacted] **Study ID:**	000000214175
SAP text filename:	18-034_SAP_Y-mAbs_Final_06Oct2021	**TFL shells filename:**	NA (to be developed later as agreed with sponsor as per timelines in the SAP preparation checklist)
Version:	Final 1.0	**Date:**	06Oct2021

[redacted] Approval(s):

Lead Statistician

Approval Signature	[redacted]
Print Name	Principal Statistician
Job Title	06Oct2021
Date	[redacted] Principal Statistician I authored this document 06 Oct 2021 3:55 PM +02:00

Sponsor Approval(s):

By signing below when the statistical analysis plan (SAP) is considered final, the signatories agree to the analyses to be performed for this study and to the format of the associated tables, figures, and listings (TFLs). Once the SAP has been signed, programming of the Analysis Dataset Model (ADaM) datasets and TFLs based on these documents can proceed. Any modifications to the SAP text and TFL shells made after signing may result in a work-scope change.

Approval Signature	[redacted]
Print Name	Senior Project Statistician
Job Title	[redacted]
Date	08 OCT 2021

Please scan/email completed form(s) to the Lead Statistician listed below:

Printed Name/Title:	[redacted] / Principal Statistician
Email:	[redacted]