Title: A multicenter, randomized, open-label phase II clinical study of anti-PD-1 antibody SHR-1210 combined with apatinib mesylate vs. doxorubicin combined with ifosfamide in treatment of soft tissue sarcoma

NCT number: NCT03711279

Date: 23 Nov., 2020

A MULTICENTER, RANDOMIZED, OPEN-LABEL PHASE II CLINICAL STUDY OF ANTI-PD-1 ANTIBODY SHR-1210 COMBINED WITH APATINIB MESYLATE VS. DOXORUBICIN COMBINED WITH IFOSFAMIDE IN TREATMENT OF SOFT TISSUE SARCOMA

Statistical Analysis Plan

(SAP)

Written by: [redacted]

Company: Jiangsu Hengrui Pharmaceuticals Co., Ltd.

Version: 1.0/Initial Draft

Date: 23 Nov., 2020

This SAP has been reviewed by the following personnel before being approved and effective.

Functional Role	Reviewer
Medical Science	[redacted]
Statistics	[redacted]

ABBREVIATIONS

Term	Definition
AE	Adverse event
BOR	Best overall response
CR	Complete response
DCR	Disease control rate
ECG	Electrocardiogram
ES	Evaluable set
FAS	Full analysis set
NE	Not evaluable
ORR	Objective response rate
PD	Progressive disease
PFS	Progression-free survival
PR	Partial response
PT	Preferred term
QD	Quaque die
Q3W	Once every 3 weeks
RECIST	Response Evaluation Criteria in Solid Tumors
SOC	System organ classification
SD	Stable disease
SAS	Statistical analysis system
SS	Safety set
TEAE	Treatment-emergent adverse event
TRAE	Treatment-related adverse event
VS	Vital signs

TABLE OF CONTENTS

ABBREVIATIONS **2**
1 REVISION **6**
2 INTRODUCTION **6**
2.1 Study Design 6
2.2 Study Objectives 8
2.2.1 Primary objective 8
2.2.2 Secondary objectives 8
2.2.3 Exploratory objectives 8
2.3 Sample Size 9
3 STATISTICAL HYPOTHESES **9**
4 STUDY ENDPOINTS **9**
4.1 Efficacy Endpoints 9
4.1.1 Primary efficacy endpoint 9
4.1.2 Secondary efficacy endpoints 10
4.2 Safety Endpoints 11
4.2.1 Adverse event 12
4.2.2 Laboratory tests 12
4.2.3 Vital signs 12
4.2.4 12-Lead ECG 12
4.2.5 Physical examination 12
4.2.6 Other safety endpoints 12
4.3 Immunogenicity and Drug Trough Concentration 12
5 STATISTICAL ANALYSIS **13**
5.1 General Considerations 13
5.1.1 Analysis sets 13
5.1.2 General rules and analysis 13
5.1.3 Derived variables 15
5.1.4 Missing data 16
5.1.5 Protocol deviations 16
5.1.6 Inclusion/exclusion violation 17

5.2 Study Subjects 17
5.2.1 Disposition of subjects 17
5.2.2 Demographics 17
5.2.3 Tumor diagnosis 17
5.2.4 Medical history 18
5.2.5 Tumor treatment history 18
5.2.6 Prior therapy and concomitant medication 18
5.3 Follow-up Systemic Anti-Tumor Therapy 18
5.4 Efficacy analysis 18
5.4.1 Primary efficacy analysis 18
5.4.2 Secondary efficacy analysis 19
5.5 Safety Analysis 19
5.5.1 Extent of exposure 19
5.5.2 Adverse event 19
5.5.3 Laboratory tests 20
5.5.4 Vital signs 21
5.5.5 12-Lead ECG 21
5.5.6 ECOG PS 21
5.5.7 Physical examination 21
5.5.8 Other safety measures 21
5.6 Pharmacokinetic Analysis 21
5.7 Other analyses 21
6 INTERIM ANALYSIS 22
7 REFERENCES 22
8 APPENDICES 23
Appendix 1 Response Evaluation Criteria in Solid Tumors 23
Appendix 2 Common Soft Tissue Sarcomas With Poor Chemotherapy Sensitivity 36
Appendix 3 Histological Coding for Common Soft Tissue Sarcomas With Poor Chemotherapy Sensitivity 36

LIST OF TABLES

Table 1. Derived variables of SHR-1210 15
Table 2. Derived variables of apatinib 15
Table 3. Derived variables of doxorubicin 15
Table 4. Derived variables of ifosfamide 16
Table 1. Time point response: subjects with target (+/- non-target) disease 31
Table 2. Time point response: subjects with non-target lesion only 31
Table 3. Best overall response when confirmation of CR and PR required 32

1 REVISION

This statistical analysis plan is the initial version, drafted based on the final version of the study protocol (version no.: 3.0, version date: 25 Feb., 2019).

The final draft of this plan will be completed before database locking and will be signed by various functional departments for confirmation.

2 INTRODUCTION

SHR-1210 (camrelizumab) is a recombinant humanized anti-programmed cell death-1 (PD-1) monoclonal antibody independently developed by Jiangsu Hengrui Pharmaceuticals Co., Ltd. It is a new class 1 therapeutic biological product and has been marketed. Preclinical studies have shown that SHR-1210 has comparable *in vivo* pharmacological and safety profiles to those of drugs of the same class abroad, and may have a better clinical potential for anti-tumor treatment.

Apatinib is a highly effective and selective inhibitor of VEGFR-2. The activity assay showed that its inhibitory effect against VEGFR-2 is stronger than that of drugs of same class (sorafenib, sunitinib, pazopanib, etc.); its selectivity for VEGFR-2 is ⩾ 30 times stronger than that of other targets (VEGFR-1/PDGFR, SRC, etc.). Apatinib can effectively inhibit the *in vitro* lumen formation of Human Umbilical Vein Endothelial Cells (HUVEC) and the formation of arterial ring capillaries in rats, and has a strong inhibitory effect against neovascularization.

Based on the past successes of VEGFR inhibitors combined with anti-PD-1/anti-PD-L1 antibodies in clinical studies and the current treatment status of soft tissue sarcoma, VEGFR-2 inhibitor apatinib combined with SHR-1210 may also exert considerable synergistic effect in the treatment of soft tissue sarcoma, thereby providing a novel and effective non-chemotherapy treatment for patients with soft tissue sarcoma.

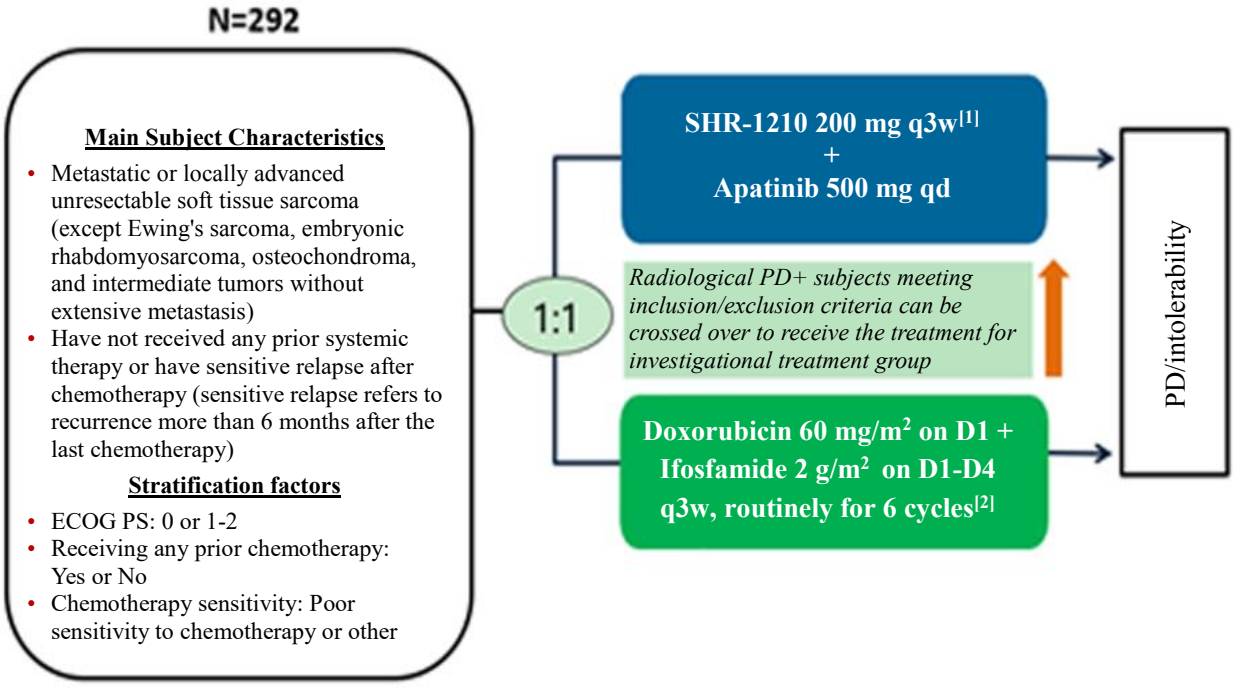
2.1 Study Design

This study is a multi-center, randomized, open-label phase II clinical study, with a total of 292 subjects planned to be enrolled.

Qualified subjects with soft tissue sarcoma will be randomized in a 1:1 ratio into the SHR-1210 + apatinib group (investigational treatment group) and the doxorubicin + ifosfamide/ifosfamide monotherapy group (control group). The stratification factors include: 1) ECOG PS, 0 vs. 1-2; 2) Whether the pathological type of the tumor is poorly sensitive to chemotherapy (see Appendix 2. Common Types of Soft Tissue Sarcoma with Poor Chemotherapy Sensitivity); 3) Whether the subjects have previously undergone chemotherapy. The administration regimens for investigational treatment group and control group are as follows:

Investigational treatment group: SHR-1210 200 mg q3w (up to 2 years of treatment) + apatinib 500 mg qd.

Control group: doxorubicin 60 mg/m^2 on D1 + ifosfamide 2 g/m^2/d on D1-D4, in 3-week cycles (i.e., the total dose per cycle is 8 g/m^2), and 6 cycles are recommended. If the investigator judges that the subject can benefit from continued doxorubicin + ifosfamide chemotherapy, additional chemotherapy cycles may be administered. When the subject's dose of anthracyclines has reached the recommended maximum cumulative dose of doxorubicin, the chemotherapy regimen should be adjusted to ifosfamide monotherapy, 2 g/m^2/d on D1-D5, in 3-week cycles (i.e., the total dose per cycle is 10 g/m^2). If the investigator judges that the subject can benefit from ifosfamide chemotherapy, additional chemotherapy cycles may be administered.



1. The treatment duration of SHR-1210 should be no more than 2 years; interruption of one drug does not require the interruption of the other; 2. If anthracycline 450 mg/m^2 is used for the control group, the chemotherapy regimen will be adjusted to ifosfamide monotherapy, 2 g/m^2 on D1-D5, q3w; more than 6 chemotherapy cycles may be administered.

In this study, the screening period should be no more than 28 days. After completing screening examinations and assessments, eligible subjects will enter the treatment period and receive the study treatment and study visits according to the protocol. Among them, tumor imaging assessment will be performed once in every 2 cycles during the first 16 cycles and once every 4 cycles thereafter. The independent review committee (IRC) will review the imaging evaluation results of each study center. The safety follow-up period of subjects in the investigational treatment group starts after the last dose, and the subjects will be followed up once every 30 days until 90 days after the last dose. Among them, the first safety follow-up will be carried out at the

study center; the second and the third follow-up visits will be made via phone calls. For subjects in the control group, if they do not receive SHR-1210 + apatinib treatment, their safety follow-up period will be 21 days after the last dose; otherwise, the same arrangement as that in the investigational treatment group will be used. The survival follow-up period will start after the end of the safety follow-up period. The survival follow-up period will end upon the subject's death, lost to follow-up, withdrawal of informed consent, or study termination by the sponsor. During this period, a follow-up shall be conducted every 2 months via phone calls or other effective methods to collect information on subject survival and subsequent anti-tumor treatment.

2.2 Study Objectives

2.2.1 Primary objective

- To evaluate whether SHR-1210 combined with apatinib is better than doxorubicin combined with ifosfamide in the treatment of soft tissue sarcoma based on the progression-free survival (PFS).

2.2.2 Secondary objectives

- To evaluate and compare the objective response rate (ORR) and disease control rate (DCR) of SHR-1210 combined with apatinib vs. doxorubicin combined with ifosfamide in the treatment of soft tissue sarcoma;
- To evaluate and compare the safety of SHR-1210 combined with apatinib vs. doxorubicin combined with ifosfamide in the treatment of soft tissue sarcoma;
- To evaluate and compare the quality of life scores in subjects with soft tissue sarcoma after treatment with SHR-1210 combined with apatinib vs. doxorubicin combined with ifosfamide.

2.2.3 Exploratory objectives

- To explore the correlation between the immunogenicity and the efficacy/safety of SHR-1210;
- To explore the correlation between the concentrations of SHR-1210 and apatinib and the efficacy/safety;
- PFS of subjects in the doxorubicin plus ifosfamide group after crossover treatment (as per RECIST v1.1);
- ORR, DCR (as per RECIST v1.1), safety, and quality of life scores of subjects in the doxorubicin plus ifosfamide group after crossover treatment.

2.3 Sample Size

In this study, IRC-assessed PFS will be used as the primary endpoint for superiority test with the control group. An interim analysis will be performed to determine the efficacy of the drugs and to determine whether to terminate or continue the study. The parameters for sample size calculation are as follows:

- Enrollment rate = 13 subjects/month (156 subjects/year), enrollment time = 20 months (the overall time will be 26 months)
- Randomization in a 1:1 ratio; one-sided $\alpha = 0.025$; the power of the test is 90%;
- Hazard ratio (treatment group/control group [HR]) = 0.64 (estimated median PFS is 4.5 months in the control group and 7.0 months in the treatment group)
- An interim analysis will be performed when 67% of PFS events (143) are collected to determine whether the drug is effective or ineffective, and to decide whether to terminate or continue the study.

Based on the above parameters, at least 214 PFS events shall be collected according to the log-rank test for PFS comparison in the two groups and the O'Brien & Fleming spending function (EAST6.4.1) in the α spending function method proposed by Lan-DeMets, i.e., approximately 262 subjects are needed. Considering a 10% dropout rate, 292 subjects are finally required for enrollment.

3 STATISTICAL HYPOTHESES

A superiority test will be conducted to compare SHR-1210 combined with apatinib (treatment group) vs. doxorubicin combined with ifosfamide (control group) in improving the PFS (assessed by the IRC as per RECIST v1.1) of patients with soft tissue sarcoma, at a significance level of one-sided $\alpha = 0.025$. The null hypothesis (H_0) and alternative hypothesis (H_1) are:

H_0: PFS hazard ratio (treatment group/control group) ≥ 1;

H_1: PFS hazard ratio (treatment group/control group) < 1.

4 STUDY ENDPOINTS

4.1 Efficacy Endpoints

4.1.1 Primary efficacy endpoint

4.1.1.1 PFS assessed by the independent review committee as per RECIST v1.1

The primary endpoint of the study is the IRC-assessed PFS (as per RECIST v1.1).

PFS: the time between the date of randomization and the date on which the first tumor progression is documented (as per RECIST v1.1, regardless of whether treatment is continued), or death from any cause, whichever comes first.

The censoring rules for PFS are as follows:

- If there is no post-baseline objective tumor response evaluation or death, then the date of randomization will be censored.
- If there is no progressive disease (PD) or death as per RECIST v1.1, then the date of the last objective tumor response evaluation will be censored.
- If no imaging PD or death is recorded before the subject uses the new anti-tumor drug, the censoring date is the date of the last objective tumor response evaluation carried out before the subject uses the new anti-tumor drug.
- Before PD or death, if a subject has missed ≥ 2 consecutive scheduled imaging assessments, then the date of the last objective tumor response evaluation prior to the missing ones will be censored.

4.1.2 Secondary efficacy endpoints

4.1.2.1 PFS assessed by the investigator (as per RECIST v1.1)

PFS assessed by the investigator (as per RECIST v1.1) is one of the secondary endpoints in this study.

4.1.2.2 Objective response rate (ORR) assessed by the investigator and independent review committee

ORR: the proportion of treated subjects with a best overall response (BOR) of complete response (CR) or partial response (PR) as per RECIST v1.1.

4.1.2.3 Disease control rate (DCR) assessed by the investigator and independent review committee

DCR: the proportion of treated subjects with a best overall response of complete response (CR), partial response (PR), or stable disease (SD) as per RECIST v1.1.

4.1.2.4 Quality of life score (EORTC QLQ-C30)

To evaluate the quality of life scores of SHR-1210 combined with apatinib and doxorubicin combined with ifosfamide in the treatment of subjects with soft tissue sarcoma

4.1.2.5 Best overall response (BOR)

The best overall response (BOR) is calculated based on the evaluation of imaging efficacy of soft tissue of all visits. The best response until the last response evaluation shall be used if no progressive disease is noted during the period between the date of first dose and the date of PD or initiation of other anti-tumor treatments. BOR is classified as: CR, PR, SD, PD, and NE. If the tumor is assessed as PD at a visit, the date of the PD should be recorded as the date of imaging assessment of the target lesion, the date of imaging assessment of the non-target lesion, or the date of discovery of the new lesion, whichever is earlier.

- Confirmed CR: Response evaluation of CR that has been reconfirmed as CR in a response evaluation 4 weeks later.
- Confirmed PR:
  - Response evaluation as CR that has been reconfirmed as PR instead of CR.
  - Response evaluation as PR that has been reconfirmed as CR/PR in a response evaluation 4 weeks later.
- SD:
  - Response evaluation as CR/PR that has been reconfirmed as SD instead of CR/PR.
  - The response is evaluated as SD at least once within the shortest time interval (not less than 6 weeks) after the start of the study.

4.2 Safety Endpoints

The following safety data will be collected and summarized following the study protocol:

- Adverse event
- Laboratory test data
- Vital signs data
- 12-Lead ECG
- Physical examination
- Other safety endpoints

4.2.1 Adverse event

An AE refers to any untoward medical condition in a clinical study subject who receives a pharmaceutical product, and the condition does not necessarily have a causality with the treatment. AEs can include any unfavorable and unintended symptoms, signs, abnormal laboratory finding, or diseases, including the following:

1) Worsening of pre-existing (prior to entering clinical study) medical conditions/diseases (including worsening symptoms, signs, or laboratory abnormalities);
2) Any new AE: Any new adverse medical conditions (including symptoms, signs, and newly diagnosed diseases);
3) Clinically significant abnormal laboratory findings.

4.2.2 Laboratory tests

Hematology, urinalysis, stool occult blood, clinical chemistry, coagulation function, thyroid function (FT3, FT4, and TSH), myocardial zymography, and other laboratory data will be collected at the protocol-specified visit time points.

4.2.3 Vital signs

Vital signs including body temperature, heart rate, respiration, blood pressure, and pulse will be collected at the protocol-preset time points.

4.2.4 12-Lead ECG

Heart rate, QT, P-R interval, and other data will be collected at the protocol-preset time points.

4.2.5 Physical examination

Physical examination includes height, weight, general condition, head, neck, chest, abdomen, perineum, and extremities. These data will be collected at protocol-specified visit/time points.

4.2.6 Other safety endpoints

Electrocardiography and pregnancy test.

4.3 Immunogenicity and Drug Trough Concentration

- Anti-drug antibody and neutralizing antibody levels of SHR-1210
- Serum concentration of SHR-1210 and plasma concentration of apatinib

5 STATISTICAL ANALYSIS

5.1 General Considerations

5.1.1 Analysis sets

Full analysis set (FAS)

It includes all subjects who are randomized and have received study drugs at least once. This analysis set will be used for the efficacy analysis.

Per-protocol set (PPS)

It is a subset of the full analysis set, and subjects with major protocol deviations judged to have a significant impact on treatment efficacy will be excluded from this set. The list of subjects included into or excluded from the PPS should be reviewed and determined by the sponsor and the investigator before the database is locked.

Safety set (SS)

It includes all enrolled subjects who have received the study drugs at least once and have post-administration evaluable safety data.

PK concentration set

All enrolled subjects who have received at least one dose of the investigational drug and have post-baseline PK data.

Immunogenicity analysis set

All enrolled subjects who have received at least one dose of the investigational drug and have baseline and at least one post-baseline evaluable ADA data.

5.1.2 General rules and analysis

Baseline

Unless otherwise stated, "baseline" in this study is defined as the last non-missing measurement value obtained before the first use of study drugs.

Study days

The day of the first dose is used as the start date of the study (Day 1).

- If the evaluation date (adverse event, laboratory tests, etc.) is on or after the study medication, the study date shall be calculated as follows:

  Study days = evaluation/event date – start date of study + 1.

- If the evaluation date (baseline characteristics, medical history, etc.) is before the study medication, the number of study days is a negative figure and shall be calculated as follows:

  Study days = evaluation/event date – start date of study;

General analysis

Unless otherwise specified, the following descriptive statistics will be summarized by the type of variables:

- The measurement data will be summarized using mean, standard deviation, median, maximum, minimum, and quartile.
- Count data will be summarized using frequency and percentage;
- For time-to-event data, Kaplan-Meier method will be used to estimate the survival function and median time to event onset, and a survival curve will be plotted.

Number of decimal places

Unless otherwise specified, number of decimal places in the analysis report will be determined as per the following rules:

- The decimal places of the minimum and maximum will remain the same as that of raw data to be acquired; there should be one additional decimal place for the mean and median, and 2 additional decimal places for standard deviation, up to 4 decimal places.
- The percentage will be rounded to 1 decimal place. If the frequency is 0, the percentage is not displayed.
- The *P* value will retain 4 decimal places. If the *P* value is < 0.0001, it will be expressed as "< 0.0001"; if the *P* value is > 0.9999, it will be reported as "> 0.9999".
- The 95% CI, if being a decimal, will retain at least 2 decimal places, up to 4 decimal places. Details are as follows: The 95% confidence interval will have one more decimal place than that of the raw data. If the raw data have no decimal place, the 95% confidence interval will retain 2 decimal places; if the raw data have 4 or more decimal places, the 95% confidence interval will retain at most 4 decimal places.
- Time to event (in months) will be rounded to one decimal place.

Results of Laboratory Tests

The results of laboratory tests are generally continuous numerical variables or character variables. If continuous numerical variables recorded in eCRF contain special characters (such as < xx and > xx), the following rules will be applied:

- When there is a situation of < xx or ≤ xx, half of the xx value will be used for analysis.
- When there is a situation of > xx or ≥ xx, the value of xx will be used for analysis.

Analysis software

All statistical analyses will be conducted using SAS® 9.4 or later version.

5.1.3 Derived variables

Table 1. Derived variables of SHR-1210

Variable	SHR-1210
Protocol-Specified Method of Administration	200 mg/dose, intravenous drip (no less than 20 min and no more than 60 min) on Day 1 of each 3-week cycle.
Duration of Drug Exposure (Month)	(Actual date of last dose − actual date of first dose + 21)/30.4375
Drug Exposure (mg)	The total administered dose of SHR-1210 in the subject
Actual Dose Intensity (mg/3 weeks)	Drug exposure/[duration of drug exposure (month) × 30.4375/21]
Relative Dose Intensity (%)	100 ×actual dose intensity (mg/3 weeks)/200 (mg/3 weeks)

Table 2. Derived variables of apatinib

Variable	Apatinib
Protocol-Specified Method of Administration	Administered orally at 500 mg/day once daily. Each cycle contains 3 weeks.
Duration of Drug Exposure (Month)	(Actual date of last dose − actual date of first dose + 1)/30.4375
Drug Exposure (mg)	The total administered dose of apatinib in the subject
Intended Dose Intensity (mg/day)	500
Actual Dose Intensity (mg/day)	Drug exposure (mg)/[duration of drug exposure (month) × 30.4375]
Relative Dose Intensity (%)	100 × actual dose intensity (mg/day)/500

Table 3. Derived variables of doxorubicin

Variable	Doxorubicin
Protocol-Specified Method of Administration	Administered via intravenous drip infusion at 60 mg/m^2 within 2 h on Day 1 of each 3-week cycle.
Duration of Drug Exposure (Month)	(Actual date of last dose − actual date of first dose + 21)/30.4375
Drug Exposure (mg/m^2)	Subject's actual cumulative dose (mg)/body surface area [(1)]
Intended Dose (mg/m^2)	(Actual date of last dose − actual date of first dose + 21) × 60 mg/m^2/21
Relative Dose Intensity (%)	100 ×drug exposure (mg/m ^{2})/scheduled dose (mg/m^2)

Table 4. Derived variables of ifosfamide

Variable	Ifosfamide	
Protocol-Specified Method of Administration	Combination with doxorubicin: Ifosfamide 2 g/m^2/d on D1-D4, intravenous drip infusion within 4-6 h, in 3-week cycles (i.e., the total dose per cycle shall be 8 g/m^2).	Ifosfamide monotherapy: 2 g/m^2/ d on D1-D5, intravenous drip infusion within 4-6 h, in 3-week cycles (i.e., the total dose per cycle shall be 10 g/m^2).
Duration of Drug Exposure (Month)	(Actual date of last dose − actual date of first dose + 21)/30.4375	
Drug Exposure (g/m^2)	Actual cumulative dose/body surface area (1)	
Intended Dose (g/m^2)	[(Date of last dose of doxorubicin − date of first dose of doxorubicin + 21)/21] × 8 g/m^2 + [CEIL × (date of last dose of ifosfamide − date of last dose of doxorubicin − 4)/21] ×10 g/m^2	
Relative Dose Intensity (%)	100 ×drug exposure (g/m^2)/intended dose (g/m^2)	

(1) The body surface area is calculated as: body surface area (m^2) = 0.0061 × height (cm) + 0.0128 × weight (kg) − 0.1529. The body weight at baseline should be used.

*CEIL: round up

5.1.4 Missing data

In this study, the missing data of the efficacy endpoints will not be treated specially, and the missing values will not be estimated in the safety evaluation.

If the specific date of death is missing, it will be imputed based on the first day of month of death.

If the specific day of the initial diagnosis is missing, it will be imputed with the first day of the month of initial diagnosis. If the month and day of the initial diagnosis are missing and the year of the initial diagnosis is before the year of the first dose, it will be imputed with 1 Jan. If the month and day of the initial diagnosis are missing and the year of the initial diagnosis is the same as the year of the first dose, it will be imputed with 1 Jan. If the date of the initial diagnosis is completely missing, it will not be imputed.

5.1.5 Protocol deviations

Before database locking, data of all subjects on the CRF will be checked for major protocol deviations. All potential major protocol deviations will be reviewed and evaluated by the investigator and the sponsor.

All major protocol deviations will be summarized and described separately by type and treatment group, and will be tabulated and analyzed.

Definitions of major protocol deviations:

- Serious violation of the inclusion/exclusion criteria;
- Incorrect treatment by study drugs (not treated using the drug or dose as per randomization);
- Other major protocol deviations jointly recognized by the investigator and the sponsor.

5.1.6 Inclusion/exclusion violation

All violations of the inclusion and exclusion criteria will be listed by subject number.

5.2 Study Subjects

The study subjects will be analyzed by treatment group.

5.2.1 Disposition of subjects

The number of screened subjects, number of enrolled subjects, number and percentage of treated subjects, number and percentage of subjects in the analysis sets (including the full analysis set, per-protocol set, safety set, PK concentration set, and immunogenicity analysis set), number of subjects who discontinue study/treatment, and number and percentage of corresponding subjects to the reasons for discontinuation.

5.2.2 Demographics

Age, gender, ethnicity, height, weight, and other demographic data of the subjects will be summarized using descriptive statistics by treatment group.

For continuous variables (such as age, height, weight, body mass index, doxorubicin dose corresponding to cumulative anthracycline, etc.), the number of observed subjects, mean, standard deviation, median, quartile, minimum, and maximum will be listed by treatment group.

For categorical variables (such as gender, ethnicity, ECOG PS, etc.), the frequency and percentage will be listed by treatment group.

Demographic and baseline characteristics will be tabulated.

5.2.3 Tumor diagnosis

The following tumor parameters will be summarized using relevant descriptive statistics: the site of primary lesion, histological type, and clinical stage at screening will be summarized by the number and percentage of subjects involved in each treatment group. The course of disease (year) will be summarized using descriptive statistics such as mean, standard deviation, median, minimum, and maximum. A detailed list of subjects for tumor diagnosis will be provided.

The course of disease (year) is defined as the time from the date of the first diagnosis to the time of randomization, which should be calculated as: (date of randomization − date of first diagnosis + 1)/ 365.25.

5.2.4 Medical history

A detailed list of subjects and their medical history will be provided.

5.2.5 Tumor treatment history

Tumor treatment history mainly includes the history of tumor surgery, the history of systemic anti-tumor treatment, the history of chemotherapy, and the history of radiotherapy.

For tumor treatment history, the following descriptive statistics will be provided:

- Number and percentage of subjects who have received tumor surgery;
- The number and percentage of subjects who have received chemotherapy;
- The number and percentage of subjects who have received radiotherapy;
- The number and percentage of subjects who have received systemic anti-tumor therapy.

In addition, a detailed list of subjects will be provided.

5.2.6 Prior therapy and concomitant medication

All previous medication, concomitant medication, and concomitant non-drug therapy will be listed.

5.3 Follow-up Systemic Anti-Tumor Therapy

Follow-up systemic anti-tumor treatments will be listed by treatment group.

5.4 Efficacy analysis

5.4.1 Primary efficacy analysis

The primary endpoint of the study is the IRC-assessed PFS (as per RECIST v1.1). The primary analysis will be based on the FAS. The survival functions of PFS of the two groups will be compared using stratified log-rank tests with and without random stratification factors. The Kaplan-Meier method will be used to estimate the PFS rate, plot the survival curve, and estimate the median PFS, and its 95% CI will be calculated (Brookmeyer-Crowley method). The analysis with stratification factors will be used as the primary analysis.

In addition, as a supporting analysis and under the assumption of proportional hazards, the Cox models with and without stratification factors will be used to estimate the hazard ratio (HR) and calculate the corresponding 95% CI (Wald method).

5.4.2 Secondary efficacy analysis

The analysis of secondary efficacy endpoints will be based on the FAS.

For the secondary endpoint PFS assessed by the investigator (as per RECIST v1.1), the analytical method will be the same as that for the primary endpoint mentioned above.

For the objective response rate (ORR) and disease control rate (DCR), the rates in the two groups and their two-sided 95% CI (Clopper-Pearson method) will be estimated, and the inter-group difference of rates and its two-sided 95% CI will be calculated (CMH method). A swimmer plot of ORR will also be produced.

Other endpoints (best overall response, quality of life score, etc.) will be statistically summarized in accordance with general principles.

5.5 Safety Analysis

All safety analyses will be conducted based on SS. All safety analyses will be presented by treatment group.

5.5.1 Extent of exposure

See Section 5.1.3 for the specific defining rules of the derived variables of study drugs SHR-1210, apatinib, doxorubicin, and ifosfamide.

The drug exposure, duration of drug exposure, actual dose intensity, and relative dose intensity of apatinib and SHR-1210, as well as the duration of drug exposure and drug exposure of doxorubicin and ifosfamide are described using the mean, standard deviation, maximum, minimum, and median.

5.5.2 Adverse event

AEs will be classified by treatment stage (before and after study treatment).

All AEs will be coded with MedDRA (version, 23.1). and graded using NCI-CTCAE v4.03. For the same SOC and/or PT, multiple cases of the same events that occur in one subject will be counted only once. For the same AE reported in one subject multiple times but varying in CTCAE grade, the AE of the greatest grade will be enumerated.

AEs will be sorted by incidence from high to low based on SOC, and PT under each SOC. If the incidence of ≥ 2 PT is equal, the AEs will be sorted alphabetically. If there is no AE under a SOC or PT, then analysis will not be conducted.

A treatment-emergent adverse event (TEAE) is defined as any adverse event that occurs after the start of study treatment. Only TEAEs will be summarized. All AEs will be listed. AEs with completely missing onset time will be treated as TEAEs.

For AEs occurring after the start of study treatment, treatment-related AEs include those whose causality with the study drug is definitely related, possibly related, or indeterminable; if the causality assessment is missing, then the AE will be deemed treatment-related for analysis.

In a listing of AEs, TEAEs will be analyzed by group, and frequencies and proportions will be statistically described. AEs will be summarized using descriptive statistics according to Hengrui's Statistical Analysis Reporting Standards, including but not limited to the following contents:

- Any TEAE
- Treatment-related AEs (TRAEs)
- Any TESAE
- Treatment-related SAEs (TRSAEs)
- Any SAE
- Any drug-related SAE
- TEAEs resulting in dose reduction, dose interruption followed by reduction, interruption, or treatment discontinuation
- TRAEs resulting in dose reduction, dose interruption followed by reduction, interruption, or treatment discontinuation
- TEAEs resulting in death
- Immune-related adverse events ("reactive capillary endothelial proliferation" will all be considered immune-related events)?
- Was hormone therapy used?

The incidence of AE is calculated based on the number of subjects having an AE, instead of the number of AE episodes.

All AEs will be listed. Drug-related AEs, AEs resulting in treatment discontinuation, AEs resulting in dose reduction or dose interruption followed by reduction or treatment interruption, and death will be listed.

5.5.3 Laboratory tests

Baseline laboratory measurement is defined as the latest non-missing test result prior to the first dose of study drugs.

Laboratory results will be listed.

5.5.4 Vital signs

All vital signs will be listed.

5.5.5 12-Lead ECG

Mean values of 12-Lead ECG results will be taken at each visit. ECG data include: HR (beats/min), PR interval (ms), QT interval (ms), and QTcF (ms).

ECG parameters measured at the study site will be summarized as follows:

- Maximum PR interval ≥ 300 ms
- Maximum QT interval ≥ 500 ms
- Maximum QTcF interval (ms) (450–<480, 480–<500, or ≥500)
- Maximum QTcF interval ≥ 480 ms or increase > 60 ms from baseline

Related data will also be listed.

5.5.6 ECOG PS

All ECOG PS will be listed.

5.5.7 Physical examination

Data of physical examination will be listed.

5.5.8 Other safety measures

Echocardiography will be listed by treatment group and time.

Pregnancy test results will be listed by treatment group and time.

Observation results of reactive capillary endothelial proliferation will be listed.

The subjects' virological examinations will be listed.

5.6 Pharmacokinetic Analysis

The drug concentrations of SHR-1210 and apatinib will be statistically described by time point.

5.7 Other analyses

A statistical analysis will be performed according to the confirmatory results of immunogenicity.

The number and percentage of subjects who are ADA positive at baseline, ADA negative at baseline and positive post baseline, and ADA positive for at least once at baseline or post baseline will be summarized using descriptive statistics.

For the population who are ADA negative at baseline and positive post baseline, the following items will be analyzed: the time to the first positive ADA result, the number and percentage of subjects with transient positive post-baseline results (transient ADA positive is defined as the subjects who are negative at baseline and have only one ADA-positive sample post baseline before the last sample and are not persistent ADA positive), the number and percentage of subjects with persistent positive post-baseline results (persistent ADA positive can be divided to two categories: 1) negative at baseline, there are at least two positive samples post baseline, where the first and last ADA-positive samples are separated by a period >16 weeks, or 2) negative at baseline, the last post-baseline sample is positive, or the last post-baseline sample is negative and the second last post-baseline sample is positive), and the number and percentage of subjects with other positive post-baseline results (other ADA positive is defined as the subjects who are negative at baseline, have at least two ADA positive samples post baseline, but not persistent ADA positive, and the last sample is negative).

The number of ADA positive and negative subjects will be summarized and analyzed, and the efficacy and safety of ADA positive and negative subjects will be analyzed to investigate the impact of ADA on efficacy and safety. ADA-positive subjects are subjects with post-baseline ADA-positive samples, mainly including: a. the subjects who have the baseline negative and post-baseline positive samples, or b. the subjects who have both baseline and post-baseline positive samples, and the titer of the post-baseline samples is ≥ 4-fold of the baseline titer. ADA-negative subjects are those who are not ADA positive.

In addition, the positive ADA results of the subjects producing neutralizing antibodies (Nab) will be summarized and analyzed. Nab-positive subjects are defined as the subjects who have post-baseline Nab positive samples (but those who have baseline and post-baseline Nab positive samples will be classified as Nab inconclusive subjects), and Nab-negative subjects are those who do not have post-baseline Nab positive samples.

Immunogenicity raw data will be reported in the form of listing.

6 INTERIM ANALYSIS

No interim analysis will be conducted for this study.

7 REFERENCES

None.

8 APPENDICES

Appendix 1 Response Evaluation Criteria in Solid Tumors

Response Evaluation Criteria in Solid Tumors Version 1.1 (Excerpt)

Note: This appendix is translated internally and is for reference only. Please refer to the English version during practice.

1 BACKGROUND

Omitted

2 PURPOSE

Omitted

3 MEASURABILITY OF TUMOR AT BASELINE

3.1 Definition

At baseline, tumor lesions/lymph nodes will be categorized measurable or non-measurable as follows:

3.1.1 Measurable lesions

Tumor lesions: Must be accurately measured in at least one dimension (longest diameter is to be recorded) with a minimum size of:

- CT scan 10 mm (the thickness of CT scan layer shall not be more than 5 mm)
- 10 mm caliper measurement by clinical exam (lesions which cannot be accurately measured with calipers should be recorded as non-measurable)
- 20 mm by chest X-ray;
- Malignant lymph nodule: pathologically enlarged and measurable, single lymph nodule must be ≥ 15 mm in short axis by CT scan (CT scan slice thickness no greater than 5 mm). At baseline and during follow-up, only the short axis will be measured and followed.

3.1.2 Non-measurable lesions

All other lesions, including small lesions (longest diameter < 10 mm or pathological lymph nodule with a short axis ≥ 10 mm and < 15 mm) as well as truly non-measurable lesions. Non-measurable lesions include: meningeal disease, ascites, pleural or pericardial effusion, inflammatory breast cancer, lymphangitis carcinomatosa of the skin or lung, abdominal masses unable to be diagnosed or followed by imaging techniques, and cystic lesions.

3.1.3 Special considerations regarding lesion measurability

Bone lesions, cystic lesions, and lesions previously treated with local therapy require particular comment:

Bone lesions:

- Bone scan, PET scan or plain films are not considered adequate to measure bone lesions. However, these techniques can be used to confirm the presence or disappearance of bone lesions;
- Lytic lesions or mixed lytic-blastic lesions, with identifiable soft tissue components, that can be evaluated by tomography techniques such as CT or MRI can be considered as measurable lesions if the soft tissue component meets the definition of measurability described above;
- Blastic lesions are non-measurable.

Cystic lesions:

- Lesions that meet the criteria for radiographically defined simple cysts should not be considered as malignant lesions (neither measurable nor non-measurable) since they are, by definition, simple cysts;
- Cystic lesions thought to represent cystic metastases can be considered as measurable lesions, if they meet the definition of measurability described above. However, if noncystic lesions are present in the same patient, these are preferred for selection as target lesions.

Lesions with prior local treatment:

- Tumor lesions situated in a previously irradiated area, or in an area subjected to other locoregional therapy, are usually considered non-measurable unless there has been demonstrated progression in the lesion. Study protocols should detail the conditions under which such lesions would be considered measurable.

3.2 Specifications by Methods of Measurements

3.2.1 Measurement of lesions

All measurements should be recorded in metric notation if clinically assessed. All baseline evaluations should be performed as close as possible to the treatment start and never more than 28 days (4 weeks) before the beginning of the treatment.

3.2.2 Method of assessment

The same method and technique should be used to assess lesions at baseline and during follow-up. Imaging based evaluation should always be done rather than clinical examination unless the lesion(s) being followed cannot be imaged but are assessable by clinical exam.

Clinical lesions: Clinical lesions will only be considered measurable when they are superficial and ≥ 10 mm diameter as assessed using calipers (e.g. skin nodules). For the case of cutaneous lesions, documentation by color photography including a ruler to estimate the size of the lesion is suggested. When lesions can be evaluated by both imaging and clinical examination, imaging evaluation should be undertaken since it is more objective and may also be reviewed at the end of the study.

Chest X-ray: Chest CT is preferred over chest X-ray, especially when tumor progression is an important clinical endpoint, since CT is more sensitive, particularly in identifying new lesions. Chest X-ray is only applicable when the measured lesion boundary is clear and the lungs are well ventilated.

CT and MRI: CT is currently the best available and reproducible method for response evaluation. This guideline has defined measurability of lesions on CT scan based on the assumption that CT slice thickness is ≤ 5 mm. When CT scans have slice thickness greater than 5 mm, the minimum size for a measurable lesion should be twice the slice thickness. MRI is also acceptable in certain situations (e.g. for whole body scans).

Ultrasound: Ultrasound should not be used as a method to measure lesion size. Ultrasound examinations are operation-dependent, and cannot be reproduced at a later date. It cannot be guaranteed that the same technique and measurements will be taken from one assessment to the next. If new lesions are identified by ultrasound in the course of the study, confirmation by CT or MRI is advised. If there is concern about radiation exposure at CT, MRI may be used instead.

Endoscopy and laparoscopy: The utilization of these techniques for objective tumor evaluation is not advised. However, they can be useful to confirm CR when biopsies are obtained, or to determine relapse in trials where recurrence following CR or surgical excision is an endpoint.

Tumor biomarkers: Tumor biomarkers alone cannot be used to assess objective tumor response. However, if the marker levels exceed the upper normal limit at baseline, they must return to the normal levels for evaluation of complete response. Because tumor biomarkers are disease specific, instructions for their measurement should be incorporated into protocols on a disease specific basis. Specific guidelines for both CA-125 response (in recurrent ovarian cancer) and PSA response (in recurrent prostate cancer) have been published. In addition, the Gynecologic Cancer Intergroup has developed CA-125 progression criteria which are to be integrated with objective tumor assessment for use in first-line studies in ovarian cancer.

Cytology/histology: These techniques can be used to differentiate between PR and CR in rare cases if required by protocol (for example, residual benign tumor tissue is commonly observed in lesions of germ cell neoplasm). When effusions are known to be a potential adverse effect of treatment (e.g., with certain taxane compounds or angiogenesis inhibitors), the cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment can be considered if the measurable tumor has met the criteria for response or stable disease in order to differentiate between response (or stable disease) and PD.

4 TUMOR RESPONSE EVALUATION

4.1 Assessment of Overall Tumor Burden and Measurable Disease

To assess objective response or future progression, it is necessary to estimate the overall tumor burden at baseline and use this as a comparator for subsequent measurements. Only patients with measurable lesions at baseline should be included in protocols where objective response is the primary endpoint. Measurable lesion is defined by the presence of at least one measurable lesion. In trials where the primary endpoint is tumor progression (either time to progression or proportion with progression at a fixed date), the protocol must specify if enrollment is restricted to those with measurable lesions or whether patients with non-measurable lesions are also eligible.

4.2 Baseline Documentation of 'Target' and 'Non-Target' Lesions

When more than one measurable lesion is present at baseline, all lesions up to a maximum of five lesions total (and a maximum of two lesions per organ), representative of all involved organs should be identified as target lesions and will be recorded and measured at baseline (this means in instances where subjects have only one or two organ sites involved, a maximum of two and four lesions respectively will be recorded).

Target lesions should be selected on the basis of their size (lesions with the longest diameter), be representative of all involved organs, but in addition should be those that lend themselves to reproducible repeated measurements. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion which can be measured reproducibly should be selected.

Lymph nodes merit special mention since they are normal tissues which may be visible by imaging even if not involved by tumor metastasis. Pathological nodes which are defined as measurable and may be identified as target lesions must meet the criterion of a short axis of ≥ 15 mm by CT scan. Only the short axis of these nodes needs to be measured at baseline. The short axis of a node is the diameter normally used by radiologists to judge if the node is involved by tumor metastasis. Nodule size is normally reported as two dimensions in the plane in which the image is obtained (for CT scan this is almost always the axial plane; for MRI the plane of acquisition may be axial,

sagittal or coronal). The smallest of these measures is the short axis. For example, an abdominal node which is reported as being 20 mm ×30 mm with a short axis of 20 mm qualifies as a malignant and measurable node. In this example, 20 mm should be recorded as the node measurement. Nodes with a short axis ≥ 10 mm but < 15 mm should be considered non-target lesions. Nodes that have a short axis < 10 mm are considered non-pathological and are thus not to be recorded or followed up.

A sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions will be calculated and reported as the baseline sum diameters. If lymph nodes are to be included in the sum, then as noted above, only the short axis is added into the sum. The baseline sum diameters will be used as reference.

All other lesions including pathological lymph nodes should be identified as non-target lesions, and while measurements are not required, they should be recorded at baseline. These lesions should be recorded as "present", "absent", or in rare cases "unequivocal progression". It is possible to record multiple target lesions involving the same organ as a single item on the case record form (e.g. "multiple enlarged pelvic lymph nodes" or "multiple liver metastases").

4.3 Response Criteria

4.3.1 Evaluation of target lesions

Complete response (CR): Disappearance of all target lesions. Any pathological lymph nodules (whether target or non-target) must have reduction in short axis to < 10mm.

Partial response (PR): At least a 30% decrease in the sum of diameters of target lesions, compared with baseline.

Progressive Disease (PD): At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition, the sum must also demonstrate an absolute increase of at least 5 mm (the appearance of one or more new lesions is also considered PD).

Stable disease (SD): Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study.

4.3.2 Special notes on the assessment of target lesions

Lymph nodes: Lymph nodes identified as target lesions should always have the actual short axis measurement recorded (measured in the same anatomical plane as the baseline examination), even if the nodes regress to below 10 mm on study. This means that when lymph nodes are included as target lesions, the sum of lesions may not be zero even if CR criteria are met, since a normal lymph node is defined as having a short axis < 10 mm. CRFs or other data collection methods may

therefore be designed to have target nodal lesions recorded in a separate section where, in order to qualify for CR, each node must have a short axis < 10 mm. For PR, SD and PD, the actual short axis measurement of the nodes is to be included in the sum of target lesions.

Target lesions that become too small to measure: While on study, all lesions (nodal and non-nodal) recorded at baseline should have their actual measurements recorded at each subsequent evaluation, even when very small (e.g., 2 mm). However, sometimes lesions or lymph nodes which are recorded as target lesions at baseline become so faint on CT scan that the radiologist may not feel comfortable assigning an exact measure and may report them as being "too small to measure". When this occurs it is important that a value be recorded on the CRF. If it is the opinion of the radiologist that the lesion has likely disappeared, the measurement shall be recorded as 0 mm. If the lesion is believed to be present and is faintly seen but too small to measure, a default value of 5 mm can be assigned. (Note: It is less likely that this rule will be used for lymph nodules since they usually have a definable size when normal and are frequently surrounded by adipose tissues as in the retroperitoneum; however, if a lymph nodule is believed to be present and is faintly seen but too small to measure, a default value of 5 mm can be assigned in this circumstance as well). This default value is derived from the 5 mm CT slice thickness (but should not be changed with varying CT slice thickness). The measurement of these lesions is potentially non-reproducible, and therefore providing this default value will prevent false evaluation based upon measurement error. To reiterate, however, if the radiologist is able to provide an actual measure, that should be recorded, even if it is below 5 mm.

Lesions that split or coalesce: When non-nodal lesions fragment, the longest diameters of the fragmented portions should be added together to calculate the target lesion sum. Similarly, as lesions coalesce, a plane between them may be maintained that would aid in obtaining maximal diameter measurements of each individual lesion. If the lesions have truly coalesced such that they are no longer separable, the vector of the longest diameter in this instance should be the maximal longest diameter for the coalesced lesion.

4.3.3 Evaluation of non-target lesions

This section provides the definitions of the criteria used to determine the tumor response for the group of non-target lesions. While some non-target lesions may actually be measurable, they need not be measured and instead should be assessed only qualitatively at the time points specified in the protocol.

Complete response (CR): Disappearance of all non-target lesions and normalization of tumor biomarker level. All lymph nodules must be non-pathological in size (short axis < 10 mm).

Non-CR/Non-PD: Persistence of one or more non-target lesion(s) and/or maintenance of tumor biomarker level above the normal limits.

Progressive disease (PD): Unequivocal progression of existing non-target lesions. Note: the appearance of one or more new lesions is also considered PD.

4.3.4 Special notes on assessment of progression of non-target disease

The concept of progression of non-target disease requires additional explanation as follows: When the patient also has measurable disease, to achieve unequivocal progression on the basis of the non-target disease, there must be an overall level of substantial worsening in non-target disease such that the overall tumor load has increased sufficiently to the point where treatment must be discontinued. A modest increase in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status. The designation of overall progression solely on the basis of change in non-target disease in the face of SD or PR of target disease will therefore be extremely rare.

When the patient has only non-measurable disease: This circumstance arises in some phase III trials when it is not a criterion of study inclusion to have measurable disease. The same general concepts apply here as noted above; however, in this instance there is no measurable disease assessment. Because worsening in non-target disease cannot be easily quantified (by definition: if all lesions are truly non-measurable), a useful test that can be applied when assessing patients for unequivocal progression is to consider if the increase in overall disease load based on the change in non-measurable disease is comparable in magnitude to the increase that would be required to declare PD for measurable disease. For example, an increase in tumor burden representing an additional 73% increase in volume (which is equivalent to a 20% increase diameter in a measurable lesion). Examples include an increase in a pleural effusion from "trace" to "large", an increase in lymphangitic disease from "localized" to "widespread", or may be described in protocols as "sufficient to require a change in treatment". Examples include an increase in a pleural effusion from trace to large, an increase in lymphangitic disease from localized to widespread, or may be described in protocols as "sufficient to require a change in therapy". If unequivocal progression is seen, the patient should be considered to have had overall PD at that point. While it would be ideal to have objective criteria to apply to non-measurable disease, the very nature of that disease makes it impossible to do so, and therefore the increase must be substantial.

4.3.5 New lesions

The appearance of new malignant lesions denotes PD; therefore, some comments on detection of new lesions are important. There are no specific criteria for the identification of radiographically detected lesions; however, the finding of a new lesion should be unequivocal. For example, it should not be attributable to differences in scanning technique, change in imaging modality, or findings thought to represent something other than tumor (for example, some new bone lesions that may be simply healing, or re-occurrence of pre-existing lesions). This is particularly important when the patient's baseline lesions show partial or complete response. For example, necrosis of a liver lesion may be reported on a CT scan report as a new cystic lesion, which it is not.

A lesion identified on a follow-up study that is not scanned at baseline will be considered a new lesion and will indicate PD. An example of this is the patient who has visceral disease at baseline and while on study has a CT or MRI brain ordered which reveals metastases. The patient's brain metastases are considered to be evidence of PD even if he/she did not have brain imaging at baseline.

If a new lesion is equivocal, for example, because of its small size, continued treatment and follow-up evaluation are required to clarify if it represents a truly new disease. If repeated scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial identification.

While FDG-PET response evaluations generally need additional study, it is sometimes reasonable to incorporate the use of FDG-PET scanning to complement CT scanning in assessment of progression (particularly possible new disease). New lesions on the basis of FDG-PET imaging can be identified according to the following process:

- Negative FDG-PET at baseline, with a positive FDG-PET at follow-up is a sign of PD based on a new lesion.
- No FDG-PET at baseline and a positive FDG-PET at follow-up:
- If the positive FDG-PET at follow-up corresponds to a new site of disease confirmed by CT, PD is confirmed.
- If the positive FDG-PET at follow-up is not confirmed as a new site of disease on CT, additional follow-up CT scans are needed to determine if there is truly progression occurring at that site (if so, the date of PD will be the date of the initial abnormal FDG-PET scan).
- If the positive FDG-PET at follow-up corresponds to a pre-existing site of disease on CT that is not progressing on the basis of the imaging examination, this is not PD.

4.4 Evaluation of Best Overall Response

The best overall response is the best response recorded from the start of the trial until the end of trial taking into account any necessary requirement for confirmation. On occasion a response may not be documented until after the end of treatment, so protocols should be clear if post-treatment assessments are to be considered in the evaluation of best overall response. Protocols must specify how any new treatment introduced before progression will affect best response evaluation. The patient's best overall response evaluation will depend on the findings of both target and non-target diseases and will also take into consideration the characteristics of new lesions. Furthermore, depending on the nature of the study and the protocol requirements, it may also require confirmatory measurement. Specifically, in non-randomized studies where response is the primary endpoint, confirmation of PR or CR is needed to determine either one is the best overall response.

4.4.1 Time point response

It is assumed that at each time point specified in protocol, an efficacy response occurs. Table 1 provides a summary of the overall response status calculation at each time point for patients who have measurable disease at baseline.

Table 1. Time point response: subjects with target (+/- non-target) disease

Target Lesion	Non-Target Lesion	New Lesion	Overall Response
CR	CR	Non	CR
CR	Non-CR/Non-PD	Non	PR
CR	Not evaluable	Non	PR
PR	Non-PD or not all evaluable	Non	PR
SD	Non-PD or not all evaluable	Non	SD
Not all evaluable	Non-PD	Non	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = not evaluable

If patient does not have measurable lesions (no target lesions), refer to Table 2.

Table 2. Time point response: subjects with non-target lesion only

Non-Target Lesion	New Lesion	Overall Response
CR	Non	CR
Non-CR/Non-PD	Non	Non-CR/Non-PD[a]
Not all evaluable	Non	Not evaluable
Equivocal PD	Yes or No	PD
Any	Yes	PD

[a]: "Non-CR/non-PD" is preferred over SD for non-target disease. Since SD is increasingly used as an endpoint for response evaluation, non-CR/non-PD response is developed to address the absence of lesion measurability.

4.4.2 Missing assessments and inevaluable designation

When no imaging/measurement is done at all at a particular time point, the patient is not evaluable at that time point. If only a subset of lesion measurements is made at an evaluation, usually the case is also considered not evaluable at that time point, unless a convincing argument can be made that the contribution of the individual missing lesion(s) has/have no effect on the assigned time point response. This would be most likely to happen in the case of PD. For example, if a patient had three measured lesions with a baseline sum of 50 mm and only two lesions with a sum of 80 mm were assessed at a subsequent follow-up, the patient has achieved PD status, regardless of the contribution of the missing lesion.

4.4.3 Best overall response: all time points

The BOR is determined once all the data for the patient are known.

BOR determination in studies where confirmation of complete or partial response is not required: BOR in these studies is defined as the best response across all time points (for example, a patient who has SD in evaluation at Cycle 1, PR at Cycle 2, and PD at the last cycle has a BOR of PR). When SD is believed to be best response, it must also meet the protocol specified minimum time calculated from baseline. If the minimum time is not met when SD is otherwise the BOR, the patient's BOR depends on the subsequent assessments. For example, a patient who has SD at Cycle 1, PD at Cycle 2 and does not meet minimum duration for SD, will have a BOR of PD. The same patient lost to follow-up after the first SD assessment would be considered not evaluable.

BOR determination in studies where confirmation of complete or partial response is required: Complete or partial responses may be claimed only if the criteria for each are met at a subsequent time point as specified in the protocol (generally 4 weeks later). In this circumstance, the BOR can be interpreted as in Table 3.

Table 3. Best overall response when confirmation of CR and PR required

Overall Response at First Time Point	Overall Response at Subsequent Time Point	Best Overall Response
CR	CR	CR
CR	PR	SD, PD, or PR[a]
CR	SD	SD (provided minimum criteria for SD duration met, otherwise, PD)
CR	PD	SD (provided minimum criteria for SD duration met, otherwise, PD)
CR	NE	SD (provided minimum criteria for SD duration met, otherwise, NE)
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD (provided minimum criteria for SD duration met, otherwise, PD)
PR	NE	SD (provided minimum criteria for SD duration met, otherwise, NE)
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = not evaluable.
[a]: If a CR is truly met at first time point, then efficacy of any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, will still be evaluated as PD at that point (since disease will reappear after CR). Best response will depend on whether minimum duration for SD is met. However, sometimes CR may be claimed when subsequent scans suggest small lesions are likely still present and in fact the subject has PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

4.4.4 Special notes on response evaluation

When nodal disease is included in the sum of target lesions and the nodules decrease to a normal size < 10 mm, they may still have a measurement reported on scans. This measurement should be recorded even though the nodules are normal in order not to overstate progression should it be based on increase in size of the nodules. As noted earlier, this means that subjects with CR may not have a total sum of zero on the CRF.

In trials where confirmation of response is required, repeated "NE" time point evaluations may complicate best response evaluation. The analysis plan for the trial must address how missing data/evaluations will be addressed in determination of response and progression. For example, in most studies it is reasonable to consider a subject with time point responses of PR-NE-PR as a confirmed response.

Subjects with an overall deterioration of health status requiring discontinuation of treatment without objective evidence of PD at that time should be reported as symptomatic deterioration. Efforts should be made to evaluate objective progression even after discontinuation of treatment. Symptomatic deterioration is not a description of an objective response: it is a reason for discontinuation of treatment. The objective response status of such subjects is to be determined by evaluation of target and non-target lesion as shown in Tables 1-3.

Conditions that are defined as early progression, early death and not evaluable are study specific and shall be clearly described in each protocol (depending on treatment duration and treatment cycle).

In some circumstances it may be difficult to distinguish residual lesions from normal tissues. When the evaluation of complete response depends upon this definition, it is recommended to perform a biopsy before evaluating the efficacy of complete remission of local lesions. FDG-PET may be used to confirm a response to a CR in a manner similar to a biopsy in cases where a residual radiographic abnormality is thought to represent fibrosis or scarring. The use of FDG-PET in this circumstance should be prospectively described in the protocol and supported by disease specific medical literature for the indication. However, it must be acknowledged that both approaches may lead to false positive CR due to limitations of FDG-PET and biopsy resolution/sensitivity.

For equivocal findings of progression (e.g., very small and uncertain new lesions; cystic changes or necrosis in existing lesions), treatment may continue until the next scheduled evaluation. If at the next scheduled evaluation, progression is confirmed, the date of progression should be the earlier date when progression was suspected.

4.5 Frequency of Tumor Re-Evaluation

Frequency of tumor re-evaluation during treatment should be protocol-specific and consistent with the type and schedule of treatment. However, in the phase II studies where the beneficial effect of treatment is not known, follow-ups for every 6-8 weeks (timed to coincide with the end of a cycle) is reasonable. Interval adjustments can be justified in specific regimens or circumstances. The protocol should specify which organ sites are to be evaluated at baseline (usually those most likely to be involved with metastatic disease for the tumor type under study) and how often evaluations are repeated. Normally, all target and non-target sites are evaluated at each assessment. In selected circumstances, certain non-target organs may be evaluated less frequently. For example, bone scans may need to be repeated only when CR is identified in target disease or when progression in bone is suspected.

After the treatment, the need for tumor re-evaluations depends on whether the study has made the response rate or the time to an event (progression/death) an endpoint. If time to an event (e.g. TTP/DFS/PFS) is the main endpoint of the study, then routine scheduled re-evaluation of protocol specified sites of disease is warranted. In randomized comparative studies in particular, the scheduled assessments should be performed as identified on a calendar schedule (for example: every 6-8 weeks on treatment or every 3-4 months after treatment) and should not be affected by delays in therapy, drug holidays or any other events that might lead to imbalance in a treatment group in the timing of disease assessment.

4.6 Confirmatory Evaluation/Duration of Response

4.6.1 Confirmation

In non-randomized trials where response is the primary endpoint, confirmation of PR and CR is required to ensure responses identified are not the result of measurement error. This will also permit appropriate interpretation of results in the context of historical data where response has traditionally required confirmation in such trials. However, in all other circumstances, i.e., in randomized trials (phase II or III) or studies where stable disease or progression are the primary endpoints, confirmation of response is not required since it will not add value to the interpretation of trial results. However, elimination of the requirement for response confirmation may increase the importance of central review to protect against bias, in particular in studies which are not blinded.

In the case of SD, measurements must have met the SD criteria at least once after study entry at a minimum interval (in general not less than 6-8 weeks) that is defined in the study protocol.

4.6.2 Duration of overall response

The duration of overall response will be measured from the time measurement criteria are first met for CR/PR (whichever is first recorded) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded on study). The duration of overall complete response will be measured from the time criteria are first met for CR until the first date that recurrent or progressive disease is truly documented.

4.6.3 Duration of stable disease

Stable disease is measured from the start of the treatment (in randomized trials, from date of randomization) until the criteria for progression are met, taking as reference the smallest sum on study (if the baseline sum is the smallest, this is the reference for calculation of PD). The clinical relevance of the duration of stable disease varies in different studies and diseases. If the proportion of patients achieving SD for a minimum period of time is an endpoint in a particular study, the protocol should specify the minimal time interval required between two measurements for determination of SD.

Note: The duration of response and stable disease as well as the progression-free survival are influenced by the frequency of follow-up after baseline evaluation. It is not in the scope of this guideline to define a standard follow-up frequency. The frequency should take into account many parameters including disease types and stages, treatment periodicity and standard practice. However, these limitations of the precision of the measured endpoint should be taken into account if comparisons between trials are to be made.

4.7 PFS/TTP

4.7.1 Phase II clinical studies

This guideline is focused primarily on the use of objective response endpoints for phase II clinical studies. In some circumstances, response rate may not be the optimal method to assess the potential anti-cancer activity of new agents/regimens. In such cases PFS/PPF at landmark time points, might be considered appropriate alternatives to provide an initial signal of biologic effect of new agents. It is clear, however, that in an uncontrolled trial, these measures are subject to criticism since an apparently promising observation may be related to biological factors such as patient selection and not the impact of the intervention. Thus, phase II screening studies utilizing these endpoints are best designed with a randomized control. Exceptions may exist where the behavior patterns of certain cancers are so consistent (and usually consistently poor) that a non-randomized trial is justifiable. However, in these cases it will be essential to document with care the basis for estimating the expected PFS or PPF in the absence of a treatment effect.

Appendix 2 Common Soft Tissue Sarcomas With Poor Chemotherapy Sensitivity

- Alveolar soft-part sarcoma
- Clear cell sarcoma of soft tissue
- Dedifferentiated liposarcoma
- Atypical lipomatous tumor
- Well differentiated liposarcoma

Appendix 3 Histological Coding for Common Soft Tissue Sarcomas With Poor Chemotherapy Sensitivity

No.	Histological Classification	Other Types	Histological Coding
1	Alveolar soft tissue sarcoma		Alveolar soft tissue sarcoma
2	Clear cell sarcoma of soft tissue		Clear cell sarcoma
3	Atypical lipomatous tumor		Liposarcoma
4	Well-differentiated liposarcoma		Liposarcoma
5	Leiomyosarcoma (not including skin)		Leiomyosarcoma
6	Pleomorphic rhabdomyosarcoma		Rhabdomyosarcoma
7	Dedifferentiated liposarcoma		Liposarcoma
8	Myxoid liposarcoma		Liposarcoma
9	Pleomorphic liposarcoma		Liposarcoma
10	Synovial sarcoma, non-specific		Synovial sarcoma
99	Others	Adult fibrosarcoma	Fibrosarcoma
99	Others	Desmoplastic small round cell tumor	Desmoplastic small round cell tumor
99	Others	Low-grade myofibroblastic tumor	Fibrosarcoma
99	Others	Low-grade fibromyxoid sarcoma	Fibrosarcoma
99	Others	Pleomorphic undifferentiated sarcoma	Undifferentiated sarcoma
99	Others	Malignant solitary fibrous tumor	Fibrosarcoma
99	Others	Malignant granular cell tumor	Nerve sheath tumor
99	Others	Malignant neurilemmoma	Nerve sheath tumor
99	Others	Malignant peripheral nerve sheath tumor	Nerve sheath tumor
99	Others	Solitary fibrous tumor	Fibrosarcoma
99	Others	Synovial sarcoma, non-specific	Synovial sarcoma
99	Others	Synovial sarcoma, biphasic	Synovial sarcoma
99	Others	Synovial sarcoma, spindle cell	Synovial sarcoma
99	Others	Giant cell tumor of soft tissues	Fibrosarcoma

No.	Histological Classification	Other Types	Histological Coding
99	Others	Angiosarcoma of soft tissue	Angiosarcoma
99	Others	Epithelioid sarcoma	Epithelioid sarcoma
99	Others	Epithelioid hemangioendothelioma	Angiosarcoma
99	Others	Spindle cell/sclerosing rhabdomyosarcoma	Rhabdomyosarcoma
99	Others	Synovial sarcoma, spindle cell	Synovial sarcoma
99	Others	Undifferentiated/unclassified sarcomas	Undifferentiated sarcoma
99	Others	Undifferentiated pleomorphic sarcoma	Undifferentiated sarcoma
99	Others	Undifferentiated spindle cell sarcoma	Undifferentiated sarcoma
99	Others	Myxofibrosarcoma	Fibrosarcoma