PROTOCOL	
Title	**A Multicenter, Randomized, Double-Blind, Placebo-Controlled Clinical Trial of Apatinib Mesylate Tablets in The Treatment of Locally Advanced/Metastatic Radioiodine (RAI) Refractory Differentiated Thyroid Cancer**
Protocol Number	**Ahead-T302**
NCT Number	**NCT03048877**
Project Investigator	**Professor Lin Yansong** **Peking Union Medical College Hospital** **Professor Qin Shukui** **The 81st Hospital of Chinese People's Liberation Army**
Version Number	**2.0**
Document Date	**2019/11/19**

2016/12/26 version number 1.2

2016/12/04 version number 1.1

2016/11/07 version number 1.0

Contents

1 Introduction

1.1 Treatment of thyroid cancer

According to the histological characteristics, thyroid carcinoma can be divided into papillary thyroid carcinoma, follicular carcinoma, medullary thyroid carcinoma (MTC) and anaplastic thyroid carcinoma (ATC). Except that medullary thyroid carcinoma originated from para follicular C cells evolved from nerve spines, the other three originated from thyroid follicular cells. Differentiated thyroid carcinoma (DTC) refers to thyroid papillary carcinoma and thyroid follicular carcinoma, which account for nearly 90% of all thyroid cancers. Patients with DTC have a good long-term survival rate. DTC has a 5-year death rate of about 2% and a recurrence rate of less than 15%; a 10-year mortality rate of 5% and a recurrence rate of 20%; a cumulative 30-year mortality rate of roughly 10% and a cumulative recurrence rate of 30%.

Surgery is the first choice for the treatment of thyroid cancer, especially for DTC, supplemented by thyroid stimulating hormone inhibition therapy and radioactive iodine therapy. There is no effective treatment for unresectable, or metastatic radioiodine resistant thyroid cancer. Most of the patients with DTC have good prognosis and long-term survival after surgery, radioiodine, and thyroid stimulating hormone inhibition therapy. However, there are still a few patients with distant metastasis, and DTC will gradually dedifferentiate in the natural state or during treatment in 1/3 of the patients with distant metastasis, and lose the ability to take iodine, and eventually develop into iodine refractory DTC. The average survival time of patients with iodine refractory DTC is only 3-5 years, and the 10-year survival is about 10%. Thyroid cancer is not sensitive to chemotherapy, only a small number of patients receive chemotherapy. At present, several antiangiogenic drugs for thyroid cancer were under clinical research and achieved good therapeutic efficacy. Sorafenib and lenvatinib were approved by the food and Drug Administration (FDA) on November 22, 2013, and February 13, 2015, respectively. Sorafenib was approved by the China Food and Drug Administration (CFDA) in March 2017 for the treatment of advanced/local

recurrence or metastasis Iodine refractory (RAI) differentiated thyroid cancer (DTC). But until November 2019, lenvatinib still has not been approved for the treatment of thyroid cancer in China.

1.2 Apatinib

Apatinib is a small-molecule VEGFR tyrosine kinase inhibitor, which is invested and developed by Hengrui Pharmaceutical Co., Ltd. with independent intellectual property rights. Its chemical name is N-[4- (1-cyanocyclopentyl) phenyl]-2-[(4-pyridylmethyl) amino-3-pyridylformamide methanesulfonate, with the molecular formula of $C_{25}H_{27}N_5O_4S$ (molecular weight: 493.58 (methanesulfonate)). Apatinib mainly plays an anti-angiogenic role in the treatment of malignant tumors by inhibiting VEGFR. Preclinical studies have shown that the anti-tumor efficacy of apatinib is better than that of PTK787

The China Food and Drug Administration (CFDA) officially approved the use of apatinib in the treatment of advanced gastric cancer on October 17, 2014.

1.3 Research advancement of apatinib in the treatment of thyroid cancer

Ten patients were treated with 750 mg q.d. of apatinib in a single-arm clinical trial for radioiodine refractory differentiated thyroid cancer (RAIR-DTC) initiated by the research team of the Nuclear Medicine Department of Peking Union Medical College Hospital, and nine patients (90 percent) achieved partial response (PR) and one patient achieved stable disease (SD) after eight weeks of treatment. The target lesion's mean diameter was 21.6 mm at baseline and had shrunk by 34% to 12.7 mm after 8 weeks of treatment. The average Tg (Thyroglobulin) level declined by 21% after 2 weeks of treatment; after 8 weeks of treatment, the average Tg level decreased from 1583.5 ng/mL to 294.6 ng/mL, indicating a rapid biochemical reaction. Hand-foot syndrome, hypertension, and hypocalcemia were the most common Grade 3 adverse events (AEs), accounting for 50%, 30%, and 20% of all cases, respectively. 70% of patients had to discontinue apatinib treatment owing to Grade 3 AEs, and 50% had to reduce apatinib dose due to tolerance issues. The study's findings preliminarily confirmed efficacy and safety of apatinib in the treatment of RAIR-DTC.

Therefore, we plan to conduct a multicenter, randomized, double-blind, placebo-controlled clinical study to further prove the efficacy and safety of apatinib in the treatment of locally advanced/metastatic radioiodine refractory differentiated thyroid cancer. The current study will use 500 mg q.d. of apatinib as the initial dose for the safety consideration, based on the previous research results, which suggested possible tolerance concerns with 750 mg q.d. of apatinib in the treatment of patients with RAIR-DTC.

2 Study Objectives

2.1 Primary Objective

- To assess the efficacy of apatinib compared with placebo, as measured by progression-free survival (PFS) according to Response Evaluation Criteria in Solid Tumors (RECIST) Version 1.1 criteria.

2.2 Secondary Objectives

- To assess the overall survival (OS) in subjects treated with apatinib, compared with placebo.
- To assess the objective response rate (ORR) in subjects treated with apatinib, compared with placebo.
- To assess the disease control rate (DCR) in subjects treated with apatinib, compared with placebo.
- To assess the duration of response (DOR) in subjects treated with apatinib, compared with placebo.
- To assess the time to response (TTR) in subjects treated with apatinib, compared with placebo.
- To assess the safety and tolerability of apatinib, compared with placebo.

3 Study Design and Methodology

3.1 Overall Study design

This is a multicenter, randomized, double-blind, placebo-controlled clinical trial to evaluate the efficacy and safety of apatinib in the treatment of locally advanced/metastatic radioiodine refractory differentiated thyroid cancer (RAIR-DTC).

The eligible subjects will be randomly assigned to either the treatment group or control group in a 1:1 ratio. Apatinib or placebo will be given 500 mg orally once daily for 28 days. Apatinib treatment will be discontinued at the time of disease progression, intolerable toxicity, withdrawal of consent, or due to investigator's evaluation. Tumor imaging with CT or magnetic resonance imaging (MRI) will be done at the end of first and second treatment cycles, then every two treatment cycles thereafter. The efficacy will be assessed per RECIST version 1.1 by investigators. Safety assessment including vital signs, physical examination, blood and urine analysis, liver, kidney and thyroid function tests, electrolyte profile, electrocardiogram and ECOG performance status will be done every two weeks for one-two cycles, and at the end of each subsequent cycle. Adverse events will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0). Survival status will be followed up every two months until death or loss to follow-up.

In the extension phase, subjects in the Apatinib group will continue the treatment when anticipated to benefit from apatinib treatment. For the subjects in the control group, they will be allowed to receive apatinib upon experiencing radiographic disease progression (PD) per RECSIT v1.1 by investigator. Extended treatments are terminated after progression disease, intolerable toxicity, and withdrawal of informed consents during extension phase.

3.2 Randomization Method

Through the central randomization system of the Department of health statistics, the Second

Military Medical University, the subjects will be randomly assigned to treatment group or control group by a ratio of 1:1. The stratified factor is gender (male, female).

The investigator at each center will log into the central randomization system and enter the subject's relevant hierarchy factor information when the subjects meet the inclusion and exclusion criteria and sign the informed consent form. According to the hierarchical factor distribution of the subjects in each center, the system will allocate a unique drug number to the subject at random and issue the matching drug number. Within 24 hours of being randomly allocated a medicine, subjects must begin taking it. All the trial's participating centers were selected through a competitive process.

3.3 Sample Size

Sample size determination is performed by PASS 11 software with group sequential log-rank test settings. Assuming a median PFS of 5.8 months in control group, and a median PFS of 12 months in treatment group; subjects randomly assigned to two group by a ratio of 1:1, with an enrollment time of 24 months and a follow-up of 12 months; 50 subjects in each group, to observe 37 PFS events in the treatment group and 46 PFS events in the control group, will provide 90% power to detect a difference between groups using a log-rank test with a two-sided significance level (α)=0.05. With a 15% drop-out rate, each group will require 59 people, for a total sample size of 118 subjects.

4 Study Population

4.1 Study Population

This is a multicenter, randomized, double-blind, placebo-controlled clinical trial to observe and evaluate the efficacy and safety of apatinib in the treatment of locally advanced/metastatic radioiodine refractory differentiated thyroid cancer (RAIR-DTC).

4.2 Inclusion Criteria

1. Age: ≥ 18 years old, both male and female.

2. Locally advanced or metastatic differentiated thyroid carcinoma (papillary, follicular, Hurthle cells, poorly differentiated carcinoma) with at least one measurable lesion after treatment (according to RECIST Version 1.1 standard, the measurable lesion with a long diameter≥10 mm or an enlarged lymph node with a short diameter≥15 mm; According to RECIST Version 1.1 standard. Previously treated lesions can be used as the targeted lesion after the progression is confirmed.

3. Progressive disease has occurred within 12 months before enrollment (by RECIST Version 1.1).

4. Radioiodine refractory (meet one of the following requirements)

a) The targeted lesion completely has lost the capacity of iodine uptake during radioiodine therapy

b) The subjects have received a single dose of radioiodine (≥ 3.7 GBq [≥ 100mCi]) within 12 months and the progressive disease has occurred.

c) The interval between two radioiodine therapies is less than 12 months, and the dose is more than 3.7 GBq [≥ 100 mCi], and the progressive disease has occurred over 12 months after iodine treatment at least one time

d) Cumulative dose of radioiodine ≥ 22.2 GBq (≥ 600 mCi).

5. The function of the main organs is normal and meets the following requirements:

Blood routine: (no blood transfusion within 14 days)

a) Hb≥90g/L；

b) ANC ≥1.5×10^9/L；

c) PLT ≥80×10^9/L；

Biochemical examination should meet the following standards

a) BIL < 1.5 folds limit of normal (ULN)

b) ALT and AST < 2.5 × ULN, ALT and AST < 5 × ULN if liver metastasis.

c) Serum Cr ≤ 1 × ULN, endogenous creatinine clearance rate ≥ 50ml / min (Cockcroft Gault formula)

6. ECOG PS: 0-2 points.

7. The estimated survival ≥ 3 months.

8. Women of childbearing age must have pregnancy test (serum) (-) within 7 days before enrollment. In addition, appropriate contraceptive methods should be used voluntarily during the observation period and within 8 weeks after the last administration of apatinib tablets; for men, it should be surgical sterilization, or agree to use appropriate contraceptive methods during the observation period and within 8 weeks after the last administration of apatinib tablets.

9. Subjects volunteered to join the study and signed the informed consent form (ICF).

10. Subjects with good compliance, who can be followed up and report AEs according to the requirements of the protocol.

4.3 Exclusion Criteria

1. Histological subtypes of thyroid carcinoma other than differentiated type (such as medullary carcinoma, lymphoma, or sarcoma).

2. Subjects who have used VEGFR-TKI, such as vandetanib, cabozantinib, lenvatinib, sunitinib, sorafenib, etc. within 1 month before the first dose of study treatment.

3. Heart diseases that cannot be well controlled, such as: (1) cardiac insufficiency (grade 2 or above) according to the New York Heart Association (NYHA) standard, or cardiac color Doppler ultrasound examination: LVEF (left ventricular ejection fraction) < 50%; (2) unstable angina pectoris; (3) myocardial infarction within 1 year before the beginning of the study; (4) clinically significant supraventricular or ventricular arrhythmia, which need treatment or intervention: (5) QTc > 450ms (male); QTc > 470ms (female) (QTc interval is calculated by Fridericia formula); the average of three measurements is calculated with an interval of 2 minutes if QTc is abnormal; (6) subjects with hypertension who cannot be reduced to the normal range after anti-hypertension drug

treatment (SBP ≥ 140mmHg or DBP ≥ 90mmHg) (based on the average value of BP readings obtained from ≥ 2 measurements), the above parameters can be achieved through anti-hypertension treatment.

4. Factors affecting oral drug absorption (such as inability to swallow, nausea and vomiting, chronic diarrhea, and intestinal obstruction, etc.).

5. Subjects with the risk of gastrointestinal bleeding should not be included in the study, including: (1) subjects with active peptic ulcer lesions and occult blood in stool (+ +); (2) subjects with a history of melena and hematemesis within 3 months.

6. Abnormal coagulation function (INR > 1.5, APTT > 1.5 × ULN) with bleeding tendency.

7. Obvious hemoptysis within 2 months before screening, or the amount of hemoptysis reached half a teaspoon (2.5ml) or more every day.

8. Arteriovenous thrombosis events occurred within 12 months before screening, such as cerebrovascular accident (including transient ischemic attack, cerebral hemorrhage, cerebral infarction), deep vein thrombosis and pulmonary embolism, etc.

9. Received radiotherapy for thyroid carcinomas within 28 days before screening.

10. Subjects with active bleeding, ulcer, intestinal perforation, intestinal obstruction, and surgery within 28 days before screening (the wound must have recovered completely if the operation is performed before 28 days or more).

11. Uncontrollable infections during screening.

12. Pregnant or lactating women.

13. Subjects with depression (HAMD score ≥ 17, see Appendix 5).

14. Subjects who are not suitable for the study.

4.4 Shedding/rejection criteria

1. The effectiveness and / or safety evaluation cannot be carried out due to the failure of medication

according to the provisions of this protocol.

2. Using other anti-tumor drugs during the study.

3. Do not meet the Inclusion/exclusion criteria.

4.5 Termination Criteria

Termination of the treatment does not mean the withdrawal from the study. Subjects must continue to complete the remaining study visits as required by the protocol. If the subject meets any of the following criteria, the treatment must be terminated:

1. Subjects withdraw their informed permission and request to be removed from the study.

2. Progressive disease is confirmed by investigator.

3. Pregnancy during the study.

4. the subjects cannot tolerate the toxicity after dose adjustment.

5. Other situations indicating the need to withdraw from the study.

5 Study Treatment

5.1 Treatment and Administration

5.1.1 Study Drug introduction

The study drug is provided by Jiangsu Hengrui Pharmaceutical Co., Ltd and the specific information is shown in the table below.

Information of the drug

Drug	apatinib mesylate tablet	apatinib mesylate tablet simulator (placebo)
Dosage form	tablet	tablet
Specification:	250 mg	-

Storage	sealed and protected from light, stored under 25℃	sealed, protected from light, stored below 25℃
Administration route	oral	oral
Manufacturer	Jiangsu Hengrui Pharmaceutical Co., Ltd.	Jiangsu Hengrui Pharmaceutical Co., Ltd.

5.1.2 Administration regimen

- Group A (apatinib group): apatinib mesylate 500mg (250mg/tablet, 2 tablets), po, q.d.; after meals (try to take at the same time every day), 4 weeks as a cycle.
- Group B (control group): 2 tablets of apatinib mesylate mimic, po, q.d.; after meals (try to take at the same time every day), 4 weeks as a cycle.

Until the patient's progressive disease, the toxicity becomes unacceptable, the patient withdraws informed consent, or the researcher determines that the medicine must be stopped, the drug should be halted.

5.1.3 Preservation, distribution, and management of the study drug

Special staff should oversee the drug's management, distribution, and recovery in this research trial. The study drugs should be kept sealed, away from light, and at a temperature below 25°C. There is a 24-month validity term. The researcher must guarantee that the study drugs are only administered to the clinical trial participants, and that the dosage and administration follow the protocol. The leftover study drugs should be returned to Jiangsu Hengrui Pharmaceutical Co., Ltd., and no non-clinical study participants should be given the drug.

At the moment of drug distribution, two people should sign the drug receipt form. One copy each should be kept by the clinical research unit and Jiangsu Hengrui Pharmaceutical Co., Ltd. The remaining study drugs and empty boxes should be collected at the end of the study, and both parties should sign the drug recycling form. Each drug's distribution and recovery should be recorded on a separate record sheet in real time.

The inspector is in charge of ensuring that the leftover study drugs in clinical trials are supplied,

used, stored, and disposed of properly.

5.1.4 Study drug destruction

The total quantity of the drug is 120% of the designed dosage, and the remaining drug should be recovered by the inspector and returned to Jiangsu Hengrui Pharmaceutical Co., Ltd.

The researcher can trash the used study drugs and empty boxes without harming the public's health after alerting the sponsor in writing. All drug handling should be documented by the investigator.

These records must show the certificate and quantity of each batch of destroyed drug, as well as the method of disposal (in accordance with the requirements of local laws), as well as the personnel handling the drug.

5.2 Study Drug Dose Modification

5.2.1 Dose

The initial dose of the drug is 500mg, oral, once a day.

If AEs of Grade 3 or higher occur during treatment, the dose should be discontinued or reduced. As a cycle, the minimum dose can be reduced to 250mg once a day for 28 days.

When the dose is modified in each administration cycle, the dose is suspended first to guarantee consistency over the whole research. If the individuals still cannot tolerate the dose after taking the dose suspension measurement, the dose is reduced.

5.2.2 Dosage adjustment

The criteria of dose suspension or reduction are as follows:

- Dose suspension and dose reduction can be carried out, only when the hematologic toxicity and non-hematological toxicity reach Grade 3 (special AEs such as hypertension, drug suspension or adjustment should be specified separately)
- in non-hematological toxicity, subjects received active symptomatic treatment for subjects with

weight loss, nausea, vomiting, alopecia, fever with definite causes (such as infection), and elevated alkaline phosphatase (AKP) (Grade 3/4) without dose suspension and/or dose reduction.

- for hypertension AEs, antihypertensive drugs should be started and/or added when blood pressure is confirmed to be elevated (systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg). If the blood pressure remains good (systolic blood pressure < 140 mmHg or diastolic blood pressure < 90 mmHg), the drug use can be continued. If systolic blood pressure ≥ 160 mmHg or diastolic blood pressure ≥ 100 mmHg still exists after the maximum dose of antihypertensive drug treatment, the drug administration should be interrupted; if the systolic blood pressure ≤ 150 mmHg and diastolic blood pressure ≤ 95 mmHg, and the subject has received the stable dose of antihypertensive drugs for at least 48 hours, the dosage can be reduced to 250mg.

5.2.3 Dose interruptions

To ensure the drug intensity of the treatment for the subjects in the trial, the time of each suspension and the total time of suspension in each administration cycle should not exceed 2 weeks; the suspension time of each cycle should not exceed 2 times.

5.2.4 Dose reduction

Dose adjustment is made at any time in each administration cycle. After the dose is reduced, it should not be brought back to the previous level. After the dose is adjusted to 250 mg q.d., continued dose adjustment is not allowed, including dose increase or decrease for any reason, but dose suspension is still allowed.

Table 1 principle of dose adjustment of apatinib

AE classification	**NCI**	**Dose adjustment**
Hematology AE	Grade 3	stop medication and continue to use the original dose when the AE ≤ Grade 2. If Grade 3 or above AEs occur again, the drug should be continued after a-dose reduction

	Grade 4	the drug is suspended, and the drug is continued after a-dose reduction and the AE recovered to ≤ Grade 2.
Non-hematological AE	Grade 3	The drug should be suspended until the AE≤ Grade 1, and the drug should be continued with the original dose; if Grade 3 or above AEs occurs again, the drug should be continued after a-dose reduction.
	Grade 4	the drug is suspended, and the drug is continued after a-dose reduction and the AE recovered to ≤ Grade 1.

Note: the above evaluation is carried out according to NCI-CTCAE4.0

For subjects with gastrointestinal perforation, wound dehiscence, fistula, severe bleeding, nephrotic syndrome or hypertensive crisis, administration should be stopped permanently. Subjects with moderate to severe proteinuria or severe hypertension who need further diagnosis should stop using the drug temporarily.

6 Concomitant medication and common AEs

6.1 Drugs prohibited or used with caution during the trial

6.1.1 Drugs interfering with liver P450 enzyme

Apatinib had strong inhibitory effect on CYP3A4, YP2C9 and CYP2C19 (IC50 < 0.5 μM). Omeprazole should be banned during treatment (except for the cases that must be used due to serious AEs). The inducers (dexamethasone, rifampicin, and phenobarbital) and inhibitors (itraconazole, erythromycin and clarithromycin) of CYP3A4, substrates of CYP3A4 (simvastatin, cyclosporin and piperazine, etc.) and other drugs metabolized by CYP3A4 (such as benzodiazepines, dihydropyridines, calcium antagonists and HMG-CoA reductase inhibitors) should be used with caution. Careful use of CYP2C9 substrates (diclofenac, phenytoin, piroxicam, S-warfarin and tolbutamide) and CYP2C19 substrates (diazepam, imipramine, lansoprazole and S-mephenytoin).

6.1.2 Drugs causing prolongation of cardiac QT interval

Since protein tyrosine kinase drugs have side effects in clinical, such as prolonging QT interval, it is required to use drugs that can prolong QT interval carefully, including antibiotics (Fluoroquinolones: sparfloxacin, gatifloxacin, levofloxacin, moxifloxacin, ofloxacin, ciprofloxacin, etc.; Macrocyclic lipids: erythromycin, clarithromycin, talimycin, azithromycin, roxithromycin, metronidazole, etc.), anti-angina drugs (ranolazine, ivabradine, etc.), anti-arrhythmic drugs (quinidine, procaine amine, propylamine, flucarni, propafenone, amiodarone, dronedarone, sotalol, dofetilide, ibutilide, etc.), anti-psychotic drugs (risperidone, flufenazine, droperidol, haloperidol, tiridazine, pimozide, olanzapine, clozapine, etc.). Anti-fungal drugs (voriconazole, posaconazole, etc.), anti-malarial drugs (mefloquine, chloroquine, etc.), gastrointestinal drugs (antiemetics: ondansetron, granisetron, dorasetron, droperidol, (0.625-1.25 mg may be a safe dose)), hydroxyzine; gastrointestinal motility drugs: cisapride, domperidone, metoclopramide, etc.), antihistamines (terfenadine, astemizole, hydroxyzine, etc.) and anti-depressants (amitriptyline, imipramine, clomipramine, dutiapine, doxepin, etc.) and other drugs.

6.1.3 Traditional Chinese medicine and immune preparations with anti-cancer effect

The use of modern Chinese medicine preparations and immunomodulators with indications for the treatment of thyroid cancer approved by CFDA is prohibited during the trial.

6.1.4 Drugs that can be combined used during the trial

All kinds of clinical AE (such as hypertension, hand foot skin reaction, liver, and kidney function damage, etc.) should be treated according to the judgment of clinicians. All drugs used in combination should be recorded in case report form (CRF) in strict accordance with GCP.

6.2 Symptomatic treatment suggestions for common AEs

6.2.1 Hand foot skin reaction

Hand foot skin reaction (HFSR): it is a kind of skin toxicity, which is more obvious in compression or stress area when it occurs. HFSR may appear in subjects with tumor during chemotherapy or molecular targeted therapy. HFSR is characterized by numbness, dullness,

paresthesia, tingling, no pain or pain, skin swelling, or erythema, desquamation, chapping, sclerosing blisters and severe pain.

6.2.1.1 HFSR classification

Grade 1：numbness/dullness/paresthesia of hand and/or foot, painless swelling, or erythema and/or discomfort that does not affect normal activity.

Grade 2: painful erythema and swelling of hand and/or foot and/or discomfort affecting the patient's daily activities.

Grade 3: wet desquamation of hand and/or foot, ulcers, blisters, or severe pain and/or severe discomfort that prevents the patient from working or performing daily activities. Severe pain and loss of skin function is rare.

6.2.1.2 Symptomatic treatment and management of HFSR

Take some necessary symptomatic supportive treatments, including strengthening skin care, keeping skin clean, avoiding secondary infection, avoiding pressure or friction, using moisturizing cream or lubricant, topical use of urea or corticosteroid emulsion or lubricant, and locally using antifungal or antibiotic therapy.

6.2.2 Hypertension

6.2.2.1 Suggestions for drug-related hypertension treatment

- Monitoring and treatment of hypertension: blood pressure monitoring should be conducted every week in the first six weeks of targeted drug treatment. Once hypertension occurs, the following standard treatment drugs can be given angiotensin-II receptor blocker (ARB), angiotensin converting enzyme inhibitor (ACEI), β-receptor blocker and diuretic, or combination of the above drugs.
- Optimization of drug selection (non-liver metabolism drugs):

  1) Valsartan (Devon), 80～320mg q.d.

2) Atenolol，50～100mg q.d.

3) HCTZ，12.5～100mg q.d.

4) Telmisartan, 20～80mg q.d.

5) For those whose blood pressure is difficult to control: amlodipine, 2.5～10mg q.d.

6.2.2.2 Clinical stages of hypertension and routine treatment

- Prehypertension: (120-139/80-89 mmHg, or systolic blood pressure 120-139 mmHg): there is no indication of using antihypertensive drugs, only blood pressure is monitored
- Grade 1 hypertension: (140-159/90-99 mmHg, or systolic blood pressure is 140-159 mmHg): use drugs to reduce blood pressure and monitor blood pressure at the same time; most of them use thiazide diuretics, but also consider using ARB, ACEI, β-receptor blockers and calcium channel blockers; continue to take apatinib, and two anti-hypertensive drugs can be considered if the blood pressure is not ideal.
- Grade 2 hypertension: (160-179/100-109 mmHg, or systolic blood pressure ≥ 160-179 mmHg): combined use of two drugs (usually thiazide diuretics and ARB or ACEI or β-receptor blockers or calcium channel blockers); monitor blood pressure.
- Grade 3 hypertension: (≥180/110 mmHg, or systolic blood pressure≥180 mmHg): combined use of two drugs (usually thiazide diuretics and ARB or ACEI or β-receptor blockers or calcium channel blockers); closely monitor blood pressure; assess other risk factors (such as target organ damage, diabetes and other clinical symptoms), and take corresponding measures
- Hypertension crisis: refers to a serious clinical state in which blood pressure is excessively elevated and diastolic blood pressure exceeds 16.0-17.3 kPa (120-130 mmHg). At present, there is no unified classification at home and abroad. Recently, it can be divided into two types from the perspective of clinical treatment: (1) Hypertension emergencies: diastolic blood pressure>16.0 kPa (120 mmHg), accompanied by acute or progressive target organ damage, such as cerebral infarction, intracranial or subarachnoid hemorrhage, hypertensive

encephalopathy, and so on. Rapidly progressive hypertension based on chronic essential hypertension is the most common (about 40%~50%). (2) Hypertension urgencies: diastolic blood pressure>16.0 kPa (120 mmHg) without or with only slight organ damage; sodium nitroprusside or nifedipine is used to reduce blood pressure rapidly; diazepam and phenobarbital are used to stop convulsion; furosemide and mannitol are used to reduce intracranial pressure.

Note: in case of hypertension crisis, the use of apatinib should be stopped immediately and the patient should be withdrawn from the clinical trial.

6.2.3 Management of gastrointestinal bleeding

Gastrointestinal bleeding, including fecal occult blood (++), hematemesis or bloody stool, should be treated actively. Subjects with upper gastrointestinal bleeding should be fasting, and given antacids, gastric mucosal protection drugs, hemostatics (hemostatic acid, reptilase, etc.), if necessary, octreotide can be used; Subjects with lower gastrointestinal bleeding should be given hemostasis, blood transfusion and support treatment; if bleeding cannot be controlled, surgical assistance is needed.

6.2.4 Suggestions for proteinuria treatment

The subjects with two consecutive urinary protein (++) need 24-hour urinary protein quantification.

Table 2 suggestions for proteinuria treatment

UPC < 3g	continue the treatment at the same dose, and monitor according to clinical indications
UPC ≥ 3g	Step 1: obtain 24-hour urine protein Step 2: if 24-hour urine protein < 3g, the treatment can be continued at the same dose If 24-hour urinary protein ≥ 3g, discontinue treatment until UPC

	returns to < 3g, start treatment again at a lower dose, and monitor UPC in the following whole treatment stage. If UPC ≥ 3g, 24-hour urine protein is obtained. Step 3: if 24-hour urine protein ≥ 3g after repeated reduction, terminate the trial, and follow up according to the protocol

Note: in case of nephrotic syndrome, the use of apatinib should be stopped immediately and the patient should be withdrawn from the clinical trial.

6.2.5 Thyroid function

During the whole treatment period, thyroid function of all subjects is closely monitored. TSH should be controlled at <0.1uIU/mL. If it exceeds the range, corresponding treatment should be given.

7 Study Procedures

7.1 Study Schedule

Item	Visit during the screening		Visit during the treatment							Visit during extension phase	Vist during follow- up every 2 months
			1st cycle				2nd cycle		3rd cycle and after		
	(-14~0 d)	(-7~0 d)	7d	14d	21d	28d	42d	56d	--	--	--
Visit	**V1**		**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8-Vn**	**V'1-V'n**	
Informed consent	x										
Past treatment history		x									
Vital signs [a]		x		x		x	x	x	#	&	
Physical examination		x				x		x	#	&	
Blood pressure [b]		x	x	x	x	x	x	x	#	&	
Myocardial enzyme spectrum and color Doppler ultrasound [c]	x			x		x		x	#	&	
Coagulation function		x				x		x	#	&	
Electrocardiogram	x			x		x		x	#	&	
ECOG PS score		x				x		x	#	&	
Blood routine		x	x	x	x	x	x	x	#	&	

Urinalysis		**x**		**x**		**x**		**x**	#	**&**	
Defecation routine		**x**		**x**		**x**		**x**	#	**&**	
Liver and kidney function, electrolytes		**x**		**x**		**x**		**x**	#	**&**	
Imaging examination (X-ray/CT/MRI/B-ultrasound)[d]	**x**					**x**		**x**	**&@**	**&@**	
Serum Tg and TgAb		**x**		**x**		**x**	**x**	**x**	**&@**	**&@**	
Thyroid function[e]		**x**		**x**		**x**	x	**x**	**&@**	**&@**	
Life quality score		**x**				**x**		**x**	#	**&**	
Hamilton Depression Scale score		**x**									
Pregnancy test (blood)		**x**									
Collection of whole blood and plasma samples		**x**[f]							@	@	
Tissue specimen collection		**x**									
AE			**x**	**x**	**x**	**x**	**x**	**x**	#	#	
Concomitant medication			**x**	**x**	**x**	**x**	**x**	**x**	#	#	
Death time											**x**

Cancer treatment									@	@	**x**

Note: (1) &: once every 2 cycles (the last day); (2) #: once every cycle (last day) and once every 2 cycles (last day) after 6 cycles of medication; (3) @: when the progressive disease or withdraw from the study in advance, if there is no such examination result within 4 weeks before the end of the trial, it needs to be collected within 7 days after the end of the study; (4) a: The subjects can increase the frequency of self-examination of vital signs according to their own conditions and fill in the relevant abnormal conditions in the research diary card. b: blood pressure monitoring: blood pressure monitoring is completed by the subjects themselves, at least three days a week in the first two cycles after medication, and once a week in the later cycles. Once abnormal, the blood pressure should be monitored to the normal level every day and recorded in the patient's diary card; in addition, the blood pressure is measured again by the researcher at each follow-up. During the blood pressure measurement, smoking and coffee drinking are prohibited within 30 minutes before the measurement, and at least 10 minutes of quiet rest are taken. During the measurement, the sitting position is taken, the elbow is placed at the same level as the heart, and each blood pressure measurement is taken on the same side. c. Myocardial enzyme spectrum and color Doppler ultrasound: supplement this examination when ECG is abnormal and has clinical significance; d: if chest, abdomen or pelvis imaging examination is negative at screening, X-ray or B-ultrasound should be used for examination during the study period; if X-ray or B-ultrasound shows positive results, CT or MRI scan should be used to confirm; e: the examination items include FT3, FT4, TSH, RT3 [RT3 can be selected as needed]; f: collect within 3 days after the first medication. (5) Tumor samples (tissue and blood) are collected when subjects include in this study. If a tumor sample is not collected, the reason should be record. (6) During the first two cycles of treatment, all examinations and life quality scores are performed within ± 3 days of the corresponding date, and after two cycles, all examinations and life quality scores are performed within ± 7 days of the corresponding date.

7.2 Visit during the screening

The screening visit should be carried out within 14 days before the start of the trial. Informed consent should be signed, imaging examination (X-ray/CT/MRI/B ultrasound), ECG examination should be carried out. In case of precordial pain, palpitation and other symptoms, myocardial enzyme spectrum should be monitored (creatine kinase, lactate dehydrogenase, aspartate aminotransferase, creatine kinase isoenzyme, α-hydroxybutyrate dehydrogenase) and color Doppler ultrasound should be performed (if necessary). Medical history should be collected within 7 days before the start of the trial, including previous treatment history, vital signs, physical examination and ECOG PS score, blood pressure test, blood routine (hemoglobin (HB), red blood cell (RBC), white blood cell (WBC), absolute neutrophil (ANC), neutrophil ratio (N), lymphocyte ratio (Lym), monocyte ratio (Mon), platelet count (PLT)), urine routine (urine protein, urine red blood cell (microscopic examination), urine white blood cell (microscopic examination)), stool routine, liver and kidney function (total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin (IBIL), ALT, AST, AKP, TP, ALB, BUN, Cr, UA, γ-GT, LDH), electrolytes (K, Na, Cl, CA) , phosphorus (P)), coagulation function (PT, APTT (thrombin time), TT, fibrinogen (FBG), international standardized ratio (INR)), serum Tg and TgAb, thyroid function (FT3, FT4, TSH, RT3 [RT3 can be selected as needed]), life quality score, Hamilton Depression Scale score and (serum) pregnancy test (if there are clinical indications indicating pregnancy is possible) etc. Tumor samples (tissue and blood) are collected when subjects include in this study. If a tumor sample is not collected, the reason should be record.

7.3 Visit during the randomization phase

1) Within the first cycle after administration:

Blood routine examination is performed once a week.

The vital signs, urine routine, stool routine, liver and kidney function examination, electrolyte test, serum Tg and TgAb levels, thyroid function test and electrocardiogram examination should be carried out once every two weeks. In case of precordial pain, palpitation and other symptoms,

myocardial enzyme spectrum should be detected immediately, and cardiac color Doppler ultrasound should be performed

Physical examination, coagulation function test, ECOG PS score, imaging examination and life quality score are performed at the end of the first week.

Blood pressure monitoring: subjects should be monitored at least three times per week. During blood pressure measurement, smoking and coffee consumption should be avoided for at least 30 minutes before measurement, and at least 10 minutes of quiet rest should be taken. During the measurement, the elbow should be put at the same level as the heart in a sitting position.

2) During the second cycle after administration

Vital signs, blood routine, serum Tg and TgAb levels and thyroid function are detected every 2 weeks.

At the end of the second cycle, imaging examination and life quality score are performed. Physical examination, coagulation function test, ECOG PS score, urine routine test, stool routine test, liver and kidney function test, electrolyte test, serum Tg and TgAb level, thyroid function test and ECG examination are conducted. If the symptoms such as precordial pain and palpitation are found, myocardial enzyme spectrum should be detected immediately, and cardiac color Doppler ultrasound should be performed.

Blood pressure monitoring: the same as the first cycle.

3) At the end of the third week after the start of administration and after

Vital signs, physical examination (including skin toxicity, nausea, fatigue, tooth function such as looseness, tooth drop, etc.), ECOG PS score, life quality score, blood routine, urine routine, stool routine, liver and kidney function test, electrolyte test, blood coagulation function test and electrocardiogram examination are performed weekly, in case of precordial pain, palpitation and other symptoms, the myocardial enzyme spectrum should be detected immediately, and cardiac color Doppler ultrasound should be performed (subjects can increase the frequency of self-examination of vital signs according to their own situation and fill in the relevant abnormal

conditions in the research diary. The above examination is changed to once every 2 cycles after 6 cycles of medication.

The efficacy of tumor is evaluated every 2 cycles (imaging examination: CT/MRI/B ultrasound), serum Tg and TgAb levels are detected, and thyroid function is examined.

Blood pressure monitoring: the subjects should self-monitor the blood pressure at least once a week, once abnormal, they need to monitor to the normal level every day, and record in the patient's diary card. During the visit, the researcher measures the blood pressure, and the measurement method is the same as that in the first cycle.

4) The concomitant medication and AE should be observed and recorded at any time from the beginning of administration to 30 days after the end of administration.

7.4 Visit during the extension phase

Subjects judged as progressive disease by investigator can be unblinded. If the subjects unblinded, the subjects who is placebo group can decide whether to enter the apatinib group. The subjects in apatinib group begin to receive apatinib treatment until the researchers determined that it is not suitable to continue treatment (including but not limited to progressive disease or intolerance). However, for the subjects who have taken apatinib before after blinding, on the premise of safety and possible effectiveness, they can enter the extension phase, continue to be treated with open-label apatinib, if the subjects themselves are willing and the investigator think it is feasible.

The subjects enter the extension phase visit:

Vital signs, physical examination (including skin toxicity, nausea, fatigue, tooth function such as looseness and tooth loss) and ECOG PS score, life quality score, blood routine, urine routine, stool routine, liver and kidney function test, electrolyte test, blood coagulation function test and electrocardiogram examination are performed every 2 cycles. In case of precordial pain, palpitation and other symptoms, the myocardial enzyme spectrum should be detected immediately, and cardiac color Doppler ultrasound should be performed (subjects can increase the frequency of self-examination of vital signs according to their own situation and fill in the relevant abnormal

conditions in the research diary). Tumor efficacy should be evaluated every 2 cycles (imaging examination: CT/MRI/B-ultrasound), and the serum Tg and TgAb levels, and thyroid function examination are monitored and conducted every 2 cycles.

Blood pressure is monitored weekly and measured by the researcher during the visit.

7.5 Visit during follow-up

Subjects with early withdrawal, intolerable AEs, progressive disease in the randomization phase or progressive disease in the extension phase can enter the survival follow-up period.

- If the subjects have been evaluated within 4 weeks before the survival follow-up period, the survival follow-up and imaging efficacy evaluation should be conducted once every 2 months until imaging progression or other anti-tumor treatment is started.
- If the subjects have not been evaluated within 4 weeks before the survival follow-up period, a follow-up and evaluation of therapeutic efficacy (imaging examination, detection of serum Tg and TgAb levels, and thyroid function test (including FT3, FT4, TSH, RT3 [RT3 can be selected as needed]) and then the survival follow-up and imaging efficacy are conducted once every 2 months until imaging progression or other anti-tumor treatment is started.

During the follow-up period, the time of death or loss of follow-up and other anti-tumor treatment is recorded.

7.6 Follow-up of AEs

Follow-up and final evaluation should be made for the AEs that have not recovered when the drug is stopped. All subjects should be followed up for 30 days after the last medication to detect any new AE.

8 Efficacy evaluation and analysis

The efficacy assessment will be conducted at the end of the first and second cycles, as well as every eight weeks after that. The assessment will take place within 3 days after the completion of

each period. From the third cycle, CT or MRI examinations will be performed within 7 days after completing of every two cycles.

Subjects who develop progressive disease during randomization phase will be unblinded and treat with apatinib until disease progression or at the discretion of the investigators.

9 Safety Evaluation

9.1 AEs

AEs refer to any adverse medical events occurred in the subjects of clinical trials, which may not be related to the treatment. Therefore, AEs can be any sign, symptom, or temporary drug-related disease, which should be considered whether it is related to medication

According to management needs, safety monitoring (reporting AEs or serious AEs) should be carried out from the beginning of subjects' enrollment to the end of the study. AEs occurred after signing the informed consent form and during the treatment are also considered as AEs.

9.2 Treatment-related Adverse Events (TRAEs)

TRAEs are defined as all harmful and unexpected medication responses associated with any dose. The drug's reaction implies that there is at least a plausible probability of a causal association between the drug and the AEs, i.e., the relationship cannot be ruled out.

9.3 Serious Adverse Events (SAEs)

SAEs refer to the medical events that need hospitalization, extended hospital stays, and is life-threatening or cause disability, teratogenesis and death during the clinical trials.

9.3.1 Events that should be dealt with as SAEs

In principle, pregnancy and lactation cannot be included. If pregnancy occurs during the trial, the patient should withdraw from the trial immediately and be followed up closely. Even if the mother and child are completely normal without any AEs, the situation should be recorded. Even if

the pregnancy does not belong to SAEs, the SAEs report form should be used for reporting.

9.3.2 Events that should not be dealt with as SAEs

Hospitalization due to progressive disease or unexpected reasons such as medical insurance reimbursement does not belong to SAEs. In this study, due to the severity of the disease, some so-called SAEs cases may need to be excluded from immediate reporting:

A. Hospitalization and surgical treatment are optional

B. Hospitalization can be chosen to simplify treatment or research measures

9.4 Recording and evaluation of AEs

AEs should be described in medical terms. All AEs should be recorded in the corresponding part of the CRF. If it is a SAE, the SAE report form (including initial, follow-up and summary reports) should be completed. All the cases involved in the trial should be included in the summary. The reasons for the cases dropped out or eliminated during the summary should be explained. In case of death or serious toxic reaction in the trial, a detailed case report should be made. The cause of death should be found out and the relationship between the drug and the death should be emphasized. All AEs during follow-up should be closely followed up until they are properly resolved or stable.

The following aspects of each event should be recorded in CRF:

(1) Occurrence time (start time), recovery time (end time).

(2) The researchers are responsible to evaluate and grade AEs, according to the definition of NCI-CTC AE version 4.0.

(3) The possible association between AEs and the drug is evaluated according to the five-grade classification of "definitely related, likely to be related, possibly related, possibly unrelated and irrelevant". The first three levels are judged to be related to the drug. The sum of the three is taken as the numerator in the calculation of the incidence of AEs, and the number of all subjects used to evaluate the safety is taken as the denominator.

Table 3 Criteria for determining the relationship between AEs and drugs

Criteria	definitely related	likely to be related	possibly related	possibly unrelated	irrelevant
Reasonable time sequence	Yes	Yes	Yes	Yes	No
Response types of known drugs	Yes	Yes	Yes	No	No
Improved after removing the cause	Yes	Yes	Yes or No	Yes or No	No
Repeated after administration again	Yes	No	No	No	No
There may be another explanation for the response	No	No	No	Yes	No

（4）Measures taken for the drug (such as continued medication, reduced dosage, continued medication after suspension, withdrawal and other situations, please specify in detail).

（5）Consequences are defined as follows: disappearance, improvement, no improvement, and aggravation

Note: if the patient has the same AE for several times, it must be recorded and reevaluated each time.

9.5 Reporting procedures for SAEs

During the trial period, if there is SAE, the researcher should report to the research leader, Peking Union Medical College Hospital and the 81st Hospital of Chinese people's Liberation Army by fax or telephone within 24 hours, and report to CFDA within 24 hours, and report to the ethics committee of the hospital in time.

9.6 Emergency unblinding

In case of emergency (such as SAE, or when the subject's medication state must be known before selecting the next treatment strategy), the investigator might break the blindness before the subject's disease progression. If it is necessary to break the blindness, the investigator should inform CRA, the sponsor and the ethics committee. Meanwhile, the time, subject, reason, and process of breaking the blindness should be recorded in the central randomization system

10 Data management and statistical analysis

10.1 Analysis Set

- Full Analysis Set (FAS): FAS include all randomized subjects who received at least 1 dose of study treatment. FAS will serve as the main analysis set of efficacy analysis.
- Per-Protocol Set (PPS): PPS will include all subjects in the FAS who did not experience significant protocol deviations which may have impact on the results.
- Safety Set (SS): SS will include all subjects who received at least 1 dose of the study treatment.

10.2 Methods of Statistical analysis

All statistical analysis will be calculated with SAS 9.3 or above. For continuous variables, descriptive statistics including sample size, mean, standard deviation, median, minimum, and maximum, will be presented. For categorical variables, frequencies and percentages will be presented. Survival data will be estimated by Kaplan-Meier method, and K-M plots will be plotted if necessary.

10.2.1 Efficacy analysis

The median PFS and its 95% Confidence Interval (CI) will be estimated by Kaplan-Meier method, and K-M plots will be plotted; the difference between groups will be tested by unstratified log-rank test. The unstratified Cox proportional hazards regression model will be used for multivariate analysis to compare the difference between groups, and to calculate the hazard ratio (HR) and its 95% CI.

For secondary efficacy endpoints, such as OS and DOR, Kaplan-Meier method will be used to estimate the median and its 95% CI, and the K-M plots will be plotted. The survival between groups will be tested by log-rank test; DCR, ORR and their 95% CI will be calculated, and the differences between groups will be compared by CMH test or Chi-square test.

The primary and secondary efficacy analysis will be performed on FAS; a sensitive analysis

for primary and secondary endpoints will be performed on PPS.

10.2.2 Safety analysis

Safety analysis will be performed on the Safety Set.

AEs will be coded as low-level terms, preferred terms and major system organ classes (SOC) using MedDRA (v20.0 or higher version).

Treatment emergent adverse events (TEAEs) is defined as any AE that is new or aggravated from baseline (before study treatment) after the initiation of medication.

Only TEAEs will be summarized. All AEs, TEAEs or other indications will be listed.

The incidence of TEAEs will be summarized according to the number and percentage of major system organs and preferred terms. Under the same system organ classification and standard terminology, if the event occurs for multiple times, the subjects only calculate the highest level of the event. The number and incidence of TEAEs are also summarized according to their relationship to treatment.

The frequencies and percentages of AEs, SAEs, ≥Grade 3 AEs, ≥Grade 3 SAEs, drug-related AEs, drug-related SAE, AEs with incidence rate ≥5%, SAEs with incidence rate ≥5%, AEs leading to dose adjustment, AEs leading to termination of treatment will be summarized by group. If necessary, for AEs with special interest, Kaplan-Meier method will be used to analyze the occurrence time and duration of events.

10.3 Sample Size Determination

The primary endpoint is PFS in subjects with iodine refractory differentiated thyroid cancer treated with apatinib. Referring to the data of phase III clinical trial of sorafenib monotherapy in subjects with DTC (Lancet 2014; 384 (9940): 319-328), assuming a median PFS of 5.8 months in control group, and a median PFS of 12 months in treatment group; subjects randomly assigned to two group by a ratio of 1:1, with an enrollment time of 24 months and a follow-up of 12 months; 50 subjects in each group, to observe 37 PFS events in treatment group and 46 PFS events in control

group, will provide 90% power to detect a difference between groups using a log-rank test with a two-sided significance level (α)=0.05. Consider a drop-out rate of 15%, 59 subjects in each group are needed, and the total sample size is 118 subjects.

10.4 Interim Analysis

An interim analysis is planned to be conducted during the study. Interim analysis will be performed when ~50 PFS events occurred in the trial (about 60% of the total PFS events required), and the final analysis is performed when ~83 PFS events occurred. According to O'Brien-Fleming's method, the cut-off values of interim analysis are shown in the table below. If $P < 0.0082$ and $HR < 1$ in the interim analysis, it is considered that the efficacy of the treatment group is better than that of the control group, and the trial can be terminated in advance if the test is successful. If the number of events exceeds the preset 50 cases in the interim analysis, the alpha will be redistributed using the Lan-DeMets alpha consumption function according to the actual number of events; whether the treatment group is superior to the control group will be decided according to the new nominal value.

Table 4 Boundary values and nominal test levels for interim and final analysis

Time point	Event number	Z	Bilateral nominal test level
Interim	50	-2.6438	0.0082
Terminal	83	-1.9825	0.0474

In this study, an independent data monitoring committee (IDMC) will be established, and the interim analysis will be completed by IDMC. The safety data and efficacy data (PFS) will be analyzed by IDMC. If the PFS results in the interim analysis do not exceed the preset threshold value, the sample size may also be re-estimated according to Cui et al (1999).

The IDMC established in this study includes experts in clinical, statistical, methodological, ethical, and other fields. It works independently from the sponsors and researchers.

11 Ethics, regulations, and administrative principles

11.1 Ethical principles

The study will be carried out in accordance with the principles established by the 18th World Federation of Medical Association (Helsinki, 1964) and all subsequent amendments.

11.2 Laws and regulations

This study will be conducted in accordance with all applicable laws and regulations.

11.3 Informed Consent

Clinical research physicians must fulfill the obligation of full disclosure to the subjects, indicating that they are completely voluntary to participate in the clinical study, and have the right to withdraw at any time without discrimination and retaliation at any stage of the trial, and their medical treatment, rights and interests will not be affected, and they can continue to receive other treatment methods. It is necessary for the subjects to know that the personal data of participating in the trial are kept confidential. The subjects should also be informed of the nature of the clinical trial, the objective of the trial, the expected possible benefits, the possible risks and inconveniences, the alternative treatment methods and the rights and obligations of the subjects stipulated in the declaration of Helsinki Time to consider whether they are willing to participate in the experiment and sign the informed consent form.

12 Principal Investigator

Site	Principal Investigator	Title
Peking Union Medical College Hospital	Lin Yansong	Professor

the 81st Hospital of the Chinese people's Liberation Army	Qin Shukui	Professor

13 Study Duration

Start time: December 2016, planned end time: June 2021

AMENDMENT HISTORY

Date	Version	Brief description of main change
2016/11/07	Version 1.0	NA
2016/12/04	Version 1.1	● Add the interim analysis
2016/12/26	Version 1.2	● Update the NCI-CTCAE version 3.0 to version 4.0
2019/11/19	Version 2.0	● Update the current status of Radioiodine (RAI) Refractory Differentiated Thyroid Cancer treatment in China. ● Add the ClinicalTrials.gov Identifier ● Update the study duration

Attachment 1 RECIST v1.1

Response Evaluation Criteria in Solid Tumors RECIST Version 1.1

The following information was extracted from ***New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)*** Eur J Cancer. 2009 Jan;45(2):228-47.

3. Tumor measurability at baseline

3.1 Definition

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

3.1.1 Measurable

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

- 10mm by CT scan (CT scan slice thickness no greater than 5 mm; see Appendix II on imaging guidance).
- 10mm caliper measurement by clinical exam (lesions which cannot be accurately measured with calipers should be recorded as non-measurable).
- 20mm by chest X-ray
- Malignant lymph nodes: To be considered pathologically enlarged and measurable, a lymph node must be ⩾15mm in short axis when assessed by CT scan (CT scan slice thickness recommended to be no greater than 5 mm). At baseline and in follow-up, only the short axis will be measured and followed (see Schwartz et al. in this Special Issue). See also notes below on 'Baseline documentation of target and non-target lesions' for information on lymph node measurement.

3.1.2 Unmeasurable lesions

All other lesions, including small lesions (longest diameter <10mm or pathological lymph

nodes with ⩾10 to <15mm short axis) as well as truly non-measurable lesions. Lesions considered truly non-measurable include: leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, abdominal masses/abdominal organomegaly identified by physical exam that is not measurable by reproducible imaging techniques.

3.1.3 Special considerations for lesion measurement

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 imaging techniques to measure bone lesions. However, these techniques can be used to confirm the presence or disappearance of bone lesions.
- Lytic bone lesions or mixed lytic-blastic lesions, with identifiable soft tissue components, that can be evaluated by cross sectional imaging 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 bone 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:

- Tumour lesions situated in a previously irradiated area, or in an area subjected to other loco-regional therapy, are usually not considered 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, using calipers if clinically assessed. All baseline evaluations should be performed as close as possible to the treatment start and never more than 4 weeks before the beginning of the treatment.

3.2.2 Method of assessment

The same method of assessment and the same technique should be used to characterise each identified and reported lesion 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 P10mm diameter as assessed using calipers (e.g. skin nodules). For the case of skin lesions, documentation by colour photography including a ruler to estimate the size of the lesion is suggested. As noted above, when lesions can be evaluated by both clinical exam and imaging, 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, particularly when progression is an important endpoint, since CT is more sensitive than X-ray, particularly in identifying new lesions.However, lesions on chest X-ray may be considered measurable if they are clearly defined and surrounded by aerated lung. See Appendix II for more details.

CT, MRI: CT is the best currently available and reproducible method to measure lesions selected for response assessment.

This guideline has defined measurability of lesions on CT scan based on the assumption that CT slice thickness is 5mm or less. As is described in Appendix II, 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 body scans). More details concerning the use of both CT and MRI for assessment of objective tumour response evaluation are provided in Appendix II.

Ultrasound: Ultrasound is not useful in assessment of lesion size and should not be used as a method of measurement. Ultrasound examinations cannot be reproduced in their entirety for independent review at a later date and, because they are operator dependent, it cannot be guaranteed that the same technique and measurements will be taken from one assessment to the next (described in greater detail in Appendix II). 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 of CT in selected instances.

Endoscopy, laparoscopy: The utilisation of these techniques for objective tumour evaluation is not advised. However, they can be useful to confirm complete pathological response when biopsies are obtained or to determine relapse in trials where recurrence following complete response or surgical resection is an endpoint.

Tumour markers: Tumour markers alone cannot be used to assess objective tumour response. If markers are initially above the upper normal limit, however, they must normalise for a patient to be considered in complete response. Because tumour markers 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 CA125 progression criteria which are to be integrated with objective tumour assessment for use in first-line trials 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 lesions in tumour types such as germ cell

tumours, where known residual benign tumours can remain). 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 tumour has met criteria for response or stable disease in order to differentiate between response (or stable disease) and progressive disease.

4. Tumor response evaluation

4.1 Assessment of overall tumour burden and measurable disease

To assess objective response or future progression, it is necessary to estimate the overall tumour burden at baseline and use this as a comparator for subsequent measurements. Only patients with measurable disease at baseline should be included in protocols where objective tumour response is the primary endpoint. Measurable disease is defined by the presence of at least one measurable lesion (as detailed above in Section 3). In studies where the primary endpoint is tumour progression (either time to progression or proportion with progression at a fixed date), the protocol must specify if entry is restricted to those with measurable disease or whether patients having non-measurable disease only 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 patients have only one or two organ sites involved a maximum of two and four lesions respectively will be recorded).For evidence to support the selection of only five target lesions, see analyses on a large prospective database in the article by Bogaerts et al.

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. To illustrate this point, see the example in Fig. 3 of Appendix II.

Lymph nodes merit special mention since they are normal anatomical structures which may be visible by imaging even if not involved by tumour. As noted in Section 3, pathological nodes which are defined as measurable and may be identified as target lesions must meet the criterion of a short axis of P15mm by CT scan. Only the short axis of these nodes will contribute to the baseline sum. The short axis of the node is the diameter normally used by radiologists to judge if a node is involved by solid tumour. Nodal 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, saggital or coronal). The smaller of these measures is the short axis. For example, an abdominal node which is reported as being 20mm· 30mm has a short axis of 20mm and qualifies as a malignant, measurable node. In this example, 20mm should be recorded as the node measurement (See also the example in Fig. 4 in Appendix II). All other pathological nodes (those with short axis P10mm but <15 mm) should be considered non-target lesions. Nodes that have a short axis <10mm are considered non-pathological and should not be recorded or followed.

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 to further characterize any objective tumour regression in the measurable dimension of the disease.

All other lesions (or sites of disease) including pathological lymph nodes should be identified as non-target lesions and should also be recorded at baseline. Measurements are not required, and these lesions should be followed as ‘present’, ‘absent’, or in rare cases ‘unequivocal progression’ (more details to follow). In addition, it is possible to record multiple nontarget 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 nodes (whether target or non-target) must have reduction in short axis to <10 mm.
- Partial Response (PR): At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
- 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 to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the appearance of one or more new lesions is also considered progression).
- 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 10mm on study. This means that when lymph nodes are included as target lesions, the 'sum' of lesions may not be zero even if complete response criteria are met, since a normal lymph node is defined as having a short axis of <10mm. Case report forms 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 achieve a short axis <10mm. 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., 2mm). 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 case report form. If it is the opinion of the radiologist that the lesion has likely disappeared, the measurement should be recorded as 0mm. If the lesion is believed to be present and is faintly seen but too small to measure, a default value of 5mm should be assigned (Note: It is less likely that this rule will be used for lymph nodes since they usually have a definable size when normal and are frequently surrounded by fat such as in the retroperitoneum; however, if a lymph node is believed to be present and is faintly seen but too small to measure, a default value of 5mmshould be assigned in this circumstance as well). This default value is derived from the 5mm CT slice thickness (but should not be changed with varying CT slice thickness). The measurement of these lesions is potentially non-reproducible, therefore providing this default value will prevent false responses or progressions 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 5mm.

Lesions that split or coalesce on treatment. As noted in Appendix II, 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 tumour 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 normalisation of tumour marker level. All lymph nodes must be non-pathological in size (<10mm short axis).

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

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

4.3.4 Special notes on assessment of progression of nontarget disease

The concept of progression of non-target disease requires additional explanation as follows:

When the patient also has measurable disease. In this setting, 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, even in presence of SD or PR in target disease, the overall tumour burden has increased sufficiently to merit discontinuation of therapy (see examples in Appendix II and further details below). A modest 'increase' in the size of one or more non-target lesions is usually not sufficient to quality 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 entry to have measurable disease. The same general concepts apply here as noted above, however, in this instance there is no measurable disease assessment to factor into the interpretation of an increase in non-measurable disease burden. 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 burden 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: i.e. an increase in tumour 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 therapy'. Some illustrative examples are shown in Figs. 5 and 6 in Appendix II. 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, therefore the increase must be substantial.

4.3.5 New lesions

The appearance of new malignant lesions denotes disease progression; therefore, some comments on detection of new lesions are important. There are no specific criteria for the identification of new radiographic lesions; however, the finding of a new lesion should be unequivocal: i.e., not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour (for example, some 'new' bone lesions may be simply healing or flare 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 in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression. 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 therapy and follow-up evaluation will clarify if it represents truly new disease. If repeat scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial scan.

While FDG-PET response assessments 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 algorithm:

- 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, this is PD. 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 anatomic images, 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 study treatment until the end of treatment taking into account any requirement for confirmation. On occasion a response may not be documented until after the end of therapy so protocols should be clear if post-treatment assessments are to be considered in determination of best overall response. Protocols must specify how any new therapy introduced before progression will affect best response designation. The patient's best overall response assignment will depend on the findings of both target and non-target disease and will also take into consideration the appearance of new lesions. Furthermore, depending on the nature of the study and the protocol requirements, it may also require confirmatory measurement (see Section 4.6). Specifically, in non-randomised trials where response is the primary endpoint, confirmation of PR or CR is needed to deem either one the 'best overall response'. This is described further below.

4.4.1 Time point response

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

When patients have non-measurable (therefore non-target) disease only, Table 2 is to be used.

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 (NE) at that time point.

If only a subset of lesion measurements is made at an assessment, usually the case is also considered NE at that time point, unless a convincing argument can be made that the contribution of the individual missing lesion(s) would not change the assigned time point response. This would be most likely to happen in the case of PD. For example, if a patient had a baseline sum of 50mm with three measured lesions and at follow-up only two lesions were assessed, but those gave a sum of 80 mm, the patient will have achieved PD status, regardless of the contribution of the missing lesion.

4.4.3 Best overall response: all time points

The best overall response is determined once all the data for the patient is known.

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

Best response determination in trials 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 best overall response can be interpreted as in Table 3.

4.4.4 Special notes on response assessment

When nodal disease is included in the sum of target lesions and the nodes decrease to 'normal' size (<10 mm), they may still have a measurement reported on scans. This measurement should be

recorded even though the nodes are normal in order not to overstate progression should it be based on increase in size of the nodes. As noted earlier, this means that patients with CR may not have a total sum of 'zero' on the case report form (CRF).

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

Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as 'symptomatic deterioration. Every effort should be made to document objective progression even after discontinuation of treatment. Symptomatic deterioration is not a descriptor of an objective response: it is a reason for stopping study therapy. The objective response status of such patients is to be determined by evaluation of target and non-target disease as shown in Tables 1–3.

Conditions that define 'early progression, early death and inevaluability' are study specific and should be clearly described in each protocol (depending on treatment duration, treatment periodicity).

In some circumstances it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends upon this determination, it is recommended that the residual lesion be investigated (fine needle aspirate/biopsy) before assigning a status of complete response. FDG-PET may be used to upgrade 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 assessment. If at the next scheduled assessment, progression is confirmed, the date of progression should be the earlier date when progression was suspected.

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

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not all evaluated	Non-PD	No	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 = inevaluable			

Table 2 Time point response: patients with non-target disease only.

Non-target lesions	New lesions	Overall response
CR	No	CR
Non-CR/non-PD	No	Non-CR/non-PD
Not all evaluated	No	NE
Unequivocal PD	Yes or No	PD
Any	Yes	PD
CR = complete response, PD = progressive disease, and NE = inevaluable. a 'Non-CR/non-PD' is preferred over 'stable disease' for non-target disease since SD is increasingly used as endpoint for assessment of efficacy in some trials so to assign this category when no lesions		

Non-target lesions	New lesions	Overall response
can be measured is not advised.		

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

Overall response First time point	Overall response 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 = inevaluable.

a If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the patient had 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.

4.5. Frequency of tumour re-evaluation

Frequency of tumour re-evaluation while on treatment should be protocol specific and adapted to the type and schedule of treatment. However, in the context of phase II studies where the beneficial effect of therapy is not known, follow-up every 6–8 weeks (timed to coincide with the end of a cycle) is reasonable. Smaller or greater time intervals than these could 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 tumour 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 complete response is identified in target disease or when progression in bone is suspected.

After the end of the treatment, the need for repetitive tumour evaluations depends on whether the trial has as a goal the response rate or the time to an event (progression/death). If 'time to an event' (e.g., time to progression, disease-free survival, progression-free survival) is the main endpoint of the study, then routine scheduled re-evaluation of protocol specified sites of disease is warranted. In randomised comparative trials 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 arm in the timing of disease assessment.

4.6 Confirmatory measurement/duration of response

4.6.1 Confirmation

In non-randomised 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 (see the paper by Bogaerts et al. in this Special Issue10). However, in all other circumstances, i.e., in randomised 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 is 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 is measured from the time measurement criteria are first met for CR until the first date that recurrent disease is objectively documented.

4.6.3 Duration of stable disease

Stable disease is measured from the start of the treatment (in randomised trials, from date of randomisation) 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 stable disease for a minimum period of time is an endpoint of importance in a particular trial, the protocol should specify the minimal time interval required between two measurements for determination of stable disease.

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.

Attachment 2: **ECOG PS score**

State	ECOG PS score
0	No difference between the activity before and after the onset of the disease
1	walk freely and engage in light physical activities, including general housework or office work, but not heavy physical activities
2	Walk freely and take care of themselves but lose the ability to work. Be able to get up at least half of the daytime
3	Only take care of themselves partially and stay in bed or wheelchair for more than half of the daytime
4	Cannot get up in bed, can’t take care of themselves