Protocol No. TS-104 (Version 3.0 08Aug2014)

A prospective randomized clinical trial on ^{90}Yttrium trans-arterial radio-Embolization (TheraSphere®) vs. Standard of care (sorafenib) for the treatment of advanced Hepatocellular Carcinoma (HCC) with Portal Vein Thrombosis (PVT)

Statistical Analysis Plan

Prepared for:

Biocompatibles UK Ltd.

Chapman House

Farnham Business Park

Weydon Lane, Farnham

Surrey GU9 8QL, UK

Final Version 1.0 Date 09Aug2017

Prepared by:

Chiltern International

1016 W 9th Ave

King of Prussia, PA 19406

Revision History

Version	Date	Revision Author	Comments
Final 1.0	09Aug2017	[redacted]	Final Version

SIGNATURE PAGE

Author:

[redacted] Date

[redacted]

Chiltern International

Reviewed by:

[redacted] Date

[redacted]

Chiltern International

Approved by:

[redacted] Date

[redacted]

Biocompatibles UK Ltd.

Table of Contents

1 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Table 1: Abbreviations and Definitions of Terms

ADE(s)	Adverse Device Effect(s)
AE(s)	Adverse Event(s)
AFP	Alphafetoprotein
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BP	Blood Pressure
bpm	Beats per Minute
C	Celsius
CI	Confidence Interval
cm	Centimeters
CR	Complete Response
CT	Computed Tomography
CTC	Common Terminology Criteria
CTCAE	Common Toxicity Criteria for Adverse Events
DpR	Depth of Response
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
ETS	Early Tumor Shrinkage
F	Fahrenheit
FACT–Hep	Functional Assessment of Cancer Therapy-Heptobiliary Questionnaire
FDA	Food and Drug Administration
g	Gram
GBq	Gigabecquerel
Gy	Gray, a measure of irradiation dose
H	Above Normal Limit

HCC	Hepatocellular Carcinoma
HR	Heart Rate
hr(s)	Hour(s)
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
INR	International Normalized Ratio for prothrombin time
IRB	Institutional Review Board
kg	Kilograms
L	Below Normal Limit
m	Meter
max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
min	Minimum
mL	Milliliter
MR, MRI	Magnetic Resonance, Magnetic Resonance Image
n	Number of Patients
N	Within Normal Limit
NCI	National Cancer Institute
OS	Overall Survival
QoL	Quality of Life
PD	Progressive Disease
PE	Physical Exam
PP	Per Protocol
PR	Partial Response
PVT	Portal Vein Thrombosis
RECIST	Response Evaluation Criteria in Solid Tumor
RR	Respiration Rate
SADE	Serious Adverse Device Effect

SAE	Serious Adverse Event
SAS®	(Statistical Analysis Software)
sd	Standard Deviation
SD	Stable Disease
SoC	Standard of Care
SOC	System Organ Class
SOP	Standard Operating Procedures
^{99m}TcMAA	Technicium-99m Magroaggregaed Albumin
TEAE	Treatment Emergent Adverse Event
TEMP	Temperature
TRR	Tumor Response Rate
TS	TheraSphere
TTDoL	Time to Deterioration in QoL
TTP	Time-to-Progression
TTSP	Time to Rymptomatic Progression
TTUP	Time to Untreatable Progression
WHO	World Health Organization

2 INTRODUCTION

Primary liver cancer is a major health problem worldwide according to the American Cancer Society. Globally, it is the fifth most commonly diagnosed cancer in men and eighth most common in women, with more than 700,000 new cases in 2007. It is the third leading cause of cancer death in men and sixth among women. Hepatocellular carcinoma is associated with known risk factors including chronic viral hepatitis (types B and C), alcohol intake and aflatoxin exposure, the presence of which vary geographically. In North America and Europe, chronic hepatitis C is a major risk factor while chronic hepatitis B is a major risk factor worldwide.

Based on published reports, TheraSphere is well tolerated when appropriate patient selection criteria are used. Early reports of serious adverse events (SAEs) possibly associated with the use of TheraSphere lead to identifying pre-treatment high risk factors associated with these SAEs. The high risk factors included death, hepatorenal failure, liver abscess, hepatic encephalopathy, hepatic decompensation, radiation hepatitis, radiation pneumonitis, duodenal ulcer, gastrointestinal bleeding and cholecystitis. These more severe events are now uncommon as patients with high risk factors associated with the occurrence of these events are typically excluded from treatment with TheraSphere. Patients in whom TheraSphere should be used with caution include those with infiltrative tumor type, bulk disease (tumor volume >70% or nodules too numerous to count), AST or ALT > five times the upper limit of normal, bilirubin >2 mg/dL, tumor volume >50% in the presence of an albumin <3 g/dL, and those in whom extra-hepatic shunting to the lungs or gastrointestinal tract cannot be managed through standard angiographic techniques.

Recent clinical data in patients with HCC and PVT indicate this group of patients may have a better outcome following TheraSphere treatment as compared to sorafenib treatment.

3 STUDY OBJECTIVES

To assess efficacy and safety of TheraSphere in comparison to standard of care therapy (sorafenib) in the treatment of patients with portal vein thrombosis associated with unresectable hepatocellular carcinoma.

4 STUDY DESIGN

4.1 General Design

This is an open-label prospective, multi-center, randomized, controlled clinical trial that will evaluate the use of TheraSphere compared to standard-of-care sorafenib alone. Up to 25 study centers will participate and recruit patients for the protocol. Participating study sites may be in Europe, Asia or North America. All patients will be followed prospectively from randomization to death.

Patients will be randomized (1:1) to either the Control Group or the Treatment group, defined as follows:

Treatment Group: TheraSphere will be administered to all disease present at randomization. Patients can receive a subsequent TheraSphere administration in the absence of radiological progression criteria at investigator discretion. During follow-up, for patients in the Treatment arm who have demonstrated hepatic progression with hepatic lesions amenable to TheraSphere, re-treatment with TheraSphere is allowed. Any re-treatment should take place at least 28 days from the last TheraSphere treatment.

Control Group: Patients randomized to the Control group will receive SOC treatment with sorafenib in accordance with the package insert. Control arm patients will start SOC sorafenib therapy within 2 weeks after randomization.

All patients who meet the criteria for progression and received the treatments as described for either arm of the trial can receive institutional best available care. Patients in the control arm who have progressed are not allowed to cross-over to the treatment arm. Best available care is based on physician judgment and current standard of care practices, including sorafenib.

Patients will have regular clinical study visits as long as they participate in the trial. During these visits, safety and efficacy data will be collected and recorded.

4.2 Method of Assignment of Patients to Treatment Groups

Patients will be randomized 1:1 to study treatment, either the Control arm or the Treatment arm.

If a patient is determined to be eligible to participate in the trial, the study site will contact the central randomization office via telephone or by internet where randomization will be determined using assignment by a computer-generated randomization scheme. Upon randomization, each patient will be assigned a subject identity code consisting of the protocol number, the site number (e.g. 01) and a patient identifier number (e.g. 001).

A centralized randomization schedule was generated by a statistician in the Chiltern International Biometrics department who is not associated with the conduct or analysis of the study, using a validated system. In order to create a balance between study arms, patients are stratified at randomization based on the following:

- AFP $\leq$400 ng/mL or >400 ng/mL
- Participating trial site

The randomization will be performed using the Balance system with Medidata. Each eligible patient will be assigned to the next sequential randomization number within the specified stratification combination and will receive the corresponding study treatment. The randomization schedule will be kept strictly confidential and accessible only to authorized persons. Only when the study has been completed, the protocol violations determined, and the study database locked will the randomization schedule be made available for analysis.

Patients randomized to the Control arm or the Treatment arm, unable to receive their planned study treatment, will continue to be followed under their assigned study arm for the purpose of the intent-to-treat analysis.

4.3 Blinding

This is an open label study and there is no blinding.

4.4 Determination of Sample Size

This is a randomized open label multi-center Phase III adaptive trial using a group sequential design with a primary end-point of OS.

The study is designed to detect a 4 months increase in median OS time from 9 months in the control arm to 13 months in the TheraSphere arm (i.e., hazard ratio = 0.69), using a log rank test.

A maximum of 250 deaths will yield 80% power to detect the target difference in median OS (i.e., HR = 0.69) with a two-sided alpha of 0.05 using a group sequential design with 2 interim analyses and stopping boundary defined by the rho family error spending function with rho=1.5 (Jennison and Turnbull, 2000). It is estimated that a maximum of 320 patients will need to be recruited over 42 months, with a 1 year additional follow-up period. This includes an adjustment to take account of an assumed 5% of patients who will be lost to follow-up and for whom a date of death is not recorded.

Sample size modification will be considered at the second interim analysis using the approach described in Mehta and Pocock[3] (2011), which employs an un-weighted test statistic at the final analysis as recommended by Burman and Sonneson[4] (2006). If the sample size is increased after the second interim analysis, the final analysis is planned when approximately, but no less than, 430 deaths have occurred, which will result in 80% power to detect an improvement in median OS from 9 to 12 months using a log rank test with a final two-sided alpha of 0.0322. It is estimated that approximately 500 patients will need to be recruited over 72 months, with a 1 year additional follow-up period, in order to observe 430 deaths. This includes an adjustment to take account of an assumed 5% of patients who will be lost to follow-up and for whom a date of death is not recorded.

5 CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

This SAP is based on protocol version 3.0 dated 08 August 2014. Due to challenges with patient recruitment the study was terminated prematurely; therefore, only a final analysis will be performed without any interim analyses..

Protocol section 10.2.3.2 states that time to progression (TTP) will be calculated as the interval between the randomization date and the date of first disease progression as defined by the respective criteria noted below, death for any cause or of last contact for patients alive. However, death will not be considered in the calculation of TTP. Section 6.4.2.1 of this SAP explains how TTP will be analyzed.

6 BASELINE, EFFICACY AND SAFETY EVALUATIONS

6.1 Schedule of Evaluations

The assessments to be conducted at each scheduled visit are displayed in the following tables.

Table 1 Schedule of Events: Control arm

Study Period & Events	Screening	Randomize	Control Arm Treatment & Response Assessment								Progression	Follow up[1]	End of Study
Timing: Study Visits (+/- 7 days) or Events	D -14 to 0[2]	D 0	w 2	w 4	w 8	w 12	w 16	w 20	w 24	w Q8		w Q8	
Informed Consent	x												
Demographics	x												
Medical History	x												
Physical Exam (PE)	x												
Categorization of PVT	x				x		x		x	x			
ECOG Performance Status	x		x	x	x	x	x	x	x	x		x	
Child Pugh Status	x				x		x		x	x			
Medication & Prior Treatment History	x												
Hematology	x			x	x	x	x	x	x				
Coagulation	x			x	x	x	x	x	x				
Chemistry panel, liver function tests	x		x	x	x	x	x	x	x				
Serum Pregnancy[3]	x												
Tumor marker (AFP)	x				x		x		x	x			
Bone Scintigraphy[4]	x						x					(Q6 mo)	
Liver Volume/Tumor mass	x												
Review Eligibility Criteria	x												
Randomize Patient		x											
FACT-Hep questionnaire[5]	x			x	x	x	x	x	x	x		x	
Initiate Sorafenib[6] therapy			x										
Spiral CT of abdomen/pelvis	x[7]				x		x		x	x			
Spiral CT of chest[8]	x[3]						x			x			
Assess/Report Adverse Events	x		x	x	x	X	x	x	x	x		x	x
Review/Record Study Treatment			x	x	x	X	x	x	x	x			
Record Concurrent Medication	x		x	x	x	X	x	x	x	x		x[9]	x
Final Endpoint Efficacy/Safety documentation & exit patient													x

[1] Visits can be conducted by telephone at the investigator's discretion, if a PS 2: 2 (symptomatic progression) is assessed and whenever in the best interest of the patient

[2] Results from tests and examinations done as part of clinical patient management within 14 days of randomization may be used to satisfy screening requirements to avoid unnecessary duplicative patient testing

[3] Female patients of childbearing potential only

[4] Required for patients with AFP >400 ng/mL to assess metastatic disease at screening, Week 16 then every 6 months

[5] Subjects to complete questionnaire at each study visit; completion not required for telephone follow-up visits

[6] According to sorafenib package insert beginning at Week 2 for Control Arm patients only

[7] Images taken up to 28 days in advance of randomization can be used for baseline images

[8] Imaging to be scheduled every 16 weeks from randomization at alternate Q8 week follow up visits

[9] Document liver treatment medications only; medications used for supportive care do not require documentation

Table 2 Schedule of Events: Treatment arm

Study Period & Events	Screening	Randomize	Treatment Arm Treatment & Response Assessment								Progression	Follow up[1]	End of Study
Timing: Study Visits (+/- 7 days) or Events	D -14 to 0[2]	D 0	w 2	w 4	w 8	w 12	w 16	w 20	w 24	w Q8		w Q8	
Informed Consent	x												
Demographics	x												
Medical History	x												
Physical Exam (PE)	x												
Categorization of PVT	x				x		x		x	x			
ECOG Performance Status	x		x	x	x	x	x	x	x	x		x	
Child Pugh Status	x				x		x		x	x		x	
Medication & Prior Treatment	x												
Hematology	x			x	x	x	x	x	x				
Coagulation	x			x	x	x	x	x	x				
Chemistry panel, liver function	x		x	x	x	x	x	x	x				
Serum Pregnancy[3]	x												
Tumor marker (AFP)	x				x		x		x	x			
Bone Scintigraphy[4] (if AFP	x						x					(Q6 mo)	
Liver Volume/Tumor mass	x												
Review Eligibility Criteria	x												
Randomize Patient		x											
FACT-Hep questionnaire[5]	x			x	x	X	x	x	x	x		x	
Hepatic Angiogram[6], TcMAA scan[6], TS Dose Calculation			x										
Order TS			x										
Administer TS treatment[7]				x									
Spiral CT of abdomen/pelvis	x[8]				x		x		x	x			
Spiral CT of chest[9]	x[8]						x			x			
Assess/Report Adverse Events	x		x	x	x	x	x	x	x	x		x	x
Review/Record Study			x	x								x	
Record Concurrent Medication	x		x	x	x	x	x	x	x	x		x[10]	x
Optional TheraSphere												x	
Final Endpoint Efficacy/Safety													x

[1] Visits can be conducted by telephone at the investigator's discretion, if a PS 2: 2 (symptomatic progression) is assessed and whenever in the best interest of the patient

[2] Results from tests and examinations done as part of clinical patient management within 14 days of randomization may be used to satisfy screening requirements to avoid unnecessary duplicative patient testing

[3] Female patients of childbearing potential only -can be repeated at physician discretion before repeat TheraSphere treatments

[4] Required for patients with AFP >400 ng/mL to assess metastatic disease at screening, Week 16 then every 6 months

[5] Subjects to complete questionnaire at each study visit; completion not required for telephone follow-up visits
[6] Pre-treatment evaluation is always done before the first TheraSphere treatment and prior to subsequent treatments as needed
[7] TheraSphere must be administered within 4 weeks of randomization
[8] Images taken up to 28 days in advance of randomization can be used for baseline images
[9] Imaging to be scheduled every 16 weeks from randomization at alternate Q8 week follow up visits
[10] Document liver treatment medications only; medications used for supportive care do not require documentation
[11] A maximum of 2 re-treatments with TheraSphere are allowed following hepatic progression for lesions amenable to TheraSphere treatment

6.2 Time Point Algorithms

6.2.1 Relative Day

Relative Day will be calculated for both efficacy and safety endpoints. For efficacy endpoints, the following relative day calculation will be used:

The date of randomization will be considered relative day 1, and the day before the randomization will be relative day -1. Relative days will be calculated as follows only when the full assessment date is known (i.e., partial dates will have missing relative days for all assessments except for adverse events, concomitant medications, date of response, date of progression, and date of death, see Section 7.3.1 on imputation of partial dates):

For days on or after randomization:

Date of Assessment – Date of Randomization + 1

For days before randomization:

Date of Assessment – Date of Randomization

For days following disease progression:

Date of Assessment – Date of Disease Progression +1

For safety endpoints, the following relative day calculation will be used:

The date of first dose of sorafenib (control arm) or date of first angiogram (treatment arm) will be considered relative day 1, and the day before the first dose of sorafenib or date of first angiogram will be relative day -1. Relative days will be calculated as follows only when the full assessment date is known (i.e., partial dates will have missing relative days for all assessments except for adverse events, concomitant medications, date of response, date of progression, and date of death):

For days on or after first dose of sorafenib or date of first angiogram:

Date of Assessment – Date of first dose of sorafenib or date of first angiogram + 1

For days before first dose of sorafenib or date of first angiogram:

Date of Assessment – Date of first dose of sorafenib or date of first angiogram

For days following disease progression:

Date of Assessment – Date of Disease Progression +1

6.2.2 Windows

For the purpose of statistical analysis, the visit windows will be calculated as shown below.

Table 3 Analysis Windows for Assessments Performed at Four Week Intervals

Week	Scheduled Study Day	Visit Window for Analysis (Days)
Screen	-14 – 0	-14 – 0
Randomize (Baseline)	0	0
Week 2	14	1 - 21
Week 4	28	22 – 42
Week 8	56	43 – 70
Week 12	84	71 – 98
Week 16	112	99 – 126
…		
End of Study		Latest assessment available

Table 4 Analysis Windows for Assessments Performed at Eight Week Intervals

Week	Scheduled Study Day	Visit Window for Analysis (Days)
Screen	-14 – 0	-14 – 0
Randomize (Baseline)	0	0
Week 8	56	1 – 84
Week 16	112	85 – 140
Week 24	168	141 – 196
…		
Disease Progression	P0	P0
FDP Week 8	P56	P1 – P84
FDP Week 16	P112	P85-P140

...		
End of Study		Latest assessment available

FDP = Following Disease Progression

Pn = Relative days following disease progression

If a patient has more than 1 assessment occurring in the same visit window, the data from the visit closest to the scheduled study day will be used. If 2 visits have the same distance from the scheduled study day, excluding time to event endpoints, the data of the visit after the scheduled study day will be used. For time to event analyses, the event that occurs first will be used.

6.3 Baseline Assessments

Baseline for the following assessments is defined as the last assessment performed on or before the day of randomization.

- Informed Consent
- Inclusion/ Exclusion Criteria
- Demographics (age, gender, race, ethnicity)
- Medical history
- Physical examination
- Vital signs (heart rate [HR], respiration rate [RR], blood pressure [BP], aural temperature, height and weight)
- Disease and treatment history
- ECOG performance status
- Laboratory tests (hematology, coagulation, chemistry)
- Serum pregnancy test
- Child Pugh Score
- Categorization of PVT
- HCC tumor biomarkers- AFP
- Liver volume/mass and tumor burden
- FACT-HEP QOL

- CT/MRI of chest
- CT/MRI of pelvis/abdomen
- Stratification factors: AFP (400 ng/mL or >400 ng/mL) and participating trial site

Time from diagnosis of HCC will be calculated as follows:

Time from diagnosis of HCC (months) to randomization = (Date of Randomization – Date of Diagnosis)/30.4375.

6.4 Efficacy Variables

For all efficacy evaluations, the baseline measurement is defined as the last measurement prior to randomization.

6.4.1 Primary Efficacy Variable – Overall Survival (OS)

The primary study endpoint is OS defined as the time from date of randomization until date of death due to any cause. For each patient that is not known to have died, OS will be censored at the time of last contact date known to be alive (contacts considered in the determination of last contact date include adverse event start and end dates, lesion assessment date, and best available care start and end dates).

Overall Survival (months)= (Date of event/censor – Date of Randomization + 1)/ *30.4375*.

6.4.2 Secondary Efficacy Variables

The secondary efficacy endpoints for this study are:

- Time to Progression (TTP) by RECIST v1.1, modified RECIST and EASL criteria
- Time to Worsening of PVT
- Time to Symptomatic Progression (TTSP)
- Tumor Response Rate (TRR), as objective response rate (ORR) and disease control rate (DCR) by RECIST v1.1, modified RECIST and EASL criteria
- Quality of Life (QoL) (FACT-Hep), including time to deterioration in QoL (TTDQoL)

6.4.2.1 Time to Progression

This secondary endpoint is TTP. TTP is defined as the time from date of randomization until date of radiological progression (including new liver lesions and extra-hepatic lesions) separately according to RECIST v 1.1, modified RECIST and EASL criteria assessed by investigator determination.

TTP (Months) = (Date of event/censor – Date of Randomization +1)/ 30.4375.

The censoring is performed in the following order:

1) If a patient does not have a baseline tumor assessment, then the TTP time will be censored at the randomization date

2) If a patient is known not to have radiological disease progression , the TTP time will be censored at the date of death or at the last valid post baseline radiological tumor assessment date or at the randomization date if the patient does not have any post-baseline radiological tumor assessments.

6.4.2.2 Time to Worsening of PVT

Time to worsening of PVT is defined as the time from date of randomization until the date of any change in the classification of PVT type by at least one sub-type based on investigator assessment. Time to Worsening of PVT (Months) = (Date of event/censor – Date of Randomization +1)/ 30.4375.

The censoring is performed in the following order:

1) If a patient does not have a baseline PVT classification, then the time to worsening of PVT will be censored at the randomization date;
2) If a patient does not have a post baseline PVT classification, then the time to worsening of PVT will be censored at the randomization date;
3) If a patient is known not to have worsening of PVT classification, the time will be censored at the last post baseline PVT assessment date.

6.4.2.3 Time to Symptomatic Progression

TTSP is defined as the time from date of randomization to date of assessment of ECOG performance status ≥2 with or without tumor progression based on investigator assessment. Deterioration in performance status is to be confirmed at one subsequent evaluation 8 weeks later. The first date at which the ECOG performance status was ≥2 will be used as the end date in the TTSP analysis (assuming the next subsequent visit confirmed the deterioration).

TTSP (Months) = (Date of ECOG ≥2 – Date of Randomization + 1)/30.4375.The censoring is performed as follows:

1) If a patient does not have a post baseline ECOG performance status, then the TTSP will be censored at the randomization date;
2) If a patient has post baseline ECOG assessment but did not have ECOG ≥2 then TTSP will be censored at the last post baseline ECOG performance status assessment date.

6.4.2.4 Tumor Response Rate (ORR and DCR)

Tumor Response is based on the radiological tumor assessment performed at specified time points. The post baseline assessments are compared to the baseline assessment and the overall response based on investigator assessment is recorded at each efficacy visit. The tumor response is categorized as Complete Response (CR), Partial Response (PR), Stable Disease (SD) or Progressive Disease (PD). There are three sets of criteria that will be assessed, RECIST 1.1, mRECIST, and EASL.

The best overall response is the best response recorded during the study. It will be determined from the overall responses derived at each efficacy visit using the rules according to the tables in Appendix 3.

The ORR is defined as the proportion of patients achieving a best overall response of confirmed CR, or PR. The disease control rate (DCR) is defined as the proportion of patients achieving a best overall response of confirmed CR, PR, or SD.

Patients who do not have any post baseline tumor assessments for any reason are considered non-responders and are included in the denominator when calculating the response rate.

6.4.2.5 Quality of Life (FACT-Hep)

The total score of the FACT-Hep QoL instrument will be calculated, the scores of each domain (Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, Functional Well-Being), Hepatobilliary Cancer subscale (HCS), FACT-hep trial outcome index and each question at each time-point and their differences from baseline will be determined. The scoring algorithm is in Appendix 4.

The time to deterioration in QoL is defined as the time from date of randomization to the assessment date when the change from baseline in FACT-Hep Total Score is ≥7 points or date of death, whichever occurs first.

TTDQoL (Months) = (Date of change from baseline in FACT-Hep ≥7 or death) – Date of Randomization +1.

The censoring is performed as follows:

- If a patient does not have a baseline FACT-Hep assessment, then the TTDQoL time will be censored at the randomization date, regardless of whether or not TTDQoL has been observed
- If a patient does not have a post-baseline FACT-Hep assessment, then the TTDQoL time will be censored at the date of randomization.

6.4.3 Additional Efficacy Variables

6.4.3.1 Child-Pugh Score Status

Severity of liver disease will be assessed according to the Child-Pugh classification of Severity of Liver Disease (see Appendix 5).

6.4.3.2 Depth of Response

Depth of Response (DpR) is defined as the maximum tumor shrinkage observed in a patient.

6.4.3.3 Early Tumor Shrinkage

Early tumor shrinkage (ETS) is defined as the percent of decrease of the sum of longest diameter of the target lesions at the first post baseline tumor assessment.

6.4.3.4 HCC Tumor Marker - AFP

AFP will be collected along with laboratory data and will be presented similarly. Change from baseline will be calculated.

AFP response, defined as a ≥50% decrease in AFP levels for patients with a baseline AFP level of ≥200 ng/mL, will also be calculated.

6.5 Safety Assessments

6.5.1 Extent of Exposure and Compliance to Study Treatment

6.5.1.1 Extent of Exposure to TheraSphere

A lobar dose is defined as TheraSphere dosing to a particular lobe, there may be more than 1 mass treated within the lobe.

Dose delivered to the lobe

The dose delivered to a particular lobe is equal to the weighted average of the dose delivered to each perfused liver tissue mass in that lobe. The target liver masses are used to weight the average.

Let M_1 and M_2 denote the total dose delivered to the first and second perfused liver masses respectively (where T_1 and T_2 are the first and second masses in kg), then the weighted average dose is computed as $[(T_1*M_1)+(T_2*M_2)]/(T_1+T_2)$

The data below show an example for one patient:

Prior or Post Progression?	Lobe	Target liver mass, kg (T)	Total dose delivered to the perfused liver tissue, Gy (M)	Dose delivered to the lobe, Gy
Prior to progression	Right	1.3	115.2	116.26
	Right	1.2	117.4	
Post-progression	Right	1.3	108.2	108.2

For the patient above, the dose delivered to the right lobe prior to progression is calculated as:

$$\frac{(1.3\times115.2)+(1.2\times117.4)}{(1.3+1.2)}=116.26 \text{ Gy}$$

Dose delivered to the lungs

The total dose delivered to the lungs is calculated as the cumulative sum of the dose delivered to the lungs from each treatment.

In addition:

Time to first angiogram will be calculated as:

Time to first angiogram (weeks) = (Date of first angiogram – Date of Randomization + 1)/7

Time to first TheraSphere treatment will be calculated as:

Time to first TheraSphere treatment (weeks) = (Treatment Date – Date of Randomization + 1)/7

Time between TheraSphere treatments will be calculated as:

Time between TheraSphere treatment (weeks) = (Treatment Date – Date of Previous Treatment + 1)/7

6.5.1.2 Extent of Exposure to Sorafenib

Sorafenib information will be presented by duration, cumulative dose and dose intensity.

Dose and duration of sorafenib includes:

- Cumulative dose (mg)
- Dose intensity (mg/month)
- Relative dose intensity (%)
- Duration of treatment (month)

Time to the first dose of sorafenib will be calculated as:

Time to first dose of sorafenib (weeks) = (first dose date – date of randomization + 1)/7

The cumulative dose of sorafenib (mg) will be the sum of all calculated administered doses per patient.

Duration of sorafenib in months will be calculated as:

Duration of treatment (months) = (date of last dose of sorafenib – date of first dose of sorafenib)/30.4375

Dose intensity of sorafenib (mg/month) per patient will be calculated as:

Dose intensity (months) = cumulative dose of sorafenib / duration of treatment (months)

Relative dose intensity of sorafenib (%) will be the dose intensity divided by the one month equivalent (2 * 400 mg * 30.4375 = 24,350 mg/month).

Relative dose intensity (%) = (monthly dose intensity / (24,350 mg/month) * 100

6.5.1.3 Extent of Study Exposure and Follow-up

The duration of survival follow-up and duration of follow-up from last study treatment will be determined.

Duration of survival follow-up (months) = (date of death/last known alive– randomization date + 1) / 30.4375

Duration of follow-up from last study treatment is defined as the time from last study treatment (TheraSphere or sorafenib) until date of death due to any cause or last contact date when patient was known to be alive.

> Duration of follow-up (from last study treatment) (months) = (date of death/last known alive – date of last dose of study treatment + 1) / 30.4375

6.5.2 Adverse Events

Treatment Emergent Adverse Events (TEAEs) are AEs which were not present at the initiation of sorafenib or first angiogram or worsened in severity following treatment. If the imputed AE start date starts on or after the first dose of sorafenib or first angiogram and prior to 30 days after last dose of sorafenib or TheraSphere then it is considered a TEAE unless the AE was an ongoing event (end date of one AE matches start date of the next) that started prior to first dose of sorafenib or first angiogram and the grade did not worsen.

The investigator's verbatim term of the AEs will be mapped to system organ classes (SOC) and preferred terms using the MedDRA Version 14.0 dictionary (Medical Dictionary for Regulatory Activities). The investigator will use the NCI Common Toxicity Criteria for Adverse Events [CTCAE] (version 4.0) or the protocol specific criteria when no NCI CTC criteria are available for the adverse event to determine the severity of the AEs.

Adverse events related to sorafenib, device, and angiographic procedures are defined as those events recorded on the eCRF with relationship of possibly, probably, or definitely related.

The incidence of TEAEs will be the number of patients who had the AE (counted only once per patient) divided by the number of patients in the safety population and represented as a percentage. The incidence of AEs will be the number of times an event with the same preferred term occurs, counting worsening events only once. For worsening events, the AE end date of the earlier AE will be the same as the start date of the same AE with a higher severity.

6.5.3 Clinical Laboratory Evaluations

Clinical laboratory results will be converted to SI units.

Laboratory data for safety analysis will be summarized as continuous summaries (ie, as absolute values, and change from baseline to each scheduled visit and end of study treatment in absolute values). Scheduled visits for hematology and coagulation

assessments will include Screening, Week 4, Week 8, Week 12, Week 16, Week 20, and Week 24, while scheduled visits for chemistry will include Screening, Week 2, Week 4, Week 8, Week 12, Week 16, Week 20, and Week 24. End of study treatment will be defined as from last dose of TheraSphere or sorafenib to 30 days after treatment.

Laboratory values will also be classified as normal (if value is within normal reference range) or lower/higher than normal (if value is either below or above the normal reference range).

Applicable laboratory values will also be classified using NCI CTCAE v4.0. Laboratory values within normal range will be categorized as either non-decreasing or non-increasing depending on the direction of the CTCAE grade (decreasing or increasing respectively).

6.5.4 Additional Safety Variables

6.5.4.1 ECOG Performance Status

The ECOG Performance Status will be assessed according to the following categories:

Score	Characteristics
0	Asymptomatic and fully active
1	Symptomatic; fully ambulatory; restricted in physically strenuous activity
2	Symptomatic; ambulatory; capable of self-care; more than 50% of waking hours are spent out of bed
3	Symptomatic; limited self-care; more than 50% of waking hours are spent in bed
4	Completely disabled; no self-care; bedridden

7 STATISTICAL METHODS

7.1 General Methodology

All statistical tests will be two-sided with a significance level of α=0.05, unless specified otherwise, and will be performed using SAS® Version 9.2 or higher. Data will be summarized using descriptive statistics (number of patients [n], mean, standard deviation

[sd], median, minimum, and maximum) for continuous variables and using frequency and percentage for discrete variables.

Patient listings of all data from the case report forms (eCRF) as well as any derived variables will be presented.

7.2 Adjustments for Covariates

The following covariates will be included, one at a time, in univariate Cox regression analysis of time-to-event efficacy endpoints, including OS

- Stratification factor (AFP ≤400 ng/mL vs >400 ng/mL

  Notes:

  Stratification factor of AFP according to the master file will be used even if the stratification factor was incorrectly recorded at randomization.

Site is also a stratification factor in this study. However, due to early termination of the study and small number of subjects per site, site will not be included as a covariate in this analysis since it would likely result in non-convergence of the model.

- ECOG status at baseline
- Age
- Gender
- Region (North America, Europe, Asia)
- Naïve or recurrent disease
- HCC diagnosis present at first oncology diagnosis
- Prior oncologic treatment for HCC (none, local regional therapy, prior resection, or systemic treatment)
- Duration from date of initial diagnosis of HCC to randomization
- Percent Tumor replacement at baseline
- HCC etiology (alcoholism, hepatitis B, hepatitis C, afloxatin-contaminated food, non-alcoholic cirrhosis, diabetes and other (including unknown))
- ALBI score (1, 2 or 3) at baseline

These covariates will also be included, one at a time, in a logistic regression analysis of binary efficacy endpoints.

All factors in the univariate models with a two-sided p-value <0.15 will be included in a multivariate analysis to determine the impact of these factors.

7.3 Handling of Dropouts or Missing Data

Dropout patients will not be replaced in this study. Censoring for the efficacy endpoints is discussed in Section 6.4.

7.3.1 Partial Date Imputation

Partial dates with day or day and month missing will be imputed as follows:

Missing day only

- The missing day of onset of an AE or start date of a concurrent therapy will conservatively be set as follows:
  - The first day of the AE/concurrent therapy start month, if the month of randomization is before the AE/concurrent therapy start month.,
  - One day after randomization, if the month of randomization is the same as the AE/concurrent therapy start month.
- The missing day of the end date of a concurrent therapy will be set to the last day of the month occurrence.
- For other variables, including date of tumor response, progression, death, partial dates that need to be imputed will use the 15th of the month to replace the missing day. The only exception is when the last visit date is after the 15th of the month then death date will use the last visit date on study.

Missing day and month

- If the onset date of an AE or start date of a concurrent therapy is missing both day and month, it will be set as follows:
  - January 1 of the AE/concurrent therapy start year, if the year of randomization is before the AE/concurrent therapy start year,
  - One day after randomization, if the year of randomization is the same as the AE/concurrent therapy start year.

If the resolution date of an AE or end date of a concurrent therapy is missing both day and month, it will be set to December 31 of the AE/concurrent therapy end year .

7.4 Interim Analyses and Data Monitoring

7.4.1 Interim Analyses

The planned interim analyses will not be performed due the early termination of the study.

7.4.2 Data Monitoring

An Independent Data Monitoring Committee (IDMC) will be established to oversee the safety and efficacy of the study, will follow the FDA's Guidance on IDMCs/DSMBs and comply with ISO 14155. The IDMC will meet to assess safety after every 100 patients randomized or every 6 months, whichever occurs first, until the end of the study. Due to early termination of the study, only two safety DMC meetings took place. No interim analysis will be performed.

Further details regarding the IDMC can be found in the IDMC charter.

7.5 Multi-center Studies and Pooling of Centers

This is a multi-center study; the data from all sites will be pooled according to treatment group.

7.6 Multiple Comparisons/Multiplicity

No adjustments for multiple comparisons or multiplicity will be made.

7.7 Examination of Subgroups

No subgroup analyses will be performed due to the limited number of patients in this early terminated study.

8 STATISTICAL ANALYSIS

8.1 Disposition of Patients

The number of patients enrolled by region (North America, Europe and Asia), country and site will be presented.

The number of patients screened, the number of patients randomized, and the number of patients treated with the treatment to which they were randomized (ie, sorafenib or TheraSphere) will be summarized.

The number of patients who completed or were ongoing at the end of the study will be summarised. Note: patients who died will be considered as having completed the study.

The number of patients who discontinued from the study will be summarized overall and separately for those who were treated with TheraSphere or sorafenib. The reasons for discontinuation will be summarized for patients by treatment group.

This information will be summarized also by the stratification factor AFP ≤400 ng/mL or >400 ng/mL).

8.2 Protocol Deviations/ Violations

Protocol deviations/violations will not be programmatically identified or entered into the database. However, protocol deviations/violations will be identified and categorized within an excel spreadsheet tracker managed by Chiltern's Clinical Department. Deviations will be categorized by BTG as major or minor. All protocol deviations/violations determinations will be finalized before database lock.

The number and percentage of patients within each category of deviation, separately for major and minor deviations, will be summarized for each treatment group and overall.

8.3 Analysis Populations

8.3.1 Treated Population

All randomized patients who received study treatment at least once will form the Treated Population and will be analyzed according to the treatment group to which they were randomized.

8.3.2 Safety Population

The Treated Population will be the same as the Safety Population, but will be analyzed according to the treatment actually received. This population will be used in all safety reporting and analysis.

8.4 Demographic and Other Baseline Characteristics

Demographic and baseline summaries will be presented for the Treated Population and Safety Population.

Gender, race, and ethnicity, age category (≥18-<65 years, ≥65 years), AFP (≤400 ng/mL or >400 ng/mL), and ECOG performance score will be summarized using counts and percentages. Continuous variables such as age, AFP, time from diagnosis of HCC to randomization will be summarized with descriptive statistics (number of patients [n], mean, standard deviation [sd], median, minimum [min], and maximum [max]). Ethnicity responses of "Not Applicable" will be summarized under "Not Reported or Not Known".

Details of categorization of PVT at screening (venous location of PVT, tumor location, categorization of PVT) will be summarized using count and percentages.

History of hepatocellular carcinoma (HCC) will be summarized as the number and percent of patients who have portal hypertension, presence of cirrhosis, whether HCC is the first oncology diagnosis, and etiology.

Weight will be displayed in kilograms (kg), height will be displayed in centimeters (cm), and temperature will be displayed in Celsius (C). BMI (kg/m^2) will be calculated as weight (kg) divided by height (m) squared. Weights, heights, or temperatures recorded in alternate units will be converted to the units being displayed using standard conversion formulas.

8.5 Physical Examination

Physical examinations are collected at screening only. The number and percent of patients with abnormal physical examination findings at screening will be summarized overall and by body system.

8.6 Baseline Tumor Burden and Liver Volume

Liver volume (mL), liver mass (kg), and tumor replacement (%) (percent of total liver volume occupied by tumor(s)) will be summarized using descriptive statistics (n, mean, sd, median, min, and max).

Liver mass will be calculated as:

Liver mass (kg) = liver volume (mL)*0.00103.

Total tumor replacement > 70% and >50% will be summarized using frequency counts and percentages. For patients whose total tumor replacement is >50%, number and percentage of those patients with albumin <3 mg/dL will be summarized.

8.7 Medical History

The number and percent of patients with medical history events will be summarized by system organ class and preferred term.

8.8 Prior and Concomitant Therapy

The World Health Organization Drug Dictionary (WHO DD) March 2011 will be used to classify medications by preferred term and WHO Anatomical Therapeutic Chemical (ATC) classification of ingredients.

The following applies to all data collected on the prior and concomitant eCRF page, medical history HCC eCRF page, and the best available care post progression eCRF pages and will be reported by each category separately.

Where a medication start date is missing, this medication will be assumed to be concomitant for reporting purposes, unless the end date is prior to the first dose of sorafenib or angiogram. Partial dates will be imputed as discussed in section 7.3.1.

Descriptive statistics, such as frequency counts and percentages will be provided to summarize the use of medications other than the study drug reported throughout the study. The number and percent of patients who took other therapy will be shown by WHO ATC classification of ingredients and by preferred term.

8.8.1 Prior Medication

A prior medication is defined as any medication given prior to first dose of sorafenib or date of angiogram.

The number and percentage of patients who had at least one prior medication will be tabulated as well as the number and percentage of patients with each medication. Patients will only be counted once for each medication.

8.8.2 Prior Oncologic Therapy for HCC

Prior HCC treatment type: Liver Directed Therapy, Prior Resection, and Systemic Treatment will be summarized from the Medical History of HCC eCRF page. Type of therapy/resection/treatment will be listed.

8.8.3 Concomitant Medication

A concomitant medication is defined as any medication given prior to and continuing after the first dose of sorafenib or date of angiogram, or any medication that is initiated on or after first dose of sorafenib or date of angiogram. Medications are considered concomitant through to the end of the study.

The number and percentage of patients who had at least one concomitant medication will be tabulated as well as the number and percentage of patients with each medication. Patients will only be counted once for each medication.

8.8.4 Best Available Care Post Progression

For best available care post progression, the number and percentage of patients receiving each medication will summarized by ATC class and medication.

Best available care procedure will be summarized as the number and percentage of patients by procedures: Resection, Ablation, and Other.

8.9 Analysis of Efficacy Parameters

Efficacy endpoints will be analysed using the Treated Population.

8.9.1 Analysis of Primary Efficacy Variable

The primary efficacy analysis is of OS. OS rates will be assessed using the Kaplan Meier approach. Quartiles of survival time will be presented and 95% CIs will also be calculated on the quartiles for each treatment group. A log-rank (two-sided) test will be used to compare OS between the two treatment groups.

The hazard ratio (HR) and corresponding two-sided 95% CI for the treatment effect will be computed using Cox proportional hazards model.

Plots of the Kaplan-Meier curves will be provided for each treatment group.

8.9.2 Analysis of Secondary Efficacy Variables

The secondary efficacy endpoints for this study are:

- Time to Progression (TTP)
- Time to Worsening of PVT
- Time to Symptomatic Progression (TTSP)
- Tumor Response Rate (ORR and DCR)
- Quality of Life (FACT-Hep)

8.9.2.1 Time to Progression

TTP rates will be derived and quartiles on the rates will be presented. 95% CIs will be calculated on the quartiles for each treatment group under RECIST 1.1, mRECIST, and EASL criteria. A log-rank (two-sided) test will be used to compare TTP between the two treatment groups at a 0.05 significance level. Plots of the Kaplan-Meier curves will be provided for each treatment group.

8.9.2.2 Time to Worsening of PVT

Time to worsening PVT rates will be derived and quartiles on the rates will be presented. 95% CIs will be calculated on the quartiles. A log-rank (two-sided) test will be used to compare time to worsening PVT between the two treatment groups at a 0.05 significance level. Plots of the Kaplan-Meier curves will be provided for each treatment group.

8.9.2.3 Time to Symptomatic Progression

TTSP rates will be derived and quartiles will be presented and 95% CIs will be calculated on the quartiles for each treatment group. A log-rank (two-sided) test will be used to compare TTSP between the two treatment groups at a 0.05 significance level. The HR and corresponding 95% CI will be computed. Plots of the Kaplan-Meier curves will be provided for each treatment group.

8.9.2.4 Tumor Response Rate

The ORR will be computed for the two treatment groups as proportion of CR+PR over the total number of patients in the specified population. The 95% CIs for the ORR for each of the treatment groups will be computed according to Wilson (1927).

ORR, as determined by the investigator will be compared between treatment groups using the continuity adjusted Newcombe-Wilson test, and the corresponding 95% CI for the difference in ORRs between the two treatment groups will be calculated. This analysis will be performed under RECIST 1.1, mRECIST, and EASL for each time point and the best overall response.

DCR (ie, CR+PR+SD) will be summarized and compared between the treatment groups in the same way as the ORR.

Details on target lesions and non-target lesions will be presented in listings only.

8.9.2.5 Analysis of Quality of Life (FACT-Hep)

The total, domain, and individual question scores of the FACT-hep QoL instrument and their differences from baseline will be summarized at each time point by treatment group. The two treatment groups will be compared by applying a mixed linear model repeated measures analysis using a REML estimation with the treatment, visit and the interaction between treatment and visit as factors, and the baseline score as a covariate. An unstructured covariance approach will be applied. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autogressive with heterogeneity, Toeplitz, and autoregressive. Means and least squares mean difference between treatment groups, along with a two-sided 95% CI and p-value for the difference between treatments will also be provided.

A deterioration in QoL is defined as a ≥7-point decline in the total FACT-hep score or death whichever occurs first.

TTDQoL quartiles will be presented and 95% CIs will be calculated on the quartiles for each treatment group. A log-rank (two-sided) test will be used to compare TTDQoL between the two treatment groups at a 0.05 significance level. The HR and corresponding 95% CI will be computed. Plots of the Kaplan-Meier curves will be provided for each treatment group.

8.9.3 Additional Analyses

A shift table comparing the baseline Child Pugh score to the Child Pugh score at each time-point will be summarized. This will be used to ascertain the number of patients with a Child Pugh score that changes after baseline and any difference between the treatment groups.

DpR will be summarized with descriptive statistics (n, mean, SD, median, min, and max) by treatment group.

ETS will be summarized with descriptive statistics (n, mean, SD, median, min, and max) by treatment group. The number of patients achieving an ETS of <10%, 10-20%, and >20% will also be summarized by treatment group.

The percentage and absolute change from baseline in the sum of the longest diameters of target lesions will be summarized separately at Week 8, 16, and 24 analysis visits (as defined in Table 4). A waterfall plot of the percentage change in the sum of longest diameters at the time of the best overall response will be presented.

The tumor marker for HCC (AFP) will be summarized with descriptive statistics (n, mean, SD, median, min, and max) for each time-point and change from baseline by treatment group. The number of patients achieving an AFP response, defined as a ≥50% decrease in AFP levels for patients with a baseline AFP level of ≥200 ng/mL, will also be summarized for each post-treatment time-point by treatment group.

8.10 Analysis of Safety

All safety analyses will be performed on the Safety Population.

8.10.1 Extent of Exposure to Study Treatment

8.10.1.1 Extent of Exposure to TheraSphere

The number of lobar treatments of TheraSphere received will be presented prior to progression, following progression and across the study.

A summary of dose delivered to a particular lobe prior to progression and following progression will be presented. In addition, whether the dose was delivered as planned, whether any prophylactic gastric inhibitor treatment was prescribed, whether there were any complications or device deficiencies will be summarized.

A summary of the dose delivered to the lungs will be summarized prior to progression, following progression and over all treatments.

8.10.1.2 Extent of Exposure to Sorafenib

The extent of patient exposure to sorafenib will be summarized using descriptive statistics (n, mean, sd, median, minimum, maximum) for the following parameters:

- Cumulative dose (mg)
- Dose intensity (mg/month)
- Relative dose intensity (%)
- Duration of treatment (months)

The reason of dose delays and changes will be summarized.

8.10.1.3 Extent of Study Exposure and Follow-up

The time to first TheraSphere treatment, time to first dose of sorafenib, duration of survival follow-up and duration of follow-up from last study treatment will be summarized by treatment using descriptive statistics (n, mean, sd., median, minimum, maximum). The time between TheraSphere administrations (first to second and second to third) will be summarized by treatment using descriptive statistics.

In addition, an alternative method for determining median duration of follow-up will be performed using the reverse Kaplan-Meier method, in which the censored values for the OS analysis will be reversed (ie, patients censored for the OS analysis will not be censored for the reverse Kaplan-Meier method, and patients not censored for the OS analysis will be censored for the reverse Kaplan-Meier method).

8.10.2 Adverse Events

The investigator's verbatim term of each adverse event will be mapped to system organ class (SOC) and preferred term (PT) using the MedDRA dictionary version 14.0.

An overall summary of adverse events will be presented including number and percentage of patients with any TEAEs, sorafenib related TEAE, adverse device effects

(ADEs), angiographic procedure-related TEAEs, TEAEs CTCAE grade ≥ 3, serious adverse events (SAEs), serious adverse device events (SADEs), TEAEs with outcome of death and TEAEs leading to discontinuation of sorafenib.

TEAEs will be summarized by SOC and PT; a patient will only be counted once per system organ class and once per PT within a treatment. Patient counts and percentages and event counts will be presented for each treatment group:

1. Overall summary of AEs
2. All TEAEs.
3. All TEAEs of CTCAE grade ≥3.
4. All TEAEs related to sorafenib.
5. All TEAEs related to device (ADE).
6. All CTCAE grade ≥3 TEAEs related to device.
7. All TEAEs related to angiographic procedure.
8. All TEAEs with outcome of death.
9. All treatment emergent serious adverse events (SAEs).
10. All treatment emergent serious adverse device events (SADEs)
11. All TEAEs leading to sorafenib discontinuation.

For the summary of AEs by CTCAE grade, if a patient has multiple events occurring in the same SOC or same PT, then the event with the highest CTCAE grade will be counted. For AEs by relationship to study device, if a patient has multiple events occurring in the same SOC or same PT, the event with the highest association to study device will be summarized (unknown relationship to study device will be considered related to study device).

All TEAEs and TEAEs grade ≥3 will be summarized by preferred term with descending frequency. In addition TEAEs will be presented by CTCAE grade (1-2, 3-4 and 5).

No statistical inference between the treatments will be performed on AEs.

Listings will be presented by patient for all AEs as well as for SAEs including SADEs, AEs with outcome of death, and AEs leading to discontinuation of sorafenib.

8.10.3 Clinical Laboratory Evaluations

Laboratory test values at each assessment and for change from baseline to each assessment will be displayed using summary statistics (n, mean, median, and standard deviation).

In addition, shift tables will be presented to display the shift in the normal range categories (L, normal [N], H) from baseline to the final evaluation. Baseline is defined as the latest non-missing value prior to randomization.

Shift table of laboratory assessments with CTCAE grade ≥3 that also worsened from baseline will be presented by visit for each laboratory parameter.

Shift table of baseline to each assessment by CTCAE grade will be summarized. Laboratory tests with CTCAE grades in two directions (increasing and decreasing) will be presented separately for each direction.

All clinical laboratory data will be presented in listings. Within each listing, laboratory values outside the normal ranges will be flagged as either high (H) or low (L).

8.11 Additional Safety Analyses

All the additional analyses described below will be performed on the Safety population.

A shift table comparing the baseline ECOG score to the ECOG score at each time-point will be summarized.

9 COMPUTER SOFTWARE

All analyses will be performed by Chiltern International using Version 9.2 or later of SAS® software. All summary tables and data listings will be prepared utilizing SAS® software.

The standard operating procedures (SOPs) of Chiltern International will be followed in the creation and quality control of all data displays and analyses.

10 REFERENCES

1 http://nccu.cancer.org/docroot/STT/content/STT 1x Global Cancer Facts and Figures 2007.asp accessed November 25, 2010 at 11:10 ET.

2 Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid

tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228-47.

3 Mehta, C. R. and Pocock, S. J. (2011), Adaptive increase in sample size when interim results are promising: A practical guide with examples. Statist. Med., 30: 3267–3284. doi: 10.1002/sim.4102

4 Burman and Sonesson (2006), Are flexible designs sound? Biometrics., 62: 664-683

5 Jennison, C. and Turnbull, B. W. (2000), Group Sequential Methods with Applications to Clinical Trials, New York: Chapman & Hall.

6 Proschan MA, Lan KKG, Wittes JT (2006), Statistical Monitoring of Clinical Trials: A Unified Approach. 1st edn. Springer: USA

7 http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_seqtest_sect027.htm

8 Schoenfeld, D (1981). "The asymptotic properties of nonparametric tests for comparing survival distributions". *Biometrika* **68**: 316–319.

9 Wilson, E. B. (1927), "Probable Inference, the Law of Succession, and Statistical Inference," Journal of the American Statistical Association, 22, 209–212.

11 APPENDICES

11.1 APPENDIX 1: VARIABLE DEFINITIONS

Age will be calculated as the randomization date minus the date of birth divided by 365.25 [Age=(ICF Date-DOB)/365.25].

Weight will be displayed in kilograms (kg), height will be displayed in centimeters (cm), and temperature will be displayed in Celsius (C). Weights, heights, or temperatures recorded in alternate units will be converted to the units being displayed using standard conversion formulas.

11.2 APPENDIX 2: STATISTICAL ANALYSIS and PROGRAMMING DETAILS

The SAS procedure LIFETEST will be used in the Kaplan-Meier analyses. Patients who did not have an event will be censored.

The following code will be used:

```
proc lifetest data=all method=km alpha=0.05 outsurv=interval;
        time aval*censr(1);
        strata trtp;
        id usubjid;
run;
```

The SAS procedure MIXED will be used for mixed effects modeling of QoL data. The following code will be used:

```
proc mixed method = reml;
        class TRT VISIT SUBJID;
        model CH=BASE TRT VISIT  TRT*VISIT/s ddfm=kr;
        repeated VISIT/type=UN subject=SUBJID;
        lsmeans TRT*VISIT /slice=VISIT diff alpha=0.05 cl;
run;
```

where TRT is the assigned treatment, VISIT is the visit based on the window mapping, CH is the change from baseline.

11.3 APPENDIX 3: Response Tables

RECIST 1.1 Timepoint 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 at all Evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

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
Unequivical PD	Yes or No	PD
Any	Yes	PD

11.4 APPENDIX 4: FACT-hep Questionnaire Scoring Rules

FACT-hep Scoring Guidelines (Version 4)

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X

2. Perform reversals as indicated, and sum individual items to obtain a score.

3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.

4. Add subscale scores to derive total scores (TOI & FACT-hep).

5. **The higher the score, the better the QOL.**

Subscale Score	**Item Code**	**Reverse item?**		**Item response**	**Item**
PHYSICAL WELL-BEING (PWB)	GP1	4	-	_____	=_____
	GP2	4	-	_____	=_____
	GP3	4	-	_____	=_____
	GP4	4	-	_____	=_____
Score range: 0-28	GP5	4	-	_____	=_____
	GP6	4	-	_____	=_____
	GP7	4	-	_____	=_____

Sum individual item scores: _____

Multiply by 7: _____

Divide by number of items answered:

_____

=**PWB subscale score**

SOCIAL/FAMILY WELL-BEING (SWB)	GS1	0	+	_____	=_____
	GS2	0	+	_____	=_____
	GS3	0	+	_____	=_____
	GS4	0	+	_____	=_____
Score range: 0-28	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____

Multiply by 7: _____

Divide by number of items answered: _____

=SWB subscale score

EMOTIONAL	GE1	4	-	_____	=_____
WELL-BEING	GE2	0	+	_____	=_____
(EWB)	GE3	4	-	_____	=_____
	GE4	4	-	_____	=_____
Score range: 0-24	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____

Multiply by 6: _____

Divide by number of items answered: _____

=EWB subscale score

FUNCTIONAL	GF1	0	+	_____	=_____
WELL-BEING	GF2	0	+	_____	=_____
(FWB)	GF3	0	+	_____	=_____
	GF4	0	+	_____	=_____
Score range: 0-28	GF5	0	+	_____	=_____
	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____

Multiply by 7: _____

Divide by number of items answered: _____

=FWB subscale score

Subscale Score	Item Code	Reverse item?		Item response	Item
HEPATOBILIARY CANCER SUBSCALE (HCS)	C1	4	-	_____	=_____
	C2	4	-	_____	=_____
	C3	0	+	_____	=_____
	C4	0	+	_____	=_____
	C5	4	-	_____	=_____
Score range: 0-72	C6	0	+	_____	=_____
	Hep1	4	-	_____	=_____
	Cns7	4	-	_____	=_____
	Cx6	4	-	_____	=_____
	HI7	4	-	_____	=_____
	An7	0	+	_____	=_____
	Hep2	4	-	_____	=_____
	Hep3	4	-	_____	=_____
	Hep4	4	-	_____	=_____
	Hep5	4	-	_____	=_____
	Hep6	4	-	_____	=_____
	HN2	4	-	_____	=_____
	Hep8	4	-	_____	=_____

***Sum individual item scores:*_____**

***Multiply by 18:* _____**

***Divide by number of items answered:* _____**

=HC Subscale score

To derive a FACT-hep Trial Outcome Index (TOI):

Score range: 0-128

_____ + _____ + _____ = _____ =**FACT-hep TOI**

(PWB score) (FWB score) (HCS score)

To Derive a FACT-hep total score:

Score range: 0-

_____ + _____ + _____ + _____ + _____ = _____ =**FACT-hep Total score**

(PWB score) (SWB score) (EWB score) (FWB score) (HCS score)

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.

11.5 APPENDIX 5: Child Pugh Classification

Assess severity of liver disease by assigning points for each of the five parameters in the table below and adding the points to obtain a total score. Record the resulting ChildPugh Grade: A (well compensated disease) 5-6 points; B (functional compromise, worsening disease) 7-9 points and C (decompensated disease) 10-15 points.

Parameter	1 point	2 point	3 point
Bilirubin	<34 µmol/L	34–50 µmol/L	>50 µmol/L

Parameter	1 point	2 point	3 point
Albumin	>35 g/L	28-35 g/L	<28 g/L
Prothrombin Time	<1.8 (INR) or <4 secs (Seconds over control)	1.8-2.2 (INR) or 4-6 secs	>2.2 (INR) or >6 secs
Ascites	absent	Slight (medically controlled)	Moderate (poorly controlled)
Encephalopathy*	None	Grade 1-2	Grade 3-4

*Grades of Encephalopathy

Grade 1 – Inverted sleep pattern; forgetfulness, agitation, irritability, apraxia
Grade 2 – Lethargy; Disorientation for time or place, Subtle personality change; Asterixis, ataxia
Grade 3 – Somnolence but rousability; Disorientation as regards place; Asterixis, hyperactive reflexes, Babinski signs, muscle rigidity
Grade 4 – Coma (unresponsive to verbal or noxious stimuli

12 TABLE SHELLS

Table, listing, and figure shells will be provided as a separate document.