Protocol No. TS-102 (Version 7.0 10May2017)

A Phase III Clinical Trial Evaluating TheraSphere® in Patients with Metastatic Colorectal Carcinoma of the Liver who have Failed First Line Chemotherapy

Statistical Analysis Plan

Prepared for:
Biocompatibles UK Ltd
Lakeview, Riverside Way
Watchmoor Park, Camberley
Surrey GU15 3YL, UK

Final Version 8.0 Date 01Feb2021

Prepared by:
Covance
1016 W 9th Ave
King of Prussia, PA 19406

Revision History

Version	Date	Comments
Final Version 1.0	02Mar2012	Approved by Nordion
Final Version 1.1	19Jan2016	Update to Sponsor and CRO information only. Further updates required as a result of protocol amendments will be made in a subsequent version of the statistical analysis plan
Final Version 2.0	13Jun2018	Updated for current protocol and eCRF.
Final Version 3.0	31May2019	Updated analyses of primary and secondary endpoints to take account of subsequent cancer therapy, clarified definition of ORR, and provided clarifications

Final Version 4.0	30Jul2019	Updated the definition of subsequent mCRC therapy in Appendix 4 to include ablation and resection as subsequent therapies, and also removed the note stating that time to event endpoints would not be censored before ablation or resection. Added the following sensitivity analyses for the primary endpoints of PFS and HPFS: 1) not censoring for subsequent mCRC therapy; 2) ablation or resection not considered as subsequent mCRC therapy;3) Inverse probability of censoring weighted analysis with ECOG status at last assessment before subsequent mCRC therapy as a time-dependent covariate.
Final Version 5.0	20Apr2020	1) Reduced the number of PFS events at which the final analysis will be performed and adjusted the boundaries at the time of final analysis 2) Updated exposure summaries for TheraSphere (including definitions provided to compute the dose absorbed by perfused volume), second-line chemotherapy, biological agents, and post-progression treatments 3) Added details on confirmation of ECOG>2 for TTSP endpoint 4) Added text on collinearity of covariates in multivariable analysis

Final Version 5.0	20Apr2020	5) Updated the definition of best overall response (correspondingly ORR and DCR), duration of response, and duration of disease control to include all radiological tumor assessments until the first PD. 6) Updated list of variables in the analyses of covariates and subgroups, and modifications to some variables 7) Added details on major protocol deviations for Per Protocol population 8) Provided clarifications for Safety population 9) Provided clarifications for exploratory efficacy endpoints such as DoR and PTTS, which use central read data based on the sum of diameters of target lesions when adjudication was not required 10) Added details for subgroup analyses and presentation of corresponding analysis results 11) Removed UADE safety summaries 12) Added imputation rules for partial dates on the medical history mCRC and best available care eCRF pages 13) Removed the restriction to keep randomization schedule confidential until study database is locked 14) Added algorithm in Appendix 5 to identify patients who received adjuvant chemotherapy 15) Updated the list of abbreviations 16) Updated the CRO name from Chiltern International to Covance

Final Version 6.0	19Aug2020	1) Updated censoring rules for primary endpoints and other time-to-event endpoints to censor patients with progression or death occurring after two or more missed visits 2) Updated p-value to be 2-sided in poolability analyses 3) Clarified that analysis windows do not apply to tumor response endpoints (except for the PTTS) and TTSP and TTDQoL 4) Clarified that the number of control arm patients who received Y90 treatment after progression will be provided, together with a summary of OS for these patients
Final Version 7.0	13Oct2020	1) Updated the method to calculate alpha level for the final analysis 2) Updated the definition of 2 or more missed visits not to consider early or late assessment in the censoring rules for primary endpoints and other time-to-event endpoints 3) Added the following sensitivity analyses for censoring rules of the primary endpoints: a) alternative definitions of 2 or more missed visits allowing early and late assessments; b) progression or death after one or more missed visits considered as PFS event at the day after the last tumor assessment before the missed visits; c) not censoring for subsequent mCRC therapy and 2 or more missed visits. 4) Added summaries detailing the reasons for censoring 5) Added summary and listing for the difference between scheduled date and actual date for tumor assessments 6) Removed the sentence that the event occurring first in an analysis window will be used in time-to-event analyses 7) Clarification of additional covariates to include in the poolability analyses

Final Version 8.0	01Feb2021	In covariate analyses, subgroup analyses and baseline characteristic summaries, added number of lesions at baseline and updated tumor replacement at baseline to use blinded central review data. In exposure summary of chemotherapies received post-progression, added prior and concurrent medication eCRF page in the derivation.

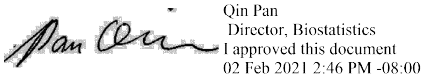
SIGNATURE PAGE

Author:

Tashana Francis
Principal Biostatistician
Covance

Date

Reviewed by:



Qin Pan, PhD
Director of Biostatistics
Covance

Date

Approved by:

DocuSigned by:
Chelsea Liu
Signer Name: Chelsea Liu
Signing Reason: I approve this document
Signing Time: 03-Feb-2021 | 05:45 PST
34457116A82F4B6888A0CDE8F6AFFDD0

03-Feb-2021 | 05:45 PST

Chelsea Liu, PhD
Fellow, Biostatistics
Biocompatibles UK Ltd.

Date

Approved by:

DocuSigned by:
Evelyne.Boucher
Signer Name: Evelyne.Boucher
Signing Reason: I approve this document
Signing Time: 03-Feb-2021 | 04:29 PST
D1BA4A32376741E0AF8A3FED7F22E389

03-Feb-2021 | 04:29 PST

Eveline Boucher, MD PhD
Medical Director, Interventional Oncology
Biocompatibles UK Ltd.

Date

Table of Contents

1 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Table 1: Abbreviations and Definitions of Terms

5-FU	5-Fluorouracil
AE	Adverse Event
ADE	Adverse Device Effect
ATC	Anatomical Therapeutic Chemical
BSA	Body surface area
CEA	Carcinoembryonic Antigen – CRC tumor marker
CI	Confidence Interval
CR	Complete Response
CRC	Colorectal Carcinoma
CRO	Contract Research Organization
CT	Computed Tomography
CTC	Common Terminology Criteria
CTCAE	Common Toxicity Criteria for Adverse Events
DCR	Disease Control Rate
DoR	Depth of Response
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EGFR	Epidermal Growth Factor Receptor
FACT –c	Functional Assessment of Cancer Therapy – colorectal
FOLFIRI	irinotecan-based chemotherapy
FOLFOX	oxaliplatin-based chemotherapy
Gy	Gray, a measure of irradiation dose
Hct	Hematocrit
Hgb	Hemoglobin
HPFS	Hepatic Progression Free Survival
HR	Hazard Ratio
IDMC	Independent Data Monitoring Committee
INR	International Normalized Ratio for prothrombin time
IPCW	Inverse Probability of Censoring Weighted
ITT	Intent-To-Treat
IVRS	Interactive Voice Response System
kg	kilograms
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LRT	Local Regional Therapy
LV	Leucovorin
m	meter
mCRC	Metastatic Colorectal Carcinoma
MedDRA	Medical Dictionary for Regulatory Activities

mg	Milligram
MRI	Magnetic Resonance Image
NCI	National Cancer Institute
NE	Not Evaluable
NLR	Neutrophil-Lymphocyte Ratio
ORR	Objective Response Rate
OS	Overall Survival
QoL	Quality of Life
PD	Progressive Disease
PFS	Progression Free Survival
PP	Per Protocol
PR	Partial Response
PT	Prothrombin Time
PTT	Partial Thromboplastin Time
PTTS	Post Treatment Tumor Shrinkage
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SADE	Serious Adverse Device Effect
SD	Stable Disease
TACE	Transarterial Chemoembolization
TEAE	Treatment-Emergent Adverse Event
TTDQoL	Time to Deterioration in Quality of Life
TTSP	Time to Symptomatic Progression
ULN	Upper Limit of Normal
UFT	Uracil and Tegafur
US	United States
VEGF(R)	Vascular Endothelial Growth Factor (Receptor)
VIF	Variance Inflation Factor
WBC	White Blood Cells
WHO	World Health Organization
WHO DE	World Health Organization Drug Enhanced
Y90	Ytrium-90

2 INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer diagnosed among both men and women in the US. Because the liver is the most frequent site of metastases, an estimated 60% of patients who are diagnosed with CRC eventually will experience liver disease as a predominate site (Sasson et al, 2002). Consequently, much of the morbidity and mortality in patients with CRC (Wagner et al, 1984) is due to unresectable liver metastases.

TheraSphere was evaluated in a cohort of seventy-two patients (Mulcahy et al 2009) with unresectable hepatic colorectal metastases who were treated at a targeted absorbed dose of 120 Gy with a median delivered dose of 118 Gy. The safety and toxicity of TheraSphere was assessed using version 3 of the National Cancer Institute (NCI) Common Terminology Criteria (CTC). Response was assessed radio-graphically and survival estimated using the Kaplan-Meier method from the diagnosis of hepatic metastases and first treatment. The tumor response rate was 40.3%. The median time to hepatic progression was 15.4 months, and the median response duration was 15 months. Based on sub-stratification analyses, tumor replacement (≤25% vs >25%) was associated with significantly greater median survival (18.7 months vs 5.2 months). The presence of extra hepatic disease was associated negatively with overall survival (7.9 months vs 21 months). Overall survival from the date of initial hepatic metastases was 34.6 months. A subset analysis of patients who had an Eastern Cooperative Oncology Group (ECOG) performance status of 0 demonstrated a median survival of 42.8 months and 23.5 months from the time of hepatic metastases and TheraSphere treatment, respectively. The data from this study also suggests that patients who have been exposed to fewer than three cytotoxic agents may have a better outcome than patients who have received all chemotherapy options prior to treatment with TheraSphere.

Based on the subset analyses of this study, it appears patients with good performance status, no extra hepatic metastases, liver disease limited to ≤25% of liver volume, who have not received all available lines of chemotherapy may benefit most from treatment with TheraSphere. It is proposed to evaluate the outcome of these patients when TheraSphere is added to second-line standard-of-care chemotherapy.

3 STUDY OBJECTIVES

The objective of this study is to evaluate the efficacy and safety of TheraSphere in the treatment of patients with metastatic colorectal cancer of the liver who have progressed with first line chemotherapy.

4 STUDY DESIGN

4.1 General Design

This is an open-label, prospective, multi-center, randomized, clinical trial.

Patients with metastatic colorectal carcinoma of the liver, who have disease progression following first line chemotherapy, and in whom the administration of standard-of-care

second-line chemotherapy with either Oxaliplatin (FOLFOX) or an oxaliplatin-based regimen or Irinotecan (FOLFIRI) or an irinotecan-based regimen is planned, are eligible to participate. Most patients receive an oxaliplatin-based regimen during first line chemotherapy; therefore, it is expected most patients will be receiving FOLFIRI or an irinotecan-based regimen as their standard-of-care second-line regimen while they participate in this trial. However, some patients are expected to receive FOLFOX or an oxaliplatin-based regimen as their second-line regimen.

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

Treatment Group: Patients randomized to the Treatment group will be treated with TheraSphere administered in place of the second cycle of their standard-of-care second line chemotherapy regimen. Subsequent chemotherapy must not begin until a minimum of two weeks (+/- 2 days) after the TheraSphere has been administered. Treatment with first cycle of chemotherapy must begin within 21 days of randomization.

Control Group: Patients randomized to the Control Group will receive only their planned standard-of-care second-line chemotherapy regimen. Treatment with chemotherapy must begin within 21 days of randomization.

The addition of approved biological agents (bevacizumab, cetuximab, panitumimab, aflibercept, ramucirumab) is permitted and should be administered at the investigator discretion per local practices, to the approved targeted population per label. Biological agents can be resumed, for patients randomized to the Treatment Group, 2 weeks after the TheraSphere administration (i.e. with cycle 2 of the standard-of-care chemotherapy). Standard of care biological agents can start along with the first cycle of second line chemotherapy for patients randomized to the Control group. Patients will have regular clinical study visits if they participate in the trial. During these visits, safety and efficacy data will be collected and recorded.

The primary efficacy endpoints of the trial will be Progression Free Survival (PFS) and Hepatic Progression Free Survival (HPFS). Once a patient has progressed, he/she will have reached the primary efficacy endpoints of the trial, but the patient will be encouraged to remain on the trial for evaluation of the secondary endpoint of survival. Following disease progression, patients in either group may receive the Best Alternative Therapy or Care for further treatment of their disease. For patients randomized to the Treatment Group, TheraSphere will be provided to patients amenable to further treatment with TheraSphere.

4.2 Method of Assignment of Patients to Treatment Groups

Patients will be randomized to study treatment, either the Control group or the Treatment group in a 1:1 ratio.

At study enrollment, each patient will be assigned a subject identity code (e.g. T020103-001) consisting of the protocol number (T02), the country number (e.g. 01), the site

number (e.g. 03), and patient number (e.g. 001).

If a patient is determined to be eligible to participate in the trial, the study site will contact the central randomization office when randomization will be determined using assignment by a computer-generated randomization scheme. Upon randomization, each patient will be assigned a 4 digits randomization number with the first digit indicating which combination of the 3 stratification factors the patient has.

A centralized randomization schedule will be generated by a statistician in the Covance Biometrics department who is not associated with the conduct or analysis of the study, using a validated system. The randomization will be stratified by the following factors:

- Unilobar vs bilobar disease
- Oxaliplatin vs Irinotecan based first-line chemotherapy
- KRAS status (wildtype vs mutant)

In order to ensure that the study treatment groups are balanced, the schedule will have randomization numbers assigned to the 2 study treatments in blocks of 4 within each combination of the 3 stratification factors to achieve a 1:1 ratio of study treatment (i.e. an equal number of patients in each treatment group). The randomization will be performed using IVRS by Perceptive, Inc. Each eligible patient will be assigned to the next sequential randomization number within the specified stratification combination and will receive the corresponding study treatment.

Patients randomized to the Treatment group who are unable to receive their planned study treatment will continue to be followed under the study group to which they were randomized for the purpose of the intent-to-treat analysis.

4.3 Blinding

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

To maintain the integrity of the study results in this open label study, the following personnel who had access to the study data before database lock, were required to sign an "Aggregate Data Declaration Form" documenting their agreement to not produce or review aggregate summaries of efficacy and death data, including AEs with an outcome of death, separated by treatment arm:

- CRO (e.g. Covance) personnel who were not involved in preparing data summaries for the independent data monitoring committee (IDMC) meetings, and
- Sponsor personnel.

4.4 Determination of Sample Size

This study is designed using a group sequential design with primary endpoints of PFS and HPFS. The study could be stopped early for efficacy at an interim analysis for superiority in PFS but not HPFS. The study is designed to detect 2.5 months increase in median PFS from 6 months in the control arm to 8.5 months in the TheraSphere arm (i.e. hazard ratio [HR] = 0.71), and a 3.5 month increase in median HPFS time, from 6.5 months in the control arm to 10 months in the TheraSphere arm (ie, HR =0.65), using log rank tests.

The analysis of PFS will be based on a group sequential design with 2 interim analyses and rho family error spending function stopping boundary with rho=1.5. It is estimated that approximately 420 patients will need to be recruited over 36 months, with a 1year additional follow-up period, allowing for 10% of patients lost to follow-up and for whom a date of progression or death is not recorded. The Hochberg procedure (Hochberg, 1988) will be used to control Type I error for the two primary endpoints at the final analysis.

A simulation study, assuming that PFS and HPFS have a correlation between 0.3 and 0.8, showed that the power to detect the target difference in either median PFS (ie, HR=0.71) or median HPFS (ie, HR=0.65) is >90%, and the power to detect the target difference in PFS or HPFS alone is >80%. The simulation study also demonstrated control of Type I error.

Although the forecasted accrual period has been increased to 60 months, this does not increase the number of patients required, or affect the statistical power of the study since both the power and the timing of the interim and final analyses are based on the number of PFS events rather than the number of patients.

5 CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

5.1 Changes in the Conduct of the Study

The following protocol versions have been implemented for this study:

- Version 2.0 dated 10Jan2012
- Version 3.1 dated 08Apr2013
- Version 5.1 dated 30May2014
- Version 6.1 dated 08Jan2016
- Version 7.0 dated 10May2017

A separate protocol (version 6.2 dated 31May2016) was implemented for sites in Germany, where different eligibility criteria were used, mainly related to tumor burden at baseline. A further updated protocol for sites in Germany (version 7.1 dated 07Mar2018) corresponding to the changes made in version 7.0, was also implemented.

5.1.1 Number of Study Centers and Patients

In version 5.1 of the protocol, the number of study centers increased from 50 to 100. In protocol version 7.0, the maximum number of patients was increased to 420.

5.1.2 Study Design

In version 5.1 of the protocol, the design of the trial was amended to be an adaptive trial using a group sequential design with 2 interim analyses, with an option to increase the sample size at the second interim analysis, based on a sample size re-estimation. In protocol version 7.0, the following changes were made to the study design before the first interim analysis was conducted:

- HPFS was added, in addition to PFS, as a second primary endpoint
- the number of PFS events required for the first interim analysis was changed from 139 to 172 events
- the number of PFS events required for the second interim analysis was changed from 195 to 241 events
- the option to increase the sample size at the second interim analysis was removed.

5.1.3 TheraSphere Administration Before and After Chemotherapy

In version 3.1 of the protocol TheraSphere was to be administered before the first cycle of chemotherapy. However, for all other versions of the protocol TheraSphere was to be administered after the first cycle of chemotherapy.

5.2 Changes in the Planned Analyses

5.2.1 Futility Stopping Rule

An assessment of futility at the two planned interim analyses, based on conditional power, was included in the study design. However, it was decided by the Sponsor, before the first interim analysis was performed, that the futility assessment would not be performed. This was primarily because patient recruitment was faster than expected towards the latter part of the study, such that approximately 85% of the maximum sample size of 420 patients had already been randomized before the first interim analysis was performed.

5.2.2 Tumor Assessments Within 6 Weeks of Randomization

All protocol versions stated that the minimum time from baseline to establish Stable Disease (SD) is 6 to 8 weeks. This is clarified in Section 6.4 below that any tumor assessments performed within 6 weeks of randomization will not be included in the analysis of imaging related efficacy endpoints.

5.2.3 Primary Endpoints

In protocol version 7.0, HPFS was added in addition to PFS as a primary endpoint. The study could be stopped early for efficacy at an interim analysis based on superiority in PFS, but not HPFS.
The current version of the statistical analysis plan reduces the minimum number of PFS events at which the final analyses will be performed from 344 events to 330 events. However, if 330 PFS events have not been reached at 31 August 2020 then the final analysis will be performed with the number of events at that time. This change was implemented because the number of patients who withdrew early from the study was greater than expected, and so the originally planned 344 events required for the final analysis will not be reached based on the status of patients who had not yet had a PFS event.

5.2.4 Secondary Endpoints

A sequential hierarchical approach was added in version 5.0 of the protocol to control the study-wise Type I error rate. Also, supportive analysis was added on the secondary efficacy time-to-event endpoints using the Cox regression model to evaluate the effect of multiple covariates, including stratification factors. Objective Response Rate was added as an additional secondary endpoint.

Protocol versions 5.1 and 6.1 stated that log rank tests converted to z-scores will be computed for the secondary endpoints of HPFS, TTSP and TTDQoL. However, since these endpoints are not included in the interim analyses, conversion to z-scores is not necessary.

The protocol states that “a sensitivity analysis for OS will be performed displaying control treatment failures who go on to receive Yttrium-90 (Y90) microspheres separately from control treatment failures that do not go on to be treated with Yttrium-90 microspheres.” However, this analysis will not be performed because at the time of preparing this version of the statistical analysis plan there was only a small number of patients in the control arm who had received Y90 after progression on second line chemotherapy. Instead, the number of control arm patients who received Y90 after progression will be provided, together with a summary of OS for these patients.

5.2.5 Per Protocol Population

The protocol states that patients in this population will be analyzed according to the treatment actually received. However, patients in this population will be analyzed according to the treatment group to which they were randomized, because patients who did not receive the treatment they were randomized to receive will be excluded from the Per Protocol population.

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 table. Note that Study Day 0 (day of randomization) in the Table below, will be referred to as Day 1 in the statistical outputs.

Table 1 Assessments Conducted at each Scheduled Visit

Evaluation/Test Timing of Visit(s)	Screen Day -14 to 0	Rand-omize Study Day 0	Chemo-therapy Q 2 weeks	1st TS work up & Admin-istration (replaces second cycle)	Study Visits to Progress-ion Q 8 weeks from random-ization (+/- 1 week)	Additional TS work up & Admin-istration Post hepatic progres-sion, TS Replaces a cycle of chemo	Study Visits Until Death or End of Study Q 8 weeks (+/- 1 week)
Informed Consent	x						
Demographics	x						
Medical History	x						
Physical Exam	x						
ECOG Performance Status	x		x[7]	x	x	x	x[7]
Medication & Prior Treatment History	x						
Review Eligibility Criteria	x						
Hematology: WBC with Differential, Hgb, Hct, platelet	x		x		x	x	
Coagulation: PT, PTT, INR	x		x[1]			x	
Chemistry panel, liver function tests	x		x		x	x	
Serum Pregnancy[2]	x			x		x	
Tumor markers for CRC (CEA)	x				x		
Liver Volume/Mass Calculation				x		x	
Estimation of Tumor Burden[3]	x						
Documentation of Type and dose of	x						

[1] Only required at chemo visits as clinically indicated (i.e. if patient is being followed for coagulopathy)
[2] Required for female patients of childbearing potential
[3] Required for Screening Purposes, may be visual or volumetric assessment
[4] TS patients only
[5] All randomized patients: all patients must receive a study treatment (Chemo) within 21 days of randomization

Evaluation/Test **Timing of Visit(s)**	**Screen** **Day -14 to 0**	**Rand-omize** **Study Day 0**	**Chemo-therapy** **Q 2 weeks**	**1st TS work up & Admin-istration** **(replaces second cycle)**	**Study Visits to Progress-ion** **Q 8 weeks from random-ization (+/- 1 week)**	**Additional TS work up & Admin-istration** **Post hepatic progres-sion, TS Replaces a cycle of chemo**	**Study Visits Until Death or End of Study** **Q 8 weeks (+/- 1 week)**
chemo and biologics							
Randomize Patient		x					
Hepatic Angiogram, ^{99m}Tc-MAA scan, TS Dose Calculation				x		x	
Order TS[4]				x		x	
Administer TS[4]				x		x	
Administer 2nd line Chemotherapy[2]			x				
Record/Administer any Chemotherapy following 2nd line[5]							x
QOL questionnaire	x				x		x[7]
Spiral CT /MRI of abdomen/pelvis/chest[8]	x				x	x	
Assess/Report Adverse Events			x	x	x	x	x
Review/Record Concurrent Medication	x		x	x	x	x	x
Final Endpoint Efficacy/Safety documentation & exit patient							x

[6] Additional approved chemotherapy for CRC may be administered only after progression of 2nd line chemotherapy
[7] Can be done remotely if patient is not coming in for clinic visit
[8] all attempts should be made to image every 8 weeks until hepatic progression, plus confirmatory scan (see main text)

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 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 chemotherapy (control arm) or date of first angiogram (treatment arm) will be considered relative day 1, and the day before the first dose of chemotherapy or date of first angiogram will be relative day -1. Relative days will be calculated as follows:

For days on or after first dose of chemotherapy or date of first angiogram:
Date of Assessment – Date of first dose of chemotherapy or date of first angiogram + 1

For days before first dose of chemotherapy or date of first angiogram:
Date of Assessment – Date of first dose of chemotherapy or date of first angiogram

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

- The missing day of onset of an adverse event (AE) or start date of a concurrent therapy will conservatively be set to:
  - The first day of the month of the AE/concurrent therapy start month, if the month of first dose of chemotherapy or first angiogram is before the AE/concurrent therapy start month,
  - One day after first dose of chemotherapy or first angiogram, if the month of first dose of chemotherapy or first angiogram is the same as the AE/concurrent therapy start month.
- The missing day of resolution of an AE or end date of a concurrent therapy will be set to the last day of month of the AE/concurrent therapy end month.
- 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.

- A missing day of death will be replaced by the 15th of the month if there are no other assessments after the 15th of the month for that patient. Otherwise, the last day of the month will be used to replace the missing day of death.
- If the onset date of an AE or start date of a concurrent therapy is missing both day and month, it will be set to:
    - January 1 of the year of AE/concurrent therapy start year, if the year of first dose of chemotherapy or first angiogram is before the AE/concurrent therapy start year,
    - One day after first dose of chemotherapy or first angiogram, if the year of first dose of chemotherapy or first angiogram 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 year the AE/concurrent therapy end year.
- For the date of diagnosis of CRC, the date of progression on first line chemotherapy and for the start date of chemotherapy agents (recorded on the Medical History mCRC Cancer and Prior Therapy electronic case report form (eCRF) page), the missing day will be set to the first day of the month, and the missing day and month will be set to January 1 of the year. If the imputed start date of the chemotherapy agent is before the date of diagnosis of CRC, it will be set to the date of diagnosis of CRC.
- For the end date of chemotherapy agents recorded on the Medical History mCRC Cancer and Prior Therapy eCRF page, the missing day will be set to the last day of the month, and the missing day and month will be set to December 31 of the year. If the imputed end date of the chemotherapy agent is before the start date of that chemotherapy agent, it will be set to the start date of the chemotherapy agent. If the imputed end date of the chemotherapy agent is after the date of randomization, it will be set to the date of randomization.
- For the start date of best available care recorded on the Post Progression Chemotherapy eCRF page and the date of best available care procedure recorded on the Post Progression Procedures eCRF page, the missing day will be set to:
    - the first day of the month if the month of progression (as determined by the investigator) is before the month of the start of best available care treatment.
    - One day after the date of progression (as determined by the investigator) if the month of progression is the same the month of the start of best available care treatment.
- For the end date of best available care recorded on the Post Progression Chemotherapy eCRF page, the missing day will be set to the last day of the month.

6.2.2 Windows

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

Table 2 Analysis Windows for Assessments Performed at Two Week Intervals

Week	Scheduled Study Day	Visit Window for Analysis (Days)
Screen	-14 – 1	-14 – 1
Randomize (Baseline)	1	1
Week 2	15	2 - 22
Week 4	29	23 – 36
Week 6	43	37 – 50
Week 8	57	51 – 64
Week 10	71	65 – 78
…		
End of Study		Latest assessment available

Table 3 Analysis Windows for Assessments Performed at Eight Week Intervals

Week	Scheduled Study Day	Visit Window for Analysis (Days)
Screen	-14 – 1	-14 – 1
Randomize (Baseline)	1	1
Week 8	57	2 – 85
Week 16	113	86 – 141
Week 24	169	142 – 197
…		
…		
End of Study		Latest assessment available

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, the data of the visit after the scheduled study day will be used.

Note that the analysis windows will not be used for tumor response endpoints, including the primary endpoints of PFS and HPFS (with the exception of the exploratory endpoint of post treatment tumor shrinkage). In addition, the analysis windows will not be used for the secondary endpoints of TTSP and TTDQoL.

6.3 Baseline Assessments

Baseline is defined as any assessment performed on or before the day of randomization.

The following baseline assessments will be conducted prior to randomization:

- Informed Consent

- Inclusion/ Exclusion Criteria
- Demographics (age, gender, race, ethnicity)
- Medical history
- Physical examination
- Vital signs (heart rate, respiration rate, blood pressure, body temperature, height and weight)
- Disease and treatment history
- ECOG performance status
- Laboratory tests (hematology, coagulation, chemistry)
- Serum pregnancy test
- CRC tumor biomarkers
- Liver volume/mass and tumor burden
- FACT-c QOL
- CT/MRI of chest
- CT/MRI of pelvis
- Stratification factors:
    - Unilobar vs bilobar
    - Oxaliplatin vs irinotecan based first-line chemotherapy
    - KRAS status (wildtype vs mutant)

Time from diagnosis of CRC will be calculated as follows:

Time from diagnosis of CRC (in months) = (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. Any tumor assessments performed within 6 weeks of randomization will not be included in the analysis of imaging related efficacy endpoints.

6.4.1 Primary Efficacy Variables – Progression Free Survival (PFS) and Hepatic Progression Free Survival (HPFS)

The primary study endpoints are PFS and HPFS by blinded central image review. The study will be considered to have met its objective if the difference in at least one of the primary endpoints between treatment groups is statistically significant. PFS is defined as the time from date of randomization until date of progression determined by blinded central image review, according to RECIST v1.1 (Eisenhauer et al, 2009), or death due to any cause, whichever is earlier. HPFS is defined as the time from randomization to the date of disease progression in the liver according to RECIST 1.1, or death due to any cause, whichever is earlier.

Radiographical disease status as indicated by the Central Imaging Review Organization will be used for the primary efficacy analysis. The independent review will be performed by ICON Medical Imaging.

6.4.2 Secondary Efficacy Variables

The secondary efficacy endpoints for this study are:

- Overall Survival (OS)
- Time to Symptomatic Progression (TTSP)
- Objective Response Rate (ORR) by blinded central image review
- Disease Control Rate (DCR) by blinded central image review
- Quality of Life (FACT-c)

6.4.2.1 Overall Survival (OS)

OS is defined as the time from date of randomization until date of death due to any cause.

6.4.2.2 Time to Symptomatic Progression (TTSP)

TTSP is defined as the time from date of randomization to date of assessment of ECOG performance status >2 that is confirmed at the first subsequent evaluation at least 8 weeks later (see examples in Table 4).

Table 4 Examples of Symptomatic Progression

Week X	Week X+2	Week X+4	Week X+6	Week X+8	Week X+10	TTSP Event
ECOG>2	ECOG>2	ECOG<2	ECOG>2	ECOG>2	ECOG<2	Yes
ECOG>2	ECOG>2	Missing	ECOG<2	ECOG>2	Missing	Yes
ECOG>2	No further ECOG assessments					No
ECOG>2	ECOG>2	ECOG>2	No further ECOG assessments			No

Week X denotes that week of the first occurrence of ECOG >2

6.4.2.3 Objective Response Rate (ORR) by blinded central image review

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 according to RECIST v1.1 is recorded at each efficacy visit. The tumor response for target lesions is categorized as Complete Response (CR), Partial Response (PR), Stable Disease (SD), Progressive Disease (PD) or Not Evaluable (NE) according to the RECIST criteria v1.1 as shown in Table 5 and Table 6 below.

Table 5: Target lesion response categories

Response	Definition
Complete Response (CR)	Disappearance of all non-nodal target lesions. Nodal target lesions must reduce in short axis to <10 mm [a]
Partial Response (PR)	At least 30% decrease in the sum of the diameters of target lesions, taking as reference the baseline sum of the diameters
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR, nor sufficient increase to qualify for PD
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
Not Evaluable (NE)	Any of the following conditions apply unless progression is otherwise noted [b]: 1) One or more target lesions cannot be assessed or measured accurately (eg, inadequate scan coverage, contrast, artifacts, or other factors) 2) Assessment methods used were not comparable with those used at baseline (eg, change of modality) 3) One or more target lesions were excised or irradiated, and have not reappeared or increased

[a] For CR, a sum of diameters of zero will not be required; a nodal target lesion can reduce in the short axis to <10 mm and become normal.

[b] For instances where there is a change in modality (eg, from enhanced CT to MRI) or surgical intervention after baseline, the patient may be evaluated for progression, but not for response.

Source: Independent Review Charter Final_v3.0_11-DEC-2019

Table 6: Non-target lesion response categories

Response	Definition
CR	Disappearance of all non-nodal non-target lesions All non-target lymph nodes must be non-pathological in size (<10 mm short axis)
Non-CR/Non-PD	Persistence of 1 or more non-target lesions
PD[a]	Unequivocal progression of existing non-target lesions
NE	Any of the following conditions apply unless progression is otherwise noted [b]: 1) One or more non-target lesions cannot be assessed (eg, inadequate scan coverage, contrast, artifacts, or other factors) 2) Assessment methods used were inconsistent with those used at baseline 3) One or more non-target lesions were excised or irradiated and have not reappeared or increased

[a] According to the RECIST 1.1 guidelines "to achieve 'unequivocal progression' on the basis of the non-target disease, there must be an overall level of substantial worsening in non-target disease such that the overall tumor burden has increased sufficiently to merit discontinuation of therapy. A modest increase in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status." The non-target status as a whole should indicate treatment failure.

[b] For instances where there is a change in modality (eg, from enhanced CT to MRI) or surgical intervention after baseline, the patient may be evaluated for progression, but not for response.

Source: Independent Review Charter Final_v3.0_11-DEC-2019

Table 7 provides a summary of the overall response status calculation at each timepoint.

Table 7: Timepoint response

Target lesions	Non-target lesions	New lesions	Timepoint response
Any	Any	Yes	PD
CR	CR	No	CR
CR	Non-CR/Non-PD or NE[a]	No	PR
PR	Non-PD or NE	No	PR
SD	Non-PD or NE	No	SD[c]
NE	Non-PD	No	NE
PD	Any	No	PD
Any	PD	No	PD
No targets[b]	CR	No	CR
	Non-CR/Non-PD	No	Non-CR/Non-PD[d]
	NE	No	NE
	PD	No	PD
	No non-targets	No	No disease (ND)[e]
CR	No non-targets	No	CR
PR			PR
SD			SD
PD			PD
NE			NE

[a] "Not all evaluated" is terminology from the RECIST 1.1 publication that describes non-target lesions as present, but not all can be adequately evaluated. It is analogous to NE. The combination of SD or PR in Target lesions and NE for non-target requires the reviewer's judgment. If unevaluated non-target disease represents a significant portion of the patient's overall tumor burden, such that changes in the missing lesions could indicate treatment failure (even in the context of stable or responding lesions elsewhere), the reviewer should select NE as the Timepoint Response.

[b] Although the inclusion criteria may require measurable disease on screening scans, the possibility exists that the central Reviewer may disagree with the site's assessment and find no measurable disease at baseline. Therefore the functionality exists.

[c] The minimum time from baseline to establish SD is 6 to 8 weeks.

[d] A label of SD for the Timepoint Response is not advisable when there was no measurable disease at baseline.

[e] Although advanced disease is indicated for this protocol or the inclusion criteria require evidence of disease at screening, the system's logic accounts for the possibility of no disease seen at baseline and is represented here.

Source: Independent Review Charter Final_v3.0_11-DEC-2019

The best overall response is based on the overall responses from each imaging assessment at efficacy visits. It is the best response a patient has had following randomization, but up to and including the first PD or the last valid post baseline imaging assessment in the absence of the first PD.

If a patient received a subsequent systemic anti-cancer regimen or local treatment with Y90 (including TheraSphere), TACE, ablation or resection of liver lesions (henceforth referred to as "subsequent mCRC therapy" in this document for ease of reference; see Appendix 4 for further details), tumor response assessments after the start of the

subsequent mCRC therapy will be excluded from the calculation of best overall response.

The ORR is defined as the proportion of randomized patients achieving a best overall response of CR or PR, as determined by blinded central image review.

Patients who do not have any post baseline tumor assessments for any reason prior to the start of the subsequent mCRC therapy, are considered non-responders and are included in the denominator when calculating the ORR.

6.4.2.4 Disease Control Rate (DCR) by blinded central image review

The DCR is defined as the proportion of randomized patients achieving a best overall response of CR, PR, or SD as defined by RECIST v 1.1, as determined by blinded central image review.

If a patient received subsequent mCRC therapy, imaging assessments after the start of the subsequent mCRC therapy will be excluded from the calculation of DCR. Patients who do not have any post baseline tumor assessments for any reason prior to the start of the subsequent mCRC therapy, are considered non-responders and are included in the denominator when calculating the DCR.

6.4.2.5 Quality of Life (FACT-c)

The total score of the FACT-c QoL instrument will be calculated, the scores of each domain (Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being), Colorectal Cancer subscale, FACT-c trial outcome index and each question at each time-point and their differences from baseline will be determined for each treatment group.

The scoring algorithm is in Section 11.3 Appendix 3.

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

6.4.3 Additional Efficacy Variables

6.4.3.1 PFS by investigator assessment

PFS by investigator assessment is defined as the time from date of randomization until date of progression determined by the investigator, according to RECIST v1.1, or death due to any cause, whichever occurs first.

6.4.3.2 HPFS by investigator assessment

HPFS by investigator assessment is defined as the time from date of randomization until date of progression in the liver determined by the investigator, according to RECIST v1.1, or death due to any cause, whichever occurs first.

6.4.3.3 ORR by investigator assessment

ORR is defined as the proportion of randomized patients achieving a best overall response of CR or PR, as defined by RECIST v 1.1, and as determined by the investigator. If a patient received subsequent mCRC therapy, imaging assessments after the start of the subsequent mCRC therapy will be excluded from the calculation of ORR. Patients who do not have any post baseline tumor assessments for any reason prior to the start of the subsequent mCRC therapy, are considered non-responders and are included in the denominator when calculating the ORR.

6.4.3.4 Duration of response

The duration of response will be determined for patients who have a best overall response of CR or PR. Duration of response is defined as the time from first date of overall response of CR or PR until date of PD, or death due to any cause, whichever occurs first.

Duration of response will be assessed separately by blinded central image review and by investigator assessment.

6.4.3.5 DCR by investigator assessment

DCR is defined as the proportion of randomized patients achieving a best overall response of CR, PR, or SD as defined by RECIST v 1.1, and as determined by the investigator.

If a patient received subsequent mCRC therapy, imaging assessments after the start of the subsequent mCRC therapy will be excluded from the calculation of DCR. Patients who do not have any post baseline tumor assessments for any reason prior to the start of the subsequent mCRC therapy, are considered non-responders and are included in the denominator when calculating the DCR.

6.4.3.6 Duration of disease control

The duration of disease control will be determined for patients who have best overall response of CR, PR, or SD. Duration of disease control is defined as the time from first date of overall response of CR, PR, or SD until date of PD or death due to any cause, whichever occurs first.

Duration of disease control will be assessed separately by blinded central image review and by investigator assessment.

6.4.3.7 Depth of response (DoR)

DoR is defined as the percentage change from baseline to nadir in the sum of the longest diameters of target lesions. If a patient received a subsequent mCRC therapy, tumor assessments after the start of the subsequent mCRC therapy will be excluded from the calculation of DoR.

DoR will be assessed separately by blinded central image review and by investigator assessment. In addition, DoR will be assessed for the following subgroups based on tumor replacement (%) at baseline:

- >5% tumor replacement by blinded central review
- >10% tumor replacement by blinded central review
- >15% tumor replacement by blinded central review

For blinded central image review, when adjudication has occurred, data from the reviewer selected by the adjudicator will be used. When adjudication was not required the average value of the sum of the longest diameters of target lesions from the two central reviewers at each visit will be used to calculate the DoR.

6.4.3.8 Post Treatment Tumor shrinkage (PTTS)

PTTS is defined as the proportion of randomized patients achieving a ≥20% decrease in the sum of the longest diameters of target lesions, separately at the Week 8, 16, and 24 analysis visits (as defined in Table 3). PTTS will be assessed separately by blinded central image review and by investigator assessment.

For blinded central image review, when adjudication has occurred, data from the reviewer selected by the adjudicator will be used. When adjudication was not required the average value of the sum of the longest diameters of target lesions from the two central reviewers at each visit will be used to calculate the PTTS.

If a patient received a subsequent mCRC therapy, tumor assessment obtained after the start of the subsequent mCRC therapy will be excluded from the determination of achieving the threshold of PTTS at each analysis visit. Patients without post baseline tumor assessments prior to the start of the subsequent mCRC therapy are considered non-responders and are included in the denominator when calculating the PTTS.

6.4.3.9 Tumor Marker for CRC (CEA)

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

TheraSphere exposure will be presented as described below for the Treatment arm. This includes summaries presented separately for TheraSphere administered prior to progression and post progression evaluated according to RECIST v1.1 and by investigator assessment.

- Number of patients who received TheraSphere during the study

- Number of patients who received TheraSphere prior to progression and post progression
- Reasons for not receiving TheraSphere prior to progression
- Number of patients with bilobar disease at baseline who received TheraSphere prior to progression
    - to both lobes or to the whole liver (i.e. TheraSphere administered to both lobes in a non-lobar approach, for example through the common hepatic artery) on the same day,
    - to both lobes on different days,
    - to one lobe (note that bilobar disease is from the stratification factor on the Randomization eCRF page, with any incorrect values at randomization replaced with the corrected value from the eCRF)
- Number of patients with unilobar disease at baseline who received TheraSphere prior to progression
    - to both lobes or to the whole liver on the same day,
    - to both lobes on different days,
    - to one lobe (note that unilobar disease is from the stratification factor on the Randomization eCRF page, with any incorrect values at randomization replaced with the corrected value from the eCRF)
- Number of patients who received first TheraSphere administration post progression
- Patients with at least one TheraSphere administration not completed as planned prior to progression and separately post progression
- Time to the first angiogram (days), defined as (angiography date – randomization date + 1)
- Time to the first TheraSphere treatment (days), defined as (treatment date of first TheraSphere administration - randomization date + 1)
- TheraSphere dose absorbed by perfused volume prior to progression and post progression
- TheraSphere dose delivered to lungs prior to progression and post progression

TheraSphere dose absorbed by perfused volume will be calculated as follows.

- Dose absorbed by perfused volume within a lobe (left lobe or right lobe) is defined using data for the corresponding lobe, as the sum of doses delivered by each vial if multiple vials are used to treat same target tissue, or as the weighted average of doses delivered by each vial (weights are target tissue masses) if multiple vials are not used to treat same target tissue.
- Dose absorbed by perfused volume within the liver is defined as the weighted average of doses delivered to each lobe (weights are the sum of target tissue masses in each lobe) for patients who had both lobes treated, and as the single dose delivered for patients who received whole liver dosing.
- Dose absorbed by perfused volume is defined as the dose absorbed by the perfused volume within the treated lobe for patients who had one lobe treated,

and dose absorbed by the perfused volume within the liver for patients who had both lobes treated or who received whole liver dosing.

TheraSphere dose delivered to lungs will be calculated as the sum of doses delivered to lungs across all TheraSphere administrations.

For patients with more than one TheraSphere treatment session prior to progression, in the calculation of dose absorbed by perfused volume and dose delivered to lungs, TheraSphere treatment on a subsequent day prior to progression will only be used if the subsequent day occurs within 5 weeks (i.e. 35 days) following the first day of TheraSphere treatment.

6.5.1.2 Extent of Exposure to Second-line Chemotherapy Regimen

The World Health Organization Drug Enhanced (WHO DE) March 2011 will be used to classify chemotherapy agents by preferred term and WHO Anatomical Therapeutic Chemical (ATC) classification of ingredients. Chemotherapy information will be presented separately for second line chemotherapy drugs and for biological agents.

The incidence of a regimen and medication will be the number of patients who had the medication (counted only once) divided by the number of patients in the analysis population and represented as a percentage.

The following variables will be calculated for each chemotherapy/biological agent and used in the presentation of exposure to second line chemotherapy (5-FU, capecitabine, irinotecan and oxaliplatin) as well as exposure to biological agents.

- Number of cycles
- Average dose per cycle (mg/m^2 of BSA or mg/kg)
- Cumulative dose (mg/m^2 of BSA or mg/kg)
- Duration on treatment (weeks)

Number of cycles is defined as the number of visit records of chemotherapy with a start date collected on the eCRF. For example, as recorded on the eCRF, the first cycle is indicated as Chemotherapy (1,2,3, etc) (1), cycle 2 is indicated as Chemotherapy (1,2,3, etc) (2), etc.

6.5.1.3 Extent of Study Exposure and Follow-up

The duration on study and duration of follow-up will be determined.

6.5.1.4 Best Available Care Post-Progression

Chemotherapies and procedures received post-progression according to RECIST v1.1 by investigator assessment will be summarized as follows:

- Systemic mCRC treatments received (from the Best Available Care Post Progression Chemotherapy eCRF page, Chemotherapy Administration eCRF page, and Prior and Concurrent Medications eCRF page with start date after progression), by preferred terms

- Procedures received, categorized according to the eCRF as Ablation, Resection, TACE, Y90 (including TheraSphere), or Other

6.5.2 Adverse Events

All AEs will be documented from the date of randomization until disease progression or 30 days after discontinuation of the study therapy (second-line chemotherapy alone or TheraSphere and second-line chemotherapy), whatever comes first. After this, only AEs related to TheraSphere will be collected.

A treatment emergent AE (TEAE) is defined as an event that was not present at baseline or worsened in severity following the start of treatment.
Adverse events will be considered to be treatment emergent according to the following algorithm:

- If the start date of an AE is known, then:
  - If the AE starts at any time prior to Day 1 (where Day 1 is the date of first dose of chemotherapy (control arm) or date of first angiogram (treatment arm) then the AE will not be considered treatment emergent.
  - If the AE starts on or after Day 1 then the AE is considered treatment emergent.
- If the start day of an AE is unknown then imputation will be done as per Section 6.2.1 and the imputed date will be used in the determination of treatment emergence.

The investigator's verbatim term of an AE will be mapped to a system organ class and preferred term 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 AE to determine the severity of the AE.

Adverse events related to chemotherapy, device, and angiographic procedures are defined as those events recorded on the eCRF with relationship of possibly, probably, or definitely relationship. Relation to TheraSphere (device) is not appropriate for the Control group.

The incidence of TEAEs will be the number of patients who had the AE (counted only once) divided by the number of patients in the safety population and represented as a percentage. For gender, age group, race and ethnicity specific AEs, the percentage will be represented as the number of patients with the AE divided by the number of patients of that group in the safety population. The incidence of AEs will be the number of times an event 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.

Adverse events will be split further into two groups:

- AEs with a start date up until disease progression (i.e. PD by investigator assessment) or 30 days after discontinuation of study therapy, whichever comes first

- AEs with a start date after this

6.5.2.1 Serious Adverse Event (SAE)

A Serious Adverse Event is any untoward medical occurrence that at any dose:

- Results in death;
- Is life-threatening ("life-threatening" refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe);
- Requires inpatient hospitalization or prolongation of existing hospitalization;
- Results in a persistent or significant disability/incapacity; or
- Is a congenital anomaly/birth defect.

6.5.2.2 Adverse Device Effect (ADE)

An Adverse Device Effect is an AE related to a medical device and includes any event resulting from insufficiencies or inadequacies in the instructions for use or the deployment, implantation, installation or malfunction of the device; any event that is the result of user error; or any potential ADE which might have occurred if suitable action had not been taken or intervention had not been made or if circumstances had been less fortunate. All AEs with a relationship to device of possibly, probably, or definitely will be considered to be ADEs.

6.5.2.3 Serious Adverse Device Effect (SADE)

A Serious Adverse Device Effect is an ADE that has resulted in any of the consequences characteristic of a SAE or might have led to any of these consequences if suitable action had not been taken; intervention had not been made or circumstances had been less fortunate. All SAEs with a relationship to device of possibly, probably, or definitely will be considered to be SADEs.

6.5.3 Clinical Laboratory Evaluations

Clinical laboratory results will be converted to SI units. Change from baseline to each visit assessed and end of study will be defined using the windowing method specified in Section 6.2.2, as the visit value minus the baseline visit. 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.

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 one-sided with a significance level of alpha=0.025, unless specified otherwise, and will be performed using SAS® Version 9.1.3 or higher. Data will be summarized using descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for continuous variables and using frequency and percentage for discrete variables.

Patient listings of all data from the 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 univariable Cox regression analysis of time-to-event efficacy endpoints, including PFS and HPFS

- Stratification factors
  - Unilobar vs bilobar disease
  - Type of first-line chemotherapy (oxaliplatin vs irinotecan)
  - KRAS status (wild type vs mutant)

  Notes: Stratification factors according to the master file, with any incorrect values at randomization replaced with the corrected value from the eCRF, will be used.
- ECOG status at baseline
- Age group (≥18 to <65 years, ≥65 to <75 years, and ≥75 years)
- Gender
- Race (White or Caucasian, Black or African American, Asian, Other [Native Hawaiian or other Pacific Islander or Native American or Alaska Native or Other])
- Region (North America, Europe, Asia)
- US and non-f region
- Duration from diagnosis of mCRC to randomization (<12 months, ≥12 months)
- Receipt of adjuvant chemotherapy (yes or no); patients who received adjuvant chemotherapy will be identified programmatically from the systemic chemotherapies collected on the Medical History mCRC Cancer and Prior Therapy eCRF page at screening as described in Appendix 5.
- Duration from start date of first line chemotherapy to date of progression on first line chemotherapy as a continuous covariate. For patients who received adjuvant chemotherapy the start date of first line chemotherapy will be the earliest start date of oxaliplatin or irinotecan, as recorded on the Medical History mCRC Cancer and Prior Therapy eCRF page, that occurs after the end date of adjuvant chemotherapy. For patients who did not receive adjuvant chemotherapy the start date of first line chemotherapy will be the earliest start

date of oxaliplatin or irinotecan, as recorded on the Medical History mCRC Cancer and Prior Therapy eCRF page.

- Duration from date of progression on first line chemotherapy to start date of second line chemotherapy (<1 month, ≥1 month)
- Duration from end date of first line chemotherapy to start date of second line chemotherapy (<3 months, ≥3 months)
- Patients with synchronous metastases; a patient has synchronous metastases if one of the following 2 conditions are satisfied, otherwise the patient has metachronous metastases
    - the stage at initial diagnosis of CRC is IV, or
    - for patients with initial diagnosis of CRC of stage I, II or III, the duration between the date of diagnosis of CRC and the date of progression on first line chemotherapy is less than 3 months if the patient received adjuvant chemotherapy, and the duration between the date of diagnosis of CRC and the start date of the first line chemotherapy is less than 3 months if the patient did not receive adjuvant chemotherapy
- Location of primary tumors at time of first diagnosis of primary CRC categorized as right-sided, left-sided, both, or not available, with right-sided and left-sided derived according to the following eCRF options:
    - Right-sided: Appendix, Cecum, Ascending colon, Hepatic flexure, Transverse colon
    - Left-sided: Splenic flexure, Descending colon, Sigmoid, Recto sigmoid, Rectum
- Primary tumor in situ at baseline (yes or no)
- Tumor replacement (as percentage of total liver volume) at baseline (<10%, ≥10 to <25%, ≥25%). This data is recorded according to investigator assessment in the eCRF, but will also be assessed by blinded central review. The blinded central review data will be used rather than the investigator data.
- Extrahepatic disease at baseline (yes or no); two definitions will be considered for this variable (note that for univariable analyses both definitions will be used and the one with lower p-value will be included in the multivariable model if both have a one-sided p-value in the univariable model <0.075):
    - <u>Definition 1</u>: A patient has extrahepatic disease if
        - the patient has primary in situ,
        - or has extrahepatic target and/or non-target lesions (identified from target lesions and non-target lesions eCRF pages at screening, based on a manual review of the free text entered in the “other location, specify” and “extrahepatic location, specify” eCRF fields)

  - Definition 2: A patient has extrahepatic disease if
    - the patient has primary in situ,
    - or has extrahepatic target and/or non-target lesions,
    - or has indeterminate lesions in lung and/or lymph nodes
- Extrahepatic metastases at baseline (yes or no); two definitions will be considered for this variable (note that for univariable analyses both definitions will be used and the one with lower p-value will be included in the multivariable model if both have a one-sided p-value in the univariable model <0.075):
  - Definition 1: A patient has extrahepatic metastases if the patient has extrahepatic target and/or non-target lesions
  - Definition 2: A patient has extrahepatic metastases if the patient has extrahepatic target and/or non-target lesions, or has indeterminate lesions in lung and/or lymph nodes
- Oxaliplatin or irinotecan based second-line chemotherapy
- Biological agent received as part of first-line treatment (yes or no)
- Type of biological agent received as part of first-line treatment (anti-VEGF(R), anti-EGFR, or none), where aflibercept, bevacizumab and ramucirumab are anti-VEGF(R) agents, and cetuximab and panitumumab are anti-EGFR agents
- Biological agent received as part of second-line treatment (yes or no)
- Type of biological agent received as part of second-line treatment (anti-VEGF(R), anti-EGFR, or none)
- Maximum liver lesion size at baseline, defined as the greatest of the longest diameters of target lesions in the liver by central readers (<40 mm, ≥40 mm)
  - When adjudication has occurred, data from the reviewer selected by the adjudicator will be used
  - When adjudication was not required the average value of maximum liver lesion size at baseline from the two reviewers will be used
- CEA at baseline (<35 ng/mL, ≥35 ng/mL)
- Alkaline phosphatase at baseline (< site ULN, ≥ site ULN)
- Albumin at baseline (< site LLN, ≥ site LLN)
- Total bilirubin at baseline (< site ULN, ≥ site ULN)
- Neutrophil-lymphocyte ratio (NLR), calculated as the ratio of neutrophils to lymphocytes (NLR<5, NLR ≥5)
- Number of lesions at baseline (<3 lesions, 3-5 lesions, 6-10 lesions, >10 lesions) by blinded central review

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

All factors in the univariable models with a one-sided p-value <0.075 will be included in a multivariable analysis to determine the impact of these factors. For both univariable and multivariable analyses, the overall p-value will be used for factors with >2 levels (i.e. the p-value corresponding to the Type 3 Wald chi-square statistic) rather than the p-values corresponding to each level of the factor.

For the multivariable analysis, collinearity of covariates will be assessed by the variance inflation factor (VIF) (Belsley et al, 1980), and further action will be taken if any covariate has a VIF value >10. Highly correlated covariates (i.e. with VIF >10) will be removed, one at a time, based on the descending order of their univariable p-values or clinical justification, until VIF values are ≤10 for all covariates remaining in the multivariable model.

7.3 Handling of Dropouts or Missing Data

Dropout patients will not be replaced in this study. The handling of missing data will be discussed throughout Section 8, where relevant. Censoring for the efficacy endpoints is discussed throughout Section 8, where applicable.

7.4 Interim Analyses and Data Monitoring

An IDMC will be established to oversee the conduct of the study. The IDMC will meet periodically during the study to review enrollment, protocol deviations and safety events for the study. In addition, the IDMC will conduct and review the interim efficacy results and will make formal recommendations to the study Sponsor at the time of the interim analysis and during the conduct of the study.

After the first 20 patients in the treatment group have received TheraSphere followed by at least 2 cycles of chemotherapy, a feasibility safety assessment will be conducted. The IDMC will review the safety results of both the control and treatment groups. The IDMC will take into consideration the established safety profiles of TheraSphere, oxaliplatin based and irinotecan based chemotherapy as described in the package inserts for each product as well as the published literature. The expected high rates of AEs and death that are associated with disease progression in patients with advanced colorectal cancer will be considered.

A consideration for adjusting the dose of cytotoxic agents, or other safety recommendations, or stopping further enrollment to trial may be made by the IDMC if there is a pattern of serious toxicity clearly related to the sequential administration of TheraSphere with oxaliplatin based or irinotecan based chemotherapy. Such a toxicity pattern must be clearly different from, or more severe than, what might be expected from independent administration of the chemotherapy regimens and TheraSphere. The potential adverse impact of any such pattern of toxicity on the survival or well-being of the patient should be considered in the context of the safety and outcome expectations of patients with advanced colorectal cancer.

This study uses an adaptive group sequential design with two interim analyses and one final analysis. The efficacy stopping boundaries are based on the rho family error spending function with the parameter value rho=1.5. The first interim analysis is planned at approximately, but no less than, 172 PFS events based on blinded central image review, with a one-sided p-value ≤0.0088 allowing the study to be stopped early for efficacy, in which case HPFS will be tested at the same boundary as PFS using a log rank test converted to a z-score. A second interim analysis is planned at approximately, but no less than, 241 PFS events based on blinded central image review, with a one-sided p-value ≤0.0099 allowing the study to be stopped early for efficacy. Based on the rho family spending function with rho=1.5 and the planned number of PFS events at each analysis, the alpha spent at the first interim analysis is 0.0088, and the cumulative alpha spent at the second interim analysis is 0.0146. Therefore, the one-sided alpha that remains to be spent at the final analysis is 0.0104 (i.e. 0.025-0.0146). If the interim analyses do not occur at exactly 172 or 241 PFS events, the corresponding efficacy boundaries will be calculated using the rho family spending function with rho=1.5. If the study is stopped early for PFS at the second interim analysis, HPFS will be tested using the boundary derived based on an incremental alpha of 0.0058 (i.e. 0.0146-0.0088). This boundary will account for the correlation between the z-score for HPFS at the first interim analysis and the z-score for HPFS at the second interim analysis, which is determined by the observed number of HPFS events at the first interim analysis and the cumulative number of HPFS events observed at the second interim analysis.

OS will also be analyzed at the interim analyses. However, no formal stopping rules will be performed based on this interim OS analysis.

The final analysis of PFS and HPFS was originally planned when approximately, but no less than, 344 PFS events based on blinded central image review have occurred. However, at the time of preparing version 5.0 of the statistical analysis plan (SAP), it became clear, based on the status of patients who had not yet had a PFS event, that it would not be possible to reach 344 PFS events because of a greater number of patients than expected who withdrew early from the study. Therefore it is now planned to perform the final analysis when approximately 330 PFS events have occurred. However, if 330 PFS events have not been reached at 31 August 2020 then the final analysis will be performed with the number of events at that time.

The boundary for the final analysis will be determined based on the following methodology which will account for the different censoring rules used in the two interim analyses and the final analysis. The first interim analysis occurred at 204 PFS events, and PFS and HPFS were analyzed using version 2.0 of the SAP where patients without progression or death were censored at the last valid tumor assessment (Censoring Method A: Original). The second interim analysis occurred at 287 PFS events, and PFS and HPFS were analyzed using version 4.0 of the SAP where patients who received subsequent mCRC therapy prior to their last valid tumor assessment or progression or death were censored at their last valid tumor assessment prior to the start of the subsequent mCRC therapy (Censoring Method B: Original + Subsequent mCRC therapy). In the final analysis, additionally, if the patient progresses or dies

immediately after 2 or more missed visits, the patient will be censored at the last valid tumor assessment prior to the 2 missed visits (Censoring Method C: Original + Subsequent mCRC therapy + 2 missed visits). Under each censoring method, the alpha level for the final analysis will be derived based on the rho family spending function with rho=1.5, and the number of PFS events recalculated for the interim and final analyses, had the same censoring method been applied in all the analyses. The most conservative boundary from the three censoring methods will be used for the final analysis. This methodology is illustrated in Table 8.

Table 8: Calculation of efficacy boundary in final analysis

Censoring Method	First interim analysis PFS Events	Second interim analysis PFS Events	Final analysis PFS Events	Alpha Level for Final analysis
A	204	$E2_A$	EF_A	α_{fA}
B	$E1_B$	287	EF_B	α_{fB}
C	$E1_C$	$E2_C$	EF_C	α_{fC}

Notes:

- $E1_B$ and $E1_C$ will be calculated based on the first interim analysis data snapshot.
- $E2_A$ and $E2_C$ will be calculated based on second interim analysis data snapshot.
- EF_A, EF_B and EF_C will be calculated based on final data with a data cut-off of 31 August 2020.
- α_{fA} will be derived using information fractions of two interim analyses under censoring method A, (204 / EF_A) and ($E2_A$ / EF_A), respectively.
- α_{fB} will be derived using information fractions of two interim analyses under censoring method B, ($E1_B$ / EF_B) and (287 / EF_B), respectively.
- α_{fC} will be derived using information fractions of two interim analyses under censoring method C, ($E1_C$ / EF_C) and ($E2_C$ / EF_C), respectively.
- Alpha level for final analysis: $\alpha_f = \min(\alpha_{fA}, \alpha_{fB}, \alpha_{fC})$.

The calculation described in Table 8 for the alpha level for the final analysis, α_f, will be carried out prior to the database hard lock so that α_f will be fixed and documented prior to the database hard lock for the final analysis. The larger p-value between PFS and HPFS at the final analysis will be compared to α_f and the smaller p-value will be compared to $\alpha_f/2$ based on the Hochberg procedure. If the larger p-value is $\leq\alpha_f$, significance can be claimed for both PFS and HPFS. On the other hand, if the larger p-value is $>\alpha_f$ and the smaller p-value is $<\alpha_f/2$, then only the endpoint with the smaller p-value can be claimed to be significant.

7.5 Multiple Comparisons/Multiplicity

For the primary analysis the Type I error is controlled at $\alpha = 0.025$ (1-sided) over the 2 planned interim analyses and final analysis. Mathematical and simulation-based demonstration of type I error control is provided in Appendix 6. For the secondary endpoints, the study-wise Type I error will be controlled using a sequential hierarchical approach as explained in Section 8.6.

7.6 Use of an "Efficacy Subset" of Patients

Patients randomized to study drug who received at least one dose of study medication and who do not have major protocol deviations will form the Per Protocol (PP) Population. The major protocol deviations will be defined at the time of evaluability evaluation, the time between the database soft lock and hard lock.

Excluding patients who have major protocol deviations will likely decrease the variability in treatment response.

7.7 Examination of Subgroups

Disposition, and primary and secondary efficacy endpoints and treatment exposure data (except TheraSphere exposure of dose delivered to lungs) will be summarized by the following subgroups:

- By each stratification factor
  - Unilobar or bilobar disease
  - KRAS status of mutant or wildtype
  - Oxaliplatin or irinotecan based first-line chemotherapy

  Note: stratification factors according to the master file with any incorrect values at randomization replaced with the corrected value from the eCRF, will be used.
- ECOG status at baseline (0 or 1)
- Age group (≥18 to <65 years, ≥65 to <75 years, and ≥75 years)
- Gender
- Race (White or Caucasian, Black or African American, Asian, Other [Native Hawaiian or other Pacific Islander or Native American or Alaska Native or Other])
- Region (North America, Europe, Asia)
- US and non-US region
- Duration from diagnosis of mCRC to randomization (<12 months, ≥12 months)
- Duration from start date of first line chemotherapy to date of progression on first line chemotherapy (<10 months, ≥10 months; <6 months, ≥6 months)
- Duration from date of progression on first line chemotherapy to start date of second line chemotherapy (<1 month, ≥1 month)
- Duration from end date of first line chemotherapy to start date of second line chemotherapy (<3 months, ≥3 months)
- Receipt of adjuvant chemotherapy (yes or no)
- Patients with synchronous or metachronous metastases
- Location of primary tumors at time of first diagnosis of primary CRC categorized as right-sided, left-sided, both, or not available
- Patients with and without the primary tumor in situ at baseline
- Tumor replacement (as percentage of total liver volume) at baseline (<10%, ≥10 to <25%, ≥25%)

- Extrahepatic disease at baseline (yes or no); separately for Definitions 1 and 2 defined in Section 7.2
- Extrahepatic metastases at baseline (yes or no); separately for Definitions 1 and 2 defined in Section 7.2
- Oxaliplatin or irinotecan based second-line chemotherapy
- Biological agent received as part of first-line treatment (yes or no)
- Type of biological agent received as part of first-line treatment (anti-VEGF(R), anti-EGFR, or none)
- Biological agent received as part of second-line treatment (yes or no)
- Type of biological agent received as part of second-line treatment (anti-VEGF(R), anti-EGFR, or none)
- Maximum liver lesion size at baseline, defined as the greatest of the longest diameters of target lesions in the liver by central readers (<40 mm, ≥40 mm)
- CEA at baseline (<35 ng/mL, ≥35 ng/mL)
- Alkaline phosphatase at baseline (< site ULN, ≥ site ULN)
- Albumin at baseline (< site LLN, ≥ site LLN)
- Total bilirubin at baseline (< site ULN, ≥ site ULN)
- Neutrophil-lymphocyte ratio (NLR), calculated as the ratio of neutrophils to lymphocytes (NLR<5, NLR ≥5)
- Number of lesions at baseline (<3 lesions, 3-5 lesions, 6-10 lesions, >10 lesions) by blinded central review

AEs and TheraSphere exposure of dose delivered to lungs will be summarized by the following subgroups:

- Age group (≥18 to <65 years, ≥65 to <75 years, and ≥75 years)
- Gender
- Race (White or Caucasian, Black or African American, Asian, Other [Native Hawaiian or other Pacific Islander or Native American or Alaska Native or Other])
- Region (North America, Europe, Asia)
- US and non-US region
- Oxaliplatin or irinotecan based second-line chemotherapy
- Biological agent received as part of second-line treatment (yes or no)
- Tumor replacement at baseline (<10%, ≥10 to <25%, ≥25%)
- Unilobar or bilobar disease
- ECOG status at baseline (0 or 1)
- Number of lesions at baseline (<3 lesions, 3-5 lesions, 6-10 lesions, >10 lesions) by blinded central review (note that this subgroup analysis only applies to AEs)

8 STATISTICAL ANALYSIS

8.1 Disposition of Patients

The number of patients enrolled will be summarized by region (North America, Europe, Asia), country, and site. The number of patients randomized, the number of patients treated with chemotherapy and TheraSphere, the number of patients treated with TheraSphere only, and the number of patients treated with chemotherapy only will be summarized. The number of treated patients who discontinued from the study (treated and untreated) and the reasons for discontinuing from the study will also be summarized.

The above information will also be summarized according to the subgroups listed in Section 7.7.

8.2 Protocol Deviations

Protocol deviations/violations will not be entered into the database. However, protocol deviations/violations will be identified and summarized within Covance's Clinical Department from which BTG can make determinations. All protocol deviations/violations determinations will be made before the database is locked for statistical analysis.

8.3 Analysis Populations

8.3.1 Intent-to-Treat (ITT) Population

All randomized patients will form the Intent-to-Treat (ITT) Population and will be analyzed according to the treatment group to which they were randomized.

8.3.2 Safety Population

Patients in the ITT Population who received at least one administration of TheraSphere or chemotherapy will form the Safety Population and will be analyzed based on the treatment actually received. This population will be used in all safety reporting and analysis.

8.3.3 Per Protocol (PP) Population

The Per Protocol population is the subset of the ITT population excluding patients with major protocol deviations which may affect the efficacy evaluation. Patients in the PP population will be analyzed according to the treatment group to which they were randomized.

Major protocol deviations resulting in a patient being excluded from the PP population will include, but may not be limited to, the following:

- Violation of at least one of the following eligibility criteria (descriptions of the criteria are taken from protocol version 7.0 and some descriptions may differ

slightly in earlier versions of the protocol; eligibility criteria numbers remained the same in all the protocol versions implemented)

- #3: Must have colorectal cancer with unresectable metastatic disease to the liver (unresectable unilobar or bilobar disease) who have disease progression in the liver with oxaliplatin or irinotecan based first-line chemotherapy
- #6: Tumor replacement ≤50% of total liver volume
- #7: Current ECOG Performance Status score of 0-1 through screening to first treatment on study
- #11: Serum bilirubin up to 1.2 x upper limit of normal
- #12: Albumin ≥3.0 g/dL
- #13: Must not have a history of hepatic encephalopathy
- #18: No cirrhosis or portal hypertension
- #19: Must not have received any prior external beam radiation treatment to the liver
- #20: Must not have received any prior intra-arterial liver-directed therapy, including TACE or Y90 microsphere therapy
- #21: Must not have any planned liver-directed therapy o radiation therapy
- #24: Must not have any clinically evident ascites (trace ascites on imaging is acceptable)
- #26: Must not have any significant life-threatening extra-hepatic disease, including patients who are on dialysis, have unresolved diarrhea, have serious unresolved infections including patients who are known to be HIV positive or have acute HBV or HCV
- #27: Must not have any confirmed extra-hepatic metastases. Limited, indeterminate extra-hepatic lesions in the lung and/or lymph nodes are permitted (up to 5 lesions in the lung, with each individual lesion <1 cm; any number of lymph nodes with each individual node <1.5 cm)

- Baseline imaging assessment not performed
- Baseline imaging assessment performed >42 days prior to date of randomization (note that although the screening period for baseline imaging assessment was 28 days, an additional 14-day window is being applied so that only baseline imaging assessments >42 days before randomization will be deemed to be a major protocol deviation that may affect the efficacy evaluation)
- Post-randomization imaging assessments not performed for 3 consecutively planned timepoints (i.e. the number of days between imaging assessments is >224 days), defined as
  - the first post-randomization imaging assessment, prior to progression as assessed by investigator according to RECIST 1.1, occurs at >32 weeks (i.e. >224 days) after randomization, or

  - any post-randomization imaging assessment, prior to progression as assessed by investigator according to RECIST 1.1, occurs at >32 weeks (i.e. >224 days) after the previous post-randomization imaging assessment
- Randomized study treatment not received (TheraSphere and/or second line chemotherapy) prior to progression as assessed by investigator according to RECIST 1.1
- TheraSphere received by patients in the control arm prior to progression assessed by investigator according to RECIST 1.1
- TheraSphere dose absorbed by perfused volume (as defined in Section 6.5.1.1) lower than the protocol stated range of 120 Gy – 10% (i.e. <108 Gy) prior to progression as assessed by investigator according to RECIST 1.1
- Bilobar disease at baseline but only one lobe treated with TheraSphere prior to progression assessed by investigator according to RECIST 1.1
- For patients enrolled under protocol version 3.1, where TheraSphere was to be administered before the first cycle of second line chemotherapy:
  - For the control arm:
    - Start of second line chemotherapy >28 days after randomization
  - For the treatment arm:
    - First administration of TheraSphere >28 days after randomization, or
    - Start of second line chemotherapy >21 days after first TheraSphere administration
- For all other patients, where TheraSphere was to be administered after the first cycle of second line chemotherapy:
  - For both arms:
    - Start of second line chemotherapy >28 days after randomization
  - For the treatment arm:

  First administration of TheraSphere >21 days after the start of second line chemotherapy

The deviations listed above will be programmatically determined. In addition, monitoring notes or data listings will be reviewed to determine any major deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification of major protocol deviations and decisions to exclude patients from the Per Protocol population will be made at the time between the database soft close and hard lock.

8.4 Demographic and Other Baseline Characteristics

All demographic and baseline summaries will be displayed for the ITT Population, Safety Population, as well as the PP Population.

Gender, race, ethnicity, and female childbearing potential will be summarized using counts and percentages. Age, height, and weight will be summarized with descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum). Age group (≥18 to <65 years, ≥65 to <75 years, and ≥75 years) will be summarized using counts and percentages.

The number and percentage of patients with abnormal physical examination findings at screening will be summarized. The number and percentage of patients with medical history events will be summarized. Vital signs collected at screening (blood pressure, heart rate, respiratory rate, and oral temperature) will be summarized with descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum).

Baseline characteristics of mCRC will be summarized using counts and percentages as follows:

- Patients who received prior therapy for mCRC
- Therapy type for mCRC
- Stage at initial diagnosis of CRC
- By each stratification factor
  - Unilobar vs. bilobar
  - KRAS status of mutant vs. wildtype
  - Oxaliplatin or irinotecan based first-line chemotherapy
- ECOG status at baseline
- Duration from diagnosis of mCRC to randomization (<12 months, ≥12 months)
- Duration from start date of first line chemotherapy to date of progression on first line chemotherapy (<10 months, ≥10 months; <6 months, ≥6 months)
- Duration from date of progression on first line chemotherapy to start date of second line chemotherapy (<1 month, ≥1 month)
- Duration from end date of first line chemotherapy to start date of second line chemotherapy (<3 months, ≥3 months)
- Patients who had indeterminate lesions in lung or lymph nodes at baseline, and the location of the indeterminate lesions
- Patients who had the primary tumor in situ at baseline
- Patients who received adjuvant chemotherapy
- Patients with synchronous vs metachronous metastases
- Extrahepatic disease at baseline according to Definitions 1 and 2 defined in Section 7.2
- Extrahepatic metastases at baseline according to Definitions 1 and 2 defined in Section 7.2
- Location of primary tumors at time of first diagnosis of primary CRC
  - According to the eCRF options, and
  - Categorized as right-sided, left-sided, both, or not available

- Patients who previously received a biological agent as part of first-line treatment
- Patients with CEA at baseline <10 ng/mL and ≥10 ng/mL
- Patients with CEA at baseline <35 ng/mL and ≥35 ng/mL
- Baseline tumor replacement <10%, ≥10 to <25%, and ≥25%. This data is recorded according to investigator assessment in the eCRF, but will also be assessed by blinded central review. The blinded central review data will be used rather than the investigator data.
- Maximum liver lesion size at baseline, defined as the greatest of the longest diameters of target lesions in the liver by central readers (<40 mm and ≥40 mm)
- Alkaline phosphatase at baseline < site ULN and ≥ site ULN)
- Albumin at baseline < site LLN and ≥ site LLN
- Total bilirubin at baseline < site ULN and ≥ site ULN
- Neutrophil-lymphocyte ratio (NLR), calculated as the ratio of neutrophils to lymphocytes (NLR<5, NLR ≥5)
- Baseline number of lesions (<3 lesions, 3-5 lesions, 6-10 lesions, >10 lesions) by blinded central review

The duration from diagnosis of mCRC to randomization, the duration from start date of first line chemotherapy to date of progression on first line chemotherapy, duration from date of progression on first line chemotherapy to start date of second line chemotherapy, duration from end date of first line chemotherapy to start date of second line chemotherapy, and baseline tumor replacement (%) will also be summarized with descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum).

Pregnancy test results will be summarized by number and percentage.

Baseline characteristics related to mCRC will also be summarized by age group, gender, race, and region.

8.5 Prior and Concomitant Therapy

The World Health Organization Drug Enhanced (WHO DE) March 2011 will be used to classify medications by preferred term and WHO ATC classification of ingredients.

The following applies to all data collected on the prior and concomitant eCRF page 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 administration of study treatment. Partial dates will be imputed conservatively as detailed in Section 6.2.1.

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 percentage of patients who took other therapy will be shown by WHO ATC classification of ingredients and by preferred term.

8.5.1 Prior Medication

A prior medication is defined as any medication stopped prior to randomization.

The number and percentage of patients who had at least one prior medication, that was not considered a colorectal therapy, will be tabulated as well as the number and percentage of patients with each medication. Patients will only be counted once for each medication.

8.5.2 Prior Therapy for mCRC

Prior mCRC treatment type and treatment will be summarized from the treatment type and treatment recorded on the Medical History of mCRC Cancer eCRF page.

8.5.3 Concomitant Medication

A concomitant medication is defined as any medication given prior to the patient being randomized and continuing after randomization, or any medication that is initiated on or after randomization. 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.

8.6 Analysis of Efficacy Parameters

8.6.1 Analysis of Primary Efficacy Variables

The primary efficacy analysis is of PFS and HPFS by blinded central review. PFS and HPFS rates will be derived from the Kaplan Meier estimates and presented with 95% CIs. Quartiles will be presented and 95% CIs will be calculated on the quartiles for each treatment group. PFS and HPFS will be compared between treatment arms using log rank tests, converted to a z-scores, at an overall one-sided alpha level of 0.025 to test the null hypothesis that the hazard rates for the treatment and control arms are equal versus the alternative hypothesis that the hazard rate for the TheraSphere arm is less than the hazard rate for the control arm. The one-sided alpha level of 0.025 will be adjusted over the 2 planned interim analyses and final analysis (as described in Sections 7.4 and 7.5). The HRs alongside their 95% CIs will also be computed from a Cox proportional hazards model. Plots of the Kaplan-Meier curves will be provided for each treatment group. This analysis will be performed on the ITT population and PP population (secondary analysis).

The assumption of proportional hazards used to compute the HRs for PFS and HPFS will be assessed. Firstly, a plot of log[-log (estimated probability of event-free survival)] versus log(time) will be examined, with nonparallel curves for the 2 treatment groups indicating non-proportional hazards. Also, a time-dependent covariate Cox regression model (i.e. adding a treatment group by time interaction) will

be fitted and if the time-dependent covariate has two-sided p-value <0.15 piecewise HRs over distinct time periods will be calculated.

Progression-Free Survival (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 PFS time will be censored at the randomization date, regardless of whether or not disease progression (i.e. PD) or death has been observed.
2) If a patient received a subsequent mCRC therapy before PD or death or in the absence of PD or death, the PFS time will be censored at the last valid post baseline radiological tumor assessment before the start of the subsequent mCRC therapy. If the patient has no post-baseline radiological tumor assessments before the start of the subsequent mCRC therapy, they will be censored at the randomization date.
3) If a patient is known not to have died or have PD and did not receive subsequent mCRC therapy, the PFS time will be censored 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.
4) If a patient had PD or died after two or more missed visits, the patient will be censored at the time of the last valid post baseline radiological tumor assessment date that occurred before the missed visits. If the patient has no post-baseline radiological tumor assessments before the missed visits, they will be censored at the randomization date.
    - For example, if a patient had a tumor assessment at Week 8 but did not have tumor assessments at Weeks 16 and 24, and then had PD at the Week 32 assessment, then the PFS time would be censored at the date of the Week 8 assessment. However, if the patient had a non-PD response at the Week 32 assessment (i.e. after 2 missed visits) then had PD at the Week 40 assessment, then the PFS would be a PD event at the date of the Week 40 assessment.
    - Given the scheduled visit scheme of tumor assessments (i.e. every 8 weeks), the definition of 2 missed visits will equate to 16 weeks since the previous tumor assessment, or since the randomization date if no previous post-baseline tumor assessment.

The units for analysis will be converted to months.

The number of deaths and the number of progressions that comprise the PFS events will be presented by treatment arm. The number of patients censored will be summarized by treatment arm for the following reasons:

- no baseline tumor assessments,

- no post-baseline tumor assessments and no subsequent therapy,
- no PD and no death and no subsequent therapy,
- subsequent therapy before PD or death,
- PD or death before subsequent therapy, or no subsequent therapy, but after two or more missed visits.

Additionally, for post baseline tumor assessments, summary statistics for the differences between scheduled day (i.e. according to the 8-weekly visit week derived based on actual study day per analysis windows in Section 6.2.2) and actual day will be presented for each treatment arm. A histogram illustrating the distribution of these differences in days by treatment arm will be generated. Since patients have varying numbers of visits, summary statistics and histogram will also be produced using the mean differences between scheduled day and actual day for each patient. All tumor assessments will be listed for each patient, including the actual assessment date (day), the scheduled date (day) and the difference between schedule day and actual day for each visit.

The following sensitivity analyses of PFS will be conducted to evaluate the robustness of the primary analysis.

- PD and death that occur after two or more missed visits post baseline tumor assessments will be considered as PFS events. All other censoring rules as described above will be used. For example, if a patient had a tumor assessment at Week 8 but did not have tumor assessments at Weeks 16 and 24, and then had PD at the Week 32 assessment, then the PFS would be a PD event at the Week 32 assessment.
- PD and death that occur after one or more missed visits post baseline tumor assessments will be considered as PFS events that occur at the day after the date of the last post baseline tumor assessment before the missed visits, or the day after randomization date if no post-baseline tumor assessments before the missed visits. All other censoring rules as described above will be used. Given the scheduled visit scheme of tumor assessments (i.e. every 8 weeks), the definition of one missed visit will equate to 8 weeks since the previous tumor assessment, or since the randomization date if no previous post-baseline tumor assessment.
- Early and late assessments will be allowed for the definition of the two or more missed visits, so two alternative definitions will be considered using 18 weeks (corresponding to the ±1 week window permitted by the protocol) and 20 weeks (corresponding to a wider ±2 week window), respectively. All other censoring rules as described above will be used.
- Patients who have dropped out of the study before PD and before the start of the subsequent mCRC therapy will be identified as PD events, unless patients

dropped out after two or more missed visits. All other censoring rules as described above will be used. The reasons for dropouts will be summarized.

- Patients who received subsequent mCRC therapy will not be censored for the subsequent mCRC therapy. The PFS event times of these patients after the start of the subsequent mCRC therapy will be included as events, unless they occur after two or more missed visits. All other censoring rules as described above will be used.
- Patients who had ablation or resection will not be considered as receiving subsequent mCRC therapy (i.e. definition of subsequent mCRC therapy in Appendix 4 will exclude ablation or resection). All other censoring rules as described above will be used.
- PD and death will be considered as PFS events, regardless of whether they occur after subsequent mCRC therapy or after two or more missed visits. Patients without PD and death will be censored 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. However, if a patient does not have a baseline tumor assessment, the PFS time will be censored at the randomization date, regardless of whether PD or death has been observed. This corresponds to the censoring rules used in the first interim analysis.
- PFS will be analysed using an inverse probability of censoring weighted (IPCW) approach (Thilakarathne et al, 2014) adjusted for the ECOG status at last assessment time before subsequent mCRC therapy as a time-dependent covariate. All censoring rules as described above will still be used. This IPCW approach will consist of the following 2 steps:
    - 1) Time-varying weights will be estimated using a multivariate logistic regression, including the censoring indicator as response variable and baseline ECOG status as a baseline covariate, and ECOG status at last assessment time before subsequent mCRC therapy as a time-dependent covariate.
    - 2) The time-dependent weights (i.e. inverse of estimated probability of censoring) will be incorporated in a Cox proportional hazards model, including the baseline ECOG status as a covariate.

Hepatic Progression-Free Survival (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 HPFS time will be censored at the randomization date, regardless of whether or not hepatic progression (i.e. hepatic PD) has been observed;
2) If a patient received a subsequent mCRC therapy before hepatic PD or death or in the absence of hepatic PD or death, the HPFS time will be censored at last

valid post baseline radiological tumor assessment before the start of the subsequent mCRC therapy. If the patient has no post-baseline radiological tumor assessments before the start of the subsequent mCRC therapy, they will be censored at the randomization date.

3) If a patient is known not to have died or to have hepatic PD and did not receive subsequent mCRC therapy, the HPFS time will be censored at the last post baseline radiological tumor assessment date or at the randomization date if the patient does not have any post-baseline radiological tumor assessments.

4) If a patient had hepatic PD or died after two or more missed visits, the patient will be censored at the time of the last valid post baseline radiological tumor assessment date that occurred before the missed visits. If the patient has no post-baseline radiological tumor assessments before the missed visits, they will be censored at the randomization date.

    - For example, if a patient had a tumor assessment at Week 8 but did not have tumor assessments at Weeks 16 and 24, and then had hepatic PD at the Week 32 assessment, then the HPFS time would be censored at the date of the Week 8 assessment. However, if the patient had non-hepatic PD at the Week 32 assessment (i.e. after 2 missed visits) then had hepatic PD at the Week 40 assessment, then the HPFS would be a hepatic PD event at the date of the Week 40 assessment.
    - Given the scheduled visit scheme of tumor assessments (i.e. every 8 weeks), the definition of 2 missed visits will equate to 16 weeks since the previous tumor assessment, or since the randomization date if no previous post-baseline tumor assessment.

The number of deaths and the number of hepatic progressions that comprise the HPFS events will be presented by treatment arm. The number of patients censored will be summarized by treatment arm for the following reasons:

- no baseline tumor assessments,
- no post-baseline tumor assessments and no subsequent therapy,
- no hepatic PD and no death and no subsequent therapy,
- subsequent therapy before hepatic PD or death,
- hepatic PD or death before subsequent therapy, or no subsequent therapy, but after two or more missed visits.

The following sensitivity analyses of HPFS will be conducted to evaluate the robustness of the primary analysis of HPFS.

- Hepatic PD and death that occur after two or more missed visits post baseline tumor assessments will be considered as HPFS events. All other censoring rules as described above will be used.

- Hepatic PD and death that occur after one or more missed visits post baseline tumor assessments will be considered as HPFS events that occur at the day after the date of the last post baseline tumor assessment before the missed visits, or the day after randomization date if no post-baseline tumor assessments before the missed visits. All other censoring rules as described above will be used. Given the scheduled visit scheme of tumor assessments (i.e. every 8 weeks), the definition of one missed visit will equate to 8 weeks since the previous tumor assessment, or since the randomization date if no previous post-baseline tumor assessment.
- Early and late assessments will be allowed for the definition of the two or more missed visits, so two alternative definitions will be considered using 18 weeks and 20 weeks respectively. All other censoring rules as described above will be used.
- Patients who have dropped out of the study before hepatic PD and before the start of the subsequent mCRC therapy will be identified as hepatic PD events, unless patients dropped out after two or more missed visits. All other censoring rules as described above will be used. The reasons for dropouts will be summarized.
- Patients who received subsequent mCRC therapy will not be censored for the subsequent mCRC therapy. The HPFS event times of these patients after the start of the subsequent mCRC therapy will be included as events, unless they occur after two or more missed visits. All other censoring rules as described above will be used.
- Patients who had ablation or resection will not be considered as receiving subsequent mCRC therapy (i.e. definition of subsequent mCRC therapy in Appendix 4 will exclude ablation or resection). All other censoring rules as described above will be used.
- Hepatic PD and death will be considered as HPFS events, regardless of whether they occur after subsequent mCRC therapy or after two or more missed visits. Patients without hepatic PD and death will be censored 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. However, if a patient does not have a baseline tumor assessment, the HPFS time will be censored at the randomization date, regardless of whether hepatic PD or death has been observed.
- HPFS will be analysed using IPCW approach adjusted for the ECOG status at last assessment time before subsequent mCRC therapy as a time-dependent covariate, as described above for PFS.

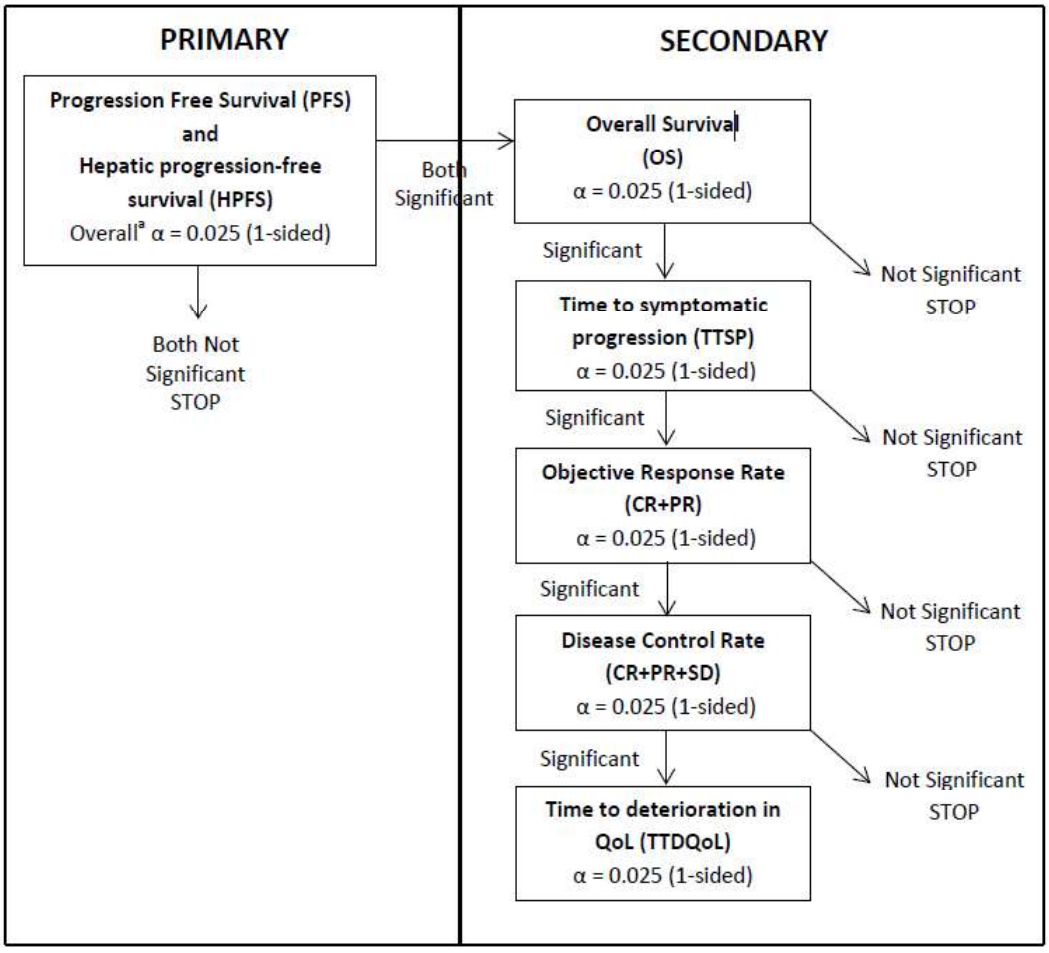
8.6.2 Analysis of Secondary Efficacy Variables

For secondary efficacy endpoints, each comparison between treatment groups will be conducted at α=0.025 (one-sided). Secondary study endpoints will be analyzed only at

the final analysis to determine the statistical significance, if any, between the treatment groups. Study-wise Type I error will be controlled using a sequential hierarchical approach, as shown in the figure below. That is, if the primary comparison is statistically significant, the secondary endpoints will be analyzed in order of the list below and will continue as long as the obtained 1-sided probability is equal to or less than 0.025. If a probability of greater than 0.025 is obtained, the inferential analysis of secondary endpoints will stop and not proceed further down the ordered list. In this manner the overall study alpha is protected and no further adjustment for multiplicity of analyses is required. If a probability of greater than 0.025 is obtained for an endpoint then the analysis of that endpoint and the endpoints further down the ordered list will still be presented but will be considered as exploratory endpoints rather than secondary endpoints.

For time-to-event secondary efficacy endpoints, the HRs alongside their 95% CIs, will be computed from a Cox proportional hazards model.

Hierarchical approach to control study-wise Type I error of primary and secondary efficacy endpoints



[a] Type I error is controlled at α=0.025 (1-sided) over the 2 planned interim analyses and final analysis.

8.6.2.1 Overall Survival (OS)

OS rates will be derived from the Kaplan-Meier estimates and 95% CIs will be calculated. A log-rank (one-sided) test will be used to compare OS between the two treatment groups at a 0.025 significance level. For each patient that has not known to have died, OS will be censored at the time of last contact date known to be alive.

Plots of the Kaplan-Meier curves will be provided for each treatment group. This analysis will be performed on both the ITT and PP populations.

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

The following subgroup analyses for OS will be performed:

- Patients who receive local regional therapies (LRTs) post-progression
- Patients who do not receive LRTs post-progression.

Kaplan Meier estimates for OS along with their 95% CIs will be presented for both of these subgroups.

The number of control arm patients who received Y90 treatment after progression according to RECIST v1.1 by investigator assessment will be provided, together with a summary of OS for these patients.

8.6.2.2 Time to Symptomatic Progression (TTSP)

TTSP rates will be derived from the Kaplan-Meier estimates and 95% confidence interval (CI) will be calculated. A log-rank (one-sided) test will be used to compare TTSP between the two treatment groups at a 0.025 significance level. Plots of the Kaplan-Meier curves will be provided for each treatment group. This analysis will be performed on both the ITT and PP populations.

Time to symptomatic progression (months) = (Date of event/censor – Date of Randomization +1) /30.4375.

The censoring is performed as follows:

1) If a patient does not have a baseline ECOG assessment, then the TTSP time will be censored at the randomization date, regardless of whether or not symptomatic progression has been observed.
2) If a patient received a subsequent mCRC therapy before symptomatic progression or in the absence of symptomatic progression (note that for this scenario a subsequent mCRC therapy may occur before the assessment of ECOG>2 or before the confirmation of ECOG>2 at the first subsequent assessment at least 8 weeks later), the TTSP time will be censored at the last post baseline ECOG assessment before the start of the subsequent mCRC therapy. If the patient has no post baseline ECOG assessments before the start of the subsequent mCRC therapy, they will be censored at the randomization date.

3) If a patient did not have a symptomatic progression and did not receive subsequent mCRC therapy, the TTSP time will be censored at the last post baseline ECOG assessment or at the randomization date if the patient does not have any post-baseline ECOG assessment.
4) If a patient had symptomatic progression after two or more missed visits (note that for this scenario two or more missed visits may occur before the assessment of ECOG>2 or before the confirmation of ECOG>2 at the first subsequent assessment at least 8 weeks later), the patient will be censored at the last post baseline ECOG assessment before the missed visits. If the patient has no post-baseline ECOG assessments before the missed visits, they will be censored at the randomization date.
    - Given the scheduled visit scheme of ECOG assessments (i.e. every 2 weeks), the definition of 2 missed visits will equate to 4 weeks since the previous ECOG assessment, or since the randomization date if no previous post-baseline ECOG assessment.

8.6.2.3 Objective Response Rate (ORR) by blinded central image review

ORR will be computed for the two treatment groups as the 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). Tumor response rate, as determined by blinded central image review using RECIST 1.1, 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 for each time point and the best overall response on both the ITT and PP populations.

8.6.2.4 Disease Control Rate (DCR) by blinded central image review

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

8.6.3 Analysis of Quality of Life Questionnaire (FACT-c)

8.6.3.1 Analysis of FACT-c Scores

The total, domain, and individual question scores of the FACT-c 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. The Kenward-Roger approximation will be used to estimate the degrees of freedom. 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, autoregressive with

heterogeneity, Toeplitz, and autoregressive. Means and least squares mean difference between treatment groups, along with a 95% CI and p-value for the difference between treatments will also be provided. This analysis will be performed on the ITT and PP populations.

8.6.3.2 Analysis of Time to Deterioration in QoL (TTDQoL)

A deterioration in QoL is defined as a 7-point decline in the total FACT-c score or death whichever occurs first. The rate of deterioration will be derived from the Kaplan-Meier estimates and 95% CIs will be calculated for each treatment group. Quartiles will be presented and 95% CIs will be calculated on the quartiles for each treatment group. A log-rank (one-sided) test will be used to compare the deterioration rate between the two treatment groups at a 0.025 significance level. Plots of the Kaplan-Meier curves will be provided for each treatment group. This analysis will be performed on the ITT and PP populations.

Time to Deterioration (months) = [(Date of change from baseline in total FACT-c score $\leq$ - 7 or death / censor) – Date of Randomization +1] /30.4375.

The censoring is performed as follows:

- If a patient does not have a baseline total FACT-c score, then the TTDQoL time will be censored at the randomization date, regardless of whether or not TTDQoL has been observed
- If a patient received a subsequent mCRC therapy before a deterioration in QoL or in the absence of a deterioration in QoL, the TTDQoL time will be censored at the last post baseline FACT-c assessment before the start of the subsequent mCRC therapy where total FACT-c score could be evaluated. If the patient has no post baseline total FACT-c scores before the start of the subsequent mCRC therapy, they will be censored at the randomization date.
- If a patient did not have a deterioration in QoL and did not receive subsequent mCRC therapy, the TTDQoL time will be censored at the last post baseline FACT-c assessment where total FACT-c score could be evaluated, or at the randomization date if the patient does not have any post-baseline total FACT-c scores.
- If a patient had a deterioration in QoL after two or more visits where total FACT-c score could not be evaluated (e.g. missed visits or visits missing items to derive total FACT-c score), the patient will be censored at the last post baseline FACT-c assessment before the missed visits where total FACT-c score could be evaluated. If the patient has no post baseline total FACT-c scores before the missed visits, they will be censored at the randomization date. Given the scheduled visit scheme of FACT-c assessments (i.e. every 8 weeks), the definition of 2 missed visits will equate to 16 weeks since the previous FACT-c assessment where total FACT-c score could be evaluated, or since the randomization date if no previous post-baseline total FACT-c scores.

8.6.4 Subgroup Analyses

Subgroup analyses of primary and secondary efficacy endpoints will be performed for the ITT population according to the subgroups listed in Section 7.7.

For each subgroup, a Cox proportional hazards model will be fitted with treatment as the only covariate for each level of the subgroup separately. The HRs and associated 95% CIs will be summarized and presented on a forest plot, along with results of the overall analysis. If an endpoint has <10 events available at a subgroup level then the relationship between that subgroup level and endpoint will not be formally analyzed, since it is unlikely to be a meaningful analysis, and only descriptive summaries will be provided.

8.6.5 Assessment of Poolability

Since this is a multi-center study, analysis will be performed by pooling data across study sites. The clinical study will be conducted under a common protocol for each investigational site, except for sites in Germany, where a separate protocol was used with different eligibility criteria, mainly related to tumor burden at baseline. It is expected that between 5 and 10% of the total number of patients randomized in the study will be from sites in Germany.

In the event that there are small sample sizes at some sites, sites may be grouped using the following procedure to create “analysis-sites” for analysis purposes. These analysis-sites will be created for North America, Europe, and Asia independently to preserve the ability to differentiate between regions. Patients from sites in Germany will not be included in this grouping mechanism. Analysis-sites are based on a target size of at least 5 patients per treatment group at each site. If investigative sites have at least 5 ITT patients per treatment group, they will retain their identities in the analysis. All investigative sites with fewer than 5 ITT patients per treatment group will be rank ordered by size and sorted secondarily by site identification number to break ties. Starting with the smallest investigative site, patients will be combined site by site by treatment group, until the first time the resulting analysis-site has at least 5 ITT subjects in each treatment group. The process continues until all investigative sites are accounted for. If the last analysis-site has fewer than 5 ITT patients per treatment group, it will be combined with the most recently created analysis-site.

To assess the poolability of data across sites a Cox regression analysis of the primary efficacy endpoints, PFS and HPFS, and secondary time-to-event endpoints (i.e. OS, TTSP, and TTDQoL) will be conducted including factors of treatment group, analysis-site, treatment group by analysis-site interaction, and the additional factors listed in Section 7.2 that are included in the multivariable analysis. Estimates of treatment effect and 95% CIs will be calculated separately by analysis-site.

Similarly, to assess the poolability of data across regions a separate Cox regression analysis will be conducted with analysis-site replaced by region. Region and study site will not be included simultaneously in the same model due to collinearity. Also, to

assess the poolability of data across genders, a Cox regression analysis will be conducted with analysis-site replaced by gender. Patients from sites in Germany will not be included in these analyses.

Logistic regression of the binary secondary endpoints of ORR and DCR will be conducted using the factors listed above for the Cox regression.

These analyses will be performed on the ITT population.

To assess the poolability of data from sites in Germany with the data from other sites Cox and logistic regression analyses will also be conducted as described above with analysis-site replaced by German-site (yes/no).

If, in the above analyses of the primary or secondary endpoints, the treatment group by analysis-site interaction, treatment group by gender interaction or treatment group by region interaction is statistically significant at a two-sided level of 0.15, the reasons for the observed differential treatment effect, such as patient demographic or clinical characteristics, will be investigated and reported. If the poolability of results is in direct question as a result of this sensitivity analysis, the endpoint(s) will also be analyzed separately by site, region, and/or gender.

8.6.6 Additional Efficacy Analyses

All the additional analyses will be performed on both the ITT and PP populations.

PFS, HPFS, ORR and DCR by investigator assessment, will be analyzed in the same way as the corresponding analysis of blinded central image review data described above, including censoring for subsequent mCRC treatment.

Duration of response and duration of disease control by both blinded central image review and investigator assessment will be summarized using descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) by treatment group. Also, Kaplan-Meier analyses of duration of objective response and duration of disease control will be conducted. A log rank-test will be performed to compare the durations of response between treatment groups.

Duration of response (months) = (Date of PD/death – Date of first overall response of CR or PR +1) /30.4375.

Duration of disease control rate (months) = (Date of PD/death – Date of first overall response of CR, PR or SD +1) /30.4375. The censoring for duration of response and duration of disease control will be performed as follows:

1) If a patient received a subsequent mCRC therapy before PD or death, or in the absence of PD or death, the duration of response and duration of disease control will be censored at the last valid post baseline radiological tumor assessment before the start of the subsequent mCRC therapy.
2) If a patient is known not to have died or have PD and did not receive subsequent mCRC therapy, the duration of response and duration of disease control will be censored at the last valid post baseline radiological tumor assessment date.
3) If a patient had PD or died after two or more missed visits, the duration of response and duration of disease control will be censored at the time of the last valid post baseline radiological tumor assessment date that occurred before the missed visits. Given the scheduled visit scheme of tumor assessments (i.e. every 8 weeks), the definition of 2 missed visits will equate to 16 weeks since the previous tumor assessment.

DoR by both blinded central image review and investigator assessment will be displayed using descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) by treatment group. Subgroup analysis will also be presented for patients with baseline tumor replacement by blinded central review at >5%, >10%, and >15%. A 2-sample t-test will be performed to compare the mean DoR between treatment groups, and the corresponding 95% CI for the mean difference between the two treatment groups will be calculated.

The observed PTTS rate by both blinded central image review and investigator assessment will be summarized by treatment group. The 95% CIs for the PTTS rate for each of the treatment groups will be computed according to Wilson (1927). The

PTTS rates will be compared between treatment groups using the continuity adjusted Newcombe-Wilson test, and the corresponding 95% CI for the difference in PTTS rates between the two treatment groups will be calculated.

The percentage and absolute change from baseline in the sum of the longest diameters of target lesions will be summarized by blinded central image review and by investigator assessment separately at Week 8, 16, and 24 analysis visits (as defined in Table 3). A waterfall plot of the percentage change in the sum of longest diameters at the time of the best overall response will be presented by treatment group. For blinded central image review, data from the reviewer selected by the adjudicator will be used when adjudication has occurred, but when adjudication was not required the average value of the sum of the longest diameters of target lesions from the two reviewers at each visit will be used to calculate percentage and absolute change from baseline.

The tumor marker for CRC (CEA) will be summarized with descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for each time-point and change from baseline by treatment group.

8.7 Analysis of Safety

All safety analyses will be performed on the Safety Population.

8.7.1 Extent of Exposure to Study Treatment

Analyses of the extent of exposure to study treatment will be performed on the ITT, PP and Safety populations.

8.7.1.1 Extent of Exposure to TheraSphere

The extent of patient exposure to TheraSphere as defined in Section 6.5.1.1 will be summarized using descriptive statistics (as appropriate, including number of patients, mean, median, standard deviation, minimum and maximum, or counts and percentages).

Summaries above will also be provided according to the subgroups listed in Section 7.7 for the ITT population, except dose delivered to lungs which will be summarized according to the subgroups listed in Section 7.7 for the Safety population.

8.7.1.2 Second-Line Chemotherapy Agents

The extent of patient exposure to second-line chemotherapy agents as well as biological agents will be summarized by treatment group using descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for the following parameters:

- Number of cycles
- Average dose per cycle (mg/m^2 of BSA or mg/kg)
- Cumulative dose (mg/m^2 of BSA or mg/kg)
- Duration on treatment (weeks)

Duration of each chemotherapy agent in weeks will be calculated as:

Duration on Treatment (weeks) = (end date of last dose of chemotherapy agent – start date of first dose of chemotherapy agent +1)/7

The number and percentage of patients receiving a second line chemotherapy agent, receiving a biological agent, and intended at baseline to be treated with a biological agent, will be summarized by treatment group.

The above information will also be summarized according to the subgroups listed in Section 7.7 for the ITT population.

8.7.1.3 Extent of Study Exposure and Follow-up

The duration of study and duration of follow-up will be summarized by treatment using descriptive statistics (number of patients, mean, standard deviation, median, minimum, maximum), and will be calculated as follows.

Duration on study (months) = (earlier of study exit date and death date – randomization date + 1) / 30.4375

Duration of follow-up is defined as the time from stop of study treatment until date of death due to any cause. Any patient not known to have died at the time of analysis will be censored based on last recorded date on which the patient was known to be alive, as defined for OS in Section 8.6.2.1.

Duration on follow-up (months) = (Date of death/last known alive – Stop Date of study treatment + 1) / 30.4375.

The above information will also be summarized according to the subgroups listed in Section 7.7 for the ITT population.

In addition, an alternative method for determining duration of follow-up will be performed using the reverse Kaplan-Meier method for the ITT population. The censored values for OS will be reversed so that 1s will be 0s and 0s will be 1s. The median follow-up time will be derived from the Kaplan-Meier method using the OS values of overall survival and the reversed censoring values.

8.7.1.4 Best Available Care Post-Progression

The number and percentage of patients who received each post-progression treatment (systemic mCRC treatments will be presented by preferred terms) will be summarized by treatment group. The summaries will also be provided according to the subgroups listed in Section 7.7 for the ITT population.

8.7.2 Adverse Events

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

Adverse events will be summarized by system organ class and preferred term; a patient will only be counted once per system organ class and once per preferred term within a

treatment. Patient counts and percentages and event counts will be presented for each treatment group for the following summaries:

1. Overall summary of TEAEs
2. All TEAEs (also presented by preferred term only in descending frequency).
3. All TEAEs of CTCAE grade ≥3 (also presented by preferred term only in descending frequency).
4. All TEAEs considered related to chemotherapy.
5. All TEAEs considered related to device (ADE).
6. All CTCAE grade ≥3 TEAEs potentially related to device.
7. All TEAEs considered related to angiographic procedure.
8. All TEAEs with outcome of fatal.
9. All treatment emergent serious adverse events (SAE) (also presented by preferred term only in descending frequency).
10. All treatment emergent serious adverse device events (SADE)
11. All TEAEs leading to chemotherapy discontinuation.

All summaries of TEAEs will be further broken into the following 2 groups:

- TEAEs with a start date up until disease progression or 30 days after discontinuation of study therapy, whichever comes first
- TEAEs with a start date after this

In addition, these AE summaries will be produced for the subgroups listed for AEs in Section 7.7.

For the summary of AEs by CTCAE grade, if a patient has multiple events occurring in the same body system or same preferred term, then the event with the highest CTCAE grade will be counted. For AEs by relationship to study drug, if a patient has multiple events occurring in the same body system or same preferred term, the event with the highest association to study drug will be summarized (unknown is considered a higher association to study drug than not related, but less of a relationship than possibly, probably, and definitely). AEs potentially related to chemotherapy, device and angiographic procedure are defined as a subset of AEs with a relationship of either possibly, probably, definitely, or unknown. No statistical inference between the treatments will be performed on AEs.

Listings will be presented by patient for all AEs as well as for serious AEs including SADE, AEs associated with death, and AEs leading to discontinuation of chemotherapy.

8.7.3 Clinical Laboratory Evaluations

Clinical laboratory results will be converted to SI units. Change from baseline to each visit assessed and end of study will be defined using the windowing method specified

in Section 6.2.2, as the visit value minus the baseline visit. 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). Hematology, chemistry, and coagulation results will each be summarized separately.

All clinical laboratory data will be presented in listings. Within each listing, laboratory values outside the normal ranges will be flagged as either high or low. In addition, shift tables will be presented to display the shift in the normal range categories (low, normal, high) from baseline to the final evaluation. Baseline is defined as the latest non-missing value prior to randomization.

Shift table of baseline to each assessment by CTC grade and a table of laboratory parameters of CTCAE Grade 3 or higher that worsened from baseline will be summarized.

8.8 Additional Safety Analyses

A shift table comparing the baseline ECOG score to the ECOG score at each time-point will be summarized for the ITT population. This will be used to ascertain the number of patients with an ECOG score that worsens after baseline and any difference between the treatment groups.

9 COMPUTER SOFTWARE

All analyses will be performed by Covance using Version 9.1.3 or later of SAS® software. All summary tables and data listings will be prepared utilizing SAS® software.

For continuous variables, descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) will be generated. For discrete/categorical variables, the number and proportion of patients will be generated. The standard operating procedures of Covance will be followed in the creation and quality control of all data displays and analyses.

10 REFERENCES

Belsley DA, Kuh E, Welsch RE. Regression Diagnostics. New York: John Wiley & Sons (1980).

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.

Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-2.

Mulcahy MF, Lewandowski RJ, Ibrahim SM, Sato KT, Ryu RK, Atassi B, Newman S, Talamonti M, Omary RA, Benson A, Salem R. Radioembolization of colorectal hepatic metastases using yttrium-90 microsperes. Cancer 2009;115:1849-58.

Sasson AR, Sigurdson ER. Surgical treatment of liver metastases. Semin Oncol 2002;29:107-18.

Thilakarathne P, Palumbo A, Diels J, Delforge M, van Sanden S, Mateos MV, Ch irita O, Dimopoulos MA, van de Velde H, San Miguel JF. Inverse Probability of Censoring Weighted Analysis to Adjust the Treatment Effect on Overall Survival for Subsequent Therapy: A Case Study in a Clinical Trial in Multiple Myeloma. Value in Health 2014; 17: 546.

Wagner JS, Adson MA, Van Heerden JA, Adison MH, Ilstrup DM. The natural history of hepatic metastases from colorectal cancer. A comparison with resective treatment. Ann Surg 1984;199:502-8.

Wilson, EB. Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association 1927;22:209–12.

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=(Randomization Date-DOB)/365.25]. Only the integer portion will be used.

Weight will be displayed in kilograms, height will be displayed in centimeters, and temperature will be displayed in Celsius. 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 outsurv=interval;
        time aval*censr(1);
        strata trtp;
        id usubjid;
run;
```

The SAS procedure PHREG will be used in the Cox regression analysis of PFS and survival. Patients who did not have an event will be censored. The SAS method of discrete will be used to handle ties.

The following code will be used for time to event analysis by treatment group HRs:

```
proc phreg data=all;
        model aval*censr(1) = trtp  /ties=discrete rl;
run;
```

The following code will be used for each univariable and multivariable analysis HRs and p-values:

```
proc phreg data=all;
        model aval*censr(1) = &covar.  /ties=discrete rl;
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 BASE is the baseline score, TRT is the assigned treatment, VISIT is the visit based on the window mapping, CH is the change from baseline.

11.3 APPENDIX 3: FACT-C Questionnaire Scoring Rules

FACT-C 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-G & FACT-C).
5. **The higher the score, the better the QOL.**

Subscale		Item Code	Reverse item?		Item response	Item Score
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	+	_____	=_____
		GS5	0	+	_____	=_____
		GS6	0	+	_____	=_____
Score range:	0-28	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____
=**SWB subscale score**

EMOTIONAL WELL-BEING (EWB)	GE1	4	-	_____	=_____
	GE2	0	+	_____	=_____
	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 WELL-BEING (FWB)		GF1	0	+	_____	=_____
		GF2	0	+	_____	=_____
		GF3	0	+	_____	=_____
		GF4	0	+	_____	=_____
		GF5	0	+	_____	=_____
		GF6	0	+	_____	=_____
Score range:	0-28	GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____

Divide by number of items answered: _____
=**FWB subscale score**

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
COLORECTAL	C1	4	-	_____	=_____
CANCER	C2	4	-	_____	=_____
SUBSCALE	C3	0	+	_____	=_____
(CCS)	C4	0	+	_____	=_____
Score range: 0-28	C5	4	-	_____	=_____
	C6	0	+	_____	=_____
C7		0	+	_____	=_____
	C8	NOT CURRENTLY SCORED			
	C9	NOT CURRENTLY SCORED			

Sum individual item scores:_____
Multiply by 7: _____

Divide by number of items answered: _____
=CC Subscale score

To derive a FACT-C Trial Outcome Index (TOI):
Score range: 0-84

_____ + _____ + _____ =_____=**FACT-C TOI**
(**PWB** score) (**FWB** score) (**CCS** score)

To Derive a FACT-C total score:

Score range: 0-136

_____ + _____ + _____ + _____ + _____ =_____=**FACT-C Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score) (**CCS** 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.

11.4 APPENDIX 4: Definition of subsequent mCRC therapy

A patient is considered to have received 'subsequent mCRC therapy', after the protocol required treatments, if they had

- a subsequent systemic anti-cancer regimen, or
- local treatment with Y90 (including TheraSphere), TACE, ablation or resection of liver lesions.

A subsequent systemic anti-cancer regimen is defined as

- Anti-cancer agents that were not previously administered during the patient treatment pathway, or
- Anti-cancer agents previously used for first line treatment

Notes:

1) For patients randomized to receive TheraSphere, a subsequent TheraSphere treatment administered before the next imaging assessment and <56 days after the first TheraSphere treatment will not be considered to be a subsequent mCRC therapy. This is justifiable because a subsequent TheraSphere treatment in this timeframe is likely to be a continuation of the initial TheraSphere treatment and not due to disease progression.
2) De-escalation of anti-cancer treatment is not considered as a subsequent anti-cancer regimen. Therefore the following would not be considered as subsequent anti-cancer regimens:
    a. A doublet chemotherapy followed by a switch to 5FU or 5FU/LV or capecitabine or UFT
    b. A doublet chemotherapy plus biological followed by a switch to the same biological only
    c. A doublet chemotherapy plus biological followed by a switch to 5FU or 5FU/LV or capecitabine or UFT + same biological
    d. A switch from 5FU or 5FU/LV to capecitabine or UFT within the same regimen.

11.5 APPENDIX 5: Programmatic identification of patients who received adjuvant chemotherapy

Statistical analyses for the EPOCH study require the identification of patients who received adjuvant chemotherapy. However, the eCRF does not have explicit data fields to collect this data. Below is a description of how data from the Medical History mCRC Cancer and Prior Therapy eCRF page will be used to programmatically identify patients who received adjuvant chemotherapy, and how the start date of first line chemotherapy will be determined for such patients. This is based on usual practice in which

- Adjuvant chemotherapy is administered up to 12 weeks after surgery of the primary tumor. Adjuvant chemotherapy is not recommended to be administered >12 weeks after surgery because of the decreased likelihood of prevention of disease recurrence
- Adjuvant chemotherapy includes at least one of the following drugs: 5-FU, capecitabine, oxaliplatin or irinotecan

Also, in clinical trials comprising of patients who are due to receive first line chemotherapy, adjuvant chemotherapy would usually be completed <6 months before a patient can be included in the trial.

After the algorithm described below has been used to programmatically identify patients who received adjuvant chemotherapy, and to identify the start date of first line chemotherapy, a manual review of the data fields used in the algorithm, for all randomized patients, will be performed by the sponsor's project physician. This manual review is to ensure that the algorithm has been correctly defined and to identify any further data issues that require data queries.

Algorithm to identify patients who received adjuvant chemotherapy

- Patients with the primary tumor in-situ at baseline are considered not to have received adjuvant chemotherapy.
- For patients with the primary tumor not in-situ at baseline:
  - Patients whose stage at initial diagnosis of CRC is IV will be considered not to have received adjuvant chemotherapy.
  - For patients whose stage at initial diagnosis of CRC is I, II or III (or the stage is confirmed after querying to be unknown):
    - If the duration between diagnosis of CRC and start of <u>all</u> systemic chemotherapy agents is >84 days (i.e. 12 weeks) the patient is considered not to have received adjuvant chemotherapy.

- For patients who received at least one systemic chemotherapy agent with duration between diagnosis of CRC and start of that chemotherapy agent ≤84 days (i.e. 12 weeks):
  - If the duration between the end date of any of 5-FU, capecitabine, oxaliplatin or irinotecan and the date of progression on first line chemotherapy is <183 days (i.e. 6 months), then the patient is considered not to have received adjuvant chemotherapy, and the chemotherapy agents entered in the eCRF are considered to be first line chemotherapy.
  - If the duration between the end date of any of 5-FU, capecitabine, oxaliplatin or irinotecan and the date of progression on first line chemotherapy is ≥183 days (i.e. 6 months), then the patient is considered to have received adjuvant chemotherapy, with the end date of adjuvant chemotherapy taken to be the latest end date of any of the agents 5-FU, capecitabine, oxaliplatin and irinotecan. The start date of first line chemotherapy for this patient is then taken to be the earliest start date of the oxaliplatin or irinotecan that the patient received after the end date of adjuvant chemotherapy (as defined above) and before randomization.
    - However, if the patient did not receive oxaliplatin or irinotecan with a start date after the end date of adjuvant chemotherapy (as defined above) and before randomization then the chemotherapy data (from the Medical History mCRC Cancer and Prior Therapy eCRF page) for the patient should be manually reviewed and queried, if necessary, to either correct the stage at initial diagnosis of CRC to stage IV, or to correct dates of chemotherapy and/or to add records for missing chemotherapy agents.

Examples to illustrate the above algorithm are given below.

Example Patient 1

Randomized on 04Dec2013; primary tumor not in-situ at baseline; diagnosis of CRC on 18Mar2013; stage III at initial diagnosis of CRC; progression on first line chemotherapy on 13Nov2013.

Chemotherapy agent	Start date	End date	Duration between diagnosis of CRC and start of chemotherapy agent[a]
Capecitabine	02Oct2013	05Nov2013	199 days
Oxaliplatin	02Oct2013	23Oct2013	199 days

[a] Start date of chemotherapy – date of diagnosis of CRC +1

This patient is considered to have not received adjuvant chemotherapy because the duration between diagnosis of CRC and start of all chemotherapy agents is >84 days.

Example Patient 2

Randomized on 07Oct2013; primary tumor not in-situ at baseline; diagnosis of CRC on 27Mar2012; stage III at initial diagnosis of CRC; progression on first line chemotherapy on 19Aug2013.

Chemotherapy agent	Start date	End date	Duration between diagnosis of CRC and start of chemotherapy agent[a]	Duration between end date of chemotherapy agent and date of progression on first line chemotherapy[b]
5-FU	07May2012	11Jun2012	42 days	435 days
Oxaliplatin	03Oct2012	07Jan2013	191 days	225 days
Leucovorin	03Oct2012	07Jan2013	191 days	225 days
5-FU	03Oct2012	07Jan2013	191 days	225 days

[a] Start date of chemotherapy – date of diagnosis of CRC +1

[b] Date of progression on first line chemotherapy – end date of chemotherapy

This patient, who had stage III disease at initial diagnosis, is considered to have received adjuvant chemotherapy because the duration between diagnosis of CRC and start of at least one chemotherapy agent (5-FU) is ≤84 days, and the duration between the end date of this agent and the date of progression on first line chemotherapy is

≥183 days. The end date of adjuvant chemotherapy is taken to be 11Jun2012, and the start date of first line chemotherapy is taken to be 03Oct2012.

Example Patient 3

Randomized on 07Mar2018; primary tumor not in-situ at baseline; diagnosis of CRC on 03Dec2015; stage III at initial diagnosis of CRC; progression on first line chemotherapy on 19Jan2018.

Chemotherapy agent	Start date	End date	Duration between diagnosis of CRC and start of chemotherapy agent[a]	Duration between end date of chemotherapy agent and date of progression on first line chemotherapy[b]
Oxaliplatin	24Feb2016	17Aug2016	84 days	521 days
5-FU	24Feb2016	17Aug2016	84 days	521 days

[a] Start date of chemotherapy – date of diagnosis of CRC +1
[b] Date of progression on first line chemotherapy – end date of chemotherapy +1

This patient, who had stage III disease at initial diagnosis, is considered to have received adjuvant chemotherapy because the duration between diagnosis of CRC and start of at least one chemotherapy agent is ≤84 days, and the duration between the end date of 5-FU (or oxaliplatin) and the date of progression on first line chemotherapy is ≥183 days. The end date of adjuvant chemotherapy is taken to be 17Aug2016. However, since there are no chemotherapy records on the Medical History mCRC Cancer and Prior Therapy eCRF page after this date the start date of first line chemotherapy cannot be determined, and so the data for this patient should be queried to ensure the stage III at initial diagnosis of CRC is correct and if so, the first line chemotherapy records should be entered in the eCRF.

11.6 APPENDIX 6: Statistical Details of the Adaptive Design for Protocol TS-102 EPOCH

CONSULTING REPORT REGARDING

A Phase III Clinical Trial Evaluating TheraSphere® in Patients with Metastatic Colorectal Carcinoma of the Liver who have Failed First Line Chemotherapy

12 October 2020

PREPARED FOR:
Nikhil Chauhan
Director, Clinical Operations
E: Nikhil.Chauhan@bsci.com
Boston Scientific

PREPARED BY:
James Matcham
Senior Principal, Cytel Consulting
T: +44 7741 243935
E: james.matcham@cytel.com

Cytel Inc.
675 Massachusetts Ave
Cambridge, MA 02139

1. Introduction

The purpose of this report is to describe the method to test the two primary endpoints (PFS and HPFS) while preserving the type I error rate of reaching a false positive claim on either of the two primary endpoints. Section 2 describes the method in detail. Section 3 discusses the strong control of Type I error for the proposed method. Section 4 presents the operating characteristics of the proposed method including power and Type I error rate.

2. Method to Test PFS and HPFS

This study uses a group sequential design with two interim analyses and one final analysis with PFS and HPFS as the two primary endpoints. The study could be stopped early for efficacy at the first or second interim analysis based on superiority in PFS but not HPFS. The study is designed to detect a 2.5 month increase in median PFS time, from 6 months in the control arm to 8.5 months in the TheraSphere arm (ie, hazard ratio HR = 0.71), and a 3.5 month increase in median HPFS time, from 6.5 months in the control arm to 10 months in the TheraSphere arm (ie, HR = 0.65), using log rank tests.

The analysis of PFS will be based on a group sequential design with 2 interim analyses and rho family error spending function stopping boundary with rho=1.5. It is estimated that approximately 420 patients will need to be recruited over 36 months, with a 1 year additional follow-up period, allowing for 10% of patients lost to follow-up and for whom a date of progression or death is not recorded. A total of 428 patients were actually randomized in this study. At the first and second interim analyses, HPFS will be tested only if the treatment benefit with PFS has been successfully demonstrated using the efficacy boundaries derived based on the rho family error spending function with rho=1.5. If the trial doesn't stop at the first or second interim analyses, the Hochberg procedure (Hochberg, 1988) will be used to control Type I error for the two primary endpoints at the final analysis.

First Interim Analysis:

The first interim analysis was planned at 172 PFS events. PFS will be compared between treatment arms using a log rank test converted to a z-score and compared to the nominal critical value of 2.372 based on the rho family error spending function corresponding to a one-sided p-value ≤0.0088 allowing the study to be stopped early for efficacy, in which case HPFS will be tested at the same boundary as PFS.

Updated boundary at IA1:

The first interim analysis occurred at 204 PFS events. The nominal critical value at the actual IA is 2.2762 which corresponds to one-sided p value boundary 0.011417.

Second Interim Analysis:

A second interim analysis was planned at 241 PFS events, where PFS will be compared between treatment arms using a log rank test converted to a z-score and compared to the nominal critical value of 2.330 based on the rho family error spending function corresponding to a one-sided p-value ≤0.0099 allowing the study to be stopped early for efficacy, in which case HPFS will be tested using the boundary derived based on an incremental alpha of 0.0057 for the second interim analysis. This boundary will account for the correlation between the z-score at the first interim analysis and the z-score at the second interim analysis which is determined by the observed

number of HPFS events at the first interim analysis and the cumulative number of HPFS events observed at the second interim analysis.

Updated boundary at IA2:

The second interim analysis occurred at 287 events. The nominal critical boundary is 2.2218 which corresponds to the p value boundary 0.0131.

Final Analysis:

The final analysis was originally planned at 344 PFS events. However, the Sponsor subsequently became aware that it will not be possible to reach 344 PFS events due to a higher than expected number of patients who withdrew early from the study. Therefore, the final analysis will now be planned at 330 PFS events. However, if 330 PFS events have not been reached at 31 August 2020 then the final analysis will be performed with the number of events that have occurred at that time. The Hochberg procedure (Hochberg, 1988) will be used to control Type I error for the two primary endpoints. Whichever of PFS or HPFS that has the larger p-value, will be compared between the treatment arms using a log rank test converted to a z-score and compared to the nominal critical value of 2.3126. A corresponding one-sided p-value ≤0.0104 is required to declare a statistically significant improvement for this endpoint at the final analysis. To ensure that Type I error is controlled for both primary endpoints, this boundary is based on the incremental alpha of 0.0104 instead of the p-value scale boundary of 0.0168 using the rho family error spending function with rho=1.5.

According to the Hochberg procedure, if the primary endpoint with the larger p-value is statistically significant then the other primary endpoint is also statistically significant. However, if the primary endpoint with the larger p-value is not statistically significant then the other primary endpoint will be compared between treatment arms using a log rank test converted to a z-score and compared to the nominal critical value of 2.562 based on the rho family error spending function, with a corresponding one-sided p-value ≤0.0104/2 = 0.0052 required to declare a statistically significant improvement in hazard rate for this endpoint.

The boundary at the interim and final analyses will be updated based on the actual interim analysis time using the rho family error spending function with rho=1.5 based on the actual number of PFS events at the first and second interim analyses.

Updated boundary at final analysis:

The boundary for the final analysis will be determined based on the following methodology which will account for the different censoring rules used in the two interim analyses and the final analysis. The first interim analysis occurred at 204 PFS events, and PFS and HPFS were analyzed using version 2.0 of the SAP where patients without progression or death were censored at the last valid tumor assessment (Censoring Method A: Original). The second interim analysis occurred at 287 PFS events, and PFS and HPFS were analyzed using version 4.0 of the SAP where patients who received subsequent mCRC therapy prior to their last valid tumor assessment or progression or death were censored at their last valid tumor assessment prior to the start of the subsequent mCRC therapy (Censoring Method B: Original + Subsequent mCRC therapy). In the final analysis, additionally, if the patient progresses or dies immediately after 2 or more missed visits, the patient will be censored at the last valid tumor assessment prior to the 2 missed visits (Censoring Method C: Original + Subsequent mCRC therapy + 2 missed visits). Under each

censoring method, the alpha level for the final analysis will be derived based on the rho family spending function with rho=1.5, and the number of PFS events recalculated for the interim and final analyses, had the same censoring method been applied in all the analyses. The most conservative boundary from the three censoring methods will be used for the final analysis. This methodology is illustrated in the Table A.

Table A: Calculation of efficacy boundary in final analysis

Censoring Method	First interim analysis PFS Events	Second interim analysis PFS Events	Final analysis PFS Events	Alpha Level for Final analysis
A	204	$E2_A$	EF_A	α_{fA}
B	$E1_B$	287	EF_B	α_{fB}
C	$E1_C$	$E2_C$	EF_C	α_{fC}

Notes:

- $E1_B$ and $E1_C$ will be calculated based on the first interim analysis data snapshot.
- $E2_A$ and $E2_C$ will be calculated based on second interim analysis data snapshot.
- EF_A, EF_B and EF_C will be calculated based on final data with a data cut-off of 31 August 2020.
- α_{fA} will be derived using information fractions of two interim analyses under censoring method A, (204 / EF_A) and ($E2_A$ / EF_A), respectively.
- α_{fB} will be derived using information fractions of two interim analyses under censoring method B, ($E1_B$ / EF_B) and (287 / EF_B), respectively.
- α_{fC} will be derived using information fractions of two interim analyses under censoring method C, ($E1_C$ / EF_C) and ($E2_C$ / EF_C), respectively.
- Alpha level for final analysis: $\alpha_f = \min(\alpha_{fA}, \alpha_{fB}, \alpha_{fC})$.

The calculation described in Table A for the alpha level for the final analysis, α_f, will be carried out prior to the database hard lock so that α_f will be fixed and documented prior to the database hard lock for the final analysis. The larger p-value between PFS and HPFS at the final analysis will be compared to α_f and the smaller p-value will be compared to α_f /2 based on the Hochberg procedure. If the larger p-value is ≤ α_f, significance can be claimed for both PFS and HPFS. On the other hand, if the larger p-value is > α_f and the smaller p-value is < α_f /2, then only the endpoint with the smaller p-value can be claimed to be significant.

3. Strong Control of Type I Error

To ensure the strong control of Type I error rate for testing the two primary endpoints PFS and HPFS, the Hochberg procedure will be applied to the final analysis using the incremental alpha 0.0104 instead of the p-value scale boundary 0.0168.

To show Type I error control, let $\boldsymbol{H_0^{PFS}}$ and $\boldsymbol{H_0^{HPFS}}$ denote the null hypotheses that there is no difference in PFS and HPFS respectively between the two treatment arms. Let $\boldsymbol{p_1^{PFS}, p_2^{PFS}, p_3^{PFS}}$ denote p-values for PFS at the first and the second interim analyses (IA1 and IA2) and the final analysis (FA), respectively. Let $\boldsymbol{c_1, c_2}$ and $\boldsymbol{c_3}$ be the p-value scale efficacy stopping boundaries at the interim and final analyses time based on the rho family spending function with rho=1.5 based on a nominal one-sided $\boldsymbol{\alpha = 0.025}$. With the planned interim analyses at 50% and 70% information fraction, we have $\boldsymbol{c_1 = 0.0089, c_2 = 0.0099}$ and $\boldsymbol{c_3 = 0.0168}$. The incremental alpha spent at IA1, IA2 and FA is given by $\boldsymbol{\alpha_1 = 0.0089, \alpha_2 = 0.0057}$ and $\boldsymbol{\alpha_3 = 0.0104}$. Let

$p_1^{HPFS}, p_2^{HPFS}, p_3^{HPFS}$ denote p-values for HPFS at the first and the second interim analyses (IA1 and IA2) and the final analysis (FA). Consider the following three scenarios under which a Type I error could potentially occur:

(1) TheraSphere improves HPFS but not PFS,
(2) TheraSphere improves PFS but not HPFS,
(3) TheraSphere improves neither PFS nor HPFS.

Under Scenario 1, Type error is given by

$$P\left(p_1^{PFS} < c_1 \text{ or } p_2^{PFS} < c_2\right)$$
$$+ P\left(p_1^{PFS} \geq c_1, p_2^{PFS} \geq c_2, \text{reject } H_0^{PFS} \text{ after Hochberg adjustment using } \alpha_3\right)$$

$$\leq P\left(p_1^{PFS} < c_1 \text{ or } p_2^{PFS} < c_2\right) + P\left(p_1^{PFS} \geq c_1, p_2^{PFS} \geq c_2, p_3^{PFS} < \alpha_3\right)$$

$$\leq P\left(p_1^{PFS} < c_1 \text{ or } p_2^{PFS} < c_2\right) + P\left(p_3^{PFS} < \alpha_3\right)$$

$$\leq \alpha_1 + \alpha_2 + \alpha_3 = \alpha$$

Under Scenario 2, a Type I error is committed if: (1) H_0^{HPFS} is rejected at IA1 or IA2, (2) the trial continues to FA and rejects H_0^{HPFS} at FA.

Note that the probability of (1) is bounded above by $P\left(p_1^{HPFS} < c_1 \text{ or } p_2^{HPFS} < c_2'\right)$. The probability of (2) is given by

$$P\left(p_1^{PFS} \geq c_1, p_2^{PFS} \geq c_2, \text{reject } H_0^{HPFS} \text{ at FA after Hochberg adjustment using } \alpha_3\right)$$

$$\leq P\left(p_1^{PFS} \geq c_1, p_2^{PFS} \geq c_2, p_3^{HPFS} < \alpha_3\right) \leq P\left(p_3^{HPFS} < \alpha_3\right)$$

Therefore Type I error under Scenario 2 is bounded above by

$$P\left(p_1^{HPFS} < c_1 \text{ or } p_2^{HPFS} < c_2'\right) + P\left(p_3^{HPFS} < \alpha_3\right) \leq \alpha_1 + \alpha_2 + \alpha_3 = \alpha$$

Where c_2' is the boundary for testing HPFS at the second interim analysis which is derived such that the incremental alpha for HPFS at the second interim analysis is equal to α_2, i.e. $P\left(p_1^{HPFS} > c_1, p_2^{HPFS} < c_2'\right) = \alpha_2$.

Under Scenario 3, a Type I error is committed if: (1) H_0^{PFS} or H_0^{HPFS} is rejected at IA1, (2) the trial continues to IA2, and rejects H_0^{PFS} or H_0^{HPFS} at IA2, (3) the trial continues to IA3, and rejects H_0^{PFS} or H_0^{HPFS} at IA3.

Note that if H_0^{HPFS} is rejected at IA1 or IA2, then H_0^{PFS} must also be rejected since H_0^{HPFS} is tested only if H_0^{PFS} crossed the boundaries at IA1 or IA2. Therefore the overall type I error is given by

$$P\left(p_1^{PFS} < c_1\right) + P\left(p_1^{PFS} \geq c_1, p_2^{PFS} < c_2\right)$$

$$+P\left(p_1^{PFS} \geq c_1, p_2^{PFS} \geq c_2, p_3^{PFS} \text{ or } p_3^{HPFS} \text{is significant at FA after Hochberg adjustment using } \alpha_3\right)$$

$$\leq P\left(p_1^{PFS} < c_1\right) + P\left(p_1^{PFS} \geq c_1, p_2^{PFS} < c_2\right)$$

$$+P\left(p_3^{PFS} \text{ or } p_3^{HPFS} \text{is significant at FA after Hochberg adjustment using } \alpha_3\right)$$

$$\leq \alpha_1 + \alpha_2 + \alpha_3 = \alpha$$

4. Simulations

This section presents the operating characteristics of the primary analysis plan with the two endpoints PFS and HPFS. The simulations were performed in R 3.1.2. in two steps: (1) generate correlated bivariate normal variables, (2) then transform the normal variables into two correlated time-to-event endpoints. The additional simulation parameters are as follows:

- Total # of patients 428 enrolled in 36 months with 10% of patients lost to follow-up
- Total number of events: 270, 280, 300, 310, 315, 320, 330
- IA1: at 172 PFS events (50% information)
- IA2: at 242 PFS events (70% information)
- Rho family spending function with rho=1.5
- Median PFS improvement from 6 months to 8.5 months
- Median HPFS improvement from 6.5 months to 10 months
- Interim Decision:
  - Test PFS and HPFS sequentially at IA1 and IA2
- Apply Hochberg procedure for final analysis: compare the larger p value between PFS and HPFS to 0.0104 and compare the smaller p-value to half of 0.0104

Table 1: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 270 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46104	0.33352	0.46104	0.33352
IA2	0.20789	0.17227	0.20789	0.17227
FA	0.24982	0.05871	0.06443	0.2441
Total	0.91875	0.5645	0.73336	0.74989

Table 2: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 270 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46108	0.35905	0.46108	0.35905
IA2	0.20769	0.17431	0.20769	0.17431
FA	0.23023	0.05744	0.06346	0.22421
Total	0.899	0.5908	0.73223	0.75757

Table 3: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 270 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46394	0.40995	0.46394	0.40995
IA2	0.20635	0.18611	0.20635	0.18611
FA	0.19215	0.05934	0.06441	0.18708
Total	0.86244	0.6554	0.7347	0.78314

Table 4: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 280 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45932	0.33197	0.45932	0.33197
IA2	0.20907	0.17338	0.20907	0.17338
FA	0.25985	0.07347	0.08069	0.25263
Total	0.92824	0.57882	0.74908	0.75798

Table 5: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 280 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45937	0.3594	0.45937	0.3594
IA2	0.20939	0.17673	0.20939	0.17673
FA	0.24107	0.0729	0.07955	0.23442
Total	0.90983	0.60903	0.74831	0.77055

Table 6: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 280 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.4582	0.40462	0.4582	0.40462
IA2	0.21029	0.18963	0.21029	0.18963
FA	0.20654	0.07407	0.08004	0.20057
Total	0.87503	0.66832	0.74853	0.79482

Table 7: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 300 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45759	0.32962	0.45759	0.32962
IA2	0.21041	0.17347	0.21041	0.17347
FA	0.27604	0.10312	0.11152	0.26764
Total	0.94404	0.60621	0.77952	0.77073

Table 8: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 300 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46119	0.35729	0.46119	0.35729
IA2	0.20914	0.1757	0.20914	0.1757
FA	0.25866	0.10178	0.10994	0.2505
Total	0.92899	0.63477	0.78027	0.78349

Table 9: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 300 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46012	0.40635	0.46012	0.40635
IA2	0.20811	0.18701	0.20811	0.18701
FA	0.22819	0.10362	0.10999	0.22182
Total	0.89642	0.69698	0.77822	0.81518

Table 10: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 310 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.4621	0.33363	0.4621	0.33363
IA2	0.20826	0.17261	0.20826	0.17261
FA	0.28171	0.11702	0.12559	0.27314
Total	0.95207	0.62326	0.79595	0.77938

Table 11: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 310 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46299	0.36042	0.46299	0.36042
IA2	0.2063	0.17277	0.2063	0.17277
FA	0.26771	0.11763	0.12633	0.25901
Total	0.937	0.65082	0.79562	0.7922

Table 12: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 310 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45969	0.40662	0.45969	0.40662
IA2	0.20842	0.18751	0.20842	0.18751
FA	0.24055	0.12031	0.12574	0.23512
Total	0.90866	0.71444	0.79385	0.82925

Table 13: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 315 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46003	0.33136	0.46003	0.33136
IA2	0.209	0.17337	0.209	0.17337
FA	0.28663	0.12415	0.13335	0.27743
Total	0.95566	0.62888	0.80238	0.78216

Table 14: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 315 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45997	0.35797	0.45997	0.35797
IA2	0.20904	0.17555	0.20904	0.17555
FA	0.27054	0.12198	0.13035	0.26217
Total	0.93955	0.6555	0.79936	0.79569

Table 15: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 315 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45972	0.40802	0.45972	0.40802
IA2	0.20868	0.18794	0.20868	0.18794
FA	0.2447	0.12567	0.13162	0.23875
Total	0.9131	0.72163	0.80002	0.83471

Table 16: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 320 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46069	0.33151	0.46069	0.33151
IA2	0.20905	0.17323	0.20905	0.17323
FA	0.28764	0.12889	0.13848	0.27805
Total	0.95738	0.63363	0.80822	0.78279

Table 17: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 320 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45882	0.35683	0.45882	0.35683
IA2	0.20823	0.17405	0.20823	0.17405
FA	0.2752	0.12863	0.13768	0.26615
Total	0.94225	0.65951	0.80473	0.79703

Table 18: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 320 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46274	0.40846	0.46274	0.40846
IA2	0.21011	0.18823	0.21011	0.18823
FA	0.24527	0.13142	0.13754	0.23915
Total	0.91812	0.72811	0.81039	0.83584

Table 19: Simulation of Power with median PFS Improvement from 6 Months to 8.5 Months and median HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.3 with 330 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45912	0.33025	0.45912	0.33025
IA2	0.21014	0.17293	0.21014	0.17293
FA	0.2935	0.14295	0.15195	0.2845
Total	0.96276	0.64613	0.82121	0.78768

Table 20: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.5 with 330 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.46421	0.35946	0.46421	0.35946
IA2	0.20755	0.17521	0.20755	0.17521
FA	0.27875	0.13842	0.1474	0.26977
Total	0.95051	0.67309	0.81916	0.80444

Table 21: Simulation of Power with PFS Improvement from 6 Months to 8.5 Months and HPFS improvement from 6.5 Months to 10 Months under Correlation of 0.8 with 330 Events

Stage	Reject H_0^{PFS} or H_0^{HPFS}	Reject H_0^{PFS} and H_0^{HPFS}	Reject H_0^{PFS}	Reject H_0^{HPFS}
IA1	0.45793	0.40456	0.45793	0.40456
IA2	0.20952	0.18711	0.20952	0.18711
FA	0.25723	0.14514	0.15103	0.25134
Total	0.92468	0.73681	0.81848	0.84301