1.0   Title Page

Statistical Analysis Plan

EORTC PROTOCOL 1410-BTG/ ABBVIE PROTOCOL M14-483

INTELLANCE 2: ABT-414 Alone or ABT-414 Plus Temozolomide Versus Lomustine or Temozolomide for Recurrent Glioblastoma: a Randomized Phase II Study of the EORTC Brain Tumor Group

Date: 30 May 2017

Version 2.0

Prepared in collaboration by

████████, **PhD, AbbVie**

And

████████, **PhD, EORTC**

SAP Version	Date of Approval/Finalized
1.0	August 15, 2016
2.0	May 24, 2017

2.0   Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP), prepared jointly by AbbVie and EORTC, will provide details to elaborate statistical methods for efficacy and safety data in adult subjects collected as outlined in the protocol version 6 for Study M14-483 dated January 18, 2017 and will describe analysis conventions to guide the statistical programming work. For pediatric sub-study, a separate SAP will provide details for statistical methodologies and data conventions.

Efficacy and safety analyses will be performed using SAS Version 9.2 or higher (SAS Institute, Inc., Cary, NC 27513) under the UNIX operating system. The SAP will be signed off before the study database is locked.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The objectives of the trial are to assess whether ABT-414 alone or in combination with temozolomide (TMZ) improves overall survival (OS), progression free survival (PFS), tumor response, quality of life, neurological deterioration-free survival (NDFS), and steroid use compared to standard treatment with lomustine single agent or TMZ re-challenge in subjects with centrally confirmed recurrent EGFR-amplified glioblastoma.

Additional exploratory objectives include analyses by stratum of EGFR pathway abnormalities (specifically EGFRvIII mutation), correlation of MGMT methylation status with clinical outcome (PFS/OS).

4.2 Design Diagram

This is a comparative randomized, open label, multicenter, multi-arm, Phase II study. The study will consist of three treatment arms; Arm 1 (ABT-414 plus TMZ), Arm 2 (ABT-414 monotherapy), and Arm 3 (mixed control arm of lomustine or TMZ). Subjects will be randomized in a 1:1:1 ratio at the EORTC headquarters after verification of the eligibility criteria discussed in Section 3 of the study protocol.

Arm 1: Subjects will be treated with ABT-414 1.0 mg/kg IV infusion over 30 to 40 minutes once every 2 weeks in combination with TMZ 150 mg/m^2 Day 1 to 5, for the first 28-day cycle, with dose escalation to 200 mg/m^2 in subsequent cycles in case of adequate tolerance and until one of the treatment withdrawal criteria has been met. In this Arm 1 the schedule of clinical and lab assessments follows the cycle of TMZ (a delay of TMZ delays the assessment schedule, withholding ABT-414 does not affect the assessment schedule).

Arm 2: Subjects will be treated with ABT-414 monotherapy 1.0 mg/kg IV infusion over 30 to 40 minutes once every 2 weeks until one of the treatment withdrawal criteria has been met.

Subjects in the control arm (Arm 3) will be treated according to the timing of relapse.

Arm 3A: Subjects relapsing during TMZ treatment or within the first 16 weeks after the first day of the last TMZ cycle will be treated with lomustine 110 mg/m^2 on Day 1 of every 42-day treatment period. Treatment will continue until one of the treatment withdrawal criteria has been met. Lomustine will be given for a maximum of 1 year.

Arm 3B: Subjects relapsing 16 weeks or more after the first day of the last TMZ cycle will be treated with TMZ 150 mg/m^2 on Day 1 to 5 for the first 28-day cycle, with dose escalation to 200 mg/m^2 in subsequent cycles in case of adequate tolerance. Treatment will continue until one of the treatment withdrawal criteria has been met.

Treatment should start within 96 hours from randomization. Treatment will be administered until one of the withdrawal criteria described in Section 5.4 of the study protocol has been met. Treatment at further progression will be according to investigators discretion.

Day 1 of the first cycle will be the first day when the study medication is taken.

A schematic of the study design is shown in Figure 1. The study procedures for each visit are outlined in Table 1.

Figure 1.   Study Schematic
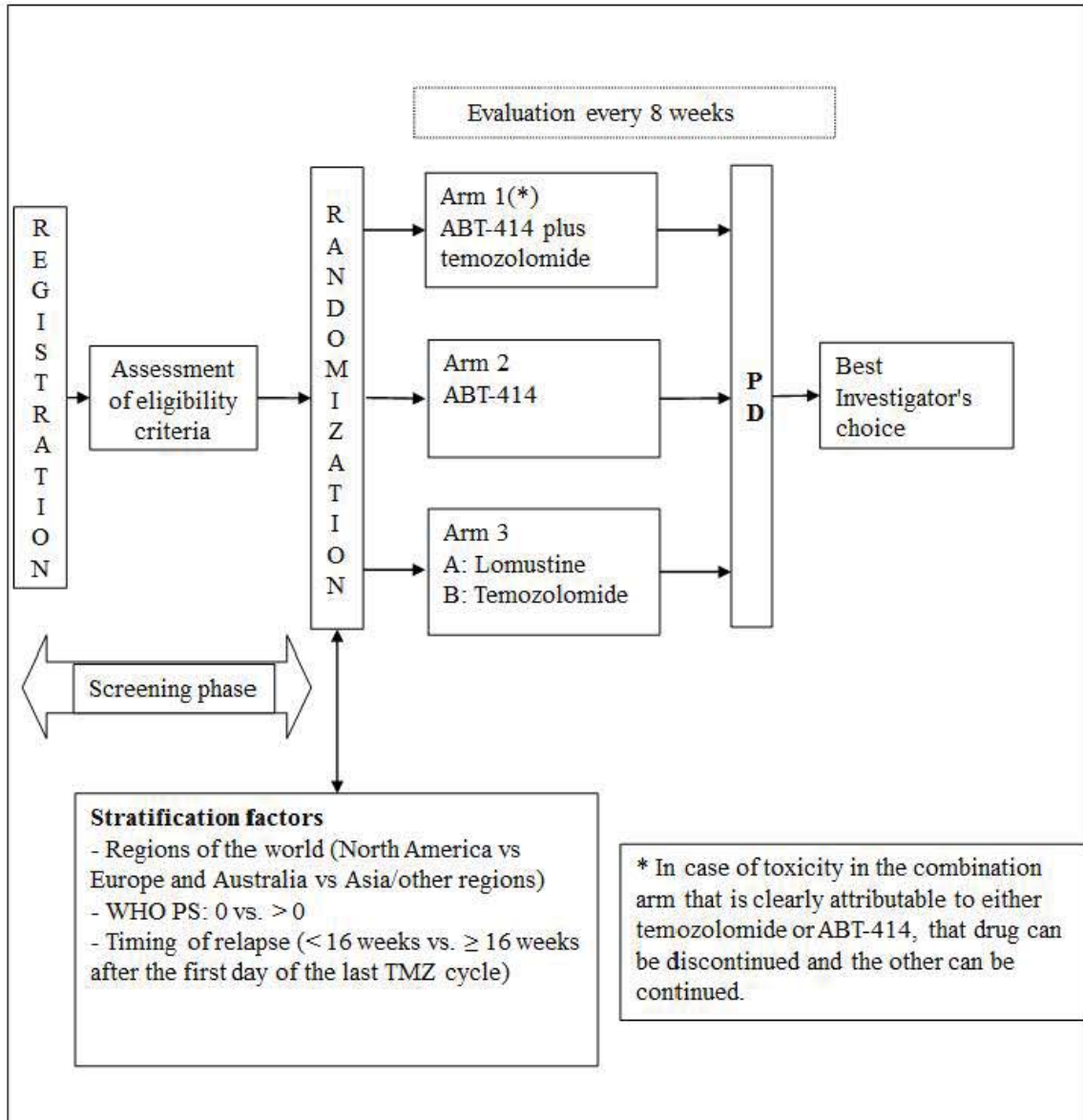
Table 1. **Study Activities**

	Before Registration	Registration	Baseline 2 Weeks Prior to Randomization	Randomization	During Protocol Treatment, Each Cycle Day 1[0]	Day 14 Arm 1 & 2	Day 21 Arm 1 and 3B	Day 28 Arm 3A[14]	Every 8 Weeks	EoT[9]	6 Months After Randomization	Follow-Up Every 12 Weeks
PIS/IC for registration (EGFR)	X											
Histology	X		X									
WHO Performance Status			X		X	X		X				X
Biological samples for central EGFR assessment	X											
EGFR gene amplification			X									
Biological samples for TR projects (optional)			X									
PIS/IC for study			X									
Weight			X		X							
Medical his. & Demographic			X									
Whole blood for germline DNA and TR projects (optional)			X									
Prior anti-cancer therapies			X									
Physical examination[1]			X			X		X				X
Vital signs			X		X	X		X				X[10]

Table 1.   Study Activities (Continued)

	Before Registration	Baseline 2 Weeks Prior to Randomization	During Protocol Treatment, Each Cycle				Every 8 Weeks	EoT[9]	6 Months After Randomization	Follow-Up Every 12 Weeks
			Day 1[0]	Day 14 Arm 1 & 2	Day 21 Arm 1 and 3B	Day 28 Arm 3A[14]				
Quality of life assessment		X					X[8]		X	
ECG		X	X[5]							
Concomitant medication collection		X	X					X		X[10]
Steroid intake		X	X					X		X
Prophylactic eye drops[15]			X	X						
Adverse events (CTCAE v.4.0)		X	X	X		X		X		X[11]
Hematology[2]		X	X	X	X[7]	X				
Biochemistry[3]		X	X							
Serum/urine pregnancy test		X[4]								
Pk samples			X[6]					X		
PD samples			X[12]							
PG samples (optional)			X[13]							
Gd-MRI		X					X			
Survival and further anti-cancer treatments										X

Table 1. **Study Activities (Continued)**

0   For Day 1 of Cycle 1, if the evaluations listed here have been already performed during the screening phase and within 2 weeks prior to this Day 1, the same results can be used for Cycle 1 and tests do not need to be repeated: physical examination including, but not limited to, WHO performance status and neurological evaluation; ocular examination if clinically indicated; hematology and biochemistry; vital signs; weight.

1   Physical examination including neurological evaluation, and ocular examination at baseline and if clinically indicated.

2   Including WBC, absolute neutrophil count, lymphocytes, hematocrit, hemoglobin and platelet count.

3   Including serum creatinine, total bilirubin, ASAT, ALAT, alkaline phosphatase, phosphate, total protein, albumin, sodium, potassium, calcium and chloride.

4   Pregnancy test must be done within 72 hours prior to randomization.

5   ECG only Day 1 of Cycle 1 and 3, immediately after the ABT–414 infusion and only for subject treated with ABT–414 (Arm 1 or 2).

6   PK samples to be collected at Day 1 of ABT–414 Cycles 1, 2, 3 and then every 2 ABT–414 cycles (Day 5 samples will only be collected during Cycles 1 and 2).

7   Hematology at Day 21 can be omitted once treatment at the present dose level caused only grade 1 or less hem. toxicity at Day 21 and only grade 2 or less at Day 28.

8   QoL only at Week 8 and Week 16 while on treatment.

9   End of Treatment Visit 35 days after last protocol drug administration.

10 If clinically indicated/if clinically relevant.

11 Assessment of AEs not resolved after EOT or AEs related to study treatment.

12 PD samples to be collected at Day 1 of Cycles 1, 2 and 3 and only for subject in an ABT–414 arm (Arm 1 or 2).

13 PG sample optional, to be collected at Day 1 of Cycle 1 or it may be collected anytime during the study.

14 Repeat physical examination including WHO performance status and neurological evaluation, hematology, vital signs and adverse event assessments at Day 28 for subjects in the lomustine arm (Arm 3A).

15 Prophylactic steroid ophthalmic solution prior to each infusion of ABT–414; 2 drops in each eye every 8 hours to start 48 hours prior to ABT–414 dosing and continue for a total of 7 days.

4.3   Randomization and Stratification

All subjects entered in the study will be centrally randomized at the EORTC Headquarters as described in Section 13.3 of the study protocol. The minimization technique used by the EORTC for random treatment allocation is based on the variance method with semi-random assignment dependent on a preset threshold as implemented by Freedman and White.[1] However, per suggestion of the ICH E9 statistical guidelines, the algorithm has been modified to incorporate a random allocation component in order to ensure an additional 15% of completely random assignments.

In this study, the threshold is set to 3, the total number of stratification factors.

Stratification factors are:

- Regions of the world (North America (USA, Puerto Rico & Canada) vs. Europe and Australia vs. Asia/other regions)
- WHO Performance Status (0 vs. > 0)
- Timing of relapse (< 16 weeks vs. ≥ 16 weeks after the first day of the last TMZ cycle)

4.4   Sample Size

This is a comparative randomized, open label, multicenter, multi-arm, Phase II study. Overall Survival (OS) is the primary endpoint.

Based on a one-sided logrank test, at an overall significance level of 2.5% and a power of 91.7% (accounting for the global testing strategy), a total of 118 deaths per comparison (i.e., monotherapy ABT-414 versus control and combination ABT-414 + TMZ versus control) are needed to detect a Hazard Ratio (HR) equal to 0.54 (i.e., a reduction in the hazard of death of 46%). A median OS of 7 months is expected on the control arm (Arm 3) based on literature review. The target treatment effect is assumed to be the same for both treatment arms (Arm 1 and Arm 2) and corresponds to an increase of median OS

to 12.9 months under proportional hazards. It is assumed that median PFS is 2.6 months on the control arm and 4.8 months on the treatment arms corresponding to an HR of 0.54.

The treatment effect on OS and PFS is generally expected to be positive correlated following historical trends.[2] Since the interim futility criterion (see Section 4.5 below) is primarily based on a PFS effect, the power of the study increases with the strength of association between PFS and OS. If PFS and OS effects are independent, the overall power of the study is estimated to be approximately 85%. The study is therefore expected to have at least 85% power overall as long as OS and PFS are not negatively correlated.

4.5   Interim Futility Analysis

An interim futility analysis is planned when at least 45 PFS events (at least 31 PFS events are observed in each comparison) are observed. At interim analysis one or both of the ABT-414 arms might be discontinued based on lack of efficacy for PFS according to RANO criteria assessed by IRC.

Treatment arms which did not show a minimum level of efficacy, defined as a computed treatment PFS HR lower than 0.91 at the interim, will be discontinued. The trial will continue with remaining treatment arm(s) and control. The decision to discontinue a treatment arm not showing a minimum level of efficacy will be based on a review of the totality of interim data and recommendations made by the IDMC. The interim analysis will be carried out according to the Statistical Considerations section in the study protocol and EORTC Policy 004 on Independent Data Monitoring Committees and Interim Analyses.

On April 28, 2016, AbbVie and EORTC contacted the IDMC chair to request a revision of the interim analysis plan. Taking into account the time that was necessary to collect all MRI scans performed, to perform their central review, to code the resulting data in the EORTC database, and to submit the interim analysis report 4 weeks before the IDMC meeting scheduled on May 23, 2016, both AbbVie and EORTC agreed that it was difficult to present the results concerning the centrally reviewed PFS endpoint at the coming

IDMC meeting. On May 20, 2016, the IDMC Chair considered this request and agreed to a postponement of the report.

In May 2016, both AbbVie and EORTC agreed to maintain the same PFS HR futility boundary of 0.91 as initially planned based on the availability of 45 PFS events. The clinical cut-off date for the revised interim analysis was set as July 01, 2016. All scans from subjects randomized before May 07, 2016 with at least one baseline and one follow-up scan were to be reviewed. The interim futility reports were presented to the IDMC on September 21, 2016 based on an interim database locked on August 16, 2016.

While there was no intent to stop the study for superior efficacy observed based on PFS during the interim analysis, an alpha spending of 0.00001 was allotted to the futility analysis per the recommendation of the US Food and Drug Administration (FDA).

Given that a total of 47 deaths had occurred across Arms 1-3 at the time of the PFS futility analysis, when a fixed Haybittle Peto alpha spending (as implemented in *EAST* 6.3) of 0.00001 is applied to the adjusted nominal significance levels for the final analysis for a range of possible values of the observed information fraction are provided in table below.

No. of OS for 2 Arms at Interim	Nominal Adjusted 1-Sided Alpha at Final OS
26 to 30	0.024996
31 to 37	0.024997

For the purposes of review by the FDA, the final adjusted significance level for each comparison (Arm 1 vs Arm3, and Arm2 vs Arm 3) will be determined at the end of the trial using the same Haybittle Peto spending approach as a function of the actual observed information (ratio of observed deaths at the time of the futility look and final analysis on the treatment and control arms). Furthermore, since the primary efficacy endpoint (OS) was never analyzed during the interim analysis and an alpha penalty is technically unnecessary for a futility interim assessment, no adjustment to the final alpha level will be made for other global regulatory purposes and final analyses will be performed at a full overall 1-sided significance level of 0.025.

4.6 Type I Error Adjustment Procedures for Multiple Testing

Although Section 8.2.3.6 of the study protocol dated November 25, 2015, describes a different testing order, the multiple testing strategy was revised per diagram (Figure 2) in version 1.0 of the SAP finalized prior to the interim analysis of efficacy data. Given the clinical importance of an overall survival (OS) benefit in recurrent GBM subjects, comparing first OS of combination arm (ABT-414 + TMZ) vs. control and next OS of ABT-414 monotherapy vs. control, followed by comparison of other secondary efficacy endpoints in a similar sequence, is considered a simple and more clinically relevant approach to test the various study hypotheses. At the final analysis, the treatment effects will be tested at the adjusted significance level mentioned in the previous section. The proposed multiple testing strategy based on the above fixed hierarchical sequence preserves the family-wise type-I error at 1-sided 2.5% level. This change was made by consensus agreement among the study sponsors (EORTC and AbbVie Inc.) without any knowledge of unblinded efficacy data from the trial and prior to interim futility analysis.

To describe the multiplicity adjustment procedure fully, consider the following null hypotheses:

H1: Arm 1 is not superior to Arm 3 in OS.

H2: Arm 2 is not superior to Arm 3 in OS.

H1a: Arm 1 is not superior to Arm 3 in PFS.

H2a: Arm 2 is not superior to Arm 3 in PFS.

H1b: Arm 1 is not superior to Arm 3 in ORR.

H2b: Arm 2 is not superior to Arm 3 in ORR.

H1c: Arm 1 is not superior to Arm 3 in OS among EGFR vIII mutated subjects.

H2c: Arm 2 is not superior to Arm 3 in OS among EGFR vIII mutated subjects.

To maintain the overall type I error control for the study, these null hypotheses will be tested in a fixed sequence of {H1, H2, H1a, H2a, H1b, H2b, H1c, H2c}. Hypothesis H1 will be tested first, and no further tests will be performed if H1 is not rejected. Thereafter, each hypothesis will be tested in the order specified if and only if H1 and all preceding null hypotheses are rejected. Otherwise testing in the hierarchical sequence will stop.

Figure 2.   Testing Procedure
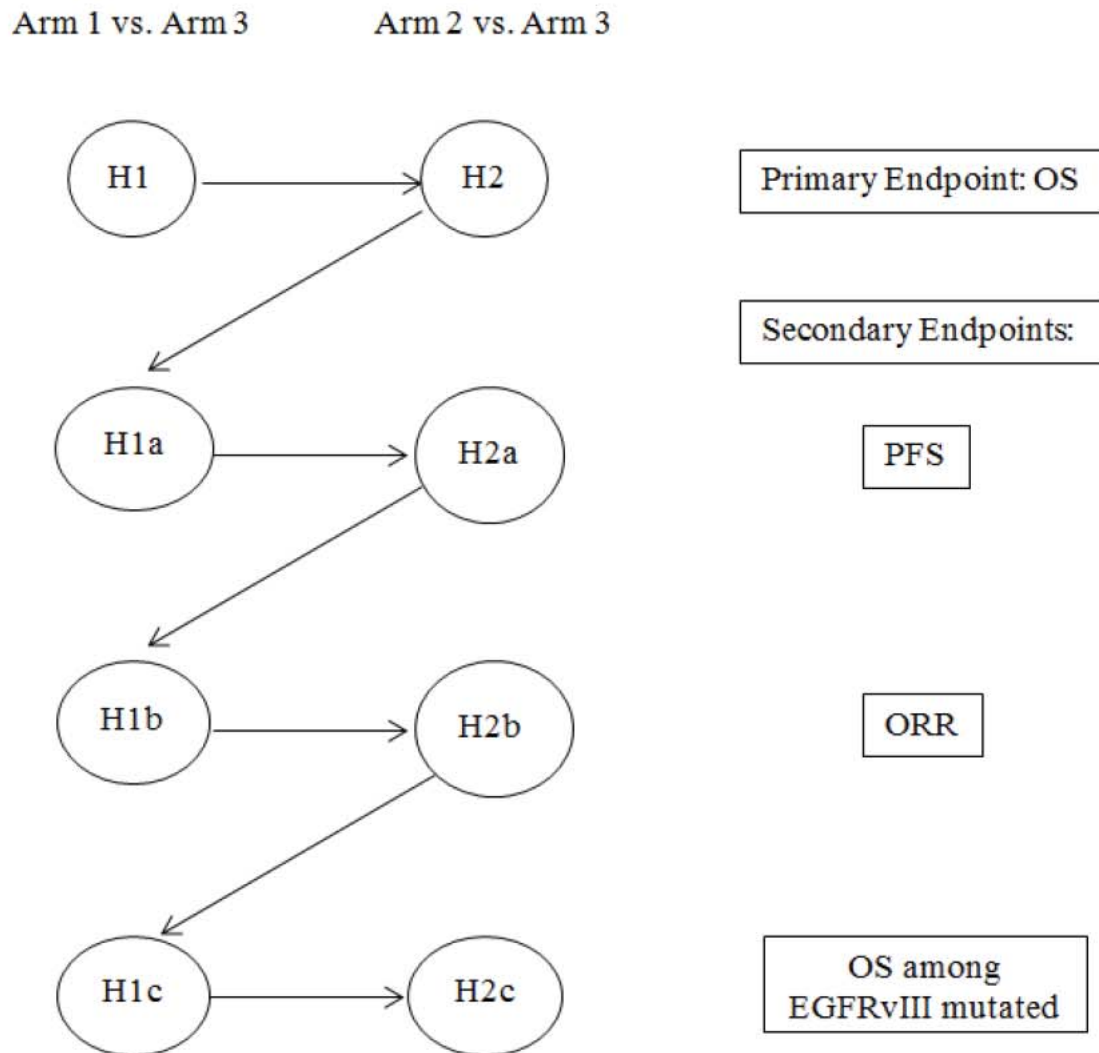
All other endpoints, e.g., PFS in EGFRvIII mutated, global health/QoL status (GHQs), and neurological disease-free survival (NDFS) are considered exploratory and will be tested at a 2-sided 5% level of significance.

5.0   Analysis Sets

The following analysis sets will be used for analysis of safety and efficacy endpoints of the study.

- Intention-to-treat (ITT) population will comprise all randomized subjects regardless of whether they received study treatment (ABT-414/TMZ/lomustine). Subjects will be classified and analyzed in the arm they were assigned at the time of randomization. The ITT will be the primary analysis set for efficacy endpoints.

- EGFRvIII-Mutated (Ev3M) population will comprise of a subset of subjects who have EGFRvIII mutated tumors in the ITT. The Ev3M will be used to address the objectives of improvement in OS and PFS among subjects with EGFRvIII mutation.

- Safety population will comprise all randomized subjects who have received at least one dose of study treatment [i.e., one dose of lomustine or TMZ (Arm 3), of ABT-414 (Arm 2), and of TMZ and ABT-414 (Arm 1)]. Subjects will be classified according to treatment received. Thus, a subject randomized to Arm 1 who received TMZ but did not receive ABT-414 will be considered in Arm 3. Subjects allocated to Arm 3 who received at least one dose of ABT-414 and those allocated to Arm 2 who received at least one dose of TMZ will be considered in Arm 1. The safety population will be the primary analysis set for analysis of safety endpoints.

- Per-protocol population will comprise all subjects who are eligible and have received at least one dose of study treatment [i.e., one dose of lomustine or TMZ (Arm 3), of ABT-414 (Arm 2), and of TMZ and ABT-414 (Arm 1)]. A subject will be considered to be eligible if he/she did not have any deviation from the subject entry criteria listed in Section 3 of the protocol. Subjects will be classified and analyzed in the arm they were assigned at the time of

randomization. Reason for ineligibility, treatment, and safety profile of excluded subjects will be reported separately.

6.0 Analysis Conventions

This section provides general considerations for data handling, summary, and analysis. This also includes general definitions of certain kinds of endpoints. More details about specific endpoints and additional analysis methods, as needed, will be specified within the description of analysis of data in this SAP.

Continuous variables will be summarized by sample size (N), mean, standard deviation (St. Dev.), median, minimum, and maximum. Frequency and percentage will be provided for categorical variables and 95% confidence intervals (CIs) will be generated for parameter estimates of interest.

Dealing with Multiple Values on the Same Day for Lab and Vital Signs Parameters

In cases where multiple values for the same variable are collected on the same day (including baseline visit and post-baseline visits), the arithmetic average will be calculated and used as the value for that day for mean changes in laboratory and vital signs parameters.

Definition of Baseline Observation

Unless otherwise specified, the baseline observation is defined as the last non-missing measurement collected prior to or on randomization. If no measurement collected prior to or on randomization, then the last non-missing observation prior to or on the first dose date of study drug (ABT-414/TMZ/lomustine) is considered.

Definition of Final Observation

For lab and vital sign parameters, the final observation is defined as the last non-missing post-baseline measurement collected during the study.

Definition of Study Rx Days (Days Relative to the First Dose of Study Drug)

Study Rx Days are calculated for each time point relative to the first dose date of study drug (ABT-414/TMZ/lomustine). They are defined as the number of days between the day of the first dose of study drug (ABT-414/TMZ/lomustine) and the specific time point. Rx days are negative values when the time point of interest is prior to the first study drug dose day. Rx days are positive values when the time point of interest is after the first study drug dose day. The day of the first dose of study drug (ABT-414/TMZ/lomustine) is defined as Study Rx Day 1, while the day prior to the first study drug dose is defined as Study Rx Day –1 (there is no Study Rx Day 0).

Definition of Analysis Windows

During the treatment period, all time points and corresponding time windows are based on Study Rx Days, unless specified otherwise. For longitudinal analyses, especially for the laboratory data, the time windows in Table 2 describe how data collected at protocol specified visits will be assigned. If more than one observation is included in a time window, the observation closest to the nominal day should be used. If there are two observations equally distant to the nominal day, the later one will be used in analyses.

Table 2.   Visit Window

Study Visit (Week)*	Nominal Study Rx Day	Time Window (Study Rx Days Range)
Baseline	-	By baseline definition
2	15	2 to ≤ 18
3	22	19 to ≤ 25
4	29	26 to ≤ 32
…	…	…
X	$7 \cdot X + 1$	$7 \cdot X - 2$ to $\leq 7 \cdot X + 4$

*   ECG at baseline and only Day 1 of Cycle 1 and 3, immediately after ABT-414 infusion and only subjects treated with ABT-414 (Arms 1 or 2).

Time-to-Event Endpoints

All time to event (TTE) endpoints are defined using a time variable and an event indicator. The time variable represents the time from a pre-defined time origin to the onset of a predefined event of interest or the last time when adequate assessments have been made to rule out the onset of the event. The latter case indicates that the endpoint has been right-censored. The event indicator is set to 1 if the event has been observed during the course of the study and 0 if it is censored. When multiple assessments are needed to ascertain the occurrence of an event, the earliest date among all of these assessments is taken to be date of the event or censoring. The time variable will be computed in days and converted into months (1 month = 30.4375 days) for analysis of TTE endpoints. When no post-randomization observations are available for a subject for any given endpoint, then the endpoint is taken to be censored on the date of randomization.

All TTE endpoints defined in this study are concerned with only the first incidence of an event of interest, and recurrence of the same event is not considered. An event, however, may be defined in a composite fashion, i.e., as the occurrence of one among several different outcomes. An example of such an endpoint is PFS where the event is either disease progression or death. The composite event is observed when at least one of the component events occurs, and the time to the earliest among the occurring component events is considered to be the TTE for the composite event.

Prior and Concomitant Medications

A prior medication is defined as any medication with an end date prior to the date of randomization and collected on the electronic case report forms (eCRFs). A concomitant medication is defined as any drug that started on the day of or after randomization, but not 35 days after the last protocol drug administration. Medications with a start date prior to the date of randomization and continuing after randomization are also classified as concomitant medications. A medication will also be considered a concomitant medication if one of the following three cases occur: (1) the start date is missing and the end date is

either after or on the day of randomization; (2) the start date is not missing but not 35 days after the last protocol drug administration and the end date is missing; (3) both the start date and the end date are missing.

End of Study

End of study occurs when all of the following criteria have been satisfied:

- At least 118 deaths have been observed for each treatment arm comparison versus control,
- Thirty five days have passed since all subjects have stopped protocol treatment, and
- The database has been fully cleaned and frozen for analyses.

7.0 Demographics, Baseline Characteristics, Medical History, Previous/Concomitant Medications, and Prior Oncology Therapies

Data for demographic, baseline characteristics, medical history, previous/concomitant medications, and prior oncology therapies will be summarized for each treatment arm and overall on the ITT population.

7.1 Demographic and Baseline Characteristics

Continuous demographic data (e.g., age, height, and weight) will be summarized with N, mean, standard deviation (St. Dev.), median, minimum, and maximum. Frequencies and percentages will be computed for the following parameters (but not limited to): sex, race, ethnicity, age group (< 40; 40 to < 60; ≥ 60 years old), weight, height, body mass index (BMI), WHO performance status, MGMT methylation status, EGFR vIII status, regions of the world, tobacco and alcohol use, timing of relapse (< 16 weeks; ≥ 16 weeks), chemotherapy (≤ 1; > 1), and surgery for glioblastoma recurrence (yes; no).

There will be no statistical comparison of demographic and baseline measurements between treatment arms.

7.2 Medical History

Medical history data will be summarized and presented using conditions/diagnoses as captured on the eCRF. The conditions/diagnoses will be presented in alphabetical order. The number and percentage of subjects with a particular condition/diagnosis will be summarized for each treatment arm. Subjects reporting more than one condition/diagnosis will be counted only once for that condition/diagnosis. There will be no statistical comparison of medical history between treatment arms.

7.3 Prior and Concomitant Medications

The frequency and percentage of subjects who took at least one dose of medication other than study drug (ABT-414/TMZ/lomustine) during the course of the trial will be summarized by the generic name coded by the World Health Organization (WHO) dictionary. Prior and concomitant medications will be summarized separately.

For each concomitant medication, a listing with subject ID numbers will be provided.

Subjects reporting the same medication generic name two or more times will be counted only once for that generic name. Subjects reporting more than one medication will be counted only once in the total number of subjects taking a concomitant medication.

A similar summary will be prepared for prior medications. In addition, the frequency and percentage of subjects taking 0, 1, 2, 3 and > 3 prior medications will be presented for each treatment arm and overall.

There will be no statistical comparison for the prior and concomitant medications between treatment arms.

7.4 Prior Anti-Cancer Treatment and Surgery

The frequency and percentage of subjects who had prior anti-cancer treatments, including prior radiation therapy and prior cancer medications, will be summarized by treatment name. Radiation therapy will be counted as one anti-cancer treatment. Prior anti-cancer

medications will be summarized using the WHO generic drug name dictionary. A subject who reported two or more uses of the same medication will be counted only once in the total for the associated WHO generic drug name. Subjects will be summarized by the number of prior anti-cancer treatments received (0, 1, 2, 3, and > 3). In addition, reason to stop and the best response to each treatment prior to study drug administration will be summarized.

The frequency and percentage of subjects who had prior brain cancer surgeries will be summarized by the number of prior brain cancer surgeries (0, 1, 2, 3, and 4). In addition, for each of initial, second, third, and fourth brain cancer surgery, extent of procedure will also be summarized.

There will be no statistical comparison for prior anti-cancer treatment/surgery between treatment arms.

8.0 Subject Disposition

Subject disposition summary (number and percentage of subjects) will be presented on all subjects who are registered, as described in Section 13.1 of the study protocol. The number of subjects will be summarized by investigator site and overall.

- Subjects who were registered
- Subjects who screen failed
- Subjects who were randomized
- Subjects who took at least 1 dose of study drug (ABT-414/TMZ/lomustine)
- Subjects who discontinued the study drug

Listing for screen failure reasons/ineligibility will be provided.

In addition, a subject disposition summary will be presented on the ITT population by treatment arm assigned at randomization – the numbers and percentages of subjects who were randomized, took at least 1 dose of study drug, and discontinued the study drug.

Major reason for study drug discontinuation will be summarized by each of ABT-414, TMZ, and lomustine within treatment arm.

No statistical tests will be performed.

9.0   Study Drug Exposure and Compliance

9.1   Dosage

Analyses for the study drug (ABT-414/TMZ/lomustine) treatment will be performed on the Safety population. Longitudinal summary statistics for average dose administered during the study will be presented. The number and percentage of subjects at each treated dose will be presented longitudinally.

An overview of dose modification will be presented by treatment arm, consisting of the number and percentage of subjects who had at least one dose modification for each of ABT-414/TMZ/lomustine throughout the study.

Dose modification of ABT-414/TMZ/lomustine will be summarized for each course/cycle. The frequency and percentage of subjects will be summarized by treatment arm, and reasons for dose modification will also be summarized.

For dose intensity and relative dose intensity of ABT-414 (Arms 1 and 2 only), the following calculation rules are applied:

● The dose intensity of ABT-414 is calculated on actual treatment duration (in weeks) and actual treatment dose received (total dose received in mg/kg). Whenever the treatment is given in cycles and the last planned day of treatment is not the last day of the cycle, the total duration of the treatment must account for this by adding the appropriate number of days to complete the last cycle. That is, if ABT-414 stopped on Day X of a 28-day cycle, the total treatment duration in days is [date of last injection] – [date of first injection] + 1 + (28-X).

- The dose intensity (mg/kg/week) of ABT-414 is the ratio of the total dose received (mg/kg, adjusted by weight) to the total treatment duration (weeks).

$$DI_{observed} \ (mg/kg/week) = \frac{Total\ dose\ (mg/kg)}{Actual\ total\ treatment\ duration\ (weeks)}$$

- The relative dose intensity is calculated as the ratio of the dose intensity as calculated above to the dose intensity indicated in the protocol, expressed in percent (%). To obtain a correct ratio, the theoretical dose intensity will be expressed in the same units of dose and treatment duration as those used for the actual dose intensity.

$$DI_{protocol} \ (mg/kg/week) = \frac{Dose\ per\ cycle\ (mg/kg)}{Theoretical\ duration\ of\ one\ cycle\ (weeks)}$$

- The relative dose intensity is usually presented using median and ranges, and by a distribution into categories (relative dose intensity $\leq 70\%$, $> 70 - 90\%$, $> 90 - 110\%$, $> 110 - 120\%$, $> 120\%$).

9.2   Study Drug Exposure

Analyses for the study drug (ABT-414/TMZ/lomustine) exposure will be performed on the Safety population.

Duration of ABT-414 exposure is defined as total number of weeks a subject received ABT-414, i.e., (last dose date – first dose date + 1)/7. The same calculation applies to TMZ and lomustine.

Descriptive statistics will be presented for the duration of exposure of ABT-414/TMZ/lomustine by treatment arm. The distribution of subjects within the following categories will be summarized: ≤ 4 weeks, > 4 to ≤ 8 weeks, > 8 to ≤ 12 weeks, etc. of ABT-414/TMZ/lomustine, respectively.

An overview of dosing schedule change will be presented by treatment arm, consisting of the number and percentage of subjects who had at least one change in dosing schedule for each of ABT-414/TMZ/lomustine throughout the study. The number and percentage of

subjects at each study drug dose in each treatment course (ABT-414 every 2 weeks; TMZ every 4 weeks; lomustine every 6 weeks) will also be presented.

Dosing schedule change (shortening or delay) of 3 or more days for any of the study drugs (ABT-414/TMZ/lomustine) will be summarized for each course. The frequency and percentage of subjects with no change, delayed, and earlier than per protocol will be summarized respectively, by treatment arm, and reasons for schedule change per course will also be summarized.

9.3 Compliance

Analyses for compliance will be performed on subjects who received at least one dose of TMZ.

The frequency and percentage of subjects will be summarized by treatment arm for (1) at least 80% compliant of TMZ at every cycle in the investigator's opinion, (2) at least 80% compliant of TMZ at all but one cycle in the investigator's opinion, (3) less than 80% compliant of TMZ at more than one cycle in the investigator's opinion.

9.4 Protocol Deviations

The number and percentage of subjects with protocol specified eligibility deviations will be summarized by treatment arm assigned at the time of randomization, and split by type of deviation.

10.0 Efficacy Analysis

10.1 General Considerations

Unless otherwise specified, all efficacy analyses described below will be performed on the ITT population according to stratification factors based on the EORTC online randomization system (ORTA).

Time-to-event endpoints (TTE), such as overall survival and progression-free survival, are defined as the time from randomization to the first occurrence of the event. For subjects

who do not experience the event, TTE will be right-censored on the date of last adequate post-baseline assessment for that event. If a subject has no post-baseline assessment performed for the event (e.g., no imaging assessments for disease progression), then the data will be censored at the date of randomization. For each TTE endpoint, the corresponding event and the date of censoring are described in individual sections below.

All TTE will be computed in days and converted into months (1 month = 30.4375 days) for analysis, as follows:

$$TTE = (\text{date of event or censoring}) - \text{date of randomization} + 1)/30.4375.$$

For the purposes of analysis, an event indicator will also be derived for each TTE for each subject, taking the value 1 if the subject is reported to have experienced the event or 0 otherwise.

Primary analysis of efficacy endpoints related to tumor assessments (e.g., ORR and PFS) will be based on assessments by a blinded, independent, central imaging review committee (IRC). For sensitivity analyses, these efficacy endpoints will also be analyzed based on assessments by local site investigators.

Tumor assessments will be based on the Response Assessment in Neuro-Oncology (RANO) criteria. Subjects for whom a complete response (CR) or partial response (PR) is observed, an additional MRI scan should be performed at least 4 weeks later to confirm the response. Any equivocal assessment of PD by the local investigator will need to be confirmed with an extra MRI performed at least 4 weeks later. If these subsequent evaluations suggest that the subject is in fact experiencing a CR, PR or PD, then the date of CR, PR or PD will be taken to be the date of the first such evaluation.

10.2   Efficacy Endpoints

The primary endpoint of this study is Overall Survival (OS), and the secondary endpoints are:

- PFS according to RANO criteria and assessed by IRC

- Objective response rate (ORR) assessed by IRC
- OS in the subgroup with EGFRvIII mutation

The following are exploratory endpoints:

- Best overall response rate (BRR), complete response rate (CRR), duration of response (DOR), and time to tumor response (TTR) assessed by IRC
- PFS in the subgroup with EGFRvIII mutation
- Neurological deterioration-free survival (NDFS)
- Quality of life
- Steroid use

Overall Survival (OS)

Overall survival (OS), or time to all-cause mortality, is defined as time from randomization to death due to any cause, regardless of whether the event occurred on or off study drug (ABT-414/TMZ/lomustine). For a subject who is not reported to have died at the time of data analysis, OS will be censored on the last documented date the subject is known to be alive.

Progression-Free Survival (PFS)

All disease progression assessed by IRC per RANO criteria will be included, regardless of whether the event occurred on or off study drug. Progression-free survival for a given subject will be defined as the time from the date of randomization till the date of first progression or the date of subject's death, whichever occurs first. If a subject neither experiences disease progression nor dies, then the subjects PFS will be censored at the last date of an adequate post-baseline tumor assessment.

Neurological Deterioration-Free Survival (NDFS)

Neurological deterioration (ND) is defined as a decrease in WHO performance status:

- for subjects with baseline WHO performance status 0 or 1: deterioration to WHO performance status 2 or worse for which no other explanation is present, and which is maintained for at least 3 weeks
- for subjects with baseline WHO performance status 2: deterioration to WHO performance status 3 or worse for which no other explanation is present and which is maintained for at least 3 weeks

NDFS will be measured as the time from randomization until first neurological deterioration or death whichever occurs first, regardless of whether radiological progression has occurred or not. If a subject neither experiences ND nor dies, then the subject's NDFS will be censored at the last date of post-baseline WHO performance status assessment.

Duration of Response (DOR)

The duration of overall response (DOR) will be analyzed only for the subset of subjects who experience a complete or partial response (CR or PR). For a given subject, DOR will be defined as the number of months from the day the criteria are met for CR or PR by IRC (whichever is recorded first) to the date of progressive disease or death, whichever comes first. If a subject is still responding (i.e., has not progressed nor died after CR or PR), then the subject's data will be censored at date of the last radiological assessment.

Time to Tumor Response (TTR)

Time to tumor response is defined as the time from the date of randomization to the observation of objective response (CR or PR).

For a subject who neither respond nor progress, TTR will be censored at the subject's last tumor assessment date. For subjects who progress and never respond, TTR will be set by convention as the longest time from randomization to the last tumor assessment observed across all subjects; i.e., max (last tumor assessment date – randomization date + 1) for all subjects. The distinction in censoring rule is made here since, at the time of censoring, the

former set of subjects would still have a possibility of experiencing a response (the event of interest), while the latter set of subjects never will.

Best Overall Response Rate (BRR)

Best overall response is the best response assessed by IRC from the date of randomization until disease progression or death, whichever comes first. Subjects who did not have post-baseline assessments will be considered "not-assessed." Subjects who had only one response that required a confirmatory MRI scan, which was not performed, will also be considered not-assessed.

Objective Response Rate (ORR)

Objective response includes best overall responses – complete response (CR) and partial response (PR) – assessed by IRC per RANO criteria from the date of randomization until disease progression or death, whichever comes first. All objective responses (CR and PR) must be confirmed by repeat MRI 4 weeks later than the first time when CR or PR is identified. Any subject who did not meet CR or PR including those who did not have post-baseline radiological assessments will be considered as non-responder.

10.3   Efficacy Analyses

Statistical analyses will be performed to compare the primary and secondary endpoints described in Section 10.2 between each treatment arm (Arm 1 or Arm 2) and control (Arm 3) following the fixed sequence described in Section 4.6.

Analysis Methods for TTE Endpoints

For all TTE endpoints, differences between the treatment arms and the control arm will be assessed by a log-rank test stratified by the randomization stratification factors described in Section 4.3. The p-value obtained from the stratified log-rank test will be considered as the primary basis for determination of statistical significance for the TTE endpoints.

Kaplan-Meier (KM) estimates of the survival distribution of each TTE endpoint in Arms 1 – 3 will be estimated and plotted. Two-sided 95% CIs for the KM curves will be computed using Greenwood's formula and plotted. The median for the TTE and the probability of being event-free at 6 months, 1 year (12 months) and 2 years (24 months), and their 95% CIs will be estimated from the KM curve and its 95% CI for each treatment arm.

The hazard ratios of the two treatment arms over the control arm, adjusting for the main effects of randomization stratification factors as covariates, and their 95% CIs will be estimated using a Cox's proportional hazards model (score test).

The above-mentioned methods will be referred to as *primary survival analysis methods* for the remainder of this SAP. Analysis of the following endpoints will be performed using these primary survival analysis methods:

- Overall Survival (OS)
- Progression-Free Survival (PFS) assessed by IRC
- OS and PFS in subjects with EGFRvIII Mutation (in the Ev3M population)
- Neurological Deterioration-Free Survival (NDFS)
- OS and PFS in the subgroup of subjects with timing of relapse (< 16 weeks vs. ≥ 16 weeks after the first day of the last TMZ cycle)
- Time to Tumor Response (TTR)
- Duration of Response among treatment responders (CR or PR). The time origin for calculating DOR is not the date of randomization.

Objective Response Rate (ORR)

The number and percentage of subjects with objective response will be presented by treatment arm. The objective response (CR and PR) rate will be reported with exact (binomial) two-sided 95% CI for each treatment arm.

The ORR will be compared between the treatment arms over the control arm using a one sided Cochran-Mantel-Haenszel (CMH) test, stratified by randomization stratification factors. Odds ratios will be presented with 95% CIs.

Best Overall Response Rate (BRR)

The number and percentage of subjects will be presented for each response category by treatment arm.

Complete Response Rate (CRR)

The same methodology as for analysis of objective response will be applied to complete response rate.

Corticosteroid Use

Systemic corticosteroids that were given orally or by injection, reported either in prior/concomitant medication form until 35 days after the last study drug (ABT-414/TMZ/lomustine) or tumor assessment form will be summarized. Baseline steroid use is defined as any use of systemic steroid prior to and within 14 days of randomization. Post-baseline steroid use is defined as any use of systemic steroid on or after randomization. Steroid dose will be converted to a dexamethasone equivalent dose.[3]

Table 3.   Conversion Factor to Dexamethasone Equivalent Dose

Corticosteroids (mg)	Conversion Factor*
Cortisone	0.75/25
Hydrocortisone	0.75/20
Prednisone, Prednisolone	0.75/5
Methylprednisolone, Triamcinolone	0.75/4
Betamethasone	0.75/0.6

*   Dexamethasone equivalent dose (mg) = Conversion factor x Corticosteroid dose (mg).

Baseline steroid dose is defined as the average dose prior to and within 14 days of randomization, i.e., sum of daily doses divided by the number of days treated with steroids. The following summary will be presented for each treatment arm:

- The number and percentage of subjects who are treated with steroids at baseline (yes; no; missing) and post-baseline (yes; no; missing) in a two-way table.
- For subjects with no steroids at baseline and at least one steroid at post-baseline, simple summary statistics for time (in weeks) to steroid use during the study since randomization.
- The number and percentage of subjects who are treated with average steroid dose group (mg/day) at baseline (≤ 2; > 2 to ≤ 4; > 4 to ≤ 6; > 6 to ≤ 8; > 8 to ≤ 20; > 20 mg/day).
- The number and percentage of subjects with objective response (CR and PR assessed by IRC per RANO criteria) and Kaplan-Meier estimates of PFS, by baseline steroid use status (yes; no; missing), respectively.

In addition, steroid dose change (sustained reduction and complete reduction) relative to baseline will be summarized. Sustained reduction of steroid dose is defined as 50% or more steroid dose reduction, relative to baseline, for 50% or more of the time on study drug (ABT-414/TMZ/lomustine). Complete reduction is defined as discontinuation of steroid use for $\geq 25\%$ of the time on study drug (ABT-414/TMZ/lomustine).

- The number and percentage of subjects with objective response status (responders; non-responders) and Kaplan-Meier estimates of PFS for subjects who have sustained steroid reduction and for subjects who have complete steroid reduction, respectively.

Quality of Life

Subject self-reported health related quality of life (HRQoL) outcomes are assessed at the baseline, Week 8 MRI visit (while on-treatment), Week 16 MRI visit (while on-treatment), and Month 6 visit (any time during the 6th and 7th month after the date of

randomization regardless of treatment or progression status) by using the QLQ-C30 version 3 (Table 4) and QLQ-BN20 (Table 5).

The main objective of HRQoL assessment is to determine the impact of ABT-414 on overall global health and HRQoL. A secondary objective is to evaluate the effect of the treatment on the remaining symptoms and functioning scales as treatment-related side effects may have a (temporary) negative influence on the health related domains of HRQoL of these subjects. The QLQ (quality of life questionnaire) comprises distinct scales, each of which represents a different aspect of QoL, and the individual questions on the QLQ are called items.

The QLQ-C30 is composed of global health status/QoL scale; five functional scales (physical, role, emotional, cognitive, and social); three symptom scales (fatigue, nausea and vomiting, and pain); and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

Table 4.   **Scoring the QLQ-C30**

QLQ-C30	Item (Question) Numbers	Item Range*
Global health status/QoL		
Global health status/QoL	29, 30	6
Functional scales		
Physical functioning	1, 2, 3, 4, 5	3
Role functioning	6, 7	3
Emotional functioning	21, 22, 23, 24	3
Cognitive functioning	20, 25	3
Social functioning	26, 27	3
Symptom scales/Items		
Fatigue	10, 12, 18	3
Nausea and vomiting	14, 15	3
Pain	9, 19	3
Dyspnea	8	3
Insomnia	11	3
Appetite loss	13	3
Constipation	16	3
Diarrhea	17	3
Financial difficulties	28	

*   Item range is the difference between the possible maximum and the minimum responses to individual items; most
     items take values from 1 to 4, giving range = 3.

Table 5.   **Scoring the QLQ-BN20**

QLQ-BN20	Item (Question) Numbers	Item Range
Symptom scales		
Future uncertainty (FU)	1, 2, 3, 5	3
Visual disorder (VD)	6, 7, 8	3
Motor dysfunction (MD)	10, 15, 19	3
Communication deficit (CD)	11, 12, 13	3
Headaches (H)	4	3
Seizures (S)	9	3
Drowsiness (Dr)	14	3
Hair loss (HL)	16	3
Itchy skin (IS)	17	3
Weakness of legs (WL)	18	3
Bladder control (BC)	20	3

The brain cancer module (QLQ-BN20) is meant for use among brain cancer subjects varying in disease stage and treatment modality (i.e., surgery, chemotherapy, radiotherapy, etc.). It is meant to be complemented by the QLQ-C30. It is composed of 20 items forming 11 symptom scales (7 single item scales and 4 multi-item scales).

All scales and single items are reported on categorical responses, and will be linearly converted to numeric values 0 through 100. Given a scale, the raw score (RS) is computed as the mean of component items over the number of items answered for that scale. Then the scale score (SS) will be computed as following.

For global health status/QoL, and symptom scales/items,

$$SS = \{(RS - 1) / \text{range}\} \times 100.$$

For Functional scales,

$$SS = \{1 - (RS - 1) / \text{range}\} \times 100.$$

If 50% or more of the items are answered for a given scale, the scale score will be computed as described above. If not, the scale score will not be computed. For example, physical functioning contains 5 items (questions 1 through 5), and the scale score is calculated if at least 3 of the items are answered. For single-item measures (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties), none of the single-item measures can be computed if not answered. If a scale score cannot be computed, the outcome for that score is left blank.

Global health status/QoL is the primary HRQoL endpoint. The other scales from the QLQ-C30 and QLQ-BN20 will be analyzed on an exploratory basis as secondary HRQoL endpoints. The financial difficulties scale (QLQ-C30) will not be analyzed.

All QoL analyses will be performed for the subset of the ITT population for whom at least one valid post-baseline QoL form has been collected. A subject will not be included in a change from baseline analysis unless a corresponding valid QoL form at baseline has also been collected.

Compliance for a group of subjects at a certain time point T is defined as:

$$\text{Compliance (T)} = \frac{\text{Valid QoL form within [L, U]}}{\text{QoL form expected at T}}$$

Where L and U are the lower and upper bound of the time windows associated with T. The protocol schedule as specified in the protocol can be applied to this definition. The resulting values for T, L and U are summarized in Table 6.

Table 6.   **Visit Window for QoL**

	T	L	U	Window Length
Baseline	Date of randomization = D0	D0 – 1 week	D0	1 week
Week 8 (on-treatment)	Week 8 MRI visit date = D1	D1 – 1week	D1 + 1 week	2 weeks
Week 16 (on-treatment)	Week 16 MRI visit date = D2	D2 – 1 week	D2 + 1 week	2 weeks
At 6 months	D0 + 6 months = D3	D3 – 30 days	D3 + 30 days	2 months

QoL forms are considered as invalid if either:

- All questions on the form are blank.
- The completion date is unknown or it cannot be assigned to a single assessment time point.
- The completion date falls outside all time windows.
- Multiple forms are received during the same time window. In the latter case, the form closest to the intended assessment time will be kept. In case of equidistance, the earlier form will be kept.

Forms are expected at T for each subject that was within the assessment window. A subject is considered to be within the assessment time window if at least one of the following conditions is met:

- At least one valid QoL form received within L and U
- Time on study > time from randomization till T

For the baseline assessment, all subjects are expected to have a completed QoL form.

The compliance rate between the arms will be compared at each post-baseline time point using a chi-square test. The following analyses will be presented per time-point:

- Compliance rate with QoL assessment by treatment group and overall
- Percentage of questions answered by treatment group (summary statistics)

- Comparison of compliance rates (chi-square test or fisher exact test (if expected cell frequencies are < 10)

In case significant differences are found, explanation of these differences should be sought and the analysis strategy adapted accordingly.

For each scale, the treatment group differences will be evaluated by analyzing the change from baseline to each of the three post-baseline time-points and to the final measurement (last QoL assessment per subject) using a one-way analysis of variance (ANOVA) model. Simple summary statistics for QoL assessments at each timepoint will also be presented. No adjustment for multiplicity is made since it is exploratory. For QoL assessment data, AbbVie will only perform and present simple summary statistics at each timepoint and change from baseline analysis. Other QoL data will be modeled and analyzed by EORTC.

Changes in HRQoL scores per time point will be evaluated by classifying them according to the 10 point change threshold into 3 categories:

- **Improved:** defined as a 10 point or more improvement (i.e., increase for functional scales and decrease for symptom scales) from baseline.
- **Stable:** defined as a less than 10 point change from baseline.
- **Deteriorated:** defined as a 10 point or more deterioration (i.e., decrease for functional scales and increase for symptom scales) from baseline. Subjects without a valid HRQoL outcome will be considered as having deteriorated.

In case of discrete numbers (N < 10), categories may be combined. The subsequent proportions will be compared between treatment arms using a chi-square test in the intent-to-treat population.

10.4   Re-Randomization Test

The robustness of the results to the use of the minimization randomization algorithm will be evaluated by performing a re-randomization test to compare overall survival between the treatment arm(s) and the control.

The re-randomization test will be performed as follows:

- The procedure will use 50,000 replications of the ITT population (all randomized subjects).
- For each replicate, a virtual sequence of treatment assignments (a virtual re-randomization schedule) will be created using the original minimization randomization algorithm based on the order in which the subjects were originally randomized into the trial and the value of their randomization stratification factors.
- For each replicate, primary survival analysis will be performed for OS.
- The re-randomization test p-values will be the fraction of replications where the log-rank p-values from the replicates are less than or equal to the corresponding p-values from the original analysis of overall survival.
- The empirical distribution function of the chi-square statistics obtained from the replicates will be plotted against the theoretical asymptotic Chi-square(1) distribution of the log-rank statistic for visual comparison to assess validity of the test procedure.

Robustness of the results will be established if the re-randomization test p-values are then found to be comparable to the corresponding p-values from the original analyses of OS, PFS assessed by IRC, and ORR assessed by IRC.

10.5   Sensitivity Analyses

Sensitivity analyses will be conducted to evaluate the robustness of efficacy results.

PFS, ORR, BRR, duration of response will be analyzed based on assessment by local investigator per RANO criteria, as supportive of the efficacy analyses with central imaging center evaluations.

An analysis of OS will be performed by fitting a Cox regression model with treatment arm and covariates identified in the references;[4,5] i.e., performance status, tumor location, patient age, baseline steroids, tumor size and load. Treatment will be forced in the model and covariates will be selected using a stepwise procedure. Bootstrap technique will be used to estimate the posterior probability of inclusion in the Cox model. EORTC will perform the analysis with various model section approaches.

Modified PFS is defined as the number of months from the date of randomization until the date of first progression assessed by IRC or death without receiving post anti-cancer therapy, whichever occurs first. Subjects for whom neither death nor progression have been documented will be censored on the date of the last radiological assessment if they have no post anti-cancer therapy. If a subject has other post anti-cancer therapy prior to progression or death, the data will be censored at the date of last radiological assessment prior to or on the start date of post anti-cancer therapy.

Accounting for subjects who died or progressed immediately after 2 or more consecutive missed assessments, modified efficacy endpoint of PFS to consider the date of the earliest missed radiological assessment prior to a subject's progression or death, as a date of progression, will be analyzed using the same statistical methodology as that for the secondary PFS endpoint.

For subjects with at least one measureable disease at baseline (defined as an enhancing tumor with two perpendicular diameters $\geq$ 10 mm), the number and percentage of subjects with objective responses will be presented using the same statistical methodology as that for the secondary ORR endpoint.

In case more than 10% are excluded from the per-protocol population, the primary efficacy analyses (both comparisons) will be performed in the per-protocol population at two-sided alpha 5%.

10.6 Subgroup Analyses

Efficacy endpoints (OS, PFS, ORR, BRR, CRR, and NDFS) will be assessed in the following subgroups on an as-stratified basis according to stratification data collected at randomization;

- Timing of relapse (< 16 weeks or ≥ 16 weeks after the first day of the last TMZ cycle)
- Region of the world (North America or Europe and Australia or Asia/other regions)
- WHO Performance status (0 or > 0)
- MGMT promoter methylation (status methylated or unmethylated)
- EGFR pathway abnormalities (abnormal or normal)

Differences between the treatment arms and the control arm (Arm 1 vs. Arm 3; Arm 2 vs. Arm 3) in time-to-event endpoints (OS, PFS, and NDFS) will be assessed by unstratified log-rank test. Kaplan-Meier curves comparing treatments will be presented in each subgroup.

For ORR, BRR, and CRR, the number and percentage of subjects with response will be reported with exact (binomial) two-sided 95% CI (CI for ORR and CRR only) in each subgroup by treatment arm. The ORR and CRR will be compared between the treatment arms over the control arm using a chi-square (Fisher's exact) test.

The hazard ratios (HRs) of the two treatment arms over the control arm and their 95% CIs will be calculated by a Cox's model with only treatment as covariate. Odds ratios (ORs) will be presented with 95% CIs. Forest plots will be presented to display these HRs and ORs. The forest plot will also contain the HR and OR for the ITT population for the purposes of visual comparing and contrasting.

11.0   Safety Analysis

11.1   General Considerations

Unless otherwise specified, safety analyses will be performed on the Safety population. No statistical comparisons will be performed between treatment arms. No p-values and CIs will be provided.

Unless specified, all summaries/analyses involving AEs will only include treatment-emergent adverse events (TEAEs). TEAEs are defined as any adverse event with onset or increase severity after the first dose of study drug (ABT-414/TMZ/lomustine) and no more than 35 days after the last dose of study drug (ABT-414/TMZ/lomustine).

AEs where the onset date is the same as the study drug start date are assumed to be treatment-emergent, unless the study drug start time and the AE start time are collected and the AE start time is prior to the study drug start time. If an incomplete onset date was collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (e.g., the AE end date was prior to the date of the first dose of study drug, a partial AE start date is available for which the month or year is prior to the first dose date of study drug).

Adverse events will be coded and summarized by system organ class and preferred term according to the Medical Dictionary for Regulatory Activities (MedDRA) version 19.0 or higher. The table of clinical toxicity grades modified from the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 4.0 will be used in the grading of adverse events and laboratory abnormalities that are reported as adverse events.

11.2   Analysis of Adverse Events

The number and percentage of subjects having treatment-emergent adverse events will be tabulated by MedDRA system organ class and preferred term. The tabulations will also be provided with further breakdowns by NCI CTCAE toxicity grade and relationship to

study drug (ABT-414/TMZ/lomustine). Serious adverse events, adverse events leading to discontinuation of treatment, and adverse events leading to death will be summarized.

An overview of AEs will be presented, by treatment arm and overall, consisting of the number and percentage of subjects experiencing the following adverse event categories.

- Any treatment-emergent adverse event
- Any treatment-emergent adverse event by NCI CTCAE grade
- Any treatment-emergent adverse event with NCI CTCAE grade 3, 4 or 5
- Any treatment-emergent adverse event NCI CTCAE grade 3 or 4
- Any treatment-emergent adverse event that is rated as a reasonable possibility of being related to ABT-414
- Any treatment-emergent adverse event that is rated as a reasonable possibility of being related to TMZ
- Any treatment-emergent adverse event that is rated as a reasonable possibility of being related to lomustine
- Any treatment-emergent adverse event NCI toxicity grade 3 or 4 that is rated as a reasonable possibility of being related to ABT-414
- Any treatment-emergent adverse event NCI toxicity grade 3 or 4 that is rated as a reasonable possibility of being related to TMZ
- Any treatment-emergent adverse event NCI toxicity grade 3 or 4 that is rated as a reasonable possibility of being related to lomustine
- Any treatment-emergent serious adverse event
- Any treatment-emergent serious adverse event NCI toxicity grade 3 or 4
- Any treatment-emergent serious adverse event that is rated as a reasonable possibility of being related to ABT-414
- Any treatment-emergent serious adverse event that is rated as a reasonable possibility of being related to TMZ
- Any treatment-emergent serious adverse event that is rated as a reasonable possibility of being related to lomustine
- Any treatment-emergent adverse event leading to discontinuation of ABT-414

- Any treatment-emergent adverse event leading to discontinuation of TMZ
- Any treatment-emergent adverse event leading to discontinuation of lomustine
- Any treatment-emergent adverse event leading to dose reduction of ABT-414
- Any treatment-emergent adverse event leading to dose reduction of TMZ
- Any treatment-emergent adverse event leading to dose reduction of lomustine
- Any treatment-emergent adverse event leading to interruption of ABT-414
- Any treatment-emergent adverse event leading to interruption of TMZ
- Any treatment-emergent adverse event leading to interruption of lomustine
- Any treatment-emergent adverse event of special interest
- Any treatment-emergent adverse event leading to death

The treatment-emergent adverse events of special interest will be based on microcystic keratopathy CMQ 80000152 (Section 11.2.1).

Additionally, selected analyses of adverse events will be performed for the subset of subjects with EGFRvIII mutation (on an as-treated basis in the Ev3M population) for the following categories.

- Any treatment-emergent adverse event (descending frequency)
- Any treatment-emergent serious adverse event
- Any treatment-emergent adverse event of special interest

11.2.1   Adverse Events of Special Interest (AESI)

The number and percentage of subjects having AESI will be tabulated by MedDRA system organ class and preferred term. Specific AESI will be identified by the following search criteria in Table 7.

Table 7.   **Adverse Events of Special Interest**

Adverse Event of Special Interest	Search Criteria	Event Definition/ Medical Concept
Microcystic Keratopathy	CMQ "80000152"; Preferred Terms: Acquired corneal dystrophy, Ciliary hyperaemia, Corneal abrasion, Corneal cyst, Corneal decompensation, Corneal defect, Corneal degeneration, Corneal deposits, Corneal disorder, Corneal dystrophy, Corneal epithelial microcysts, Corneal epithelium defect, Corneal erosion, Corneal irritation, Corneal lesion, Corneal oedema, Corneal opacity, Deposit eye, Dry eye, Eye pain, Foreign body sensation in eyes, Keratitis, Keratopathy, Lacrimation increased, Limbal swelling, Ocular toxicity, Photophobia, Slit-lamp tests abnormal, Topography corneal abnormal, Vision blurred	Any adverse event report identified by search criterion

11.2.1.1   Adverse Event of Special Interest

The number and percentage of subjects experiencing at least one AESI for each of the following events will be summarized:

- Any AESI
- Any AESI by NCI CTCAE grade
- Any AESI with NCI CTCAE grade 3 or 4
- Any serious AESI
- Any AESI leading to discontinuation of ABT-414
- Any AESI leading to dose interruption of ABT-414
- Any AESI leading to dose reduction of ABT-414

11.2.1.2   Time to Onset of Adverse Event of Special Interest

For specified AESIs, time to the first occurrence (onset) the event will be assessed using survival analysis methodology (Kaplan-Meier estimates) separately for all AESIs, for all AESIs with NCI CTCAE grade ≥ 2, and for all AESIs with NCI CTCAE grade ≥ 3.

Time to onset will be measured in days relative to the day of a subject's first dose of study drug (ABT-414/TMZ/lomustine) to the start date of the first occurrence of AESI. If a subject has not experienced an AESI, the subject will be censored on the day of the subject's last assessment (i.e., the day of the subject's last known laboratory assessment, last known vital sign assessment, last known physical exam, last known ocular exam, last known tumor assessment, or last known follow-up visit, whichever is the latest) or 35 days from the subject's last ABT-414 treatment, whichever is earliest; if the subject has not experienced an AESI and had no post-baseline assessment (i.e., none of laboratory assessment, vital sign assessment, physical exam, ocular exam, or tumor assessment), the subject will be censored on the day of the subject's first study drug (ABT-414/TMZ/lomustine).

Median onset time will be estimated using a Kaplan-Meier method with corresponding 95% CI for the median for each treatment arm.

11.3   Deaths

The number and percentage of subject deaths will be summarized 1) for deaths occurring while the subject was still receiving study drug (ABT-414/TMZ/lomustine) in this study; 2) for deaths occurring off-treatment within 35 days after the last dose of study drug (ABT-414/TMZ/lomustine); and 3) for all deaths in this study regardless of the number of days after the last dose of study drug (ABT-414/TMZ/lomustine).

Similarly, the number and percentage of subject deaths will be summarized for deaths occurring while the subject was still receiving 1) ABT-414; 2) within 35 days after the last dose of ABT-414; and 3) for all deaths regardless of the number of days after the last dose of ABT-414.

11.4   Analysis of Laboratory, Vital Signs, and ECG Data

Biochemical parameters (serum creatinine, total bilirubin, AST, ALT, alkaline phosphatase, phosphate, total protein, albumin, sodium, potassium, calcium, and chloride) will be collected at baseline (all Arms), and Day 1 of each cycle (all Arms).

Hematological parameters (WBC, absolute neutrophil count, lymphocytes, hematocrit, hemoglobin, and platelet count) will be collected at baseline (all Arms), Day 1 of each cycle (all Arms), Day 14 of each cycle (Arms 1 and 2 only), Day 21 of each cycle (Arms 1 and 3B only), and Day 28 of each cycle (Arm 3A only).

Vital sign parameters (diastolic/systolic blood pressure, pulse rate, and body temperature (°C)) will be collected at baseline (all Arms), Day 1 of each cycle (all Arms), Day 14 of each cycle (Arms 1 and 2 only), and Day 28 of each cycle (Arm 3A only), and follow-up visit if clinically indicated/relevant.

Twelve-lead ECG data will be collected at baseline, immediately after ABT-414 dosing on Day 1 of Cycle 1 and Cycle 3 (Arms 1 or 2 only).

For hematological, biochemical, and vital sign parameters, changes from baseline are analyzed for each post-baseline visit (as specified in Section 6.0 Analysis of Conventions) and the final visit. If more than one measurement exists for a subject on a particular day, an arithmetic average will be calculated. The average will be considered to be the subject's measurement of that day. Descriptive statistics will be presented for baseline, each post-baseline, and the final visit by each treatment arm. Mean change from baseline to each post-baseline visit and the final visit within each treatment arm will also be presented. The highest and lowest values of each hematological parameter will be identified for each subject, and descriptive statistics for mean change from baseline within each treatment arm will be presented.

Biochemistry and hematology determinations will be categorized according to NCI CTCAE version 4.0 grades, and shifts from baseline grades to maximum and final post-baseline grades will be assessed. The baseline and final grades will be defined respectively as the grade of the last measurement collected prior to or on randomization, or if missing, then prior to or on the first dose of study drug (ABT-414/TMZ/lomustine), and as the grade of the last post-baseline measurement collected no more than 35 days after the last dose of study drug (ABT-414/TMZ/lomustine). If multiple values are available for a post-baseline measurement, then the value with the highest NCI CTCAE

grade will be used in the assessment of shift. Frequency and percentage of subjects will be presented by treatment arm: 1) baseline grade vs. maximum post-baseline grades, and 2) baseline grade vs. final post-baseline grade. Frequency and percentage of subjects experiencing a shift from baseline grades of 0 to 2 to maximum post-baseline grades of 3 to 4, and from baseline grades of 0 to 2 to final post-baseline grades of 3 to 4 will be presented by treatment arm.

Detailed listings of data for subjects experiencing NCI CTCAE grade 3 to 4 biochemistry and hematology values will be provided. All measurements collected, regardless of the number of days after the last dose of study drug, will be included in these listings.

Descriptive summary statistics for categorical ECG data will be presented for baseline and Day 1 of Cycle 1 and Cycle 3 (Arms 1 or 2 only). Visit window in Table 2 will not be applied to the ECG data. Shift tables will also be provided.

11.4.1   Assessment of Potentially Clinically Significant Laboratory Values

For selected laboratory variables, the number and percentage of subjects who had at least one post-baseline observation meeting the pre-defined criteria for potentially clinically significant values will be provided for each variable. If a subject did not have the laboratory measurement at baseline but had a post-baseline measurement which met the pre-defined criteria, this subject is considered as meeting the criteria for potentially clinically significant laboratory values for the measurement.

Pre-defined criteria for potentially clinically significant laboratory values are given in Table 8 and Table 9 below based on NCI CTCAE criteria:

Table 8. **Criteria for Potentially Clinically Significant Laboratory Values – Hematology Variables**

Hematology Variables	Units	Definition of Potentially Clinically Significant Values (NCI CTCAE Grade ≥ 3)
Hemoglobin	g/L	< 80
White blood cell count	10^9/L	< 2
Neutrophil count	10^9/L	< 1
Lymphocyte count	10^9/L	< 0.5
Platelet count	10^9/L	< 50

Table 9. **Criteria for Potentially Clinically Significant Laboratory Values – Chemistry Variables**

Chemistry Variables	Units	Definition of Potentially Clinically Significant Values (NCI CTCAE Grade ≥ 3)	
		Low	High
Total bilirubin			> 3.0 × ULN
Albumin	g/L	< 20	
Aspartate aminotransferase (AST/SGOT)			> 5.0 × ULN
Alanine aminotransferase (ALT/SGPT)			> 5.0 × ULN
Alkaline phosphatase			> 5.0 × ULN
Creatinine			> 3.0 × ULN
Sodium	mmol/L	< 130	> 155
Potassium	mmol/L	< 3.0	> 6.0
Total calcium	mmol/L	< 1.75	> 3.1

11.4.2 Assessment of Hepatotoxicity

The number and percentage of subjects with maximum treatment-emergent laboratory values meeting the following criteria were summarized to assess potential hepatotoxicity.

Potential hepatotoxicity criteria defined in Hy's law:

- ALT ≥ 3 × ULN and total bilirubin ≥ 2 × ULN

- AST ≥ 3 × ULN and total bilirubin ≥ 2 × ULN
- ALT and AST ≥ 3 × ULN and total bilirubin ≥ 2 × ULN

The maximum ratio relative to the ULN will be used to determine if subjects met the criteria listed above. The ALT, AST, and total bilirubin values do not need to be concurrent in order to meet the defined criteria as long as collected up to 35 days following the last dose date of study drug (ABT-414/TMZ/lomustine). For ALT, AST, and total bilirubin, a subject will be counted if the post-baseline laboratory value up to 35 days following the last dose date of study drug (ABT-414/TMZ/lomustine) met the above criteria regardless of the baseline laboratory value (i.e., the post-baseline laboratory value does not need to be worse than the baseline laboratory value).

A listing of all ALT, AST, total bilirubin, and alkaline phosphatase values will be provided for all subjects who meet any of the criteria defined above.

11.5   Analysis of Vital Signs Using Criteria for Potentially Clinically Significant Vital Sign Values

Vital signs values will be assessed for potential clinical significance through the application of criteria developed by the sponsor as detailed in Table 10.

Table 10.   Criteria for Potentially Clinically Significant Vital Sign Values

Vital Signs Variables	Criterion	Definition of Potentially Clinically Significant Values
Systolic blood pressure	High	Value ≥ 160 mmHg
Diastolic blood pressure	High	Value ≥ 100 mmHg
Heart rate	Low	Value ≤ 50 bpm
	High	Value ≥ 120 bpm
Temperature	Low	Value ≤ 32°C
	High	Value > 40°C

The number and percentage of subjects with post baseline values meeting Criteria for Potentially Clinically Significant Vital Signs values will be summarized. The baseline will be defined as the last non-missing measurement collected prior to or on the first dose

of study drug (ABT-414/TMZ/lomustine). A separate listing will be provided that presents all of the subjects and values that meeting the criteria. No comparisons of the rates of subjects met the above criteria between the treatment arms will be performed.

12.0   Summary of Differences Between EORTC and AbbVie Analyses

EORTC and AbbVie Inc. may follow slightly different approaches for the analyses of certain types of study data. EORTC will perform analyses primarily for the purposes of certain clinical publications. Such analyses, unless also performed by AbbVie (perhaps using different approach), may be included in AbbVie's Clinical Study Report of this study.

While there may be other differences, the following is a summary of differences in EORTC planned analysis from the pre-specified analysis methods described above that AbbVie and EORTC have been able to identify at the time of drafting this SAP (paragraph numbering refers to sections of the present document):

3.0   Introduction

EORTC will perform efficacy and safety analyses using SAS Version 9.4 or higher (SAS Institute, Inc., Cary, NC 27513) under the Windows operating system.

5.0   Analysis Sets

Safety population: For EORTC analysis, subjects are classified and analyzed in the arm they were assigned at the time of randomization if they started the corresponding treatment. The safety profile of subjects who did not receive the allocated treatment is reported separately instead of being reported in the arm actually received, as done by AbbVie. The safety population for EORTC is thus strictly included in that of AbbVie and the approaches become identical if no subject switches between protocol treatment arms.

7.3   Prior and Concomitant Medications

EORTC does not perform this analysis.

9.1   Study Drug Treatment

EORTC does not perform longitudinal analysis of drug dose administered.

9.4   Protocol Deviations

EORTC will present the trial compliance per treatment arm as evaluated during the medical review. Frequencies and percentages of protocol deviations as described in the medical review plan will be tabulated overall and split by type of deviation and severity.

10.3   Efficacy Analyses

For QoL assessment data, Abbvie will only perform and present simple summary statistics at each timepoint and change from baseline analysis.

EORTC will not perform Time to Tumor Response (TTR) analysis.

10.4   Re-Randomization Test

EORTC will not perform the analysis for PFS assessed by IRC or ORR assessed by IRC.

10.5   Sensitivity Analyses

EORTC will only present PFS, ORR, BRR, duration of response analyzed based on assessment by local investigator per RANO criteria.

EORTC will perform OS analysis by fitting a Cox regression model with treatment arm and potential prognostic covariates using a stepwise procedure.

In case more than 10% are excluded from the per-protocol population, the primary efficacy analysis (both comparisons) will be performed in the per-protocol population at two-sided alpha 5%.

11.2   Analysis of Adverse Events

EORTC will not present frequencies and percentages for

- Any treatment-emergent adverse event leading to discontinuation of ABT-414
- Any treatment-emergent adverse event leading to discontinuation of TMZ
- Any treatment-emergent adverse event leading to discontinuation of lomustine
- Any treatment-emergent adverse event leading to dose reduction of ABT-414
- Any treatment-emergent adverse event leading to dose reduction of TMZ
- Any treatment-emergent adverse event leading to dose reduction of lomustine
- Any treatment-emergent adverse event leading to interruption of ABT-414
- Any treatment-emergent adverse event leading to interruption of TMZ
- Any treatment-emergent adverse event leading to interruption of lomustine
- Any treatment-emergent adverse event of special interest

EORTC will not perform the analyses of adverse events for the subset of subjects with EGFRvIII mutation

11.2.1   Adverse Events of Special Interest (AESI)

EORTC will not perform this analysis.

11.2.1.1   Analysis of Adverse Events of Special Interest (AESI)

EORTC will not perform this analysis.

11.2.1.2   Time to Onset of Adverse Event of Special Interest

EORTC will not perform this analysis.

11.4.2   Assessment of Hepatotoxicity

EORTC will not perform this analysis.

11.5 Analysis of Vital Signs Using Criteria for Potentially Clinically Significant Vital Sign Values

EORTC will not perform this analysis.

13.0 References

1. Freedman LS, White SJ. On the use of Pocock and Simon's method for balancing treatment numbers over prognostic factors in the controlled clinical trial. Biometrics. 1976;32(3):691-4.

2. Han K, Ren M, Wick W, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. Neuro Oncol. 2014;16(5):696-706.

3. Vredenburgh JJ, Cloughesy T, Samant M, et al. Corticosteroid use in patients with glioblastoma at first or second relapse treated with bevacizumab in the BRAIN study. Oncologist. 2010;15(12):1329-34.

4. Gorlia T, Stupp R, Brandes AA, et al. New prognostic factors and calculators for outcome prediction in patients with recurrent glioblastoma: a pooled analysis of EORTC Brain Tumour Group phase I and II clinical trials. Eur J Cancer. 2012;48(8):1176-84.

5. Carson KA, Grossman SA, Fisher JD, et al. Prognostic factors for survival in adult patients with recurrent glioma enrolled onto the new approaches to brain tumor therapy CNS consortium phase I and II clinical trials. J Clin Oncol. 2007;25(18):2601-6.

14.0   Summary of Major Changes

- Modify Section 5.0, Interim Futility Analysis to adjust observed information fraction.

  Rationale: Clarification and details regarding the nominal significance level for the final OS analysis.

- Modify Section 5.0, Analysis Sets to add EGFRvIII-Mutated Population.

  Rationale: To specify EGFRvIII-Mutated (Ev3M) Population is a subset of subjects who have EGFRvIII mutated tumors.

- Modify Section 6.0 Analysis Conventions to clarify baseline definition.

  Rationale: Clarification and details regarding missing measurement collected prior to randomization.

- Modify Section 9.0 Dosage to include dose intensity and relative dose intensity.

  Rationale: Additional assessment to summarize dose intensity and relative dose intensity of ABT-414, adjusting actual treatment duration.

- Modify Section 10.1 General Considerations to provide clarification of stratification factors.

  Rationale: Clarify that stratification factors based on ORTA are to be used.

- Modify Section 10.1 General Considerations to provide overall descriptions of analysis method and censoring rule for TTE.

  Rationale: Details were added for clarification.

- Modify Section 10.2 Efficacy Endpoints to add time to tumor response (TTR).

  Rationale: TTR was added as an exploratory endpoint to assess time to response.

- Modify Section 10.3 Efficacy Analyses to revise analysis of corticosteroid use.

  Rationale: Kaplan-Meier estimates of PFS by baseline steroid use status were added, taking PFS-censored subjects into account.

- Modify Section 10.3 Efficacy Analyses to clarify analysis of QoL.

Rationale: To clarify that Abbvie only performs and presents simple summary statistics and change from baseline analysis. Various QoL data modelings and compliance analysis will be performed by EORTC.

- Modify Section 10.4 Re-Randomization Test to include PFS and ORR assessed by IRC

  Rationale: PFS and ORR re-randomization analyses were added to support robustness of the efficacy results.

- Remove Section 10.7 Exploratory Analyses.

  Rationale: Removed the analyses from SAP, as they are to be performed in a data dependent manner.

- Added selected AE summaries to Section 11.2 Analyses of Adverse Events.

  Rationale: Details were added for selected AE summaries in Ev3M population.

- Added AESI to Section 11.2.1 Adverse Events of Special Interest (AESI).

  Rationale: Details were added for clarification.

  Removed Time to Resolution Analysis from Section 11.2.1.2 (Section 11.3 previously) Time to Onset of Adverse Event of Special Interest.

  Rationale: Analysis plan and methodological details will be available to integrated safety summary SAP.

- Added Sections 11.4.1 and 11.4.2 Assessment of Potentially Clinically Significant Laboratory Values and Assessment of Hepatotoxicity.

  Rationale: Details were added for clarification.

- Modified Section 11.5 Analysis for Vital Signs Using Criteria for Potentially Clinically Significant Vital Sign Values.

  Rationale: Table 10 was revised for clarification.

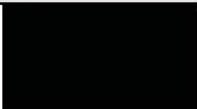- Removed Section 12.0 Pediatric Sub-Study.

  Rationale: A separate SAP for the pediatric sub-study will be provided.

Document Approval

Study M14483 - Statistical Analysis Plan Version 2 - 30May2017 (E3 16.1.9)

Version: 1.0   **Date:** 31-May-2017 09:13:18 PM   **Company ID:** 05312017-00F9F6815323FD-00001-en

Signed by:	Date:	Meaning Of Signature:
█████████	31-May-2017 01:31:27 PM	Approver
	31-May-2017 09:13:18 PM	Approver