STATISTICAL ANALYSIS PLAN

A Phase 1b/3, Multicenter, Trial of Talimogene Laherparepvec in Combination With Pembrolizumab (MK-3475) for Treatment of Unresectable Stage IIIB to IVM1c Melanoma (MASTERKEY-265)

Protocol Number: *20110265*

Version: *Version 4.0*

Date: *16 January 2020*

Authors: [redacted]

NCT Number: NCT02263508
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

Table of Contents

[redacted]

[redacted]

[redacted]

List of Tables

List of Figures

List of Appendices

[redacted]

Table of Abbreviations

Abbreviation/Acronym	Definition
AE	adverse event
BICR	blinded independent central review
BRAF^{V600}	serine/threonine protein kinase B-Raf V600
BOR	best overall response
BQL	below quantification limit
CI	confidence interval
CR	complete response
CRF	case report form
CT	computerized tomography
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DCR	disease control rate
DILI	drug induced liver injury
DLRT	dose level review team
DLT	dose limiting toxicity
DMC	Data Monitoring Committee
DMP	data management plan
DOR	duration of response
DpR	**depth of response**
DRT	data review team
DRR	durable response rate
DTP	data transfer plan
eCRF	electronic case report form
ECI	events of clinical interest
EOI	event of interest (Amgen)
EQ5D	EuroQoL-5D
HLA	human leukocyte antigen
HR	hazard ratio
HSU	**health state utility**
HSV, HSV-1	herpes simplex virus, herpes simplex virus type 1
IA	interim analysis
iBOR	best overall response (by modified irRC-RECIST)
iCR	complete response (by modified irRC-RECIST)
iCRR	complete response rate (by modified irRC-RECIST)
iDCR	disease control rate (by modified irRC-RECIST)
iDOR	duration of response (by modified irRC-RECIST)
iDRR	durable response rate (by modified irRC-RECIST)
iORR	objective response rate (by modified irRC-RECIST)

Abbreviation/Acronym	Definition
IP	investigational product
iPD	progressive disease (by modified irRC-RECIST)
IPD	Important protocol deviation
iPFS	progression-free survival (by modified irRC-RECIST)
iPR	partial response (by modified irRC-RECIST)
irRC	Immune-related Response Criteria
irRC-RECIST	Immune-related Response Criteria (irRC) simulating Response Evaluation Criteria in Solid Tumors (RECIST)
iSD	stable disease (by modified irRC-RECIST)
iUE	unevaluable (by modified irRC-RECIST)
IVRS	Interactive Voice Response System
KM	Kaplan-Meier
MedDRA	Medical Dictionary for Regulatory Activities
ND	not done
NN	non-CR/non-PD
OR	objective response (CR or PR)
ORR	objective response rate
OS	overall survival
PA	primary analysis
PD	progressive disease
PFS	progression free survival
PFU	plaque-forming unit
PRO	patient reported outcome
PT	preferred term
qPCR	real-time polymerase chain reaction
RECIST	response Evaluation Criteria in Solid Tumor
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SMQ	standardized MedDRA Queries
SOC	system organ class
SOP	standard Operating Procedure
SUSAR	Suspected Unexpected Serious Adverse Reaction
QLQ-C30	Quality of Life Questionnaire Core 30
UE	unevaluable for tumor evaluation
[redacted]	
WHODRUG	World Health Organization Drug Dictionary

1. Introduction

The purpose of this statistical analysis plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol for Talimogene Laherparepvec Study *20110265* (Amendment **5** dated **15 January 2020**). The scope of this plan includes the interim, primary and the final analyses that are planned and will be executed by the Biostatistics department unless otherwise specified.

The clinical study report (CSR) will be written based on the results of the primary analysis of Overall Survival (OS) for phase 3. Analysis of changes in PD-L1 expression by immunohistochemistry (eg, negative to positive) introduced by talimogene laherparepvec (for phase 1b) will also be included. A separate final analysis CSR will be drafted with the updated results from the final analysis at the completion of the study. Analysis of patient reported outcomes (PRO) endpoints will be described separately in a supplemental SAP (SSAP). [redacted]

Data collected and analyzed in Amgen-owned databases and systems will adhere to approved Data Element Standards and International Case Report Form (CRF) Standards established by Biomedical Data Stewardship Governance (BDSG).

2. Objectives

2.1 Primary

The primary objectives are:

Phase 1b: To evaluate the safety, as assessed by incidence of dose limiting toxicity (DLT), of talimogene laherparepvec in combination with pembrolizumab in subjects with previously untreated, unresectable, stage IIIB to IVM1c melanoma.

Phase 3: To evaluate the efficacy of talimogene laherparepvec with pembrolizumab versus placebo with pembrolizumab, as assessed by progression-free survival (PFS) (response evaluation by blinded independent central review (BICR) using modified Response Evaluation Criteria in Solid Tumors [RECIST] 1.1) and overall survival (OS).

2.2 Secondary

The secondary objectives are:

Phase 1b:

- To evaluate the efficacy of talimogene laherparepvec in combination with pembrolizumab as assessed by:
  - Confirmed objective response rate (ORR), best overall response (BOR), durable response rate (DRR), duration of response (DOR), disease control rate (DCR), and PFS (response evaluation by investigator using modified immune-related Response Criteria [irRC])
  - OS
- To evaluate the safety of talimogene laherparepvec in combination with pembrolizumab as assessed by incidence of treatment-emergent and treatment-related adverse events (AEs), and abnormal laboratory tests.

Phase 3:

- To evaluate the efficacy of talimogene laherparepvec with pembrolizumab versus placebo with pembrolizumab, as assessed by:
  - Complete response rate (iCRR) by blinded independent central assessed modified immune-related response criteria simulating response evaluation criteria in solid tumors (modified irRC-RECIST)
  - iPFS by blinded independent central assessed modified irRC-RECIST
  - OS in subjects excluding stage IVM1c per case report form (CRF)
  - ORR, BOR, DRR, DOR, and DCR (response evaluation by blinded independent central review assessed using modified RECIST 1.1), and iORR, iBOR, iDRR, iDOR, and iDCR (response evaluation by blinded independent central review assessed using modified irRC-RECIST)
- To evaluate the safety of talimogene laherparepvec with pembrolizumab versus placebo with pembrolizumab, as assessed by incidence of treatment-emergent and treatment-related adverse events, and abnormal laboratory tests.
- To evaluate patient reported outcomes (PRO) in phase 3 as assessed by the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30) Global Health Status/Quality of Life (GHS/QoL) subscale.

2.3 Exploratory

The exploratory objectives are:

Phase 1b:

- To evaluate whether talimogene laherparepvec induces changes in PD-L1 expression by immunohistochemistry (eg, negative to positive) in injected and non-injected lesions after starting talimogene laherparepvec in subjects whose tumors are PD-L1-negative prior to receiving talimogene laherparepvec in phase 1b.

[redacted]

Phase 3:

- iORR, iBOR, iDRR, iDOR, iDCR and iPFS (response evaluation by investigator-assessed modified irRC-RECIST)

[redacted]

- To evaluate the incidence of anti-pembrolizumab antibodies when pembrolizumab is administered in combination with talimogene laherparepvec or placebo.
- To evaluate the pharmacokinetics (PK) of pembrolizumab when administered in combination with talimogene laherparepvec or placebo.
- To evaluate PROs as assessed by the EORTC QLQ-C30 subscales.
- To estimate health state utility (HSU) values using the EuroQoL-5D-3 Level (EQ-5D-3L).

[redacted]

3. Study Overview

3.1 Study Design

This is a phase 1b/3, multicenter, clinical trial. The study will be conducted in 2 parts (phase 1b and phase 3) as described below.

3.1.1 Phase 1b

Phase 1b is an open-label, multicenter, single-arm study to evaluate the safety of talimogene laherparepvec in combination with pembrolizumab. Two doses (5 weeks) of talimogene laherparepvec will be administered first prior to addition of pembrolizumab to

approximately 20 subjects with previously untreated, unresectable, stage IIIB to IVM1c melanoma until disappearance of injectable lesions, CR, confirmed PD per modified immune-related Response Criteria (irRC; Wolchok et al, 2009; protocol Appendix D), intolerance of study treatment, 24 months from the date of the first dose of pembrolizumab, or end of study, whichever occurs first. In addition, phase 1b will evaluate potential blood and [redacted] that associate with response or resistance to talimogene laherparepvec in combination with pembrolizumab.

The DLRT was to initially meet when 6 subjects are evaluable for DLT. The DLRT was to declare the combination of talimogene laherparepvec and pembrolizumab tolerable if the incidence of DLT is < 33% during the DLT evaluation period. Refer to protocol section 3.1.1.1. for more details.

3.1.2 Phase 3

The phase 3 is a double-blind, placebo-controlled, multicenter randomized trial to evaluate the efficacy, as assessed by PFS and OS of talimogene laherparepvec with pembrolizumab versus placebo with pembrolizumab in subjects with unresectable, stage IIIB to IVM1c melanoma.

Approximately 660 subjects will be randomized 1:1 to receive the following:

- Arm 1: talimogene laherparepvec plus pembrolizumab
- Arm 2: placebo plus pembrolizumab

Randomization will be stratified by stage of disease: less advanced stages (IIIB, IIIC, and IVM1a) versus more advanced stages (IVM1b and IVM1c) and by prior serine/threonine protein kinase B-Raf (BRAF) inhibitor therapy: no prior BRAF inhibitor versus prior BRAF inhibitor with or without MEK inhibitor.

Talimogene laherparepvec/placebo and pembrolizumab treatment will be initiated simultaneously.

Talimogene laherparepvec/placebo will be administered until disappearance of injectable lesions, iCR, documented confirmed progressive disease (iPD) per modified irRC-RECIST (protocol Appendix F), intolerance of study treatment, 24 months from the date of the first dose of talimogene laherparepvec/placebo, or end of study, whichever occurs first.

Pembrolizumab will be administered until confirmed iPD per modified irRC-RECIST (protocol Appendix F), intolerance to treatment, 24 months from the date of the first dose

of pembrolizumab, or end of study, whichever occurs first. Discontinuation of treatment may be considered for subjects who have attained a confirmed iCR and have been treated for at least 24 weeks with pembrolizumab and had at least 2 treatments with pembrolizumab beyond the date when the initial CR was declared.

At least **seven** interim analyses for safety, efficacy and futility are planned to be reviewed by a Data Monitoring Committee (DMC). Refer to Section 8 for details on interim analyses.

3.1.3 Second Course (Retreatment)

Subjects enrolled in phase 1b who ended talimogene laherparepvec and/or pembrolizumab because they reached a CR but then subsequently progressed may be eligible for a second course treatment (retreatment) with talimogene laherparepvec and/or pembrolizumab are eligible to receive for up to 12 additional months at the investigator's discretion as long as the phase 3 remains open for treatment (ie, all subjects enrolled in phase 3 have not yet completed treatment).

3.1.4 Follow-up

3.1.4.1 Safety Follow-up

Subjects enrolled in phase 1b and phase 3 will be followed for safety approximately 30 (+7) days after the last dose of study treatment. If an end of treatment decision occurs > 30 (+7) days after the last treatment date, then the Safety follow-up should be performed as soon as possible (eg, within a week of the end of treatment decision). Applicable safety follow-up procedures should occur even if the subject has initiated new anticancer therapy. Serious adverse events observed by the investigators or reported by the subjects that occur within the earlier time of 90 (+7) days after the cessation of all study treatment or 30 (+7) days following cessation of treatment if the subject initiates new anticancer therapy will be reported.

3.1.4.2 Long-term Follow-up

Subjects enrolled in phase 1b and phase 3 will be followed for survival and subsequent anticancer therapies, and injected investigational product-related adverse events for every 12 weeks (± 28 days) for approximately 60 months after the last subject is randomized in phase 3.

3.2 Sample Size

3.2.1 Sample Size Considerations for Phase 1b

See protocol Section 10.2.1.

3.2.2 Sample Size Considerations for Phase 3

See protocol Section 10.2.2.

4. Study Endpoints and Covariates

4.1 Study Endpoints

4.1.1 Primary Endpoint(s)

4.1.1.1 Phase 1b

Incidence of DLT (protocol section 3.1.1.2).

4.1.1.2 Phase 3

- Progression-free Survival by BICR using modified RECIST 1.1
- Overall Survival (OS)

4.1.2 Secondary Endpoints

Phase 1b:

- Confirmed ORR (response evaluation by investigator using modified irRC).
- BOR, DCR, DRR, DOR and PFS (response evaluation by investigator using modified irRC)
- OS.
- Incidence of treatment-emergent and treatment-related adverse events (all AEs, grade ≥ 3 AEs, SAEs, fatal AEs, and AEs defined as events of interest), and abnormal laboratory tests.

Phase 3:

- iCRR by BICR using modified irRC-RECIST
- iPFS by BICR using modified irRC-RECIST
- OS in subjects excluding stage IVM1c per CRF
- ORR (CR + PR), BOR, DRR, DOR, and DCR (response evaluation by BICR using modified RECIST 1.1) and iORR (iCR + iPR), iBOR, iDRR, iDOR, and iDCR response evaluation by BICR using modified irRC-RECIST).
- Incidence of treatment-emergent and treatment-related adverse events (all AEs, grade ≥ 3 AEs, SAEs, fatal AEs, and AEs defined as events of interest, and abnormal laboratory tests.
- Changes in EORTC QLQ-C30 GHS/QoL subscale.

4.1.3 Exploratory Endpoints

Phase 1b:

- Changes in PD-L1 expression by immunohistochemistry (eg, negative to positive) in injected and non-injected lesions after starting talimogene laherparepvec in subjects whose tumors are PD-L1-negative prior to receiving talimogene laherparepvec in phase 1b.

[redacted]

Phase 3:

- iORR, iBOR, iDRR, iDOR, iDCR and iPFS (response evaluation by investigator using modified irRC-RECIST)

[redacted]

- Changes in the EORTC QLQ-C30 subscales.
- **Changes in** HSU estimates derived via EQ-5D-3L.

[redacted]

- Incidence of anti-pembrolizumab antibodies
- Descriptive statistics of trough and peak levels of pembrolizumab (MK-3475)

[redacted]

4.2 Planned Covariates

The following baseline covariates, including the stratification factors (phase 3) will be used to examine efficacy individually:

- Region, if applicable (USA or non-USA)
- Age: < 50, ≥ 50; < 65, ≥ 65; < 75, ≥ 75 years
- Individual disease stage: IIIB vs IIIC vs IVM1a vs IVM1b vs IVM1c
- Grouped disease stage: IIIB/C-IVM1a vs IVM1b/c (CRF)
- Grouped disease stage: IIIB/C-IVM1a vs IVM1b/c (IVRS)
- Grouped disease stage: IIIB/C-IVM1b vs IVM1c
- LDH: ≤ ULN vs > ULN
- Sex (female vs male)
- ECOG performance status (0 vs 1)
- Herpes simplex virus type 1 (HSV-1) serostatus (positive vs negative)
- The sum of longest, or shortest for nodal, diameters of baseline target lesions (SLD) determined by BICR (will be dichotomized based on median)

- **Baseline** PD-L1 status (positive vs **not positive**)
- Prior BRAF inhibitor (none vs BRAF inhibitor alone vs BRAF and MEK inhibitors in combination): if subjects included in each of the 2 BRAF related categories represent ≥ 10% of the study population
- Prior BRAF inhibitor (yes versus no BRAF inhibitor with or without MEK inhibitors) (IVRS)
- Baseline *BRAF*^{V600} mutation (yes vs no **vs missing/unknown**)

5. Hypotheses and/or Estimations

Phase 1b:

Talimogene laherparepvec in combination with pembrolizumab will be safe and well tolerated in subjects with previously untreated, unresectable, stages IIIB to IVM1c melanoma.

Phase 3:

Two clinical hypotheses will be evaluated independently in the phase 3. The first clinical hypothesis is that talimogene laherparepvec with pembrolizumab compared to placebo with pembrolizumab will improve PFS as evaluated by BICR using modified RECIST 1.1. The second clinical hypothesis is that talimogene laherparepvec with pembrolizumab compared to placebo with pembrolizumab will improve OS.

6. Definitions

Definitions described in this section apply to data points derived by the study team including investigator-determined endpoints, unless stated otherwise. The BICR Charter details the specifications for the data points provided by BICR.

Actual Follow-Up Time

Actual follow-up time for a subject is calculated from the randomization date (Phase 3) or first dose date (Phase 1b) to the last on-study date (ie death date, or date last known to be alive for subjects still alive).

Baseline

Baseline in general refers to study day 1. The baseline value of a parameter (eg, vital signs, laboratory tests, and tumor measurement) is considered to be the latest value prior to receiving any study drug (ie, on or prior to the first date of dosing) in the respective phase of the study. If the measurement is done on the same day as the first dose of any study drug and the exact measurement time relative to the first dose of any study drug is unknown, it will be assumed that the measurement is done before the administration of the first dose of any study drug. If a subject does not receive any study drug, baseline is the closest recorded measurement on or before the

enrollment/randomization date. *BRAF* mutational status, which is not sensitive to treatment, will be considered baseline regardless of the timing of ascertainment. For phase 1b, talimogene laherparepvec starts 5 weeks before pembrolizumab, therefore Week -5 is used as baseline. [redacted]

Best overall response (BOR) per modified irRC (Phase 1b)

BOR of complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD) or unevaluable (UE) will be derived based on investigator assessment per modified irRC. Overall visit response assessments occurring **on or** after the start of the first subsequent anticancer therapy, including complete or partial removal/reduction of any index lesion which contained melanoma on pathology evaluation or pathology results were unknown, will not be included. **The timepoint responses after initial confirmation of PD will not be used to derive BOR.** Confirmation of CR, PR, and PD is required as noted in the individual definitions for CR, PR and PD per modified irRC (refer to protocol Appendix D). For the derivation of BOR, the overall visit response will be considered SD if it is an unconfirmed CR or PR, and UE if it is either SD earlier than 84 days after the date of enrollment or an unconfirmed PD when confirmation of PD is required (ie, initial PD without clinical instability). As indicated in Table 1, BOR is defined as the best overall visit response in the following order: CR, PR, SD, PD, or UE.

Table 1. Matrix of Determining BOR per Modified irRC

Visit Overall Response Sequence	Examples	Best Overall Response	Specifications
, CR, CR,	PR, CR, CR CR, CR, PD	CR	A confirmatory CR must be at least 4 weeks (28 days) later; a subsequent CR within 28 days will not be valid for confirmation and will be ignored; the CR will also not be confirmed if there is a subsequent PR/SD/PD at any time prior to the next CR.
*, PR, PR, * *, PR, CR/PR, non-CR, *	PR, PR, PD PR, CR, PD CR, PR, PD	PR	Criteria for BOR=CR not met. A confirmatory PR/CR must be no less than 4 weeks (28 days) later; a subsequent PR/CR within 28 days will not be valid for confirmation and will be ignored; the PR will also not be confirmed if there is a subsequent SD/PD at any time prior to the next PR/CR.
*, SD, * *,CR, non-PR/CR, * *,PR, non-PR/CR, *	CR PD, CR PR PD, PR, SD SD PD,SD,PD	SD	Criteria for BOR=CR or PR not met. SD must be ≥ 84 days from date of enrollment; however, this is not required for an unconfirmed CR/PR.
*,PD, PD, * *, PDci, *	PD, PD PD, SD, PD, PD PD, UE, PD PDci	PD	Criteria for BOR= CR, PR, or SD not met. A consecutive confirmatory PD must be no less than 4 weeks (28 days) later unless the initial progression qualified for PDci.; PDci = **the last available** PD with "Nonconfirmed Disease Progression with Clinical Instability" as the reason for ending radiographic follow-up.
*, SD, * PD	SD UE, SD PD UE, PD	UE	Criteria for BOR=CR, PR, SD, or PD not met. SD must be < 84 days from date of enrollment.

Best overall response (iBOR) per modified irRC-RECIST (Phase 3)

iBOR of complete response (iCR), partial response (iPR), stable disease (iSD), progressive disease (iPD) or unevaluable (iUE) will be derived per modified irRC-RECIST based on investigator assessment and BICR assessment, respectively. Overall visit response assessments occurring **on or** after the start of the first subsequent anticancer therapy, including complete or partial removal/reduction of any target lesion which contained melanoma on pathology evaluation or pathology results were unknown, will not be included. **The timepoint responses after initial confirmed iPD will not be used to derive BOR.** Confirmation of iCR, iPR, and iPD is required as noted in the individual definitions for iCR, iPR and iPD per modified irRC-RECIST (refer to protocol Appendix F). For the derivation of iBOR, the overall visit response will be considered iSD if it is an unconfirmed iCR or iPR, and iUE if it is either iSD earlier than 84 days after the date of randomization or an unconfirmed iPD when confirmation of iPD is required (ie, initial iPD without clinical instability). As indicated in Table 2, iBOR is defined as the best overall visit response in the following order: iCR, iPR, iSD, iPD, or iUE.

Table 2. Matrix of Determining iBOR per Modified irRC-RECIST

Visit Overall Response Sequence	Examples	Best Overall Response	Specifications
, iCR, iCR,	iPR, iCR, iCR iCR, iCR, iPD	iCR	A confirmatory iCR must be at least 4 weeks (28 days) later; a subsequent iCR within 28 days will not be valid for confirmation and will be ignored; the iCR will also not be confirmed if there is a subsequent iPR/iSD/iPD at any time prior to the next iCR.
*, iPR, iPR, * *, iPR, iCR/iPR, non-iCR, *	iPR, iPR, iPD iPR, iCR, iPD iCR, iPR, iPD	iPR	Criteria for iBOR=iCR not met. A confirmatory iPR/iCR must be no less than 4 weeks (28 days) later; a subsequent iPR/iCR within 28 days will not be valid for confirmation and will be ignored; the iPR will also not be confirmed if there is a subsequent iSD/iPD at any time prior to the next iPR/iCR.
*, iSD, * *,iCR, non-iPR/iCR, * *,iPR, non-iPR/iCR, *	iCR iPD, iCR iPR iPD, iPR, iSD iSD iPD,iSD,iPD	iSD	Criteria for iBOR=iCR or iPR not met. iSD must be ≥ 84 days from date of randomization; however, this is not required for an unconfirmed iCR/iPR.
*,iPD, iPD, * *, iPDci, *	iPD, iPD iPD, iSD, iPD, iPD iPD, iUE, iPD iPDci	iPD	Criteria for iBOR= iCR, iPR, or iSD not met. A consecutive confirmatory iPD must be no less than 4 weeks (28 days) later unless the initial progression qualified for iPDci.; iPDci = **the last available** iPD with "Nonconfirmed Disease Progression with Clinical Instability" as the reason for ending radiographic follow-up.
*, iSD, * iPD	iSD iUE, iSD iPD iUE, iPD	iUE	Criteria for iBOR=iCR, iPR, iSD, or iPD not met. iSD must be < 84 days from date of randomization.

Best overall response (BOR) per modified RECIST 1.1 (Phase 3)

BORs will be derived based on modified RECIST 1.1 (Eisenhauer *et al.* 2009) utilizing response evaluations obtained from BICR. Per RECIST 1.1, a confirmation of CR or PR by a consecutive subsequent tumor assessment will not be required for the derivation of BOR. Overall visit response assessments occurring on or after the start of the first subsequent anticancer therapy, including complete or partial removal/reduction of any target lesion which contained melanoma on pathology evaluation or pathology results were unknown, will not be included. The timepoint responses after initial PD will not be used to derive BOR. A timepoint response report of Non-CR/Non-PD (NN), a valid timepoint response for a subject who has been identified by BICR to have only non-target disease per RECIST 1.1, will be reported as is for analysis purposes. For the derivation of BOR, the overall visit response will be considered UE if it is SD earlier than 84 days after the date of randomization. BOR is defined as the best overall visit response up to and including the first overall visit response of PD in the following order: CR, PR, SD, NN, PD or UE.

Complete response rate (iCRR) per modified irRC-RECIST (Phase 3)

Complete response rate (iCRR) per modified irRC-RECIST is defined as the incidence of **an** iBOR of iCR per modified irRC-RECIST in the respective analysis set.

Disease control rate (DCR) per modified irRC (Phase 1b)

Disease control rate (DCR) per modified irRC is defined as the incidence of a BOR of CR, PR or SD in the respective analysis set.

Disease control rate (iDCR) per modified irRC-RECIST (Phase 3)

Disease control rate (iDCR) per modified irRC-RECIST is defined as the incidence of **an** iBOR of iCR, iPR or iSD in the respective analysis set.

Disease control rate (DCR) per modified RECIST 1.1 (Phase 3)

Disease control rate (DCR) per modified RECIST 1.1 is defined as the incidence of a BOR of CR, PR or SD in the respective analysis set.

Duration of response (DOR) per modified irRC (Phase 1b)

Duration of response (DOR) is defined as the time from the date of an initial response that is subsequently confirmed to the earlier of PD (see definition of BOR per modified irRC) or death. Subjects who have not ended their response at the time of analysis will

be censored at their last evaluable tumor assessment before start of the first subsequent anticancer therapy.

Duration of response (iDOR) per modified irRC-RECIST (Phase 3)

Duration of response (iDOR) is defined as the time from the date of an initial response that is subsequently confirmed to the earlier of iPD (see definition of BOR per modified irRC-RECIST) or death. Subjects who have not ended their response at the time of analysis will be censored at their last evaluable tumor assessment before the start of the first subsequent anticancer therapy.

Duration of response (DOR) per modified RECIST 1.1

Duration of response (DOR) is defined as the time from the date of an initial response of CR or PR to the earlier of PD (see definition of BOR per modified RECIST 1.1) or death. Subjects who have not ended their response at the time of analysis will be censored at their last evaluable tumor assessment date before start of the first subsequent anticancer therapy.

Durable response rate (DRR) per modified irRC (Phase 1b)

DRR is defined as the percent of subjects with a CR or PR per modified irRC with a DOR ≥ 6 months. One month will be calculated based on 365.25 days per year.

Durable response rate (iDRR) per modified irRC-RECIST (Phase 3)

iDRR is defined as the percent of subjects with **an** iCR or iPR per modified irRC-RECIST with a DOR ≥ 6 months. One month will be calculated based on 365.25 days per year.

Durable response rate (DRR) per modified RECIST 1.1 (Phase 3)

DRR is defined as the percent of subjects with a CR or PR per modified RECIST 1.1 with a DOR ≥ 6 months. One month will be calculated based on 365.25 days per year.

Measurable Tumor Burden

Measurable tumor burden is defined as the sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions and up to 10 (maximum 5 per organ) new measurable lesions. The measurable tumor burden will be censored at the start of the first subsequent anticancer therapy, or at the time of complete or partial removal/reduction of any target lesion which contained melanoma on pathology evaluation or pathology results were unknown, whichever occurs first.

Depth of Response (DpR)

DpR is defined based on the maximum post-baseline measurable tumor burden decrease from baseline measurable tumor burden where maximum post baseline-measurable tumor burden decrease is categorized as: G0 = no shrinkage or increase, G1 = >0 ≤25%, G2 = >25-50%, G3 = >50-75%, G4 = >75- < 100%, and G5 = 100%.

End of Study

Primary Completion: The primary completion date is defined as the date when the last subject is assessed or receives an intervention for the purposes of final collection of data for the primary endpoint(s).. The primary completion date is the date when data for the primary endpoints are last collected for the purpose of conducting the primary analysis.

If the study concludes prior to the primary completion date originally planned in the protocol (ie, early termination of the study), then the primary completion will be the date when the last subject is assessed or receives an intervention for evaluation in the study (ie, last subject last visit).

End of Study: The end of study date is defined as the date when the last subject

across all sites is assessed or receives an intervention for evaluation in the study (ie, last subject last visit), following any additional parts in the study (eg, long term follow-up), as applicable.

Evaluable tumor assessment

An overall visit response other than **i**UE **or UE**.

Event of Interest (EOI) and Events of Clinical Interest (ECI)

EOIs are defined in version-controlled search strategies maintained by Global Regulatory Affairs and Safety (GRAAS) at Amgen for talimogene laherparepvec and those for ECIs for pembrolizumab will be provided by Merck before the analysis.

Investigational product (IP)

Investigational product (IP) refers to talimogene laherparepvec/placebo (also known as “Amgen IP”) and Pembrolizumab (also known as “non-Amgen IP”) in this study. IP is also referred as study drug, or study therapy.

Last date known to be alive

For subjects not known to have died, their last date known to be alive will be determined as the latest date associated with clinic visits including, for example, but not limited to the following:

1. Survival status: Last subject status date if status of subject is “Alive”
2. End of study date if the primary reason for ending study is not “Lost to follow-up” or “Death”
3. Any Subsequent anti-tumor therapy initialization date
4. AE start or end date
5. Procedure and concomitant medication start/end date
6. Date of tumor assessments
7. Date of herpetic lesion swabbing
8. Visit date of vital signs; physical exam; ECOG; local lab and central lab
9. Date of last investigational product administration

Objective response rate (ORR) per modified irRC (Phase 1b)

ORR is defined as the incidence of a BOR of CR or PR per modified irRC in the respective analysis set. Subjects who do not have any post-baseline tumor assessments will be regarded as non-responders.

Objective response rate (iORR) per modified irRC-RECIST (Phase 3)

iORR is defined as the incidence of a**n** iBOR of iCR or iPR per modified irRC-RECIST in the respective analysis set. Subjects who do not have any post baseline tumor assessments will be regarded as non-responders.

Objective response rate (ORR) per modified RECIST 1.1 (Phase 3)

ORR is defined as the incidence of a BOR of CR or PR per modified RECIST 1.1 in the respective analysis set. Subjects who do not have any post-baseline tumor assessments will be regarded as non-responders.

Overall Survival (OS)

Overall survival is defined as the interval from randomization (phase 3) or first dose (phase 1b) to death from any cause. Subjects without an event will be censored at the last known to be alive date.

Progression-free survival (PFS) per modified irRC (Phase 1b)

PFS per modified irRC is defined as the interval from first dose (phase 1b) to the earlier event of PD or death from any cause. Subjects without an event will be censored at their

last evaluable tumor assessment **if available; otherwise will be censored on Study Day 1**.

Progression-free survival (iPFS) per modified irRC-RECIST (Phase 3)

iPFS per modified irRC-RECIST is defined as the interval from randomization (phase 3) to the earlier event of iPD or death from any cause. Subjects without an event will be censored at their last evaluable tumor assessment **if available; otherwise will be censored on randomization date.**

Progression-free survival (PFS) per modified RECIST 1.1 (Phase 3)

PFS per modified RECIST 1.1 is defined as the interval from randomization to the earlier event of PD per modified RECIST 1.1 or death from any cause. Subjects without an event will be censored at their last evaluable tumor assessment **if available; otherwise will be censored on randomization date**.

Safety follow-up visit

Safety follow-up visit will be performed approximately 30 (+7) days after the last dose of IP. In addition, serious adverse events will be reported that occur through 90 (+7) days after the cessation of all study treatment or 30 (+7) days following cessation of treatment if the subject initiates new anticancer therapy, whichever is earlier.

Serious adverse events (SAE)

Serious adverse events (SAE) are defined as any adverse event categorized as serious as determined by investigator per protocol defined criteria.

Study day

For phase 3, study day is calculated from the first day IP is administered (ie, non-zero dosing) to the subject. Study day = visit date – first dose date +1 if visit date is on or after the first dose date. Study day = visit date – first dose date, if visit date is before the first dose date.

For phase 1b, study day is calculated from the first day pembrolizumab is administered (ie, non-zero dosing) to the subject. Study day = visit date – first pembrolizumab dose date +1 if visit date is on or after the first dose date. Study day = visit date – first pembrolizumab dose date, if visit date is before the first dose date.

Study Week 0

For phase 3, the start of IP administration (ie, non-zero dosing) to the subject is study week 0. Study day 1 is corresponding to study week 0.

For phase 1b, the start of pembrolizumab is study week 0. Study day 1 is corresponding to study week 0.

Study Week -5

In phase 1b, the start of talimogene laherparepvec is study week -5.

Treatment period

Treatment period is defined as the period between the first date of study therapy administration and 30 days after the last study therapy administration.

Treatment-emergent Adverse Events (TEAE)

Treatment-emergent adverse events are defined as any adverse event occurring after initiation of the first dose of study therapy through 30 days after the last administration of study therapy. Adverse events that occur on the same day as the first dose date of study therapy will be treated as treatment emergent events unless indicated otherwise (for example, if an event occurs on the same date as the first administration of study therapy and the check box indicating prior to the first dose of study therapy is checked on eCRF, then the event will not be counted as a treatment-emergent AE).

Treatment-emergent serious adverse events (TESAE)

Treatment-emergent serious adverse events are defined as any serious adverse event occurring after initiation of the first dose of study therapy through 90 days after the last administration of study therapy or 30 days after the last administration of study therapy if the subject initiates new anticancer therapy, whichever, is earlier. Serious adverse events that occur on the same day as the first dose date of study therapy will be treated as treatment emergent serious events unless indicated otherwise (for example, if an event occurs on the same date as the first administration of study therapy and the check box indicating prior to the first dose of study therapy is checked on eCRF, then the event will not be counted as a treatment-emergent serious adverse event).

[redacted]

7. Analysis Subsets

7.1 DLT Analysis Set

The DLT analysis set will include DLT-evaluable subjects enrolled in phase 1b who have had the opportunity to be on treatment for at least 6 weeks from the initial dosing of pembrolizumab and received at least 2 doses of talimogene laherparepvec and 2 doses of pembrolizumab in combination (ie, on the same day) or those who otherwise experience a DLT within 6 weeks after starting the combination therapy.

7.2 Full Analysis Set (FAS)

The primary analysis of all efficacy endpoints for phase 3 part of the study, unless noted otherwise, will be conducted on the full analysis set defined as all randomized subjects. All subjects will be analyzed according to treatment to which they are randomized.

7.3 Safety Analysis Set

The safety analysis set will include all subjects who received at least 1 dose of talimogene laherparepvec (or placebo in phase 3) or pembrolizumab. The safety analysis set will be defined separately for the phase 1b and phase 3. All subjects will be analyzed according to the treatment they received. Subjects will be classified in the talimogene laherparepvec and pembrolizumab arm for safety analyses if they received at least one dose of talimogene laherparepvec at any time during the treatment period.

7.4 PD-L1 Negative Analysis Set

The PD-L1 negative analysis set will include all subjects that are PD-L1 negative prior to randomization (phase 3) /first IP dose (phase 1b).

7.5 PD-L1 Positive Analysis Set

The PD-L1 positive analysis set will include all subjects that are PD-L1 positive prior to randomization (phase 3) /first IP dose (phase 1b).

7.6 qPCR Suspicious Lesion Swab Analysis Set

The qPCR Suspicious Lesion Swab Analysis Set will include subjects who are enrolled (phase 1b) or in the FAS (phase 3), receive at least one dose of talimogene laherparepvec / placebo, and have at least one swab sample evaluable result from lesions suspicious to be herpetic in origin during the study. Evaluable samples are sample with either positive, below quantification limit (BQL), or not detected results.

7.7 Per Protocol Analysis Set

The Per Protocol Analysis Set is a subset of the Full Analysis Set (phase 3) which includes subjects who do not have important protocol deviations (IPDs) that are considered to have an impact on efficacy outcomes.

Subjects with any of the following eligibility criteria IPDs will be removed from the Per-Protocol Analysis Set:

- Subject does not have histologically confirmed diagnosis of melanoma
- Subject does not have unresectable stage IIIB, IIIC, IVM1a, IVM1b, or IVM1c melanoma
- Subject is not a candidate for intralesional therapy

Additionally, subjects assigned to the talimogene laherparepvec arm, but who did not receive the first two consecutive doses of talimogene laherparepvec, will be excluded from the Per Protocol Analysis Set

For purposes of the Per-Protocol sensitivity analyses, subjects who (a) received the investigational product not matching the subject's randomized treatment group, or (b) have been identified to have an IPD under the category of prohibited concomitant medications during treatment, all event data collected after the first occurrence of the violation will be excluded from the analysis, and the endpoints analyzed as time-to-event will be censored at the time of the first violation for subjects who have not had an event prior to the violation date.

7.8 Retreatment Analysis Set

The retreatment analysis set will include subjects receiving second-course of treatment in retreatment phase. The option of retreatment is available only for subjects in phase 1b as documented in country-specific protocols.

7.9 Subgroup Analyses

Efficacy analyses specified for the Full Analysis Set for phase 3 will be repeated within the following subgroups:

- PD-L1 status (positive, not positive) at baseline
- Disease stage (less advanced [stage IIIB/C-IVM1a], more advanced [stage IVM1b/c]) at baseline (IVRS)
- Prior treatment with BRAF inhibitor (IVRS)
- Region, if applicable (USA or non-USA)
- Age: < 65, ≥ 65; < 75, ≥ 75 years
- LDH: ≤ ULN vs > ULN
- Sex (female vs male)
- ECOG performance status (0 vs 1)
- HSV-1 serostatus (positive vs negative vs missing/unknown)
- The sum of longest, or shortest for nodal, diameters of baseline target lesions (SLD) determined by BICR **(≤ median of SLD vs > median of SLD)**
- Baseline *BRAF*^{V600} mutation (yes vs no vs missing/unknown)

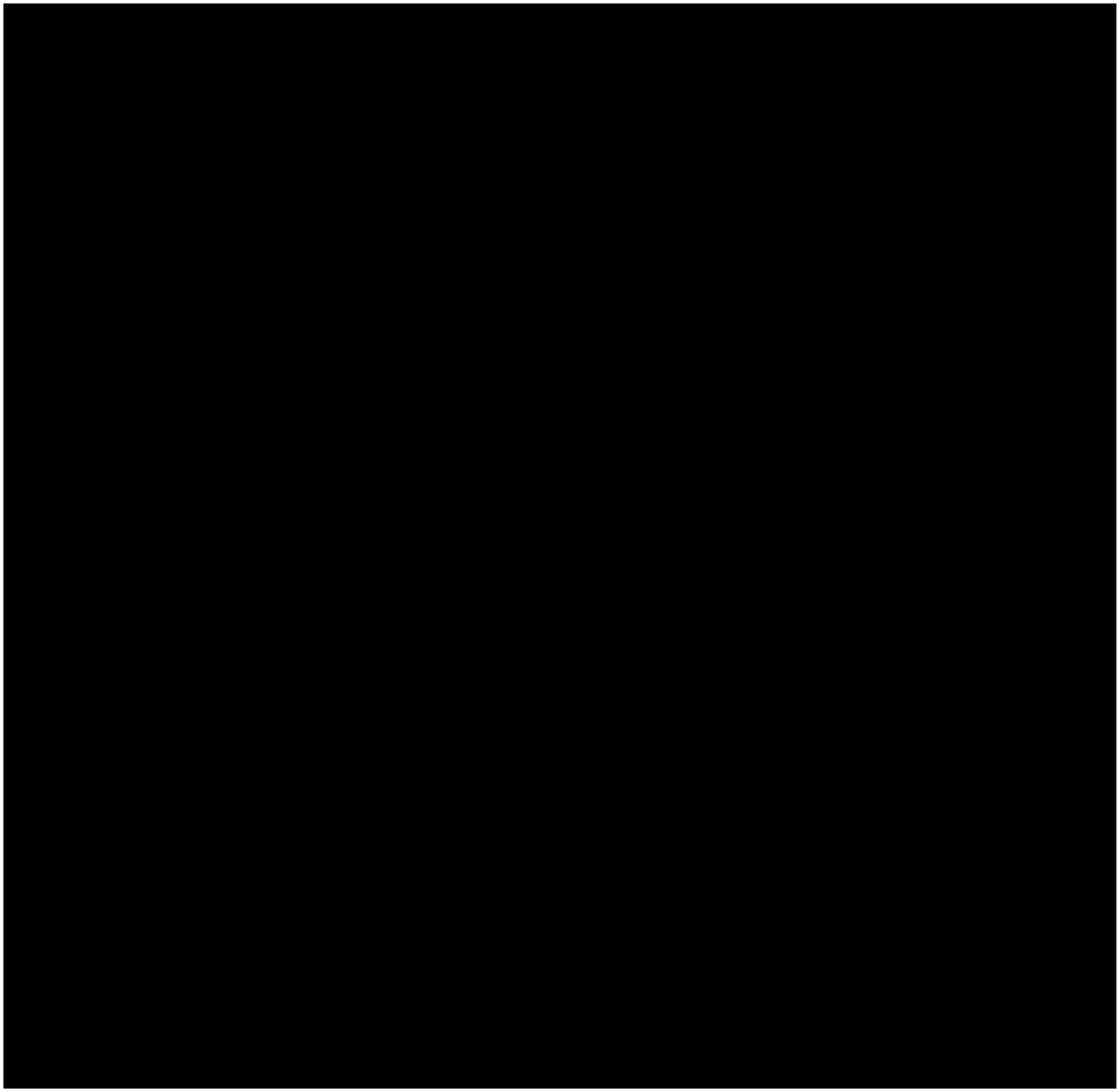
The sum of longest, or shortest for nodal, diameters of baseline target lesions (SLD) will be dichotomized into two groups based on the median calculated at the time of analysis.

These subgroups will be re-examined for appropriateness and may be re-categorized or omitted (due to small sample size, for example, if there are < 10% of subjects within a subgroup) before unblinding. The analyses of these subgroups will be exploratory in nature.

The treatment effects in subgroups will be estimated using stratified odds ratio for binary endpoints; and stratified HR for time to event endpoints. Stratification factors in the analysis model will include the randomization stratification factors as well as PD-L1 status at baseline (positive vs not positive). In the case of a subgroup analysis of one of the stratification factors, only the other two stratification factors will be included in the models. To evaluate the efficacy of talimogene laherparepvec in combination with pembrolizumab in subgroups defined by the stratification factors and key covariates, the treatment effect will be investigated using the Gail and Simon quantitative and qualitative interaction tests (1985). Post hoc power for a nominal 2-sided 5% significance level test

may be calculated to help aid in the assessment (see code fragment in Appendix C). The adjusted hazard ratio of the treatment effect with the corresponding 95% confidence interval and p-value adjusting for each of the subgroup variables listed above, using the stratified Cox proportional hazards (PH) model, as well as adjusting for all subgroup variables simultaneously in a multiple regression analysis, will be presented.

Subgroup analysis of exposure by age (< 65 vs ≥ 65, < 75 vs ≥ 75 years), gender and race (white vs non-white) will be presented. Subgroup analysis of adverse events by age, gender, and race will also be presented.



8. Interim Analysis and Early Stopping Guidelines

8.1 Interim Analyses

A DLT interim safety analysis was planned and conducted for phase 1b. Data were reviewed by a DLRT.

All analyses prior to the primary OS analysis will be conducted by an external independent biostatistics group (IBG) and reviewed by the DMC per the DMC charter. The analyses will include safety monitoring, iORR/iDCR futility, primary and possibly interim PFS, and interim OS (including OS futility) analyses. The study team is responsible for conducting the primary analysis of OS and all subsequent analyses.

8.1.1 DLT Safety Analysis (Phase 1b)

An interim safety analysis in phase 1b to evaluate DLT included 6 subjects who were eligible for DLT evaluation. A DLRT consisting of the Amgen study team, including at least one clinician, safety representative, and biostatistician, at least one representative from the Merck study team, and at least one investigator participating in the study who recruited subjects into the phase 1b, was to review the safety data to evaluate possible DLTs. This team was to recommend either to enroll more subjects for DLT evaluation in the phase 1b, to prematurely stop enrollment into the phase 1b, or to declare that the combination is tolerable and open the phase 3. The DLRT would review additional safety data from subjects enrolled in the phase 1b beyond the 6 DLT evaluable subjects to further assess the overall safety profile of talimogene laherparepvec in combination with pembrolizumab.

8.1.2 Interim Safety Analysis (Phase 3)

An independent Data Monitoring Committee (DMC) will review interim safety analyses to evaluate the safety of talimogene laherparepvec in combination with pembrolizumab in the target population for the phase 3. The first two safety analyses occurred after approximately 20 and 100 total subjects had enrolled in phase 3, respectively, who had

the opportunity to be followed for at least 6 weeks after initiating study treatment. Additional safety review meetings may occur every 6 months after the interim iORR and iDCR futility review meeting (Section 8.1.3) until the primary analysis of OS. Unblinded safety tables and listings by treatment arm will be prepared for the DMC by IBG according to the DMC charter. All available safety data from the phase 1b part of the study was provided to DMC at the time of the first planned interim safety analysis (ie, at 20 subjects trigger) in phase 3.

8.1.3 Interim iORR and iDCR Futility Analysis (Phase 3)

An interim futility analysis of iORR and iDCR per BICR using modified irRC-RECIST is planned when approximately 160 subjects (ie, 80 per arm) have been enrolled and have had a chance to be followed for at least 24 weeks (+2 weeks). This interim futility analysis will be conducted by the DMC with support from an IBG. The analysis will be limited to those subjects who had the potential to be followed for at least 24 weeks (+ 2 weeks). All tumor response data will be included up to 26 weeks from the date of first dose of study therapy for the 160th enrolled subject (or randomization date if subject is not dosed).

Beta (8.75, 16.25) and Beta (12.5, 12.5) prior distributions are assumed for the true placebo plus pembrolizumab arm (arm 2) iORR and iDCR in the trial population with means equal to 0.35 and 0.50, respectively, and both with the precision equal to a sample size of 25. Beta (10.08, 10.92) and (14.91, 6.09) prior distributions are assumed for the iORR and iDCR for the combination of pembrolizumab and talimogene laherparepvec (arm 1) with means equal to 0.48 and 0.71, respectively, and both with a concentration parameter of 21. The priors for the combination correspond to the results from a 24-week minimum follow-up analysis of the 21 subjects in the phase 1b part of the study.

Beta posterior distributions for the iORR and iDCR will be calculated for each arm based on the corresponding prior distribution and the observed study iORR and iDCR results. The iORR futility criterion will be a posterior probability < 0.65 of at least a 6% absolute iORR increase (Arm 1 – Arm 2), and the iDCR futility criterion will be a posterior probability < 0.75 of at least a 10% absolute iDCR increase. The DMC will declare the combination therapy futile if both the iORR and iDCR futility criteria are met; however, the DMC's recommendation will be non-binding and therefore the sponsor will make the final decision to terminate the study prematurely.

8.1.4 Interim Efficacy and Futility Analysis (Phase 3)

There are **5** planned analysis time points for hypothesis testing, including primary and/or interim analyses of the 5 hypotheses (H): H_1, associated with OS; H_2, PFS by BICR assessed modified RECIST 1.1; H_3, iPFS by BICR assessed modified irRC-RECIST; H_4, iCRR by BICR assessed modified irRC-RECIST; and H_5, OS subgroup excluding IVM1c. The analysis times are defined by the following event-driven timepoints with the DMC responsible for Analyses #1**, #2, and #3.**

1. **OS interim analysis and** PFS primary analysis at 407 **PFS** events (DMC).
2. OS interim analysis at 282 events (including a futility analysis) (DMC).
3. **OS interim analysis at 315 events (DMC).**
4. OS primary analysis at 346 events.
5. OS subgroup primary analysis at earlier of 208 events in subgroup or 5 years minimum follow-up.

Note that there is no sequential order in these **5** analysis times (with the exception that Analysis #2 is before Analysis #3 **and Analysis #3 is before analysis #4)** and the analyses may be combined. Each analysis time may also include interim or primary analyses for the other endpoints. Endpoints to be analyzed, and the type of analysis (ie, interim or primary), are presented in Table 3 and are dependent on which hypotheses have been rejected. For example, testing for iCRR and iPFS at an analysis timepoint will be conditional upon previously rejecting either H_1 or H_2. O'Brien-Fleming group sequential boundaries based on the Lan-DeMets approach (O'Brien and Fleming, 1979) will be used for testing endpoints with interim analyses. Multiple testing procedures will utilize the graphic-based Maurer-Bretz approach (Section 10.5.1.1).

The information fractions will be based on the **observed** events included at the time of analysis for testing of all time to event endpoint null hypotheses (ie, H_1, H_2, H_3, and H_5) and it will be based on the number of subjects with potential follow-up time of at least 36 weeks for testing the iCRR null hypothesis (ie, H_4). The maximum information for testing iCRR is all randomized subjects. Each analysis time will have a corresponding database snapshot; however, there will not be a separate snapshot for each endpoint tested at the same analysis time.

1. The initial alpha allocation for testing the hypothesis of PFS (H_2) is 0.5% and 4.5% for OS (H_1). There will not be a nominal alpha passed back to H_2 after rejecting H_1 or other hypotheses at any analysis.
2. An interim OS analysis will be performed at Analysis #1 with all **observed** death events, regardless if H_2 is rejected. In the event when Analysis #2 occurs prior to Analysis #1, then it will include an interim analysis to test H_2 for PFS (using all PFS events) and, if H_2 is not rejected, it will be tested again at the planned time for Analysis #1. However, if the projected time for achieving 407 PFS events is within 4 months from the projected 282 OS events time, then Analysis #2 will be combined with Analysis #1 and be performed at Analysis #1 using all death events with PFS times censored after the 407th PFS event.
3. At each of the **5** planned analysis times, it will be considered an interim analysis of iCRR and iPFS, rather than the primary analysis, if the maximum information has not been reached for the respective endpoint at that time. For example, if H_3 is tested in an analysis that includes fewer than 256 iPFS events, then it will be an interim analysis of iPFS that includes all observed events; otherwise, it will be the primary analysis with all iPFS times censored after the date of the 256th event. If all randomized subjects do not have at least 36 weeks follow-up in an analysis, then it will be an interim analysis for iCRR only including those subjects with at least 36 weeks follow-up; otherwise, it will be the primary analysis and include all randomized subjects. With the exception of PFS, the hypothesis for an endpoint may be re-tested after its primary analysis using its primary analysis data if alpha is subsequently propagated to it from the rejection of another hypothesis.
4. Assuming Analysis #1 is before Analysis #2, and at Analysis #1 the null hypothesis of PFS (H_2) is rejected, the hypotheses H_3 and H_4 will be tested. H_1 will be tested at Analysis #1 regardless of whether H_2 is rejected. The iPFS (H_3) and iCRR (H_4) can be tested if H_2 is not rejected but H_1 is significant. Likewise, OS subgroup (H_5) can be tested at Analysis #1 if any of the following set of hypotheses are rejected: (1) H_2 and H_3; (2) H_2 and H_4; (3) H_1 and H_3; or (4) H_1 and H_4.
5. If projections indicate 256 iPFS events will occur in the same timeframe as either Analysis #1 or Analysis #2, then the option will be considered to extend the data cutoff to include all 256 iPFS events. In this event, Analysis #1 will have PFS times censored after the 407th PFS event and Analysis #1 or #2 will include all OS events.
6. The OS hypothesis (H_1) will be tested at Analyses #1-**4** with all (Analysis #1) or required number (Analysis #2**,** Analysis #3**, and Analysis #4**) of events in the absence of its prior rejection (or futility at Analysis #2).
7. The OS subgroup hypothesis (H_5) will be tested at the first and all subsequent analysis times when alpha can be propagated to it from the rejection of H_1 or H_2 followed by H_3 and/or H_4. All OS events in the subgroup at the end of 5 years follow-up will be used if 208 events have not been observed. If H_1 is not rejected by Analysis #**4,** it will be tested again using events at Analysis #**4** if alpha can be propagated from the rejection of H_5 at Analysis #**5.**
8. Likewise, H_3 and H_4 may be re-tested if not rejected by their primary analysis using their respective primary analysis data if H_1 is subsequently rejected.

Table 3 illustrates possible hypothesis tests at the **5** planned analyses.

Table 3. Study Analyses and Possible Hypothesis Tests

		Hypothesis Tested (Analysis Type)				
Analysis	**Timing**	**H_1 (OS)**	**H_2 (PFS)**	**H_3 (iPFS)**	**H_4 (iCRR)**	**H_5 (OS subgroup)**
#1	407 PFS events	(IA)	(PA)	If reject H_1 or H_2 (IA if < 256 events)	If reject H_1 or H_2 (IA if includes fewer than all randomized)	If reject H_2 and either H_3 or H_4, or reject H_1 and either H_3 or H_4 (IA if < 208 events)
#2	282 OS events	(IA)	n/a (IA if Analysis #2 is performed first)	If reject H_1 or H_2 (PA on first 256 events)	If reject H_1 or H_2 (PA if includes all randomized)	If reject H_2 and either H_3 or H_4, or reject H_1 and either H_3 or H_4 (IA if < 208 events)
#3	**315 OS events**	**(IA)**	**n/a**	**If reject H_1 or H_2 (PA on first 256 events)**	**If reject H_1 or H_2 (PA if includes all randomized)**	**If reject H_2 and either H_3 or H_4, or reject H_1 and either H_3 or H_4 (IA if < 208 events)**
#4	346 OS events	(PA)	n/a	See Analysis #2 **& #3** above	See Analysis #2 **& #3** above	See Analysis #2 **& #3** above
#5	Earlier of 208 OS events in subgroup or 5 years minimum follow-up	If reject H_2 previously and H_5 (PA at first 346 events)	n/a	See Analysis #2 **& #3** above	See Analysis #2 **& #3** above	See Analysis #2 **& #3** above (PA)

iCRR = complete response rate (by modified irRC-RECIST); iPFS = progression free survival (by modified irRC-RECIST); H = hypothesis; IA = interim analysis; n/a = not applicable; OS = overall survival; PA = primary analysis; PFS = progression free survival

Futility Analysis for OS

At the 282 OS events interim analysis (ie, Analysis #2), if hypothesis H_1 is not rejected, then an OS futility analysis will be evaluated by the DMC. The interim OS futility

analysis will be non-binding and define futility as a conditional power < 10% assuming a constant treatment effect (Denne, 2001); the corresponding OS futility boundary is 0.89, assuming a constant HR = 0.70. OS futility will also be evaluated by the DMC considering a possible non-constant treatment effect that increases over time (eg, a 2 interval model with a cutpoint between 4 to 8 months and constant treatment effect in each interval, where the second interval has a larger effect than the first one). To limit the probability of futility < 0.02, the futility boundary for the observed HR at 282 OS events analysis will be 0.93. Guidelines for this possible late effect are provided in the DMC charter amendment. Operating characteristics of the primary analysis of PFS and sequential tests of OS are shown in Table 4 based on the OS events expected at Analysis #1.

Table 4. Analysis Timing, Nominal Significance Levels, and Boundary Properties

Analysis	Data cut-off (mon)	Endpoint (scenario)	Event (% goal)	Efficacy nominal 2-sided α	Futile HR [b]	Prob. Futile[b]		Power total
						H_0	H_1	
PFS PA	**44**	PFS	407 (100)	0.0050				90%
OS IA_1	**44**	OS (PFS-)	**252 (73)[a]**	**0.0150**				**66%**
		OS (PFS+)		**0.0173**				**67%**
OS IA_2	**52**	OS (PFS-)	282 (82)	**0.0187**	0.89	84%	2%	76%
		OS (PFS+)		**0.0212**	0.89	83%	2%	77%
OS IA_3	**61**	**OS (PFS-)**	**315 (91)**	**0.0261**				**84%**
		OS (PFS+)		**0.0292**				**85%**
OS PA	**75**	OS (PFS-)	346 (100)	**0.0336**				**90%**
		OS (PFS+)		**0.0372**				**90%**

IA = interim analysis; H_0 = null hypothesis; H_1 = alternative hypothesis; HR = hazard ratio; OS = overall survival; PFS = progression free survival; PA = primary analysis.
O'Brien-Fleming spending α for OS; futility if conditional power < 10%; H_0: HR = 1 and H_1: HR = 0.67 (PFS) and 0.70 (OS). Testing OS at an overall 2-sided 0.045 if PFS is not significant at 2-sided 0.005 (eg, PFS-); testing OS at 0.05 when all other null hypotheses are rejected (eg, PFS+). [redacted]

[a] The events for OS IA_1 are not pre-determined.
[b] The futile HR and probabilities are shown for the conditional power criterion only which assumes a constant OS effect.

9. Data Screening and Acceptance

9.1 General Principles

The objective of the data screening is to assess the quantity, quality and statistical characteristics of the data relative to the requirements of the planned analyses.

9.2 Data Handling and Electronic Transfer of Data

[redacted]

9.3 Handling of Missing and Incomplete Data

Every effort will be made to obtain complete data in the clinical study. Partial or missing dates of adverse events and concomitant medications will be imputed. Adverse events with missing IP relatedness, seriousness, or CTCAE severity grades are included in TEAE as long as the events qualify for the reporting period. Events with missing relatedness, seriousness, and severity grades will be excluded from treatment-related, serious, and with a CTCAE grade of 3 or higher AE analysis, respectively. [redacted]

9.4 Detection of Bias

Lack of protocol compliance may introduce potential bias in the estimation of protocol endpoints. All important protocol deviations (IPDs) will be reported, documented and stored in eClinical (a clinical trial management system). IPD reports will be produced [redacted] by the study manager and will be regularly reviewed in the study team's IPD review meetings as well as before analysis.

Protocol compliance will be examined by tabulating subjects with IPDs. Bias will be evaluated, in part, by conducting sensitivity analyses using the Per Protocol Analysis Set (see Section 7.7).

9.5 Outliers

Descriptive statistics will be used to identify potential outliers in key variables. Suspected outliers will be included in all analyses unless there is sufficient scientific justification to exclude them.

9.6 Distributional Characteristics

All binary endpoints will follow a binomial distribution. A beta prior is assumed in the Bayesian model for the interim iORR/iDCR futility analysis. The KM estimates for the probability of time-to endpoints are based on non-parametric methods. Treatment comparisons in time-to endpoints based on the Cox proportional hazards (PH) model are semi-parametric. Assumption of proportionality in hazards will be examined graphically with the Schoenfeld residuals plot (see code fragment in Appendix C). Comparisons based on the log-rank test are fully non-parametric.

Should there be inferential statistics for continuous variables that rely on the normal distribution assumption, the assumption will be checked using a normal probability plot. If severe departure from normality is observed, a non-parametric method such as the Wilcoxon rank sum test will be employed.

9.7 Validation of Statistical Analyses

Programs will be developed and maintained and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software, for example the SAS System version 9.4 or later and qualified R.

10. Statistical Methods of Analysis

10.1 General Principles

In general, the efficacy and safety analyses will be conducted separately for phase 1b and phase 3. Data collected after the analysis data cutoff date at each analysis time will be excluded.

For phase 1b, descriptive statistics for the confirmed ORR, BOR, DCR, DRR, DOR, PFS, and OS will be provided where tumor response evaluations are based on investigator assessments using modified irRC. The analysis will include data from phase 1b subjects in the Safety Analysis Set.

For phase 3, the efficacy analyses will be conducted on the Full Analysis Set. Treatment effects on efficacy endpoints will be evaluated and compared between the 2 treatment arms according to the treatment as randomized. Formal treatment comparisons will be performed on the dual primary endpoints and the 3 key secondary endpoints.

Mean, standard deviation, median, first and third quartiles, minimum and maximum will be calculated for continuous variables; frequency count and percent will be calculated for binary and categorical variables. For binary endpoints, proportions and the corresponding 95% confidence intervals, and p-values will be based on normal approximations and the treatment comparison (ratio of proportions, treatment over control) will be based on Cochran-Mantel-Haenszel test (see code fragment in Appendix C) stratified by randomization stratification factors and baseline PD-L1 status as the primary method. Additionally, p-values from the stratified Mantel-Haenszel test for the common risk difference will also be calculated (see code fragment in Appendix C). Exact confidence intervals (Clopper & Pearson 1934) (see code fragment in Appendix C) and Fisher exact tests (see code fragment in Appendix C) for binary endpoints will be considered for phase 1b and all subgroup analyses in phase 3. Logistic regression will be used to estimate the odds ratio, 95% CI for the odds ratio, and p-value (see code fragment in Appendix C).

[redacted] time-to-event endpoints will be analyzed as follows. Time to-event endpoints will be summarized descriptively by treatment group using the Kaplan-Meier (K-M) method. Standard errors of estimated treatment differences by time point will be calculated using Greenwood's formula (Kalbfleisch and Prentice, 1980). KM median estimates will be provided for each treatment arm along with KM estimates at appropriate times for each endpoint (eg, month 3, month 6, month 9, month 12, etc for PFS and 1 year, 2 year, 3 year etc. for OS). In addition, the point estimate of the adjusted risk difference (difference in Kaplan-Meier estimates at the time point of interest, treatment arm – control arm) and the corresponding 95% confidence interval will be provided using the inverse variance-weighted method (see code fragment in Appendix C). Stratified log-rank test statistics and associated p-values will also be used as the primary analysis method to assess the hypothesis of no treatment difference. Hazard ratios will be estimated using stratified Cox proportional hazards (PH) models (see code fragment in Appendix C).

For stratified analyses, the stratification factors will include randomization stratification factors per IVRS and PD-L1 status at baseline (positive vs not positive). The criteria for PD-L1 status is as follows:

- Any available PD-L1 positive result makes the subject a positive (irrespective of tissue analyzed and irrespective of the timing).
- If there are more than one PD-L1 result, and none of them are positive, then choose the one from a fresh biopsy.

Listings (eg, baseline characteristics, adverse events and exposure) of subjects in the Retreatment Analysis Set will be provided.

10.2 Subject Accountability

The number of subjects enrolled will be tabulated by country and investigator site overall and by treatment group (when applicable). Subject disposition (including the number enrolled, treated, ended treatment, ended radiographic follow-up and that completed the safety follow-up visit and the study) will be summarized separately for all enrolled subjects in phase 1b, and phase 3. Subject accountability will be tabulated by respective treatment group when applicable. Reasons for not receiving treatment, for ending treatment, ending radiographic follow-up and not completing the 30-day safety follow-up visit will be provided.

Key study dates for the first subject enrolled, last subject enrolled, first subject treated in each respective portion of the study, and last subject's end of study will be presented.

In phase 3, the number and percent of subjects randomized (enrolled) will be tabulated by the randomization stratification factors overall and by randomized treatment.

10.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category codes and descriptions will be used during the course of the study.

Eligibility deviations are defined in the protocol. Eligibility deviations that are defined as IPDs will be summarized in both the IPD and Eligibility Deviation table and IPD and Eligibility Deviation listings.

10.4 Demographic and Baseline Characteristics

Summary statistics of the following demographic and baseline disease characteristics will be tabulated using the Safety Analysis Set for phase 1b and Full Analysis Set for phase 3.

- Region (USA or Non-USA)
- Age at enrollment: <50, ≥ 50; < 65, ≥ 65; <75, ≥75 years
- Sex (male or female)
- ECOG performance status (0 vs 1)
- LDH: ≤ ULN vs > ULN**; ≤ 2 x ULN vs > 2 x ULN**
- HSV-1 serostatus (positive vs negative vs missing/unknown)
- Individual disease stage per eCRF (stage IIIB, IIIC, IVM1a, IVM1b, and IVM1c)
- Grouped disease stage per IVRS and CRF (less advanced vs more advanced): IIIB/C-IVM1a vs IVM1b/c (CRF and IVRS)
- Grouped disease stage: IIIB/C-IVM1b vs IVM1c
- **Baseline** PD-L1 status (positive, negative, vs indeterminate/unknown)
- Interval from initial melanoma diagnosis and enrollment / randomization
- The sum of longest, or shortest for nodal, diameters of target lesions (SLD) determined by BICR (both continuous and dichotomized group based on median will be presented)
- *BRAF* mutation status (V600E/K mutation present, no mutation, or unknown)
- Prior BRAF inhibitor (yes versus no BRAF inhibitor with or without MEK inhibitors) (IVRS)
- Prior BRAF inhibitor (none, BRAF inhibitor alone, BRAF and MEK inhibitors in combination): if subjects included in each of the 2 BRAF related category represent ≥ 10% of the study population (CRF)
- Baseline $BRAF^{V600}$ mutation: yes vs no vs missing/unknown
- **Treated brain metastasis (yes vs no)**
- **Liver metastasis (yes vs no)**
- **In-transit metastases (yes vs no)**
- **IVM1c with no visceral lesions (yes vs no)**
- ***BRAF*^{v600} status (*BRAF* wild type vs *BRAF*^{v600} mutant, previous directed treatment vs *BRAF*^{v600} mutant, no previous directed treatment)**
- **Previous adjuvant or neoadjuvant therapy (yes vs no)**

10.5 Efficacy Analyses

The following sections describe analyses of efficacy endpoints for phase 1b and phase 3, separately. Analysis sets used for efficacy endpoints will be specified accordingly.

10.5.1 Phase 3

Analysis of efficacy endpoints will be based on the Full Analysis Set.

10.5.1.1 Analyses of Primary Efficacy Endpoint(s)

The primary analyses of the dual primary endpoints of PFS and OS will be triggered when 407 PFS events and 346 OS events have occurred, respectively, as well as OS interim analysis at 282 events (including a futility analysis) **and at 315 events.** The final analysis will occur approximately 5 years (60 months) after the last subject has been enrolled in phase 3.

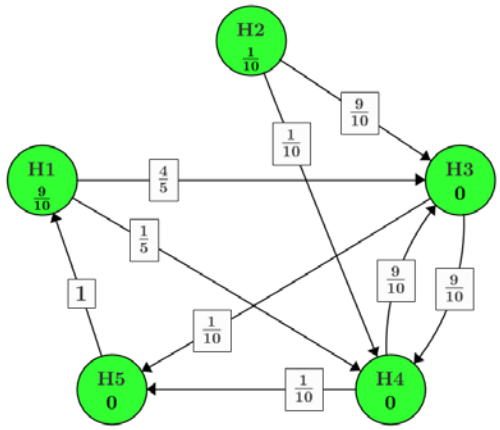
- The analyses of primary efficacy endpoints will be based on stratified log-rank tests as a primary method for testing the null hypothesis of no treatment difference.

In order to allow for the use of PD-L1 status in a stratified analysis, indeterminate and missing PD-L1 status will be classified together with the “negative” PD-L1 group and analyze in the categories of “positive” vs “not positive” PD-L1 group.

Hazard ratios and corresponding two-sided 95% CI will be estimated using Cox PH models stratified by randomization stratification factors per IVRS (disease stage [IIIB-IVM1a vs IVM1b/c] and prior BRAF inhibitor therapy [no vs yes]) and the baseline PD-L1 status (positive vs not positive). KM time to event curves will be presented by the randomized treatment group.

Multiple statistical hypothesis tests will be conducted which include testing for primary endpoints of PFS and OS and the 3 secondary endpoints (iCRR, iPFS, and OS in the subgroup excluding disease stage IVM1c) at multiple analysis time points. To control the study-level overall Type I error at an overall 2-sided 0.05 level, a graphic-based multiple testing procedure for group sequential tests for multiple endpoints will be utilized (Maurer and Bretz, 2013) (see code fragment in Appendix C). No updated information will be used when a hypothesis is re-tested at a later time point after its primary analysis. A hypothesis can be re- tested repeatedly with a different nominal level that is propagated from rejecting other hypothesis test(s). The following figure illustrates the **Maurer-**Bretz multiple testing procedure with initial 2-sided α allocation specified as a fraction used for each hypothesis test. Each hypothesis is represented by a circle. A 0 fraction in the circle indicates a hypothesis that is tested conditional on rejection of another hypothesis. The fractions in the square on the directed arrows connecting 2 hypotheses indicate the proportion of α propagated to the next hypothesis test(s) when the current hypothesis is rejected.

Figure 1. Initial Graph of Maurer-Bretz Multiple Testing Procedure



H_1 α = 0.05 × 9/10 = 0.045
H_2 α = 0.05 × 1/10 = 0.005

Due to the nature of the Maurer-Bretz multiple testing procedure, nominal levels used for testing the OS subgroup can range between 0.05% to 5%.

Further details regarding the derivation of nominal significance levels for each hypothesis test at the planned analyses are provided in Appendix B.

OS follow-up time will be assessed overall and by treatment arm. The distribution of actual and potential follow up will be provided (eg, median, range, etc.). Clark's completeness index (C) will be calculated (Clark, 2002) to quantify the actual relative to potential follow up (see code fragment in Appendix C). Plots of actual follow-up time versus time of study entry will be provided overall and by treatment arm for all subjects to assess observed follow-up times relative to potential follow up, and will be repeated for the subset with censored OS.

10.5.1.2 Analyses of Secondary Efficacy Endpoint(s)

Analysis of secondary efficacy endpoints will be based on the Full Analysis Set. Hypothesis testing for iPFS, iCRR, and OS subgroup are described in multiple testing procedures in section 10.5.1.1.

iPFS and OS excluding disease stage IVM1c subgroup: The analysis for these endpoints will follow that described for PFS.

ORR, DCR, DRR, iCRR, iORR, iDCR and iDRR: These endpoints will be analyzed as described in Section 10.1. In addition, Logistic regression with the same stratification variables used by the PFS analysis will be used to estimate a p-value and a 95% CI for the treatment odds ratio. A 95% CI for the between-arm difference will be estimated

using Wilson's score method with continuity correction. (Newcombe 1998) (see code fragment in Appendix C).

DOR and iDOR: DOR and iDOR among responders will be estimated using the KM method.

BOR categories will be summarized (number and percent) by treatment arm as randomized.

10.5.1.3 Agreement of Blinded Independent Central Review (BICR) and Investigator Reported Endpoints

To assess the agreement of the BICR and the investigator reported endpoints for iPFS and iCRR, two by two tables will be created (iPD vs non-iPD, and iCR vs non-iCR respectively) and the kappa coefficient with the 95% CI will be provided (see code fragment in Appendix C).

10.5.2 Phase 1b

10.5.2.1 Analyses of Primary Endpoint

Subject incidence of DLT

The subject incidence of DLT will be summarized as a binary variable using the DLT analysis set.

10.5.2.2 Analyses of Secondary Efficacy Endpoint

Analysis of secondary efficacy endpoints will be based on the Phase 1b Safety Analysis Set. Analysis of phase 1b efficacy endpoints will be performed after the last subject has had a chance to be followed for 24 weeks from the initiation of pembrolizumab and at the time of the **P**hase 3 OS primary analysis. All tumor responses will be assessed by investigator per modified irRC for **P**hase 1b subjects.

ORR, DCR, and DRR will be summarized with associated exact binomial 95% CI (Clopper & Pearson 1934).

BOR categories will be summarized (number and percent).

DOR, PFS, and OS: KM estimates of DOR among responders, PFS, and OS will be provided.

10.5.3 Analyses of Exploratory Endpoints

10.5.3.1 PD-L1 Induction

PD-L1 expression from tumor biopsy tissue samples will be reported as positive or negative. PD-L1 induction (ie, change of PD-L1 expression from negative to positive from paired biopsy samples taken before and after initiating talimogene laherparepvec)

will be tabulated overall and for injected and uninjected lesions. The pairing will be based on lesion number. Should there be multiple biopsies on the same lesion after initiating talimogene laherparepvec, and if there is any positive observation, the post-dose observation for PD-L1 will be considered positive. The analysis will include subjects in the PD-L1 Negative Analysis Set in the Safety Set of **Phase** 1b.

[redacted]

10.5.3.3 iORR, iBOR, iDRR, iDOR, iPFS (response evaluation by investigator using modified irRC-RECIST)

These endpoints will be analyzed in the similar manner to corresponding primary and secondary endpoints assessed by BICR.

[redacted]

10.5.4 Sensitivity Analyses

Sensitivity analyses will be performed for the Phase 3 portion of the study utilizing the Full Analysis Set to assess the robustness of select primary and secondary endpoints as noted below. Primary analysis method for primary efficacy endpoints as well as the secondary efficacy endpoints will be repeated for the Per Protocol Analysis Set to assess the impact of IPDs. All p-values will be descriptive.

10.5.4.1 PFS and iPFS Sensitivity Analyses

The following set of analyses will be used to assess the robustness of the PFS and iPFS primary analysis given the potential for ascertainment or time bias due to skipped or unscheduled tumor assessments, or deaths in the absence of prior radiologic evidence of progression that are remote from the last evaluable tumor assessment, and also to assess the impact of subsequent anti-cancer therapy.

- Stratified log-rank and Cox model analyses will be repeated in which PFS/iPFS times not associated with death in the absence of PD/iPD, that are outside of a + 1 week scheduled assessment window, are moved to the next scheduled post-randomization assessment day on both treatment arms. Times will not be moved if the next scheduled assessment is beyond the date of death or the analysis data cutoff date This analysis will also be performed moving the same PFS/iPFS times to the closest assessment day.
- Stratified log-rank and Cox model analyses will be repeated in which PFS/iPFS death events that occur more than 13 weeks (12 weeks + 1 week window) after the last evaluable radiologic tumor assessment will be censored at the time of the last evaluable radiologic tumor assessment (Dodd, 2008).
- A complementary log-log model (Allison, 1997) will be used to estimate the hazard ratio of talimogene laherparepvec in combination with pembrolizumab compared to pembrolizumab alone adjusting for the randomization stratification factors as well as PD-L1 status at baseline (positive vs not positive) . The complementary log-log model is based on a Cox model for continuous time in that the data will be grouped into time intervals. Twelve-week time intervals will be used to match the tumor assessment schedule.
- The impact of subsequent anti-cancer therapy on PFS/iPFS will be assessed by summarizing the incidence by treatment arm of selected anti-cancer therapies with potential PFS/iPFS benefit, the timing of such therapies among subjects that received it, and treatment HR estimates of time to first use among all subjects. Stratified log-rank and Cox model analyses will be repeated in which PFS/iPFS censoring at the last evaluable tumor assessment prior to the start of the first selected anti-cancer therapy.

In the event of inadvertent Suspected Unexpected Serious Adverse Reaction (SUSAR) unblinding, an additional sensitivity analysis for PFS will performed to assess the potential effect of the unblinding, by removing all event data after the date of the SUSAR unblinding from the analysis.

For the iPFS endpoint, analyses will be repeated if applicable imputing an event at the time of a last overall visit response of **i**PD that was not confirmed as required per modified irRC-RECIST.

Additionally, to assess the robustness of the PFS and iPFS primary and supportive analyses given the potential for bias related to potential informative censoring, the analyses noted for OS potential censoring in Section 10.5.4.2 will be repeated for PFS and iPFS. Censoring times will be classified as (1) administrative (random) if within 91 days (corresponding to maximum lag time allowable between tumor assessments) of the analysis data cutoff date or otherwise (2) informative.

10.5.4.2 OS Sensitivity Analyses

The following set of analyses will be used to assess the robustness of the OS primary analysis given the potential for bias related to informative censoring, and also to assess the impact of subsequent anti-cancer therapy.

- OS potential informative censoring will be assessed with a competing risks analysis utilizing the Fine-Gray proportional subdistribution hazards model (Fine, 1999; Gray, 1988) (see code fragment in Appendix C). Censoring times will be classified as (1) administrative (random) if within 140 days (corresponding to maximum lag time allowable between long term follow up assessments) of the analysis data cutoff date or otherwise (2) informative. Type (2) censoring will be defined as a competing risk event. Gray's test will be performed and the Fine-Gray treatment HR with 95% CI will be calculated stratified as per dual primary efficacy endpoints.
- The impact of subsequent anti-cancer therapy on OS will be assessed by summarizing the incidence by treatment arm of selected anti-cancer therapies with potential OS benefit, the timing of such therapies among subjects that received it, and K-M and treatment HR estimates of time to first use among all subjects. OS analyses will be repeated in which OS is censored at the first selected anti-cancer therapy.
- Since this analysis may introduce bias due to dependent censoring, the OS effect will also be estimated with an inverse probability censoring weighting (IPCW) method (see code fragment in Appendix C) that incorporates a pre-specified covariate history (eg, grouped stage of disease (not IVRS), LDH, SLD, change in ECOG performance score, *BRAF* mutation (not IVRS), PD-L1, **and** time to progression per irRC-RECIST.) and outcomes up to the start of subsequent anti-cancer therapy (Robins et al., 2000)

In the event of inadvertent SUSAR unblinding, an additional sensitivity analysis for OS will performed to assess the potential effect of the unblinding, by removing all event data after the date of the SUSAR unblinding from the analysis.

[redacted]

10.5.5 Additional Summaries

CR per BICR, and separately DpR per BICR, will be summarized for the Full Analysis Set. The proportion of subjects in each response level (CR analysis response levels: CR vs non-CR; DpR analysis response levels: G4/G5 vs G0/G1/G2/G3) will be provided by treatment arm. The between-treatment arm differences in proportions will calculated and 95% CIs will be estimated using Wilson's score method for continuity correction.

Additionally, OS will be summarized by CR response level (CR vs non-CR) per BICR and treatment arm as well as by DpR response level (G4/G5 vs G0/G1/G2/G3) per BICR and treatment arm for the Full Analysis Set. The OS KM median estimates with the associated 95% CIs will be summarized for each level of the response within treatment arm.

10.6 Safety Analyses

10.6.1 Adverse Events

Safety analyses will be conducted separately for phase 1b and phase 3. The Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 or later will be used to code all adverse events (AE) to a system organ class and a preferred term. The CTCAE version 4.0 **or later** will be used to grade severity of adverse events. In general, events with missing IP relatedness, seriousness, or CTCAE severity grades are included in the analysis of treatment emergent AE as long as the event is satisfied for a TEAE. However, analyses of treatment-related, SAE, or grade 3 or higher, or combination thereof will exclude events with missing relatedness, seriousness, and severity grades, respectively.

The subject incidence of AEs will be summarized for all treatment-emergent AEs, serious AEs, AEs leading to withdrawal of investigational product, grade 3 or 4 AEs, fatal AEs, talimogene laherparepvec events of interest (EOI) and pembrolizumab events of clinical interest (ECI). Subject incidence of EOIs and ECIs will be summarized by the EOI categories and ECI categories. Preferred terms within each EOI and ECI category will also be summarized.

The subject incidence of all treatment-related AEs, serious AEs, AEs leading to withdrawal of investigational product, grade ≥ 3 AEs, and fatal AEs will be tabulated by system organ class and preferred term in descending order of frequency.

Summaries of treatment-emergent and serious AEs occurring in at least 5% of the subjects by preferred term in any treatment arm will be provided in descending order of frequency.

A listing of fatal AEs will be provided. A listing of SAEs/EOIs/ECIs reported in the clinical database with an event onset after 30 days since last dose of IP will be provided. Listing of SAEs from consent and before first dose of study therapy will also be provided separately. Listing of SAEs reported with event onset 90 days after the last dose of study therapy or 30 days after last dose of study therapy if the subject initiates new anticancer therapy, whichever is earlier, will be provided. Additional listings of SAEs/EOIs/ECIs for subjects who received treatment by error that occur during the incorrect treatment period will be provided.

Potential or known unintended exposure to talimogene laherparepvec, related suspected signs or symptoms, and detection of talimogene laherparepvec in a subject's household member, caregiver, or healthcare provider will be summarized. Listings of reported cases with available qPCR testing results will be provided.

10.6.2 Laboratory Test Results

Shifts in grades of safety laboratory values between the baseline and the worst on-study value will be tabulated by treatment group. Subject incidence of potential hepatotoxicity as identified by the Hy's Law criteria (FDA guidance for Industry Drug Induced Liver Injury: pre-marketing evaluation, July 2009) as well as confirmed DILI events as reported by investigators will be presented.

10.6.3 Vital Signs

Analysis of vital signs will not be performed as routine monitoring and collection of vital sign data is not sufficient for analysis.

10.6.4 Electrocardiogram (ECG)

The ECG measurements from this clinical study were performed as per standard of care for routine safety monitoring, rather than for purposes of assessment of potential QTc effect. Since these evaluations may not necessarily be performed under the rigorous conditions expected to lead to meaningful evaluation of QTc data; neither summaries nor

statistical analyses will be provided, and these data would not be expected to be useful for meta-analysis with data from other trials.

10.6.5 Antibody Formation

The incidence and percentage of subjects who develop anti-pembrolizumab antibodies will be analyzed by Merck.

10.6.6 Exposure to Investigational Product

Summary statistics for exposure to talimogene laherparepvec/placebo, including total doses administered, total volume administered, duration from the first to the last administration of talimogene laherparepvec/placebo, and the average volume received by subject per visit will be provided and will be separated by the first (concentration of 10^6 PFU/ml) and subsequent doses (concentration of 10^8 PFU/ml). Subject incidence rate and reasons for IP delay, dose change, withdrawal and/or injections with <4 ml will be tabulated.

Exposure to pembrolizumab including total doses (in mg) administered, relative dose intensity, duration from the first to the last administration, and the average dose received by subject per visit will be provided. Subject incidence rate and reasons for IP delay, dose change, withdrawal and/or receiving ≥ 300 mg of pembrolizumab will be examined.

10.6.7 Exposure to Concomitant Medication

The number and proportion of subjects receiving concomitant medications will be summarized by preferred term for each treatment group as coded by the World Health Organization Drug (WHODRUG) dictionary.

10.6.8 ECOG Performance Status

ECOG performance status scores will be summarized for each treatment arm at each assessed time point. The change in scores from baseline to each assessed time point will also be summarized.

10.7 Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Analysis

The **summary of the** PK analysis **performed by Merck** will be **provided in a** separate document.

10.8 Analyses for Retreatment Phase

Listings of all data collected in the retreatment course (eg, baseline characteristics, adverse events, exposure, tumor assessments, *etc.*) for subjects in the Retreatment Analysis Set will be provided.

10.9 Incidence of Detection of Talimogene Laherparepvec DNA in Lesions Suspected to be Herpetic in Origin

Phase 1b and Phase 3

The number and proportion of subjects with positive qPCR for talimogene laherparepvec DNA detection in any swab of a lesion suspected to be herpetic in origin will be calculated based on the qPCR Suspicious Lesion Swab Analysis Set.

Individual subjects having a positive qPCR results for talimogene laherparepvec DNA detection will be reviewed.

11. Changes From Protocol-specified Analyses

Not applicable.

12. Literature Citations / References

Allison, PD. (1997) Survival Analysis Using the SAS System: A Practical Guide. SAS Institute Inc.

Clark, T.G., Altman, D.G., and De Stavola, B.L. (2002) Quantification of the Completeness of Follow-up. Lancet, 359, 1309–1310.

Clopper, C. and Pearson, E. (1934) "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404–413

Denne JS.(2001) "Sample size recalculation using conditional power." Statistics in Medicine, 20,2645-2660.

Dodd, L.E, Korn, E.L., Freidlin, B., Jaffe, C.C., Rubinstein, L.V., Dancey, J., and Mooney, M.M. (2008) "Blinded Independent Central Review of Progression-Free Survival in Phase III Clinical Trials: Important Design Element or Unnecessary Expense?" J Clin Oncol, 26(22), 3791–3796.

Eisenhauer, A., et al (2009) "New Response Evaluation in Solid Tumors: Revised RECIST Guideline (version 1.1)", EJC, 45, 228-247.

Fisher, L.D. and van Belle, G. (1993): New York: John Wiley & Sons. Biostatistics - A Methodology for the Health Sciences. P 322

Fine, J.P. and Gray, R.J. (1999) "A Proportional Hazards Model for the Subdistribution of a Competing Risk.", Journal of the American Statistical Association, 94, 496–509

Gail, M. and Simon, R. (1985) "Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets." Biometrics, 41, 361-372.

Gray, R.J. (1988) "A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk." The Annals of Statistics, 16(3), 1141-1154.

Kalbfleisch, J.D. and Prentice, R.L. (1980) The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.

Maurer, W., Bretz, F. (2013) Multiple Testing in Group Sequential Trials Using Graphical Approaches, Statistics in Biopharmaceutical Research, 5:4, 311-320

Newcombe RG. (1998) "Interval estimation for the difference between independent proportions: Comparison of eleven methods." Statistics in Medicine, 17(8), 873-890.

O'Brien PC, and Fleming TR.(1979) "A multiple testing procedure for clinical trials." Biometrics,35(3),549-556.

Osgood C, Mulkey F, Mishra-Kalyani PS, et al. (2019) "FDA analysis of depth of response (DpR) and survival across 10 randomized controlled trials in patients with previously untreated unresectable or metastatic melanoma (UMM) by therapy type". Presented at 2019 American Society of Clinical Oncology (ASCO) Annual Meeting; May 31-June 4, 2019; Chicago, IL. Abstract 9508.

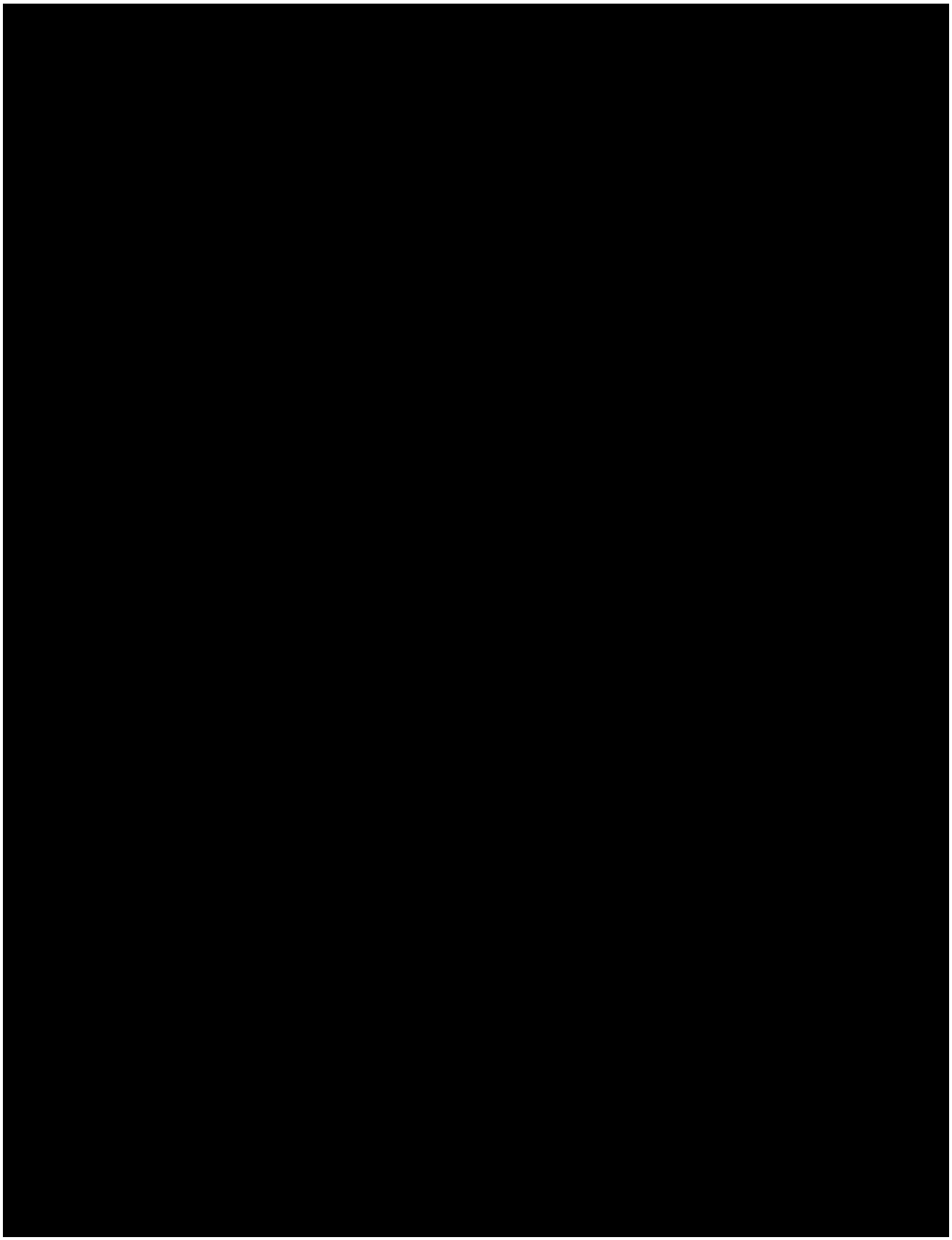
Robins, R., Hernan, M., and Brumback, B. (2000) Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology, 11(5), 550-560.

Wolchok JD, Hoos A, and O'Day S, et al. (2009) "Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria." Clin Cancer Res, 15(23), 7412-7420.

13. Data not Covered by This Plan

Analysis plans for [redacted] and PRO will be provided in separate SSAP**s**.

14. Appendices



Appendix B. Nominal Significance Levels for Maurer-Bretz Multiple Testing

Multiple testing will be conducted on primary endpoints (OS and PFS) and three secondary endpoints (iPFS, iCRR and OS subgroup excluding disease stage IVM1c) at **5** planned analyses for the study: PFS PA or OS IA1 (407 PFS events), OS IA2 (282 deaths), **OS IA3 (315 deaths),** OS PA (346 deaths), and OS in subgroup (208 deaths). At the time of 1st interim analysis, the nominal significant level depends on the information fractions observed where the spending function mimics the O'Brien-Fleming (OBF) boundaries:

$$\alpha(\gamma, y) = 2\left(1 - \Phi\left(\Phi^{-1}(1-\gamma/2)/\sqrt{y}\right)\right)$$

Where

γ = *1-sided* overall significance level

y = information fraction at the analysis time point

Φ = cumulative normal.

Based on the progress of the study, the PFS PA will be the 1st analysis and thus, the information fraction is 100% for this endpoint. At the time of PFS PA, it is likely that the information for hypotheses testing on OS, iPFS and OS subgroup will be partial and iCRR will be full information because all enrolled subjects will have been followed for more than 6 months.

Figure 2 to Figure 5 below illustrate the Maurer-Bretz successive sequential testing procedure for (1) the initial test of primary endpoints; (2) testing secondary endpoints (iPFS and iCRR); (3) testing secondary endpoint (OS subgroup); (4) possible re-test of one of the primary endpoints, OS, when rejection of both PFS and OS subgroup endpoints occur (note PFS cannot be re-tested at PFS PA or any subsequent analyses).

Table 5 provides the nominal significance levels for all possible scenarios at the PFS PA as illustrated in Figure 2 to Figure 5 under a set of information fractions assumed as: **252**/346 for OS**, 100%** for iPFS **and** iCRR, and 168/208 for OS subgroup. Note because nominal significance levels are very small, to avoid rounding error, the observed number of events out of the target number of events should be used to calculate information fractions for OS, iPFS, and OS subgroup. Upon the PFS PA snapshot (lock), this table will be re-calculated for all possible significance levels for these endpoints by replacing the observed number of events for the 3 endpoints (OS, iPFS, and OS subgroup).

Figures will be re-created illustrating testing procedures and a table containing significance levels prior to each subsequent analysis (OS IA2, **OS IA3,** OS PA, and OS subgroup PA) based on the outcome of the previous analysis. The rationale for this that (a) the previous analysis outcomes (eg, the rejection of PFS hypothesis) will significantly reduce the number of possible testing scenarios; and (b) a more precise set of information fractions can be applied for the remaining endpoints that did not achieve full information at the previous analysis.

Figure 2. Maurer-Bretz Sequential Rejection Testing Procedure: The First Tier Endpoints (OS and PFS)

Test: PFS and Initial OS

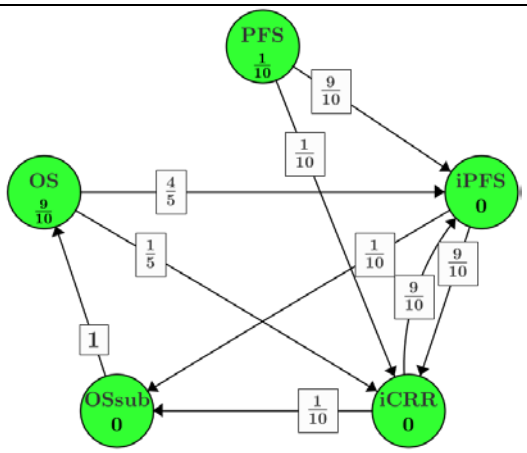


Figure 3. Maurer-Bretz Sequential Rejection Testing Procedure: The Second Tier Endpoints (iPFS and iCRR)

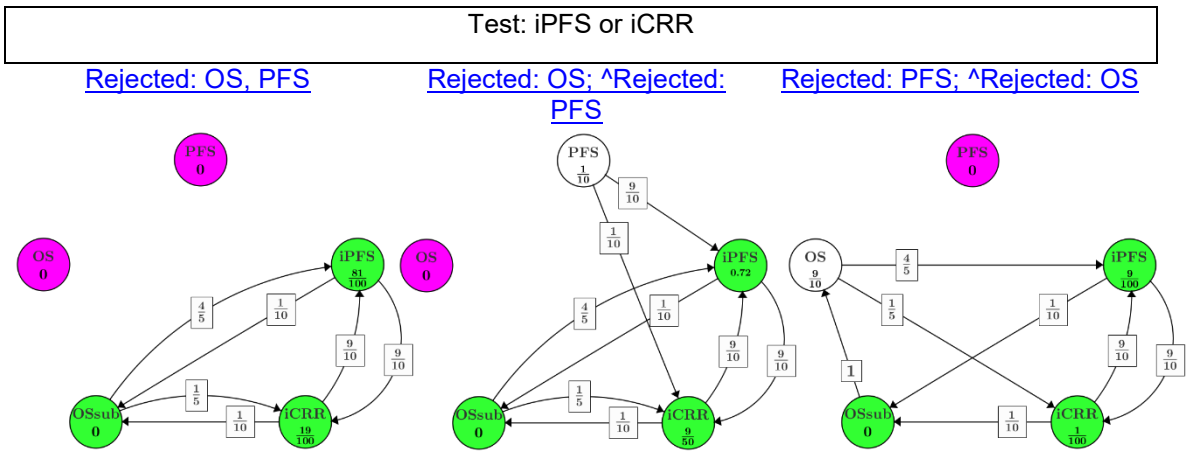


Figure 3. Maurer-Bretz Sequential Rejection Testing Procedure: The Second Tier Endpoints (iPFS and iCRR)

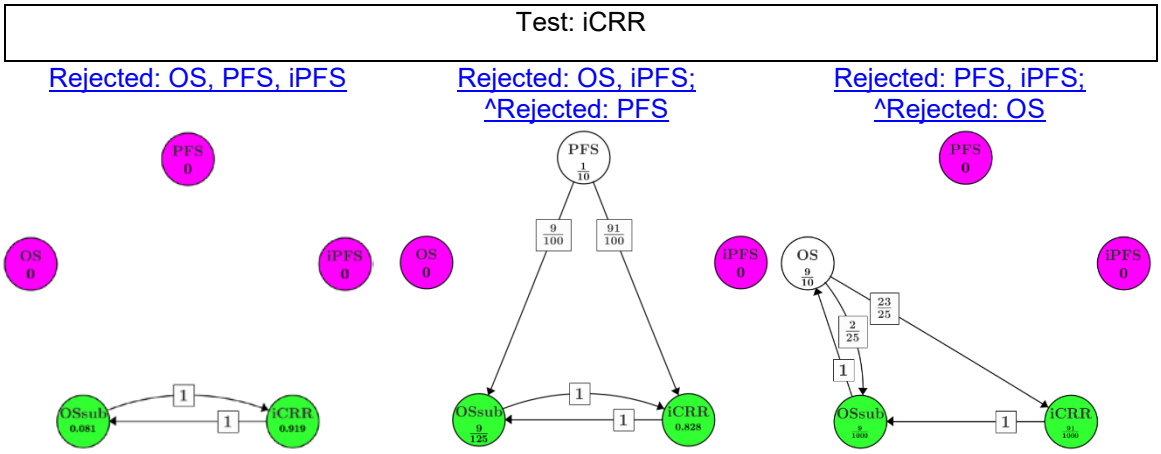


Figure 3. Maurer-Bretz Sequential Rejection Testing Procedure: The Second Tier Endpoints (iPFS and iCRR)

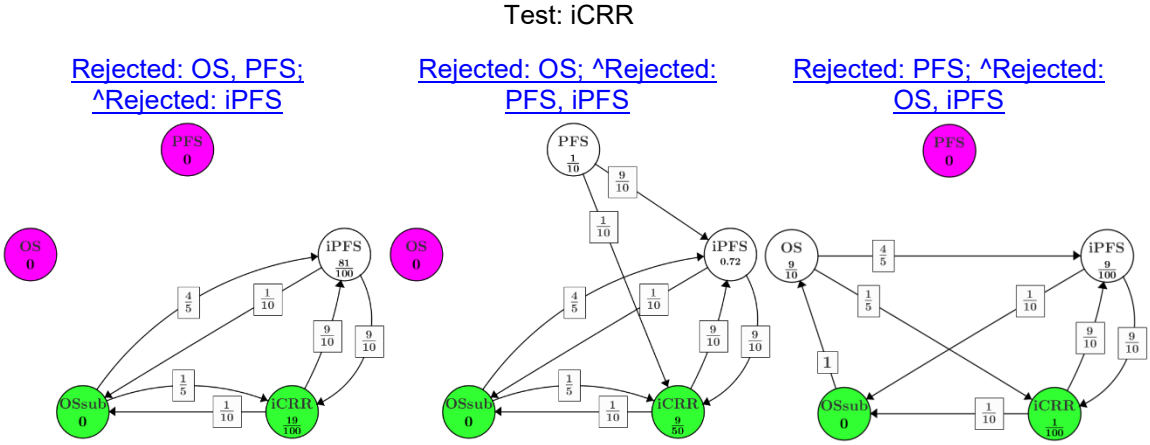


Figure 3. Maurer-Bretz Sequential Rejection Testing Procedure: The Second Tier Endpoints (iPFS and iCRR)

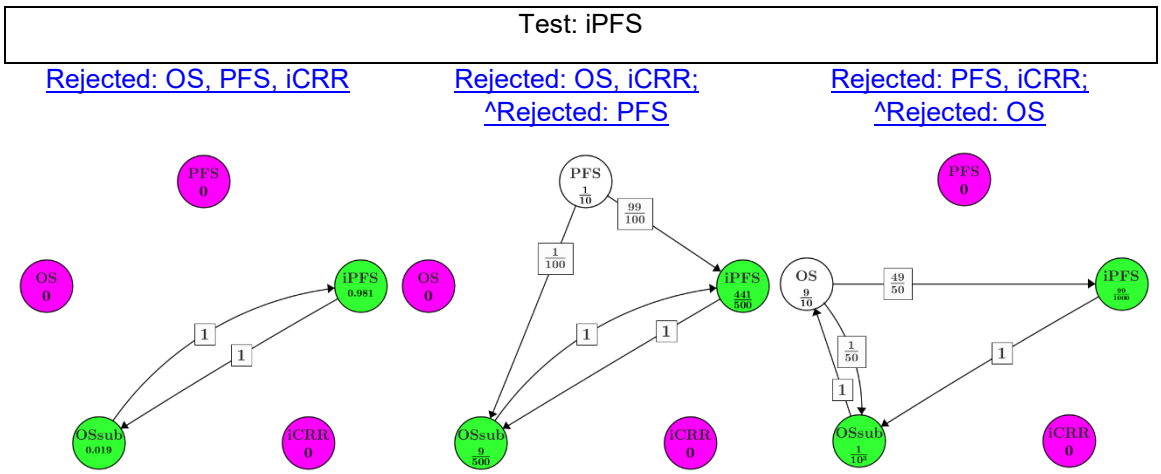
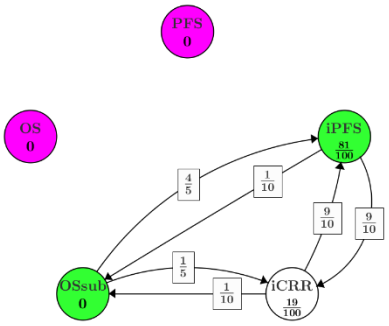


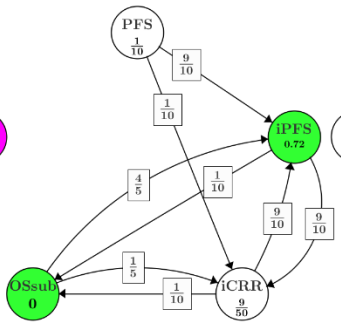
Figure 3. Maurer-Bretz Sequential Rejection Testing Procedure: The Second Tier Endpoints (iPFS and iCRR)

Test: iPFS

Rejected: OS, PFS; ^Rejected: iCRR



Rejected: OS; ^Rejected: PFS; iCRR



Rejected: PFS; ^Rejected: OS, iCRR

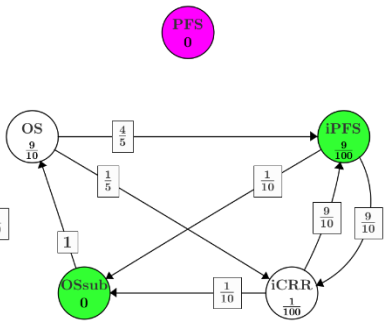


Figure 4. Maurer-Bretz Sequential Rejection Testing Procedure: The Third Tier Endpoint (OS subgroup)

Rejected: OS, PFS, iPFS, iCRR

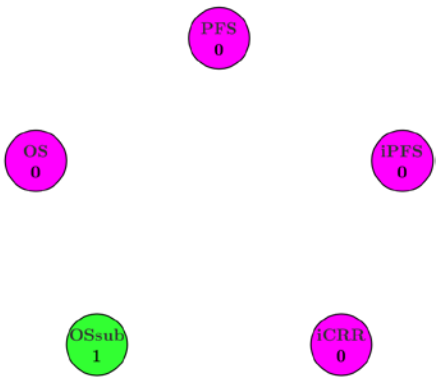
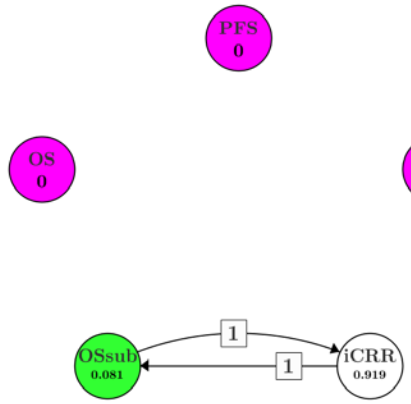


Figure 4. Maurer-Bretz Sequential Rejection Testing Procedure: The Third Tier Endpoint (OS subgroup)

Rejected: OS, PFS, iPFS; ^Rejected: iCRR



Rejected: OS, PFS, iCRR; ^Rejected: iPFS

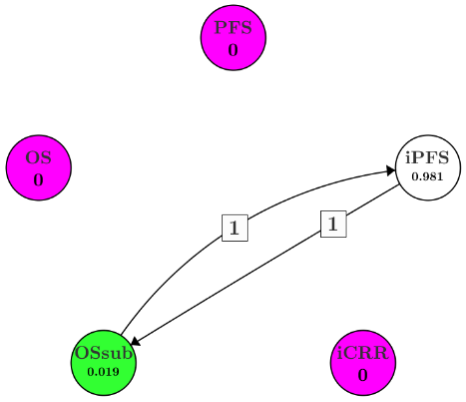
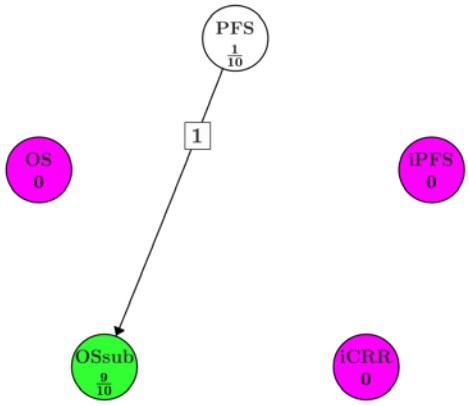
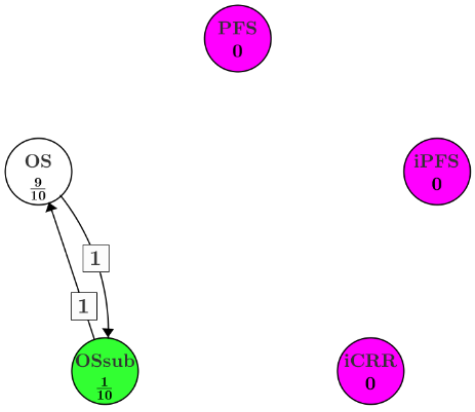


Figure 4. Maurer-Bretz Sequential Rejection Testing Procedure: The Third Tier Endpoint (OS subgroup)

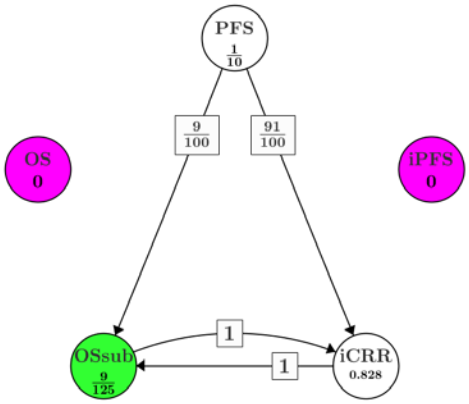
Rejected: OS, iPFS, iCRR; ^Rejected: PFS



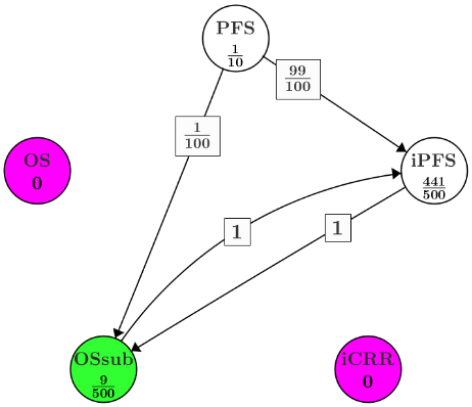
Rejected: PFS, iPFS, iCRR; ^Rejected: OS



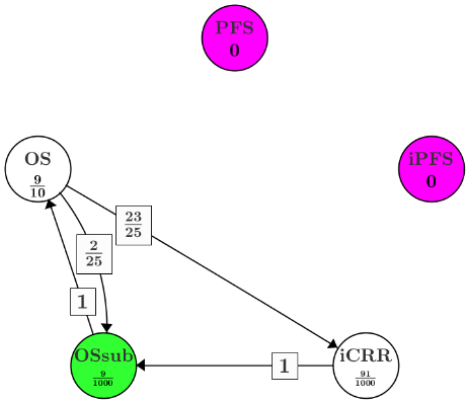
Rejected: OS, iPFS; ^Rejected: PFS, iCRR



Rejected: OS, iCRR; ^Rejected: PFS, iPFS



Rejected: PFS, iPFS; ^Rejected: OS, iCRR



Rejected: PFS, iCRR; ^Rejected: OS, iPFS

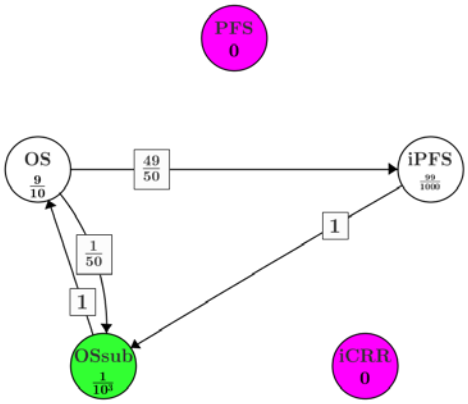


Figure 5. Maurer-Bretz Sequential Rejection Testing Procedure: Potential Re-testing of the Hypothesis Given Alpha Is Subsequently Propagated to It From the Rejection of Another Hypothesis

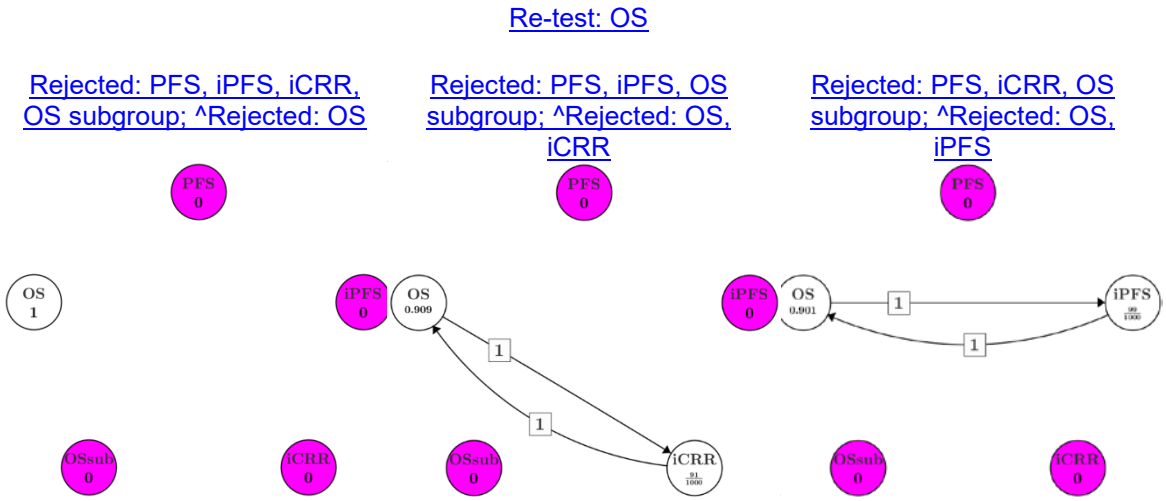


Table 5 summarizes nominal significance levels (2-sided) $\alpha_{i,1}^{*}(w_i(\cdot)\alpha)$ at the time of PFS PA (ie, with 407 PFS events) based on sequential testing, assuming overall (one-sided) significance level α = 0.025, where a possible scenario of information fraction (y) is explored as described below.

Table 5. Example 2-sided Nominal Significance Levels at the OS Interim Analysis 1 or PFS PA (ie, 407 PFS) With Partial or Full Information for All Remaining Hypotheses

	OS	PFS	iPFS	iCRR	OS Subgroup
Maximum information (events or subjects)	346	407	256	692	208
Number of events (subjects)	**252**	407	**256**	692	168
Information fraction, y	**252/346**	1.00	**1.00**	1.00	168/208
Rejected: None	***Initial test: OS***	***Test: PFS***			
$w_i(\cdot)$	9/10	1/10	0	0	0
$w_i(\cdot)\alpha$	0.045	0.005	0	0	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	**0.0150**	0.0050	NA	NA	NA
Rejected: OS, PFS			***Test: iPFS or iCRR***		
$w_i(\cdot)$	-	-	81/100	19/100	0
$w_i(\cdot)\alpha$	-	-	0.0405	0.0095	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0405**	0.0095	NA
Rejected: OS; ^Rejected: PFS			***Test: iPFS or iCRR***		
$w_i(\cdot)$	-	-	0.72	9/50	0
$w_i(\cdot)\alpha$	-	-	0.036	0.0090	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0360**	0.0090	NA
Rejected: PFS; ^Rejected: OS			***Test: iPFS or iCRR***		
$w_i(\cdot)$	-	-	9/100	1/100	0
$w_i(\cdot)\alpha$	-	-	0.0045	0.0005	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0045**	0.0005	NA
Rejected: OS, PFS, iPFS			***Test: iCRR***		
$w_i(\cdot)$	-	-	-	0.919	0.081
$w_i(\cdot)\alpha$	-	-	-	0.04595	0.00405
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.04595	NA
Rejected: OS, PFS; ^Rejected: iPFS			*Test: iCRR*		
$w_i(\cdot)$	-	-	-	19/100	0
$w_i(\cdot)\alpha$	-	-	-	0.0095	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.0095	NA

Table 5. Example 2-sided Nominal Significance Levels at the OS Interim Analysis 1 or PFS PA (ie, 407 PFS) With Partial or Full Information for All Remaining Hypotheses

	OS	PFS	iPFS	iCRR	OS Subgroup
Rejected: OS, iPFS; ^Rejected: PFS			*Test: iCRR*		
$w_i(\cdot)$	-	-	-	0.828	9/125
$w_i(\cdot)\alpha$	-	-	-	0.0414	0.0036
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.0414	NA
Rejected: OS; ^Rejected: PFS, iPFS			*Test: iCRR*		
$w_i(\cdot)$	-	-	-	9/50	0
$w_i(\cdot)\alpha$	-	-	-	0.0090	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.0090	NA
Rejected: PFS, iPFS; ^Rejected: OS			*Test: iCRR*		
$w_i(\cdot)$	-	-	-	91/1000	9/1000
$w_i(\cdot)\alpha$	-	-	-	0.00455	0.00045
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.00455	NA
Rejected: PFS; ^Rejected: OS, iPFS			*Test: iCRR*		
$w_i(\cdot)$	-	-	-	1/100	0
$w_i(\cdot)\alpha$	-	-	-	0.0005	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	0.0005	NA
Rejected: OS, PFS, iCRR			*Test: iPFS*		
$w_i(\cdot)$	-	-	0.981	-	0.019
$w_i(\cdot)\alpha$	-	-	0.04905	-	0.00095
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.04905**	-	NA
Rejected: OS, PFS; ^Rejected: iCRR			*Test: iPFS*		
$w_i(\cdot)$	-	-	81/100	-	0
$w_i(\cdot)\alpha$	-	-	0.0405	-	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0405**	-	NA
Rejected: OS, iCRR; ^Rejected: PFS			*Test: iPFS*		
$w_i(\cdot)$	-	-	441/500	-	9/500
$w_i(\cdot)\alpha$	-	-	0.0441	-	0.0009
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0441**	-	NA

Table 5. Example 2-sided Nominal Significance Levels at the OS Interim Analysis 1 or PFS PA (ie, 407 PFS) With Partial or Full Information for All Remaining Hypotheses

	OS	PFS	iPFS	iCRR	OS Subgroup
Rejected: OS; ^Rejected: PFS, iCRR			***Test: iPFS***		
$w_i(\cdot)$	-	-	0.72	-	0
$w_i(\cdot)\alpha$	-	-	0.036	-	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0360**	-	NA
Rejected: PFS, iCRR; ^Rejected: OS			*Test: iPFS*		
$w_i(\cdot)$	-	-	99/1000	-	1/1000
$w_i(\cdot)\alpha$	-	-	0.00495	-	0.00005
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.00495**	-	NA
Rejected: PFS; ^Rejected: OS, iCRR			*Test: iPFS*		
$w_i(\cdot)$	-	-	9/100	-	0
$w_i(\cdot)\alpha$	-	-	0.0045	-	0
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	**0.0045**	-	NA
Rejected: OS, PFS, iPFS, iCRR			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	1
$w_i(\cdot)\alpha$	-	-	-	-	0.05
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0253
Rejected: OS, PFS, iPFS; ^Rejected: iCRR			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	0.081
$w_i(\cdot)\alpha$	-	-	-	-	0.00405
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0012
Rejected: OS, PFS, iCRR; ^Rejected: iPFS			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	0.019
$w_i(\cdot)\alpha$	-	-	-	-	0.00095
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0002
Rejected: OS, iPFS, iCRR; ^Rejected: PFS			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	9/10
$w_i(\cdot)\alpha$	-	-	-	-	0.045
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0222

Table 5. Example 2-sided Nominal Significance Levels at the OS Interim Analysis 1 or PFS PA (ie, 407 PFS) With Partial or Full Information for All Remaining Hypotheses

	OS	PFS	iPFS	iCRR	OS Subgroup
Rejected: PFS, iPFS, iCRR; ^Rejected: OS			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	1/10
$w_i(\cdot)\alpha$	-	-	-	-	0.005
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0015
Rejected: OS, iPFS; ^Rejected: PFS, iCRR			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	9/125
$w_i(\cdot)\alpha$	-	-	-	-	0.0036
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0010
Rejected: OS, iCRR; ^Rejected: PFS, iPFS			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	9/500
$w_i(\cdot)\alpha$	-	-	-	-	0.0009
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0002
Rejected: PFS, iPFS; ^Rejected: OS, iCRR			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	9/1000
$w_i(\cdot)\alpha$	-	-	-	-	0.00045
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.0001
Rejected: PFS, iCRR; ^Rejected: OS, iPFS			*Test: OS subgroup*		
$w_i(\cdot)$	-	-	-	-	1/1000
$w_i(\cdot)\alpha$	-	-	-	-	0.00005
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	-	-	-	-	0.000005
Rejected: PFS, iPFS, iCRR, OS subgroup; ^Rejected: OS	*Re-test: OS*				
$w_i(\cdot)$	1	-	-	-	-
$w_i(\cdot)\alpha$	0.05	-	-	-	-
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	**0.0173**	-	-	-	-
Rejected: PFS, iPFS, OS subgroup; ^Rejected: OS					
$w_i(\cdot)$	0.909	-	-	-	-
$w_i(\cdot)\alpha$	0.04545	-	-	-	-
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	**0.0152**	-	-	-	-

Table 5. Example 2-sided Nominal Significance Levels at the OS Interim Analysis 1 or PFS PA (ie, 407 PFS) With Partial or Full Information for All Remaining Hypotheses

	OS	PFS	iPFS	iCRR	OS Subgroup
Rejected: PFS, iCRR, OS subgroup; ^Rejected: OS					
$w_i(\cdot)$	0.901	-	-	-	-
$w_i(\cdot)\alpha$	0.04505	-	-	-	-
$\alpha^*_{i,1}(w_i(\cdot)\alpha)$	**0.0150**	-	-	-	-

Page 5 of 5

Table 6 provides an example of testing 5 endpoints at the time of PFS PA assuming the observed information fractions are: **252**/346 for OS, 100% for PFS, iPFS, **and** iCRR, and 168/208 for OS subgroup and observed nominal p-values are 0.03 for OS, 0.001 for PFS, 0.003 for iPFS, 0.011 for iCRR, and 0.01 for OS subgroup. Because $p_{2,1}$ (PFS endpoint) is less than 0.005, the PFS endpoint is significant and therefore, the two secondary endpoints (iPFS and iCRR) can be tested. Regardless of what happens to iCRR, because $p_{3,1}$ is less than 0.0040, which is the significance level for iPFS when only PFS is significant, this endpoint is significant. Because both PFS and iPFS are significant, iCRR is tested at 0.00455, which is not significant, and OS subgroup is tested at 0.0001, which is not significant.

Table 6. Example of Nominal P-values and Significance Levels at PFS PA

Endpoint / Hypothesis	i	Nominal $p_{i,1}$	$\alpha^*_{i,1}(w_i(1,2,3,4,5)\alpha$	$\alpha^*_{i,1}(w_i(3,4,5)\alpha$	$\alpha^*_{i,1}(w_i(5)\alpha)$	Significance at PFS PA
OS	1	0.03	**0.0150**	-	-	Not significant
PFS	2	0.001	0.0050	-	-	Significant
iPFS	3	0.003	0	**0.0045**	-	Significant
iCRR	4	0.011	0	0.00455	-	Not significant
OS subgroup	5	0.01	0	NA	0.0001	Not significant

Appendix C. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

C.1 Cochran-Mantel-Haenszel (CMH) Statistics

```
proc freq data=data1;
    tables <stratum> * trtgrp * response /cmh;
run;
```

C.2 Mantal-Haenszel Common Risk difference

```
ods output CommonPdiff=cpdiff2;
proc freq data=data1;
   tables Gender*Treatment*Response /riskdiff(common);
   weight Count;
run;

data cpdiff_3;
   set cpdiff2;
   if method='Mantel-Haenszel';
   p_value=min(1, 2*(1-probnorm(abs(value/stderr))));
run;
```

C.3 Clopper-Pearson Exact Confidence Interval

```
proc freq data=data1;
   tables var1/ binomial(exact);
   by trtgrp;
run;
```

C.4 Fisher's Exact Test

```
proc freq data=data1;
   tables var1 * trtgrp /exact;
run;
```

C.5 Logistic Regression

```
/* event='1': value 1 indicating an event/response observed */
proc logistic data=data1;
   class trtgrp cov1 cov2 cov3;
   model response (event='1') = trtgrp cov1;
run;

/* Stratified logistic regression */
proc logistic data=data1;
   class <stratum> trtgrp cov1 cov2 cov3;
   model response (event='1') = trtgrp cov1;
   strata <stratum>;
run;
```

C.6 Kappa Coefficient

```
proc freq data=data1;
   tables var1 * trtgrp /agree;
run;
```

C.7 Wilcoxon Rank-sum Test on Continuous Response

```
proc npar1way wilcoxon data=<data>;
   class trtpxcd;
   var <continuous endpoint>;
run;
```

C.8 Fine-Gray Subdistribution Proportional Hazards Model

```
/* The variable STATUS has three values: 0 for censored observations, 1
for subjects with event of interest (e.g. local recurrence for LRFS),
and 2 for patients who experienced competing risk. */

proc phreg data=dat1 ;
   class TRTGRP (order=internal ref=first) param=glm;
   model LRFS*STATUS(0)= TRTGRP x1 x2 / eventcode=1;
   hazardratio 'Subdistribution Hazards' TRTGRP;
   strata randomization_strata;
   * Delete strata statement to assess overall proportional hazards;
   * x1, x2 are covariates.
run;
```

C.9 Assumption of Non-Proportionality Hazards: Schoenfeld Residuals Plot

```
/* cnsr(1): value 1 indicating censored observations*/
proc phreg data=data1;
    model OS * CNSR(1) = TRTGRP covariate1 covariate2 covariate3
/ties=exact;
    strata randomization_strata;
    output out=outdata ressch= s_TRTGRP s_covariate1 s_covariate2
s_covariate3;
run;

proc sgplot data=outdata;
    scatter y=TRTGRP x=OS/ name="Schoenfeld_&CHECK";
    loess y=TRTGRP x=OS /SMOOTH=.8 DEGREE=1 INTERPOLATION= CUBIC ;
run;

*Note:  Repeat proc gplot for other covariates (covar1, covar2, etc.)*;
```

C.10 Kaplan Meier Method: Stratified Log-Rank Test and Standard Errors Using Greenwood's Formula

```
/* cnsr(1): value 1 indicating censored observations */
proc lifetest data=data1;
   time Days*cnsr(1);
   strata <stratum> / group=TRTGRP;
run;
```

C.11 Stratified multivariate Cox Proportional Hazards Model

```
/* cnsr(1): value 1 indicating censored observations */
proc phreg data=dat1;
   model OS * CNSR(1) = TRTGRP covariate1 covariate2 / ties=exact
rl=pl;
   strata <stratum>;
run;
```

C.13 Inverse Probability Censoring Weighting (IPCW)

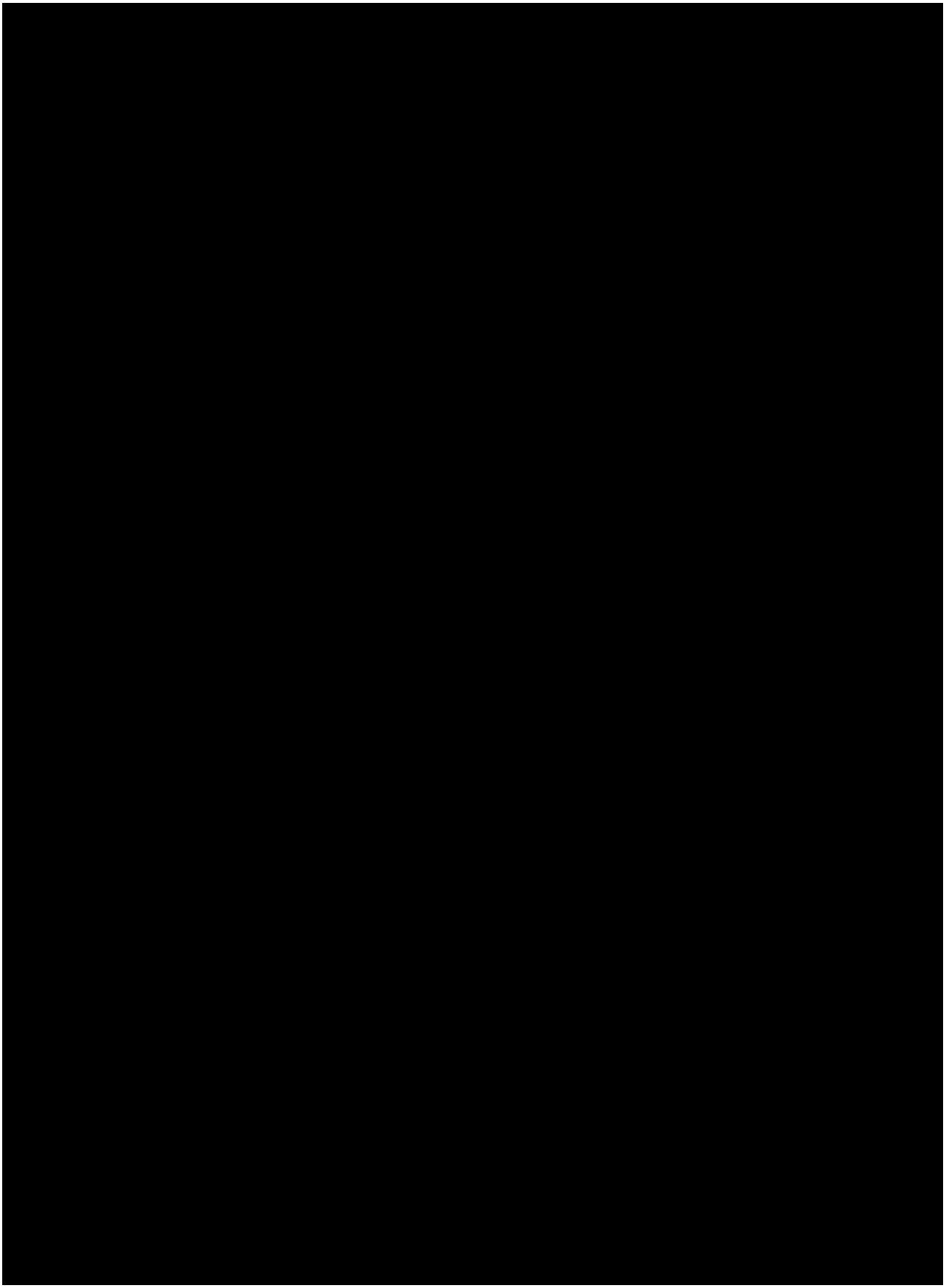
```
*** Model 1: time-independent model
*** model the probability of without subsequent anti-cancer therapy per
each subject
*** include the baseline covariates;
proc sort data=adsl out=indep;by subther;run;
proc genmod data=indep order=data;
  class disstgi LDH brafmut PDL1;
  model subther = disstgi LDH SLD brafmut PDL1 / dist=binomial
link=logit ;
  output out=num prob=num;
run;

*** Model 2: time-dependent model
*** model the probability of without subsequent anti-cancer therapy per
each subject-timeinterval
*** include both baseline covariates and time-dependent covariates;
proc sort data=all2 out=dep;by subther;run;
proc genmod data=dep order=data;
  class disstgi LDH brafmut PDL1   <time-dep categorical vars> ;
  model subther = disstgi LDH SLD brafmut PDL1   <time-dep vars> /
dist=binomial link=logit ;
  output out=den prob=den;
run;

proc sort data=num ;by usubjid;run;
proc sort data=den ;by usubjid;run;

data wgt;
retain <stratification factors> armcd usubjid subther
       num den
;
  merge num den ;
  by usubjid;
run;

*** calculate weights;
data wgt2;
  set wgt;
  by usubjid;

retain w_num w_den ;
  if first.usubjid then do;
      w_num = 1;
      w_den = 1;
  end;
  w_num = w_num*num ;
  w_den = w_den*den ;

  if subther = 'y' then weights = (1-w_num)/(1-w_den);
  else if subther = 'n' then weights = w_num/w_den;
run;

data os_wgt;
  merge wgt2(in=a)
        os(in=b)
;
```

```
  by usubjid;
  if a and b;
run;

*** Inverse Probability Censoring Weighted Cox model ;
proc phreg data=os_wgt;
  class armcd(ref='PB_PEM');
  model aval*cnsr(1)=armcd;
  strata <stratification factors>;
  freq weights / notruncate;
  hazardratio armcd / alpha=0.05 cl=wald diff=ref;
  ods output hazardratios=hr_ci_ipcw;
run;
```

C.14 Gail and Simon Quantitative and Qualitative Interaction Tests and Post Hoc Power Test

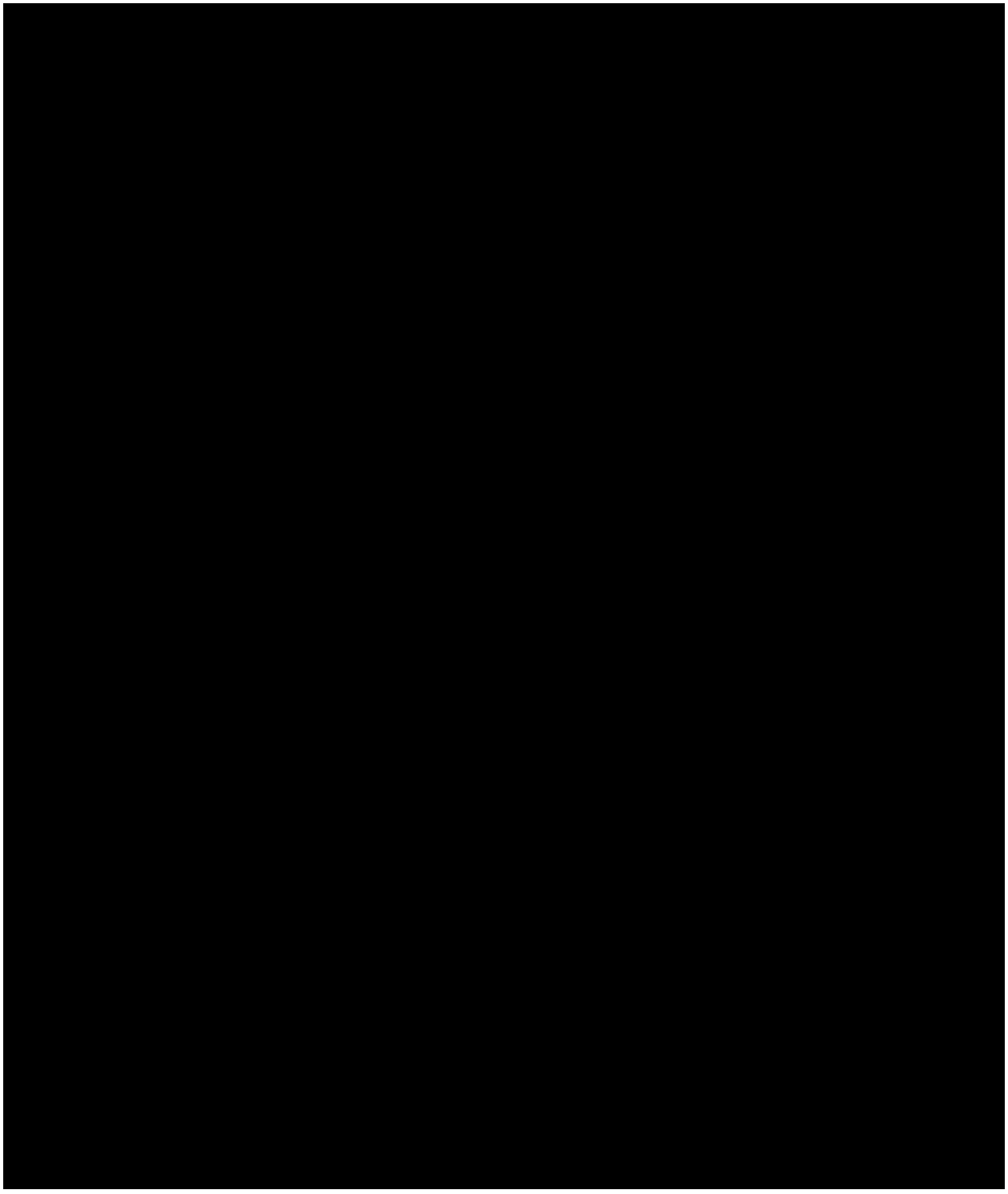
```
%macro int(ds     =,  /* Dataset                           */
           txvar  =,  /* Treatment indicator               */
           sgroup =,  /* Subgroup variable                 */
           time   =,  /* Time to event variable            */
           censor =,  /* Censor var 1=observed 0=censored */
           covar  =,  /* Covariates for model              */
           alpha  =,  /* 2-sided nominal alpha for power   */
           outf   =   /* SASname of results file           */ );

proc sort data=&ds out=&ds._d1;
  by &sgroup;
run;

/*
proc printto new print=&outf; run;
*/
ods output ParameterEstimates=phfit;
proc phreg data=&ds._d1;
  model &time*&censor(0)=&txvar &covar
  / rl ties=exact;
  by &sgroup;

data ph (keep=&sgroup estimate v rename=(estimate=b));
  set phfit (keep=&sgroup estimate stderr);
  v=stderr**2;

proc transpose data=ph prefix=b out=b (drop=_name_);
  var b;
proc transpose data=ph prefix=v out=v (drop=_name_);
  var v;

data t; merge b v;
  label H='(GS)|Quant.|Test|Value'
    ph='(GS)|Quant.|p-val' ph_p='(GS)|Quant.|Post Hoc|Power'
    Q='(GS)|Qual.|Test|Value' pq='(GS)|Qual.|p-val'
    pp='(PG)|Qual.|p-val' Qmin='(PG)|Qmin' Qmax='(PG)|Qmax'
    alpha='Alpha|2-tail';
  alpha = &alpha+0;
  array b b1-b30;
  array v v1-v30;
  g=0; do i=1 to 30; if b{i} ne . then g+1; end;

  * (GS) Quantitative interaction test ;
  n=0; d=0;
  do i=1 to g; n=n + b{i}/v{i}; d=d + 1/v{i}; end;
  Dbar=n/d;
  H=0;
  do i=1 to g; H=H + ((b{i}-Dbar)**2)/v{i}; end;
  ph=1-probchi(H,g-1);

  * Post hoc power ;
  c_alpha = quantile('CHISQ', 1-alpha, g-1);
  ph_p = 1 - probchi(c_alpha, g-1, H);

  * (GS) Qualitative Interaction ;
```

```
  Q=0; Qn=0; Qp=0; pq=0;
  do i=1 to g;
    Qn=Qn + ((b{i}**2)/v{i})*resolve(b{i}>0);
    Qp=Qp + ((b{i}**2)/v{i})*resolve(b{i}<0);
  end;
  Q=min(Qn,Qp);
  do i=1 to g-1;
    pq=pq + (probbnml(.5,g-1,i)-probbnml(.5,g-1,i-1)) *
            (1-probchi(Q,i));
  end;

  * (PG) Qualitative Interaction ;
  array r r1-r30;
  do i=1 to g;
    r{i}=b{i}/sqrt(v{i});
    if i=1 then do; Qmin=r{i}; Qmax=r{i}; end;
    if r{i} lt Qmin then Qmin=r{i};
    if r{i} gt Qmax then Qmax=r{i};
  end;
  if Qmin ge 0 then pp=.5;
  else do;
    c=min(abs(Qmin),abs(Qmax));
    pp=1 - (1-probnorm(-c))**(g-1);
  end;
run;

proc print data=t noobs label split='|';
  var h ph ph_p q pq Qmin Qmax pp;
run;
%mend int;
```



[redacted]

[redacted]



C.16 Wilson's Score Method With Continuity Correction

Macro and Sample Code for Wilson's score method with continuity correction

- S1 is the numerator and N1 is the demoninator for T-VEC; S2 is the numerator and N2 is the denominator for GM-CSF
- There are variables P1, P2 , LL and UL.
- The percentage difference is the P1-P2 and 95% CI is (LL, UL)

```
%macro method11(ds=,
                dsout=,
                by=,
                cc=0.80);

  data &dsout;
    length s1 n1 p1 s2 n2 p2 LL UL cc 8;
    set &ds;
    retain cc &cc;
    array nn{2} n1 n2;
    array s{2} s1 s2;
    array p{2} p1 p2;
    array q{2} _temporary_;
    array np{2} _temporary_;
    array nq{2} _temporary_;
    array l{2} _temporary_;
    array u{2} _temporary_;
    ca2=probit(0.5 + cc/2);
    ca2sq=ca2*ca2;
    do i=1 to 2;
      p{i}=s{i}/nn{i};;
      q{i}=1-p{i};;
      nq{i}=nn{i}*q{i};
      np{i}=nn{i}*p{i};
      if p{i}=0 then l{i}=0;
      else l{i}=(((2*np{i})+(ca2sq-1))-(ca2*(sqrt((ca2sq-(2+(1/nn{i}))+
((4*p{i})*(nq{i}+1)))))))/(2*(nn{i}+ca2sq)) ;
      if p{i}=1 then u{i}=1;
      else u{i}=(((2*np{i})+(ca2sq+1))+(ca2*(sqrt((ca2sq+(2-(1/nn{i}))+
((4*p{i})*(nq{i}-1)))))))/(2*(nn{i}+ca2sq)) ;
    end;
    d=sqrt((p{1}-l{1})**2+(u{2}-p{2})**2);
    e=sqrt((p{2}-l{2})**2+(u{1}-p{1})**2);
    LL=p{1}-p{2}-d;
    UL=p{1}-p{2}+e;
  *  keep s1 n1 p1 s2 n2 p2 LL UL cc;
    label s1='Number of Sample from Population 1'
          n1='Number of Population 1'
          p1='Proportion 1 = s1 / n1 '
          s2='Number of Sample from Population 2'
          n2='Number of Population 2'
          p2='Proportion 2 = s2 / n2'
          ll='Lower limit of interval estimation for difference'
          ul='Upper limit of interval estimation for difference'
          cc='Confidence coefficient';
  run;

  %if %length(&by)>0 %then %do;
    proc sort data=&dsout;
      by &by;
    run;
  %end;

  title1 "Comparison of two independent proportions using continuity
correction" ;
  title2 'NEWCOMBER METHOD 11 : Example H ';
  title3 "LL UL &cc% CI for Difference between Independent Proportions";
```

```
  proc print data=&dsout label;
    var &by n1 s1 p1 n2 s2 p2 LL UL cc;
  run;
%mend method11;


data dsin;
  input n1 n2 s1 s2;
cards;
25 50 5 4.25;
run;

%method11(ds=dsin, dsout=dsout, by=, cc=0.95);
```

C.17 Clark's Quantification of the Completeness of Follow-up

```
%let cutdt = 02-OCT-18;

proc univariate data=adamp.adtte noprint;
where paramcd = "PFS" and cohort ne "PHASE 1B";
  var randdt;
  output out=regmin min=randdt_min_
                    max=randdt_max_;

data fuci1;
  set adamp.adtte(keep=subjid saffl cohort trt01p aval cnsr paramcd
randdt adt);
  format obs_fu_dt date9.;
  where paramcd = "PFS" and cohort ne "PHASE 1B";
  * Observed follow-up time = death time ;
  obs_fu=aval;
  obs_fu_dt=randdt + aval;

  * Potential follow-up time ;
  if cnsr=0 then po_fu=aval;
  else po_fu="&cutdt"d - randdt;

  * Time since survival time ;
  lag = "&cutdt"d - obs_fu_dt;
run;

data fuci2;
  drop randdt_min_ randdt_min;
  merge fuci1
        regmin;
  retain randdt_min;
  if _n_=1 then randdt_min=randdt_min_;
  entry_day=randdt-randdt_min;

  obs_fu = obs_fu/365.25;
  po_fu = po_fu/365.25;
run;

proc univariate data=fuci2 noprint;
  var obs_fu po_fu;
  output out=f_all n=n1 n2 sum=t_obs_fu t_po_fu;
run;

data fu;
```

```
  set f_all;
  ci=100 * t_obs_fu / t_po_fu;
  call symput('ci',trim(left(put(ci,8.0))));
run;

proc print data=fu;
  title 'Completion Index (CI)';
run;
```

C.18 Complementary log-log model with data grouped into 12-week time intervals

```
data pfs_12wk;
   set pfsdur;

   /* convert duration into 12-week interval */
   duration_in12wk = ceil(duration/12);

   do interval_12wk=1 TO duration_in12wk;
   if interval_12wk=duration_in12wk and event=1 then PD=1;
      else PD=0;
   output;
   end;
run;

/* variable event:  1=event occurred 0=no event */
proc logistic data= pfs_12wk;
   class interval_12wk;
   model PD(event='1')=trtgrp strata1 strata2 strata3 interval_12wk
/link=cloglog;
run;
```