STATISTICAL ANALYSIS PLAN

Title: A Phase 2, Multicenter, Randomized, Open-label Trial Assessing the Efficacy and Safety of Talimogene Laherparepvec Neoadjuvant Treatment Plus Surgery Versus Surgery Alone for Resectable, Stage IIIB to IVM1a Melanoma

Protocol Number:	**20110266**
	NCT02211131
Version:	**3.0**
Date:	**19 September 2014**
Amendment 1 Date:	**23 September 2016**
Amendment 2 Date:	**23 March 2018**
Amendment 3 Date:	**14 May 2019**
Authors:	[redacted]

NCT Number: NCT02211131
This NCT number has been applied to the document
for purposes of posting on clinicaltrials.gov

Table of Contents

Table of Abbreviations 5

1. Introduction 6

2. Objectives 6
   - 2.1 Primary 6
   - 2.2 Secondary 6
   - 2.3 Exploratory 7

3. Study Overview 7
   - 3.1 Study Design 7
   - 3.2 Sample Size 9

4. Study Endpoints and Covariates 10
   - 4.1 Study Endpoints 10
     - 4.1.1 Primary Endpoint 10
     - 4.1.2 Secondary Endpoint 11
       - 4.1.2.1 Efficacy Endpoints 11
       - 4.1.2.2 Safety Endpoints 13
     - 4.1.3 Exploratory Endpoints 13
   - 4.2 Planned Covariates 13

5. Hypotheses and/or Estimations 14

6. Definitions 14

7. Analysis Subsets 19
   - 7.1 Intent to Treat Analysis Set 19
   - 7.2 Safety Analysis Set 20
   - 7.3 Efficacy Analysis Set 20
   - 7.4 Subgroup Analyses 20

8. Planned Analyses 20
   - 8.1 Interim Analysis and Early Stopping Rules 20
     - 8.1.1 Data Review Team (DRT) 21
   - 8.2 Primary Analysis 21
   - 8.3 Final Analysis 22
   - 8.4 Additional Analysis 22

9. Data Screening and Acceptance 22
   - 9.1 General Principles 22
   - 9.2 Data Handling and Electronic Transfer of Data 22
   - 9.3 Handling of Missing and Incomplete Data 22
     - 9.3.1 Efficacy 22
     - 9.3.2 Safety 22
   - 9.4 Detection of Bias 22

9.4.1 Important Protocol Deviations (IPD) Impact 22
9.4.2 Timing Bias Impact 23
9.4.3 Subsequent Anti-Cancer Therapy Impact 23
9.5 Outliers 23
9.6 Distributional Characteristics 24
9.7 Validation of Statistical Analyses 24

10. Statistical Methods of Analysis 25
10.1 General Principles 25
10.2 Subject Accountability 26
10.3 Important Protocol Deviations 26
10.4 Demographic and Baseline Characteristics 26
10.5 Efficacy Analyses 27
10.5.1 Analyses of Primary Efficacy Endpoint(s) 31
10.5.2 Analyses of Secondary Efficacy Endpoint(s) 31
10.6 Safety Analyses 33
10.6.1 Adverse Events 33
10.6.2 Laboratory Test Results 34
10.6.3 Vital Signs 34
10.6.4 ECOG Performance Status 35
10.6.5 Electrocardiogram (ECG) 35
10.6.6 Antibody Formation 35
10.6.7 Exposure to Investigational Product 35
10.6.8 Exposure to Concomitant Medication/Adjuvant Therapy/Subsequent Anti-cancer Therapy 36

11. Changes From Protocol-specified Analyses 37

12. Literature Citations / References 38

13. Data Not Covered by This Plan 39

14. Appendices 40

List of Tables

Table 1. Study Design Characteristics 10
Table 2. Study Design Characteristics Example Endpoint Summary Table 28

List of Figures

Figure 1. Example of Treatment Period and T-VEC Monotherapy Period 15

List of Appendices

Appendix A. Conventions for Clinical Data That Require Imputation for Partial or Missing Dates 41

Appendix B. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs 43

Table of Abbreviations

In addition to the study glossaries and abbreviations defined in the protocol, here is a table of abbreviations used in this document:

Abbreviation/Acronym	Definition
ADaM	Analysis Data Model data sets
BDSG	Biomedical Data Stewardship Governance
CRF	**Case Report Form**
IPCW	Inverse Probability Censoring Weighting
IPD	Important Protocol Deviation
ITT	Intent to Treat
SAP	Statistical Analysis Plan
SDTM	Study Data Tabulation Model data sets
SOA	Schedule of Assessments
TFL	Table, Figure, Listing

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 amendment 3 for the Talimogene Laherparepvec, Study 20110266 dated 23 March 2018. The scope of this plan includes the three interim analyses, the primary analysis, the final analysis, and one additional ad hoc analysis that may also be conducted before the planned primary and/or final analysis if interim data are required for submission to regulatory authorities. These analyses will be executed by the biostatistics department unless otherwise specified.

[redacted]

This study will have data collected and analyzed in Amgen-owned databases and systems and will adhere to approved Data Element Standards and International Case Report Form (CRF) Standards, if available, established by Biomedical Data Stewardship Governance (BDSG). Amgen-owned databases and systems include vendor and service provider databases built for Amgen-only use.

[redacted]

2. Objectives

2.1 Primary

To estimate the treatment effect of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone on recurrence-free survival (RFS).

2.2 Secondary

The secondary objectives are:

- To estimate the effect of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone on landmark RFS by year
- To estimate the effect of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone on rate of histopathological tumor-free margin (R0) surgical resection
- To estimate the effect of neoadjuvant talimogene laherparepvec on rate of pathological complete response (pCR) (Arm 1 only)

- To estimate the effect of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone on local recurrence-free survival (LRFS**),** regional recurrence-free survival (RRFS) and distant metastases-free survival (DMFS)
- To estimate the effect of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone on landmark overall survival (OS) by year, and overall survival (OS)
- To estimate response to neoadjuvant talimogene laherparepvec overall and separately in injected and uninjected lesions during treatment (Arm 1 only)
- To evaluate the safety of neoadjuvant talimogene laherparepvec plus surgery compared to surgery alone

2.3 Exploratory

The exploratory objectives are:

- To explore the correlation between baseline intratumoral CD8+ cell density and clinical outcomes
- To explore the correlation between changes in an intratumoral CD8+ cell density during talimogene laherparepvec treatment and clinical outcomes
- To investigate the correlation between the changes in the population of tumor-specific cytotoxic T-cells during treatment and clinical outcomes
- To assess blood and tumor for potential biomarkers which correlate with or predict clinical outcomes to talimogene laherparepvec

3. Study Overview

3.1 Study Design

This is a phase 2, multicenter, randomized, open-label study to estimate the efficacy of talimogene laherparepvec as a neoadjuvant treatment followed by surgery compared to surgery alone in subjects with resectable stage IIIB, IIIC, or IVM1a melanoma. Approximately 150 subjects with completely resectable melanoma and at least one injectable lesion will be randomized 1:1 to receive the following:

- Arm 1: Talimogene laherparepvec for 6 doses followed by surgical resection of melanoma tumor lesion(s).
  - Talimogene Laherparepvec:

Talimogene laherparepvec will be administered by intralesional injection into the injectable cutaneous, subcutaneous, and nodal tumors initially at a dose of 10^6 plaque forming units (PFU)/mL at day 1 of week 1 followed by a dose of 10^8 PFU/mL at day 1 (± 3 days) of weeks 4, 6, 8, 10 and 12 or until all injectable tumors have disappeared, or intolerance of study treatment, whichever occurs first. Due to the mechanism of action of talimogene laherparepvec, subjects may experience transient growth in existing tumors or the appearance of new tumors prior to maximal clinical benefit of talimogene laherparepvec. Subjects who experience growth in existing tumors

or the appearance of new tumors will be allowed to remain on talimogene laherparepvec treatment until week 12 of therapy unless, in the opinion of the investigator, immediate surgical resection or any other treatment for melanoma is warranted. If talimogene laherparepvec treatment will end prior to week 12, the investigator or designee should notify the sponsor's medical monitor as soon as possible.

- Surgical Resection of Melanoma Tumor Lesions(s):

  For subjects who complete 12 weeks of talimogene laherparepvec treatment, surgical resection of melanoma lesion(s) will be performed at any time during weeks 13 to 18. Subjects who stop talimogene laherparepvec prior to week 12 due to disappearance of all injectable tumor lesions, intolerance to talimogene laherparepvec, or any other reason will undergo adequate surgical resection of melanoma lesion(s) or tissue where melanoma was present before achieving CR, within 1 to 6 weeks after the last dose of talimogene laherparepvec. If surgery will be performed prior to week 12, the investigator or designee should notify the sponsor's medical monitor as soon as possible. Refer to protocol Appendix F for surgery guidelines.

- Arm 2: Immediate surgical resection of melanoma tumor lesion(s)
  - Surgical resection of melanoma tumor lesion(s) will be performed after enrollment any time during weeks 1 to 6.

  Randomization will be stratified by disease stage (IIIB nodal versus IIIB in-transit versus IIIC nodal versus IIIC in-transit with nodal versus IVM1a) and planned adjuvant therapy (adjuvant systemic therapy [eg,INFα, ipilimumab] with or without radiotherapy versus radiotherapy without adjuvant systemic therapy versus none). Following surgery, adjuvant systemic therapy and/or radiotherapy may be administered at the investigator's discretion and per the institutional standard of care. Subjects will be followed for safety approximately 30 (+ 15) days after surgery and for disease recurrence (local, regional, or distant), subsequent anticancer therapy for melanoma, adverse events thought to be potentially related to talimogene laherparepvec (Arm 1 only), and survival every 3 months (± 30 days) for first 3 years after the end of the safety follow-up period and then every 6 months (± 30 days) until death, subject withdraws full consent, or up to 5 years after the last subject is randomized, whichever is first. Thereafter, subjects randomized to Arm 1 who received at least a single dose of talimogene laherparepvec will be followed under an ongoing separate registry protocol (study 20120139) for the long-term survival

follow-up of subjects treated with talimogene laherparepvec. The registry protocol will also monitor for late and long-term adverse events thought to be potentially related to talimogene laherparepvec and use of subsequent anticancer therapy for melanoma. Subjects who, after the long-term follow-up period of this study (Study 20110266), elect to participate in the registry study (Study 20120139) must sign new informed consent form before any registry protocol-specific activities.

3.2 Sample Size

All analyses will be descriptive with no formal hypothesis testing. The primary objective of the study is to estimate the treatment effect, as measured by the hazard ratio of RFS, of talimogene laherparepvec neoadjuvant therapy followed by surgery (Arm 1) compared to surgery alone (Arm 2) in subjects with resectable, stage IIIB to IVM1a melanoma. Approximately 150 subjects will be randomized 1:1 to receive Arm 1 vs Arm 2 stratified by: (i) disease stage (IIIB nodal vs IIIB in-transit vs IIIC nodal vs IIIC in-transit with nodal vs IVM1a), and (ii) planned adjuvant therapy (adjuvant systemic therapy [eg, INFα, ipilimumab] with or without radiotherapy vs radiotherapy without adjuvant systemic therapy vs none). Primary efficacy analysis of RFS will be based on the Intent to Treat Analysis Set. An overall between-group difference in RFS will be evaluated with a log-rank test and corresponding proportional hazard**s** model without stratification. The Kaplan-Meier (KM) method will be used to estimate 1-year, 2-year, 3-year, 5-year, and overall RFS. Subjects who are not confirmed to be disease-free post-surgery (i.e., who do not have an R0 surgical outcome) or who withdraw prior to surgery will be considered a failure at randomization for RFS, RRFS, LRFS and DMFS. The primary analysis for the primary endpoint will occur at the later time of either the occurrence of approximately 64 events (local, regional, or distant recurrence of melanoma or death) or approximately 2 years after the end of randomization. An 80% CI will be estimated for the hazard ratio of RFS. In addition, an 80% CI will also be estimated for the between-arm difference (Arm 1 – Arm 2) in 2-year RFS, i.e., Δ 2-year RFS. It is assumed that: (i) RFS is exponential over the first 2 years, (ii) the 2-year RFS for Arm **2** is about 0.60 (Eggermont et al, 2008), and that there will be a 10% exponential probability of drop-out by the primary analysis. Simulations were performed for various scenarios from 0.05 to 0.15 to evaluate the following parameters assuming possible true Δ 2-year RFS values of 0.10, 0.125, and 0.15: (i) the average Δ 2-year RFS 80% CI and its width, (ii) the probability that the Δ 2-year RFS 80% CI is above 0, and (iii) the average number of events at the

primary analysis. The possible true Δ 2-year RFS values were translated to a RFS hazard ratio (HR) (Arm 1 relative to Arm 2) and the HR 80% CI and the probability of it being < 1 were calculated based on the simulated average number of events. The average width of the Δ 2-year RFS 80% CI was 0.20 (Table 1).

Table 1. Study Design Characteristics

2-year RFS			RFS HR	Ave. Δ 80% CI			Prob. Δ 80% CI > 0	Ave. HR 80% CI		Prob. HR 80% CI < 1	Ave. Events
Arm 1	Arm 2	Δ		LL	UL	Width		LL	UL		
0.700		0.100	0.70	0.00	0.20	0.20	0.51	0.51	0.96	0.56	64
0.725	0.60	0.125	0.63	0.02	0.22	0.20	0.62	0.45	0.87	0.71	62
0.750		0.150	0.56	0.05	0.25	0.20	0.74	0.40	0.78	0.83	60

RFS, Recurrence-free Survival; HR, Hazard Ratio; CI, Confidence Interval; LL, Lower Limit; UL, Upper Limit; Ave, Average; Prob., Probability; Δ = Arm 1 – Arm 2-year RFS

4. Study Endpoints and Covariates

4.1 Study Endpoints

4.1.1 Primary Endpoint

Recurrence-Free Survival (RFS): The event for RFS is defined as the first of local, regional, or distant recurrence of melanoma or death due to any cause. RFS is defined as time from randomization to the date of the event. The following **four** rules will be used:

1. For subjects having not experienced an event by the cutoff date, RFS will be censored at the last valid tumor evaluation (refer to Section 6 for the definition of last valid tumor evaluation).
2. Subjects who are not confirmed to be disease-free post-surgery (i.e., who do not have an R0 surgical outcome) will be considered to have an event at randomization for RFS.
3. Subjects who withdraw prior to surgery will be considered to have an event at randomization for RFS.
4. Subjects with a non-melanoma tumor or a second malignancy identified before or at surgery will be censored at the day after randomization.

Alternative RFS definitions will be evaluated in sensitivity analyses as follows.

- **RFS Sensitivity # 1: RFS is defined as the first of local, regional, or distant recurrence of melanoma or death due to any cause, regardless of surgical outcome, achieved by omitting rule 2.**

Additional RFS rules related to death will be evaluated in sensitivity analyses as follows.

- **RFS Sensitivity # 2:** The primary analysis of RFS will be repeated with the consideration that in the absence of prior documented recurrence, death beyond 90 days after the last valid tumor evaluation or randomization date (whichever is later) will not be considered an event, and instead will result in a censored RFS time.

- **RFS Sensitivity # 3:** The primary analysis of RFS will be repeated with the consideration that in the absence of prior documented recurrence, deaths not determined to be related to melanoma will not be considered an event. Deaths will be will considered as related to melanoma if the death reason is disease progression or unknown.

RFS sensitivity analysis # 2 and/or # 3 will be conducted if RFS is revised for more than 10% of subjects in either arm due to revised specification, or if the absolute between-arm frequency difference is greater than 5%.

4.1.2 Secondary Endpoint

4.1.2.1 Efficacy Endpoints

4.1.2.1.1 Tumor Response (Arm 1 Only)

The overall tumor response is evaluated up to the earlier of surgical resection or subsequent anti-cancer therapy. Lesion-level response is evaluated up to the earliest event of start of subsequent anti-cancer therapy, merge with another lesion, or resection. Subject-level response is evaluated up to the earliest event of start of subsequent anti-cancer therapy or a partial or complete lesion resection.

Following endpoints will be analyzed for tumor response analysis,

- The overall tumor response is per the modified WHO criteria of tumor.
- Tumor response in injected/uninjected lesion(s)
  - The subject-level injected and uninjected responses will be based on the tumor burden of only injected index lesions including injected new measurable lesions, or the tumor burden of only uninjected index lesions including uninjected new measurable lesions, respectively.
  - The lesion-level injected and uninjected response will evaluate the change of the area of each individual lesion.
- Similar to tumor response in injected/uninjected lesions, lesion-level and subject-level response may also be conducted for lesion subtypes, eg, cutaneous / subcutaneous, nodal.

For the subject-level tumor response by tumor burden type (injected, uninjected lesions), a response is achieved if the percentage of the specific tumor burden type reduction from baseline for both baseline and new measurable lesions (if applicable) is at least 50%. Overall tumor burden is calculated at a visit as the sum of tumor areas of all lesions of the same type. The incidence of subject-level tumor response will be reported as the proportion of subjects with an evaluable tumor burden of the corresponding type that achieves a response.

The incidence of lesion-level response will be reported as the proportion of evaluable lesions in response. A lesion is in response if the decrease in tumor area is ≥ 50%.

All overall tumor burden and lesion response analyses will be based on bi-dimensional lesion measurements reported at a tumor assessment and will not consider injected lesion longest diameters reported at the time of investigational product administration forms.

4.1.2.1.2 Other Endpoints

- Landmark RFS by year: The Kaplan-Meier (K-M) estimate of RFS rate at 1 year, 2 years, 3 years**,** and 5 years
- Rate of histopathology tumor-free margin (R0) surgical resection
- Rate of pathological complete response (pCR) (Arm 1 only)
- Local recurrence-free survival (LRFS): Time from randomization to the earlier date of the first of local disease recurrence or death due to any cause. Local recurrence is defined as histologically or cytologically confirmed reappearance of melanoma in the area of up to 2 cm from the scar from the surgical excision or at the edge of the skin graft if that was used for closure. For subjects not having experienced an event by the cutoff date, LRFS will be censored at the last valid tumor evaluation before the data cutoff date. An event will be imputed at randomization for subjects who do not have an R0 surgical outcome. Subjects with a non-melanoma tumor identified at surgery will be censored at the day after randomization. Sensitivity analyses **as per RFS Sensitivity #1 (see Section 4.1.1) will be performed for LRFS.**
- Regional recurrence-free survival (RRFS): Time from randomization to the date of the first of regional disease recurrence or death due to any cause. Regional recurrence excludes local recurrence and is defined as histologically, cytologically, or radiographically confirmed reappearance of melanoma in the regional lymph node basin. New in-transit melanoma metastases in the regional lymphatic drainage will be reported as regional recurrence. Histological or cytological confirmation of new in-transit metastases is recommended but is not required. For subjects not having experienced an event by the cutoff date, RRFS will be censored at the last valid tumor evaluation before the data cutoff date. Subjects with a non-melanoma tumor identified at surgery will be censored a day after randomization**. Sensitivity analyses as per RFS Sensitivity #1 (see Section 4.1.1) will be performed for RRFS.**

- Distant metastases free survival (DMFS): Time from randomization to the earlier date of the first of distant metastases or death due to any cause. Distant metastases exclude local and regional recurrence and will include distant cutaneous/subcutaneous metastases, distant nodal metastases, or visceral, central nervous system, brain, or bone metastases. For subjects not having experienced an event by the cutoff date, DMFS will be censored at the last valid tumor evaluation before the data cutoff date. Subjects with a non-melanoma tumor identified at surgery will be censored at the day after randomization. Sensitivity analyses **as per RFS Sensitivity #1 (see Section 4.1.1) will be performed for DMFS**.
- Overall survival (OS): Time from randomization to death due to any cause
- Landmark OS by year: The Kaplan-Meier (K-M) estimate of the overall survival rate at 1 year, 2 years, 3 years and 5 years.

4.1.2.2 Safety Endpoints

- Subject incidence of treatment-emergent and treatment-related adverse events
- The incidence of subjects with herpetic lesions found to be positive for talimogene laherparepvec DNA per qPCR in swab samples collected from lesions suspected to be herpetic in origin and the rate of positive lesions (if any); 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

4.1.3 Exploratory Endpoints

The following endpoints are included as exploratory endpoints and their analysis plan may be described in a separate supplemental SAP if appropriate and they may be reported separately from the primary CSR.

- Correlation between baseline intratumoral CD8+ cell density in melanoma lesions and clinical outcomes
- Correlation between changes in an intratumoral CD8+ cell density in melanoma lesions during talimogene laherparepvec treatment and clinical outcomes (for Arm 1 only)
- Correlation between the changes in the population of tumor-specific cytotoxic T-cells in blood samples during treatment and clinical outcomes (for Arm 1 only)
- Assessment of blood and tumor for potential biomarkers which correlate with or predict clinical outcomes to talimogene laherparepvec

4.2 Planned Covariates

The following covariates, including the stratification factors, will be used, with possible exceptions where associated subgroups are too small (eg, < 20 subjects) or have too few events (< 5 events), to examine efficacy **individually and multivariately**:

- Region, if more than 10% subjects in ITT population are from non-USA countries (USA or non-USA)
- Age at baseline: < 50, ≥ 50; < 65, ≥ 65; < 75, ≥ 75 years

- Number of measurable tumors at baseline **(will be dichotomized based on the median)**
- The sum of the products of the two largest perpendicular diameters of baseline measurable lesions (SPD) **(will be dichotomized based on the median)**
- Baseline absolute lymphocyte count: ≤ 1000 vs > 1000
- **Recurrent disease ≥1 year from primary diagnosis vs recurrent disease < 1 year from primary diagnosis vs no prior melanoma**
- Current disease stage (IVRS): IIIB nodal vs IIIB in-transit vs IIC nodal vs IIIC in-transit with nodal vs IV M1a
- Baseline ECOG performance status: 0 vs 1
- BRAF mutation status: mutation present vs mutation not present
- Previous systemic anti-cancer therapy for melanoma: yes vs no
- Previous systemic immune check-point inhibitor therapy for melanoma: yes vs no
- **Planned adjuvant therapy (IVRS): adjuvant systemic therapy with or without radiotherapy vs radiotherapy without adjuvant systemic therapy vs none**

5. Hypotheses and/or Estimations

Formal hypotheses will not be tested in this trial. The study will provide an estimate of the treatment effect, as measured by the difference in RFS of talimogene laherparepvec neoadjuvant therapy followed by surgery compared to surgery alone in subjects with resectable, stage IIIB to IVM1a melanoma.

6. Definitions

The definitions of efficacy endpoints are given in Section 4.

- Investigational Product

Investigational product refers to talimogene laherparepvec in this study.

- Study Drug

Study drug is defined as talimogene laherparepvec in this study.

- Screening Phase

The screening phase is the time period after signing the informed consent and before enrollment when study specific laboratory tests and procedures are performed, and medical history is reviewed to confirm subject eligibility for the study.

- Study Day 1

Study day 1 for a subject is defined as the first day that the study drug is administered to the subject or the surgical resection of melanoma tumor lesion(s), whichever is earlier. Study day 1 is the date of randomization if a subject does not

receive talimogene laherparepvec or the surgical resection of melanoma tumor lesion(s). The day before the study day 1 is referred to as study day −1.

- Baseline

Baseline in general refers to study day 1. The baseline value of a parameter (eg, vital signs, laboratory tests and efficacy endpoints) is considered to be the latest value prior to receiving talimogene laherparepvec or the onset of the surgical resection of melanoma tumor lesion(s), whichever is earlier. If subjects that do not receive study therapy or the surgical resection of melanoma tumor lesion(s), the baseline value will be the last record on or prior to the randomization date.

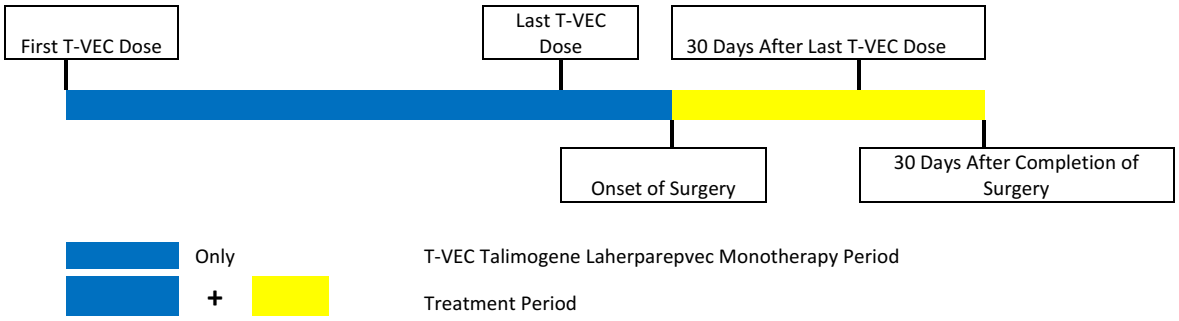
- Treatment Period (also Study Treatment Period, Overall Treatment Period)

Study day 1 through 30 days after the last administration of talimogene laherparepvec or 30 days after the surgical resection of melanoma tumor lesion(s), whichever is later.

- Talimogene Laherparepvec Monotherapy Period (Arm 1 Only)

Talimogene laherparepvec monotherapy period starts at the initiation of the first talimogene laherparepvec dose administration and ends before the surgical resection of the melanoma tumor lesion(s), or 30 days after the last administration of talimogene laherparepvec, whichever is earlier.

Figure 1. Example of Treatment Period and T-VEC Monotherapy Period



- Treatment-Emergent Adverse Events

Treatment-emergent adverse events are events occurring during the Treatment Period. Adverse events occurring on the same day of the start or end day of the treatment period will be counted as treatment-emergent adverse event unless there is evidence to indicate otherwise (eg, the checkbox in CRF, which indicates a AE

onset is prior to first dose of talimogene laherparepvec or surgical resection, is checked).

- Treatment Emergent Death

Treatment emergent death is defined as any death that occurs during the treatment period. Any treatment emergent death will be reported as an SAE.

- Safety Follow-Up

Safety follow-up will be performed approximately 30 (+ 15) days after the last dose of talimogene laherparepvec or the completion of the surgical resection of melanoma tumor lesion(s), whichever is later.

- End of Study

Primary Completion: the time when the last subject is assessed or receives an intervention for the purposes of final collection of data for the primary analysis. The primary completion is anticipated to occur at the later time of either the occurrence of approximately 64 events (local, regional or distant recurrence of melanoma or death) or approximately 2 years after the end of randomization.

End of Trial: the time when the last subject is assessed or receives an intervention for evaluation in the study. The end of trial will occur when the last subject has had the opportunity to complete the long-term follow-up. The end of trial is anticipated to occur approximately 5 years after the end of randomization.

- Long-term Follow-up

Long-term Follow-up will occur every 3 months (± 30 days) for 3 years following the safety follow-up visit and then every 6 months (± 30 days) until death, subject withdraws full consent, or up to 5 years after the last subject is randomized, whichever is first.

- Concomitant Medications

The concomitant medications include all non-anticancer medication during treatment period.

- On-treatment Procedure/Surgery

The on-treatment surgery includes all surgery subjects receive during treatment period except the protocol specified surgical resection of melanoma lesion(s)

- Prior Systemic Anti-cancer Therapies

Prior systemic anti-cancer therapies for melanoma include chemotherapy, chemoradiotherapy, immunotherapy, targeted biologics, targeted small molecules, and chemo-embolization which are not in the setting for adjuvant or neoadjuvant treatment.

- Previous Systemic Immune Check-Point Inhibitor Therapy for Melanoma

Previous systemic immune check-point inhibitor anti-cancer therapies (eg, anti-CTLA4, anti-PD1) not in the setting for adjuvant or neoadjuvant treatment.

- Subsequent Anti-Cancer Therapy for Melanoma after Recurrence

Subsequent anti-cancer therapy for melanoma is defined as any anti-cancer therapy (or treatment) received for melanoma after recurrence. Subsequent anti-cancer therapy for melanoma will be collected from CRF form “Subsequent Anti-Cancer Therapies for Current Malignancy”.

- Adjuvant therapy for Melanoma

Adjuvant therapy is defined as anti-cancer therapy subjects receive after surgical resection of melanoma lesion(s) but before recurrence. Adjuvant therapy will be collected from CRF form “Subsequent Anti-Cancer Therapies for Current Malignancy”.

- Landmark OS by year

Estimated proportion of subjects alive at each year by the Kaplan-Meier method.

- Landmark RFS by year

Estimated proportion of subjects experiencing overall recurrence (local, regional, distant) at each year by the Kaplan-Meier method.

- Response (Arm 1 only)

Refer to Section 4.1.2.1.1.

- Best overall response (BOR) per investigator (Arm 1 only)

Best response for a subject is the best overall response observed across all time points in the following order: CR, PR, SD, PD or UE.

- Histopathology Tumor-free margin (R0) Surgical Resection Rate

Histopathology tumor-free margin (R0) surgical resection is defined by pathologist as absence of ink on the tumor for all disease. R0 surgical resection rate is defined as proportion of subjects with R0 resection. For each subject, the R0 resection status (R0 or non-R0) will be derived from lesion level resection status collected from CRF. In order for a subject to count as an R0 resection event, all lesion observations will be R0 within the given subject. Any observation of a non-R0 (R1, R2) for a subject will make the subject a non-R0 case for this endpoint.

- Pathological Complete Response (pCR) Rate (Arm 1 Only)

Pathological Complete Response (pCR) is defined as no evidence of viable tumor cells on complete pathological evaluation of the surgical specimen per institutional standards of care. pCR rate is defined as proportion of subjects which achieves a pCR. For each subject, the pCR status (pCR or non-pCR) will be determined based on whether or not viable tumor was detected in the surgical specimen.

- Last Valid Tumor Evaluation

For overall tumor response per modified WHO response criteria the last valid tumor evaluation refers to the last valid tumor response assessment (PR, CR, SD or PD) per investigator. For the subject-level tumor response based on the tumor burden of a subset of lesions (injected, uninjected lesions, or other subtype) or the lesion-level response the last valid tumor evaluation refers to the assessment where the corresponding tumor burden or the corresponding individual tumor area is not missing.

- Baseline Lesion

Lesions measured at the most recent visit on or before receiving the first administration of talimogene laherparepvec (i.e., study day 1) will be flagged as baseline lesions. Baseline lesions must have positive tumor area. Lesions reported in the baseline measurable lesion CRF page but with a tumor area of zero or missing value at the last measurement on or before study day 1 will not be considered baseline lesions.

- New Lesion

New lesions are those that first appear after study day 1 with a positive tumor area or lesions without a positive lesion area in the last assessment on or before study day 1

but a positive lesion area from follow up tumor measurements. Baseline tumor area of a new lesion is obtained from the earliest date with positive bi-dimensional tumor measurements after study day 1. If a new lesion(s) that appears in between scheduled tumor response assessments disappear(s) before the next scheduled assessment, there is no need to include this lesion for tumor response evaluation

- Evaluable Lesion

A baseline or new measurable lesion is defined as evaluable if it has an initial positive area followed by at least 1 assessment with a non-missing area.

- Injected/Uninjected Lesion

Baseline or new lesions that are reported as injected with talimogene laherparepvec at any time during the treatment period are considered injected lesions. Baseline or new lesions that are not reported as injected with talimogene laherparepvec are considered uninjected lesions.

- Baseline Tumor Burden

The baseline tumor burden for a subject is the sum of the initial area of baseline lesions of the specific lesion type.

- Tumor Burden

The tumor burden for a subject at a visit is the sum of the area at that visit of baseline lesions of the specific lesion type.

- Evaluable Tumor Burden

A subject is defined as having an evaluable tumor burden of a specific type (injected, uninjected) if they have at least one evaluable tumor burden assessment of the specific type. An evaluable assessment is one with a tumor burden that includes all lesions of the specific type with a non-missing area that were included in a baseline tumor burden assessment.

7. Analysis Subsets

7.1 Intent to Treat Analysis Set

The primary analysis of all efficacy endpoints of the study will be conducted on the intent to treat analysis set defined as all randomized subjects who were randomized to either treatment arm. All subjects will be analyzed according to their treatment randomization.

7.2 Safety Analysis Set

The safety analysis set will include all randomized subjects who received talimogene laherparepvec or surgical resection of melanoma tumor lesion(s). The safety analyses will be performed by treatment received (i.e., neoadjuvant talimogene laherparepvec + surgical resection versus surgical resection alone). Subjects will be classified in the talimogene laherparepvec arm for safety analyses if they received at least one administration of talimogene laherparepvec at any time during the treatment period.

7.3 Efficacy Analysis Set

The efficacy analysis set will include all subjects in Arm 1 who received at least one dose of talimogene laherparepvec and surgical resection of melanoma tumor lesion(s) and all the subjects in Arm 2 who received surgical resection of melanoma tumor lesion(s). In addition, all subjects in the efficacy analysis set must have achieved histopathologic tumor-free margin (R0) status after surgery and should not have any important protocol deviations considered to affect efficacy outcomes listed in Section 10.3.

Subject with any of the following eligibility criteria IPDs will be removed from the Efficacy Analysis Set:

- **Subject does not have histologically proven stage IIIB, IIIC, or IVM1a melanoma eligible for complete surgical resection**
- **Subject is not a candidate for intralesional therapy**

For purposes of the Efficacy Analysis Set sensitivity analysis, subjects who did not receive the assigned treatment arm will be excluded from the analysis.

The **Efficacy Analysis Set will** be used for a sensitivity analysis of **the primary** efficacy endpoint.

7.4 Subgroup Analyses

Efficacy analyses will be conducted utilizing the subgroups as defined in Section 4.2.

Safety analyses will be conducted on the following subgroups: age group (< 65, ≥ 65 and < 75, ≥ 75), sex, and race (white vs not white).

8. Planned Analyses

8.1 Interim Analysis and Early Stopping Rules

Two interim analyses with no formal stopping rules are planned to evaluate safety when approximately 40 and 75 subjects randomized to talimogene laherparepvec have had the opportunity to complete the safety follow-up visit. **The database will be cleaned**

and a database snapshot will be used in each analysis. A Data Review Team (DRT) independent of the talimogene laherparepvec product team will review the first interim analysis (see Section 8.1.1). The DRT will also review some efficacy endpoints (eg, R0 resection rate, pCR rate, response to neoadjuvant treatment).

The **Amgen study** team will review the second interim analysis because it will occur after all subjects have been already treated, have completed safety follow-up, and no additional interventions that may influence safety or efficacy of the investigational product are planned after this time point. The **Amgen study** team will also review some efficacy endpoints (eg, R0 resection rate, pCR rate, response to neoadjuvant treatment).

An additional third interim analysis with no formal stopping rule is planned to evaluate RFS approximately 1 year after the end of randomization. This analysis will be conducted by the Amgen study team. **The database will be cleaned and a locked database will be used for this analysis.**

Ad hoc analyses for safety or some efficacy endpoints (eg, response to neoadjuvant treatment, R0 resection rate, pCR rate) may be conducted before the planned primary and/or final analyses if interim data are required for submission to regulatory authorities. These ad hoc analyses will be executed by the **Amgen** study team.

8.1.1 Data Review Team (DRT)

A Data Review Team (DRT) independent of the talimogene laherparepvec product team will be formed. The DRT will consist of one Amgen biostatistician and two Amgen clinicians (one from Clinical Development and one from Global Safety) who collectively have experience in oncology clinical research and in the conduct and monitoring of randomized clinical trials. In addition, the DRT will include one external clinical expert (eg, a surgical oncologist in melanoma) not directly involved in the conduct of the study. The DRT will be supported by a statistician internal to Amgen but independent of the talimogene laherparepvec product team. This DRT will review unblinded safety data at the first interim analyses. This independent DRT will be governed by a study-specific DRT charter.

8.2 Primary Analysis

The primary analysis for RFS will occur at the later time of either the occurrence of approximately 64 events (local, regional, or distant recurrence of melanoma or death) or approximately 2 year after the end of the randomization. The database will be cleaned and a locked database will be used in the analysis.

8.3 Final Analysis

The final analysis will occur approximately 5 year after the end of randomization. The database will be cleaned and a locked database will be used in analysis.

8.4 Additional Analysis

An additional analysis will also occur approximately 3 years after the end of randomization. The database will be cleaned and a locked database will be used in analysis.

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

Amgen's Clinical Data Management department will provide all data to be used in the planned analyses. This study will use the RAVE database. Antibody will be processed by vendor for Amgen BSM and then transferred to Amgen's Medidata Rave database to be accessed for analysis.

9.3 Handling of Missing and Incomplete Data

9.3.1 Efficacy

The method for handling missing data is described in the definition for each of the efficacy endpoints (Section 4).

9.3.2 Safety

Partial or missing dates of adverse events and concomitant medications will be imputed.

[redacted]

9.4 Detection of Bias

9.4.1 Important Protocol Deviations (IPD) Impact

All IPDs will be reported, documented, and stored in eClinical (Clinical Trial Management System). An IPD report will be produced using Cognos by the Clinical Study Manager for regularly scheduled Study Team IPD Review meetings and before each analysis.

Lack of protocol compliance and the potential for biased statistical analyses will be examined by tabulating subjects with important protocol violations by study arm. A

listing of all IPDs will also be provided. High rates of lack of compliance or differing rates between study arms could be an indication of bias. A sensitivity analysis will be performed on the primary efficacy endpoint for the impact of important protocol deviations by repetition of analysis using the **E**fficacy **A**nalysis **S**et (**see** Section 7.3).

9.4.2 Timing Bias Impact

Differences between the arms in the Schedule of Assessments (SOA) may result in a possible bias in the detection of overall recurrence (local, regional, or distant recurrence).

If 10% or more of the overall post-surgery recurrences (both arms combined), regardless of surgical outcome, occur within the first 6 months post-surgery, then a stratified piece-wise Cox model will be used to allow estimation of within interval treatment hazard ratios of talimogene laherparepvec plus surgery compared to surgery alone adjusting for the randomization factors for the overall ITT population (Collett 2003). The time intervals are defined as: ≤ 6 months post-surgery, then every 6 months thereafter. Please see Appendix B for code fragments.

9.4.3 Subsequent Anti-Cancer Therapy Impact

To evaluate the potential bias introduced by the use of subsequent anti-cancer therapy on RFS, **three separate sensitivity analyses will be performed based on the type of anti-cancer therapy received: (a) any subsequent anti-cancer therapy, (b) adjuvant anti-cancer therapy for melanoma, (c) non-adjuvant anti-cancer therapy for melanoma. For analyses (a) and (b), subjects who receive the respective subsequent anti-cancer therapy type prior to an RFS event will be censored at the last evaluable tumor assessment prior to receiving the respective subsequent anti-cancer therapy type. For analysis (c), subjects who receive non-adjuvant anti-cancer therapy for melanoma prior to an RFS event will be an event at the last evaluable tumor assessment prior to receiving the non-adjuvant anti-cancer therapy for melanoma.**

These sensitivity analyses will be repeated using the alternative definition for RFS as defined in RFS Sensitivity #1 (Section 4.1.1).

9.5 Outliers

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

9.6 Distributional Characteristics

Time to event endpoints will be evaluated with the non-parametric log-rank test. The semi-parametric proportional hazards model will be used to estimate the treatment hazard ratio with and without stratification which assumes the hazard for events are proportional over time between the treatment arms within strata.

Log-log curves will be plotted to assess the hazard proportionality between treatment arms. The curves will be repeated by randomization stratum. If the hazards are proportional between treatment arms, the curves must be parallel. Also, KM curves should not cross (unless the hazards are similar which means no treatment difference).

The assumption of proportional hazards will be investigated adding a treatment*log(time) interaction in the Cox model. If the hazards are proportional between treatment arms, the treatment*log(time) interaction must be non-significant (see code fragment in Appendix B).

A Kolmogorov-type supremum test will be used to test the proportional hazards assumption for treatment comparison based on an analysis of cumulative residuals overall and within the randomization strata (Lin, Wei, and Ying, 1993, see code fragment in Appendix B).

The diagnostic approaches described before will be conducted by the study statistician and no official output will be provided unless necessary.

If the proportional hazards assumption is strongly violated, then an alternative analysis will be considered. A stratified piece-wise Cox model will be used to allow estimation of an overall weighted-average hazard ratio as well as within interval treatment hazard ratios (Collett 2003). Three intervals will be used in the piecewise model, one from time 0 to 12 months, one from time 12 to 24 months, and the other beyond 24 months (see code fragment in Appendix B), other intervals may be evaluated to assess sensitivity.

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.

10. Statistical Methods of Analysis

10.1 General Principles

All analyses will be descriptive with no formal hypothesis testing.

In principle, mean, standard deviation, median, first and third quartiles, minimum and maximum will be calculated for continuous variables; frequency and percent will be calculated for binary and categorical variables. Graphical summaries of the data may also be presented.

Analyses of tumor tissue or serum biomarkers may be performed after collection of all samples during the conduct of the study and therefore may be reported after the primary analysis of efficacy endpoint.

The primary analysis for efficacy endpoints will be performed on the Intent to Treat Analysis Set by randomized treatment. In addition, some of the analysis may be repeated on the Efficacy Analysis Set by treatment actually received. Safety analyses will be conducted on the Safety Analysis Set.

Data analysis is scheduled to occur at the following time points:

- Two interim analyses with no formal stopping rules will be performed when approximately 40 and 75 subjects randomized to talimogene laherparepvec have had the opportunity to complete the safety follow-up visit. (Please refer to section 8.1 for details).
- An additional third interim analysis with no formal stopping rule is planned to evaluate RFS approximately 1 year after the end of randomization. This analysis will be conducted by the Amgen study team. (Please refer to section 8.1 for details).
- The primary analyses for RFS will occur at the later time of either the occurrence of approximately 64 events (local, regional, or distant recurrence of melanoma or death) or approximately 2 years after end of randomization. The primary analysis for certain secondary endpoints (eg, response to neoadjuvant treatment, R0 resection rate, pCR rate, and safety) will be performed using the data from the second interim analysis.
- The final analyses will occur approximately 5 years after the end of randomization. An additional analysis will also occur approximately 3 years after the end of randomization.
- Ad hoc analyses for safety or some efficacy endpoints (eg, response to neoadjuvant treatment, R0 resection rate, pCR rate) may be conducted before the planned primary and/or final analyses if interim data are required for submission to regulatory authorities. These ad hoc analyses will be executed by the **Amgen study** team.

10.2 Subject Accountability

The number of subjects in each of the strata at randomization will be summarized by randomized treatment.

A subject disposition summary table will be produced. The number of subjects screened, randomized who are treated (talimogene laherparepvec or/and surgery) and not treated will be tabulated by randomized treatment. Reasons for discontinuation of talimogene laherparepvec and not receiving surgery, end of study status (completed vs. did not complete), reason for not completing the study, the status of completing safety follow-up, and the reason for not completing safety follow-up will also be summarized by randomized treatment. The number of screen failure subjects and screen failure reasons will also be provided.

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, deviation codes and descriptions will be used during the course of the study. The final IPD list is used to produce the summary of IPDs table and the list of subjects with IPDs by deviation category, subcategory, and randomized treatment. A separate summary table and listing will also be provided for inclusion/exclusion deviations.

Subjects with the IPDs considered affecting efficacy outcomes will be excluded from the **E**fficacy **A**nalysis **S**et (**see Section 7.3**).

10.4 Demographic and Baseline Characteristics

Summary statistics of the **following demographic and baseline disease characteristics** will be tabulated by randomized treatment arm on the Intent To Treat Analysis Set:

- **Region: USA or non-USA**
- **Age at baseline: < 50, ≥ 50; < 65, ≥ 65; < 75, ≥ 75 years**
- **Sex: male or female**
- **Number of measurable tumors at baseline**
- **Presence of uninjected tumors at baseline and during the treatment: yes vs no (Arm 1 only)**
- **The sum of the products of the two largest perpendicular diameters of baseline measurable lesions (SPD)**

- **Histogenetic subtype: acral lentiginous melanoma vs desmoplasitic vs lentigo maligna melanoma vs nodular melanoma vs superficial spreading melanoma vs unclassifiable vs missing**
- **Current disease stage (IVRS): IIIB nodal vs IIIB in transit vs IIC nodal vs IIIC in-transit with nodal vs IV M1a**
- **Current disease stage (CRF): IIIB vs IIIC vs IV M1a**
- **Time from initial histological diagnosis of melanoma to study entry**
- **Baseline absolute lymphocyte count: ≤ 1000 vs > 1000**
- **Recurrent disease ≥ 1 year from primary diagnosis vs recurrent disease < 1 year from primary diagnosis vs no prior melanoma**
- **Baseline ECOG performance status: 0 vs 1**
- **BRAF mutation status: mutation present vs mutation not present vs missing**
- **Baseline HSV-1 status: negative vs positive vs equivocal vs unknown**
- **Previous systemic anti-cancer therapy for melanoma: yes vs no**
- **Previous systemic immune check-point inhibitor therapy for melanoma: yes vs no**
- **Prior surgical procedures: yes vs no**
- **Prior radiotherapy: yes vs no**
- **Planned adjuvant therapy (IVRS): adjuvant systemic therapy with or without radiotherapy vs radiotherapy without adjuvant systemic therapy vs none**

10.5 Efficacy Analyses

The following table (Table 2) summarizes the efficacy endpoints and planned analysis methods.

Table 2. Study Design Characteristics Example Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Primary Endpoint		
• RFS	• Log-rank test and 80% CI for the treatment hazard ratio from a proportional hazards model without stratification in the ITT analysis set. • Kaplan-Meier (K-M) estimation by randomized treatment arm. • Simultaneous 80% confidence band for each arm KM estimate from 1 to 5 years per equal-precision band method. • 80% CI for K-M quartiles by arm (where estimable) per Brookmeyer & Crowley. • 80% CI for between-arm quartile difference in KM estimate via bootstrap.	• Repeat with stratification by the two randomization factors • If 10% or more of the overall recurrence (both arms combined) occur within the first 6 months post-surgery, the stratified analysis will be repeated with stratified piece-wise Cox model to estimate within interval hazard ratios adjusting for randomization factors. • Estimate the Cox model hazard ratio with Robins IPCW weighting censoring RFS at the start of all or selected adjuvant therapy for melanoma. • If warranted (eg, the planned adjuvant therapy stratification factor is discordant with the actual for > 10% of all subjects), the stratified analysis will be repeated replacing planned adjuvant therapy with actual adjuvant therapy and/or adjuvant therapy containing selected types of systemic therapy, including immunotherapy. • Cox model and KM analysis of RFS according to randomized treatment and actual adjuvant systemic therapy (yes vs no) with 3 comparisons: all subjects, those with adjuvant systemic therapy, and those without adjuvant systemic therapy. • Repeat with and/or without stratification by the two randomization factors in the efficacy analysis set, if necessary. • Univariate and multivariate proportional hazards models without stratification may be used to explore the prognostic and/or predictive value of baseline covariates using primary endpoint definition and alternative endpoint definition #1 in Section 4.1.1.. • Subgroup analyses will be evaluated with a log-rank test and corresponding proportional hazard model without stratification using primary endpoint definition and alternative endpoint definition #1 in Section 4.1.1. • Repeat subgroup analyses with a log-rank test and proportional hazards model with stratification by two randomization factors. • Repeat using sensitivity analysis alternative endpoint definitions (see Section 4.1.1). • Repeat based on the type of anti-cancer therapy received: any subsequent anti-cancer therapy, adjuvant anti-cancer therapy for melanoma, non-adjuvant anti-cancer therapy for melanoma using primary endpoint definition and alternative definition # 1 in Section 4.1.1. Subjects who receive the respective subsequent anti-cancer therapy type prior to an RFS event will be censored at the last evaluable tumor assessment prior to receiving the respective subsequent anti-cancer therapy type.

Page 1 of 3

Table 2. Study Design Characteristics Example Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Key Secondary Endpoints		
• OS • DMFS • LRFS • RRFS	• The same methods described for the primary endpoint.	• Excluding OS, repeat with stratification by two randomization factors. • For LRFS, assess the equality between treatment groups using Gray's test, and estimate treatment HR with 95% CI stratified by the randomization factors using subdistribution proportional hazards model, with distant recurrence as competing risk. • DMFS will be compared with Gray's test and the subdistribution proportional hazards model with censoring considered a competing risk if the subject has a primary reason for ending the study after a local or regional recurrence of withdrawal of consent or lost to follow-up. • RRFS will be compared with Gray's test and the subdistribution proportional hazards model with censoring considered a competing risk if the subject has a primary reason for ending the study after a local recurrence of withdrawal of consent or lost to follow-up.
• Landmark RFS, LRFS, RRFS, DMFS by year; • Landmark OS by year	• 80% CI for the K-M estimate in the ITT analysis set by randomized treatment arm with s.e. per Greenwood. • 80% CI for the between-arm difference with s.e., per Greenwood.	• Excluding OS, repeat using the alternative endpoint definition #1 in Section 4.1.1.

Page 2 of 3

Table 2. Study Design Characteristics Example Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Key Secondary Endpoints		
• Rate of R0 surgical resection; • Rate of pCR	• 80% exact CI by randomized treatment arm in the ITT analysis set per Clopper-Pearson. • 80% approximate exact CI for the between-arm difference per Wilson's score method with continuity correction.	• Repeat using the efficacy analysis set, if necessary. • Multivariate logistic models may be used to explore the prognostic and/or predictive value of baseline factors.
Tumor response (Arm 1 Only): • Overall Tumor Response (per Investigator) • Individual injected lesion response overall and by subtype (cutaneous / subcutaneous vs lymph node) • Individual uninjected lesion response overall and by subtype (cutaneous / subcutaneous vs lymph node) • Subject response on overall injected lesions • Subject response on overall uninjected lesions	• For response rate use the Clopper-Pearson method to calculate exact 80**%** CIs; • For time to response the Kaplan-Meier (K-M) estimate • The tumor response analysis is performed in the ITT analysis set.	• Repeat using the efficacy analysis set, if necessary.

Page 3 of 3

10.5.1 Analyses of Primary Efficacy Endpoint(s)

An overall between-group difference in RFS will be evaluated with a log-rank test and corresponding proportional hazards model without stratification. Sensitivity analyses, using alternate definitions of RFS as defined in section 4.1.1, will be evaluated in the same manner. Additionally, for RFS, a log-rank test and proportional hazards model with stratification by the two randomization factors will also be conducted.

A sensitivity analysis will be performed censoring RFS at the start of all or selected actual adjuvant therapy for melanoma. However, censoring at start of all or selected adjuvant therapy for melanoma may introduce dependent censoring as the ability to receive further adjuvant therapy will likely be correlated with RFS. An inverse probability censoring weighting (IPCW) method (Robins, et al) which may reduce the potential bias will also be used to obtain an estimate of the treatment effect on RFS and may include baseline covariates in Section 4.2.

In order to assess the impact of imbalances in the type of adjuvant therapy, **a stratified** Cox model will be repeated stratifying actual adjuvant systemic therapy (yes vs no). A KM time to event plot for RFS will be generated by 4 groups (Arm 1 with adjuvant systemic therapy , Arm 1 without adjuvant systemic therapy, Arm 2 with adjuvant systemic therapy, Arm 2 without adjuvant systemic therapy), and 2 separate KM plots will be generated by randomized treatment group in subjects with and without adjuvant systemic therapy respectively. If warranted due to actual adjuvant therapy (eg, the planned adjuvant therapy stratification factor is discordant with the actual for > 10% of all subjects), the stratified Cox model analysis will be repeated stratifying on **actual** adjuvant therapy type (**adjuvant systemic therapy with or without radiotherapy vs radiotherapy without adjuvant systemic therapy vs none).**

To explore the prognostic and/or predictive value of the covariates listed in Section 4.2, univariate and multivariate proportional hazards models without stratification will be used. Subgroup analyses will be evaluated with a log-rank test and corresponding proportional hazards model without stratification.

10.5.2 Analyses of Secondary Efficacy Endpoint(s)

KM estimates will be calculated for RFS, LRFS, RRFS, DMFS, RFS landmarks by year, OS landmarks by year and quartiles. Greenwood's formula (Kalbfleisch and Prentice, 1980) for standard error will be used to calculate CIs for each group and between-group differences in landmark KM rate estimates. CIs for quartiles of each group will be estimated per Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) using a

log-log transformation and bootstrap methods will be used to estimate CIs for between-group differences (see Section B.3 for SAS code fragments). For each group the equal-precision band method will be used to calculate a simultaneous confidence band for RFS, LRFS, RRFS and DMFS over the interval from **1** to 5 years (Nair, 1984) (see Section B.2 for SAS code fragments).

An overall between-group difference in LRFS, RRFS, DMFS, and OS will be evaluated with a log-rank test and corresponding proportional hazard model with or without stratification by the two randomization factors.

Informative censoring due to distant recurrence is expected for LRFS. For LRFS, after considering distant recurrence as competing event, Gray's test (Gray, 1988) will be performed to assess the equality of cumulative incidence functions between treatment groups, and the subdistribution proportional hazards model (Fine and Gray, 1999) will be used to calculate treatment HR with 95% CI with or without stratification by the randomization factors. The same sensitivity analysis will be performed for DMFS with censoring considered a competing risk if the subject has a primary reason for ending the study after a local or regional recurrence withdrawal of consent or lost to follow-up. The same analysis will be performed for RRFS with censoring considered a competing risk if the subject has a primary reason for ending the study after a local recurrence of withdrawal of consent or lost to follow-up. Please see Appendix B for SAS code fragments.

A sensitivity analysis will be performed for LRFS censoring at the start of all or selected actual adjuvant therapy (systemic therapy radiotherapy) for melanoma. However, censoring at start of selected adjuvant therapy (systemic therapy radiotherapy) for melanoma may introduce dependent censoring as the ability to receive further adjuvant therapy will likely be correlated with LRFS. An inverse probability censoring weighting (IPCW) method (Robins, et al) which may reduce the potential bias will also be used to obtain an estimate of the treatment effect on LRFS and may include baseline covariates in Section 4.2. Similar analyses will be performed on DMFS and RRFS.

If warranted due to actual adjuvant therapy (eg, the planned adjuvant therapy stratification factor is discordant with the actual for > 10% of all subjects), the stratified Cox model analysis will be repeated stratifying on **actual** adjuvant therapy type

(**adjuvant systemic therapy with or without radiotherapy vs radiotherapy without adjuvant systemic therapy vs none).**

Multivariate proportional hazards models may be used to explore the prognostic and/or predictive value of baseline factors.

The Clopper-Pearson method (Clopper and Pearson, 1934) will be used to calculate exact CIs for binary endpoints (eg, overall tumor response, response in injected and uninjected lesions, R0 rate, pCR) (see Section B.1 for SAS code fragments). Wilson's score method with continuity correction (Newcombe, 1998) will be used to calculate an approximate exact CI for between-group differences in binary rates (see Section B.4 for SAS code fragments). Multivariate logistic models may be used to explore the prognostic and/or predictive value of baseline factors.

10.6 Safety Analyses

10.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 19.1 or later will be used to code AEs to a system organ class (SOC) and a preferred term within the SOC. The CTCAE version 3.0 will be used to grade severity of AEs.

Only treatment-emergent adverse events (see Section 6 for definition) will be summarized by treatment received. Subject incidence rates of all treatment-emergent, grade ≥ 3, serious, non-serious, treatment related, serious treatment related, non-serious treatment related, events leading to talimogene laherparepvec withdrawal/delay/less than 4ml volume per administration, events leading to cancel the scheduled surgical removal of the melanoma tumor lesion(s), fatal adverse events and local effects on the lesion (i.e., pain, inflammation and ulceration) as well as adverse events of interest will be tabulated by SOC and PT in descending order of frequency within an SOC. For all treatment-emergent adverse events additional summary tables will also be provided by (1) SOC, PT in descending order of frequency within an SOC and worst CTCAE grade, and (2) PT in descending order of frequency.

In addition, certain categories of adverse events occurring within the talimogene laherparepvec monotherapy period and within the period after the start of the first dose of talimogene laherparepvec of 10^6 PFU/mL but before the first dose talimogene laherparepvec of 10^8 PFU/mL will be summarized separately by system organ class and preferred term in descending order of frequency. These categories of adverse events include all treatment-emergent adverse events, serious treatment-emergent adverse

events, non-serious treatment-emergent adverse events, talimogene laherparepvec related treatment-emergent adverse events.

Subgroups group analyses for age group (< 65, ≥ 65 and < 75, ≥ 75), sex, and race (white vs not white) will be presented for all treatment-emergent adverse events by system organ class and preferred term by treatment received in descending order of frequency.

Narratives of deaths through either the 30 days after the last dose of talimogene laherparepvec or the surgical resection of melanoma tumor lesion(s), whichever is later, will be provided.

Adverse events after the completion of the surgical resection of melanoma tumor lesion(s) through the 30 days after the last dose of talimogene laherparepvec or the completion of the surgical resection of melanoma tumor lesion(s), whichever is later, will be also be summarized.

A listing of treatment-related AEs reported in long term follow-up will be provided for the final CSR and may be analyzed earlier as part of a program-wide analysis.

10.6.2 Laboratory Test Results

Laboratory values will be summarized descriptively. In addition, grade shifts in laboratory value from baseline to worst on-study value (according to NCI Common Toxicity Criteria (CTC) version 3.0 grading) and grade 3 or higher laboratory toxicities will be presented for selected laboratory tests with CTC grade. Textbook lower and upper limits of the normal ranges will be used for CTC grading, where applicable. For serum lactate dehydrogenase (LDH) local laboratory range will be used.

10.6.3 Vital Signs

Descriptive statistics will be presented for systolic blood pressure, diastolic blood pressure, pulse, respiration rate and temperature for baseline, each post-baseline visit, and change from baseline.

10.6.4 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.

10.6.5 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.

10.6.6 Antibody Formation

HSV-1 serostatus will be summarized for each visit for patients who receive talimogene laherparepvec. IgG reactivity with HSV-1 will be determined. The data summaries and baseline serostatus analyses will be performed for IgG. Treatment-emergent adverse events will be summarized for subjects who receive talimogene laherparepvec by baseline HSV-1 serostatus by system organ class and preferred term. In addition, treatment-emergent immune-related adverse events of interest will also be summarized by baseline HSV-1 serostatus and preferred term for patients treated with T-VEC.

The incidence of subjects with herpetic lesions found to be positive for talimogene laherparepvec DNA per qPCR in swab samples collected from lesions suspected to be herpetic in origin and the rate of positive lesions (if any) will be calculated.

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 reported in the final CSR. Listings of reported cases with available qPCR testing results will be provided. Additionally, unintended exposures will be analyzed periodically as part of a program-wide analysis.

10.6.7 Exposure to Investigational Product

Summary statistics for the total number of talimogene laherparepvec doses administered, total volume of talimogene laherparepvec administered, duration of first talimogene laherparepvec to last talimogene laherparepvec administration, and average volume of talimogene laherparepvec per visit will be provided for subjects who receive talimogene laherparepvec separately for the first dose of talimogene laherparepvec of

10^6 PFU/mL and the subsequent doses. The incidence and reasons for drug withdrawal and injection less than 4 ml will be tabulated for subjects who receive talimogene laherparepvec.

A listing of unique manufacturing lot number(s) for talimogene laherparepvec and subject listings of manufacturing lot number(s) of talimogene laherparepvec will be provided.

The protocol specified surgical resection of melanoma tumor lesion(s) will be summarized. For subjects who do not receive the surgery the reason why the surgery is not conducted will also be summarized.

10.6.8 Exposure to Concomitant Medication/Adjuvant Therapy/Subsequent Anti-cancer Therapy

Prior systemic or regional anti-cancer therapies for current malignancy of melanoma, prior surgeries for current malignancy of melanoma and prior radiation therapy for current malignancy of melanoma will be summarized separately.

The concomitant medications/on-treatment surgery include all /surgery except the protocol specified surgical resection of melanoma lesion(s) after randomization and all medications after randomization before receiving surgery except for talimogene laherparepvec. The number and proportion of subjects receiving any concomitant medication during the treatment period will be summarized by preferred term as coded by the World Health Organization Drug (WHODRUG) dictionary.

The subject incidence and duration of subsequent adjuvant therapy for melanoma will be summarized overall, by type (systemic therapy with radiotherapy, systemic therapy without radiotherapy, radiotherapy without systemic, no subsequent adjuvant therapy at all), and by specific therapy components**. Time to first use of subsequent adjuvant anti-cancer therapy for melanoma will be summarized.**

The subject incidence and time to first use of subsequent non-adjuvant anti-cancer therapy for melanoma will be summarized.

11. Changes From Protocol-specified Analyses

The changes from protocol-specified analyses are as following,

- In the statement regarding subjects who are not confirmed to be disease-free post-surgery or who withdraw prior to surgery, failure was changed from “a day after randomization” to “at randomization” for RFS, RRFS, LRFS and DMFS in section 3.2.
- Safety endpoints related to talimogene laherparepvec DNA shedding and potential or known unintended exposure to talimogene laherparepvec were added as safety endpoints. The current protocol mentioned the analysis of these endpoints without specifically claiming them as safety endpoints.

12. Literature Citations / References

Amgen, Protocol : “A Phase 2, Multicenter, Randomized, Open-label Trial Assessing the Efficacy and Safety of Talimogene Laherparepvec Neoadjuvant Treatment Plus Surgery Versus Surgery Alone for Resectable, Stage IIIB to IVM1a Melanoma.” Amendment 1, 09 June 2014

Brookmeyer R, Crowley J. A confidence interval for median survival time. *Biometrics*. 1982;38:29-41.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934:26:404-413.

Fine, J. and Gray, R. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94: 496-509.

Gray, R. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. The Annals of Statistics. 1988;16: 1141–1154.

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York, NY: John Wiley;1980.

Nair VN. Confidence bands for survival functions with censored data: a comparative study. *Technometrics*. 1984;26:265-275.

Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*. 1998;17:873-890.

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

13. Data Not Covered by This Plan

The analysis on exploratory endpoints described in Section 4.1.3 is not covered by this SAP.

14. Appendices

Appendix A. Conventions for Clinical Data That Require Imputation for Partial or Missing Dates

[redacted]

[redacted]

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

Code Fragments

B.1 Exact Confidence Interval

```
proc freq data=dat1;
   tables R0/ binomial(exact) alpha=.2;
   by TRTGRP;
run;
```

B.2 Nair Equal-Precision Band

```
proc lifetest data=dat1 plots=survival(cb=ep test);
   time OS * CENSOR(1);
   strata TRTGRP;

run;
```

B.3 Bootstrap for Confidence Interval of Quantile Difference per K-M Estimate

```
%let rep = 100;
proc surveyselect data= os out=bootsample
     seed = 1347 method = urs
       samprate = 1 outhits rep = &rep;
run;
ods listing close;

ods output  Quartiles=TimeToDeath;
proc lifetest data=bootsample alpha=0.2 ALPHAQT=0.2 plots=(s);
  time OSM*CENSOR(1);
  strata TRTGRP;
  by Replicate;
run;
ods output close;
```

B.4 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 denominator 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);
```

B.5 Assumption of non-proportional Hazard: Treatment*log (time) Interaction in the Cox Model

```
proc phreg data=dat1;
  model OS * CENSOR(1) = TRTGRP TRTGRP_T / ties=exact rl=pl;
  TRTGRP_T = TRTGRP *log(OS);
  strata randomization_strata;
  proportionality_test: test TRTGRP_T;
run;
```

B.6 Assumption of non-proportional Hazard: Kolmogorov-type Supremum Test

```
ods select cumulativeresiduals scoreprocess;
proc phreg data=dat1;
  model OS*CENSOR(1) = TRTGRP / ties=exact rl=pl;
* Delete strata statement to assess overall proportional hazards;
  strata randomization_strata;
  assess ph / resample seed=1234;
run;
```

B.7 Piecewise Cox Model

```
proc phreg data= dat1;
  model OS*CENSOR(1) = TRTGRP1 TRTGRP2 TRTGRP3 / ties=exact rl=pl;
  TRTGRP1 = TRTGRP *(OS lt 12);
  TRTGRP2 = TRTGRP *(12 le OS lt 24);
  TRTGRP3 = TRTGRP *(24 le OS);
  strata randomization_strata;
run;
```

B.8 Estimates Crude Cumulative Incidence Function and Test Equality of Cumulative Incidence (Gray's Test)

```
proc lifetest data=dat1;
    time LRFS*Status(0)/eventcode=1;
    strata TRTGRP / order=internal;
    *test for overall difference between groups;

    *strata randomization_strata/group=TRTGRP order=internal;
    * stratified test;
run;

/* 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 .*/
```

B.9 Fine-Gray Subdistribution Proportional Hazards Model

```
Proc phreg data=dat1 ;
   class TRTGRP (order=internal ref=first) param=galm;
   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;

/* 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. */
```

B.10 IPCW

```
/*Model probability of not receiving all or selected actual adjuvant
therapy for melanoma including time-independent covariate(s)*/
proc logistic data=time_indep;
   class <categorical time-independent covariate(s)>;
   model subther(event='0') = <time-independent covariate(s)>;
   output out=out1 prob=num;
run;

proc sort data = time_dep;
   by descending subther;
run;

/*Model probability of not receiving all or selected actual adjuvant
therapy for melanoma including time-independent covariate(s) and time-
dependent covariate(s)*/
proc genmod data = time_dep descending order=data;
   class <categorical time-independent covariate(s)> <categorical time-
dependent covariate(s)> usubjid;
   model subther = <time-independent covariate(s)> <time-dependent
covariate(s)> / d=binomial link=logit;
   repeated subject=usubjid;
   output out=out2 prob=den;
run;

proc sort data=out1;by usubjid;run;
proc sort data=out2;by usubjid;run;

/*Calculate weights*/
data dat1;
   merge out1 out2;
   by usubjid;
run;

data dat2;
   set dat1;
   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 = 1 then weights = (1-w_num)/(1-w_den);
   else if subther = 0 then weights = w_num/w_denom;
run;
```

```
data dat3;
   merge dat2(in=a) adam.adtte(in=b where=(paramcd='RFSI' keep=usubjid
paramcd aval cnsr);
   by usubjid;
   if a and b;
run;

/*HR and corresponding CI using IPCW*/
proc phreg data = dat3;
   class trt01p(ref="SURGERY");
   model aval*cnsr(1) = trt01p;
   freq weights / notruncate;
   hazardratio trt01p / alpha=0.05 cl=wald diff=ref;
  strata <stratification factors>;
run;
```