Statistical Analysis Plan

NCT03703310

Sponsor:	PellePharm, Inc.
Protocol No:	Pelle-926-301
Protocol Title:	A Multicenter, Randomized, Double-blind, Vehicle-controlled, Phase 3 Efficacy and Safety Study of Patidegib Topical Gel, 2%, for the Reduction of Disease Burden of Persistently Developing Basal Cell Carcinomas (BCCs) in Subjects with Basal Cell Nevus Syndrome
Document Date:	17-Jun-2020
Document Version:	6.0

APPROVAL SIGNATURES

Synteract Personnel

Title	Printed Name	Signature	Date
Senior Biostatistician	[redacted]		

PellePharm Personnel

Title	Printed Name	Signature	Date
Vice President, Clinical Development	[redacted]		

DOCUMENT HISTORY

Version	Date	Description
0.8	15-May-2018	Original document
2.1	10-Oct-2018	Updated SAP based on Pelle-926-301 Protocol Amendment 1 (06 September 2018)
3.0	15-May-2019	Updated SAP to include further clarifications on derivations, added further sensitivity analyses for the primary endpoint, and included derivation details for questionnaires.
4.0	13-Jan-2020	Updated SAP based on Pelle-926-301 Protocol Amendment 2 (18 July 2019) and associated country specific amendments, added additional sensitivity analyses and time to first nSEB as an exploratory endpoint, adjusted the flagging of which record to choose if multiple assessments within the window to always choose the latest assessment, and added the lab data normalization.
5.0	31-Mar-2020	Updated SAP to update analysis visit windows to allow for greater flexibility in scheduling due to COVID-19, further clarified primary endpoint definition to ensure clarity about a lesion that was marked as an nSEB by an Investigator for subjective criteria when in fact the lesion does not meet subjective criteria, added a compliance calculation based on interruptions, and clarified that exploratory endpoints were to be summarized descriptively.
6.0	17-Jun-2020	[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

aBCCdex	Advanced Basal Cell Carcinoma Index
AE	Adverse Event
ANCOVA	Analysis of Covariance
BCC	Basal Cell Carcinoma
BCNS	Basal Cell Nevus Syndrome
BMI	Body Mass Index
CI	Confidence Interval
CMH	Cochran–Mantel–Haenszel
CSR	Clinical Study Report
DLQI	Dermatology Life Quality Index
DMC	Data Monitoring Committee
EQ-5D-5L	EuroQol (Quality of Life) using 5 levels
FCS	Fully Conditional Specification
HHI	Hedgehog Signaling Pathway Inhibitor
ITT	Intent-to-Treat
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
nSEB	New Surgically Eligible Basal Cell Carcinomas
PellePharm	PellePharm, Inc.
PP	Per-Protocol
SAE	Serious Adverse Event
SD	Standard Deviation
SEB	Surgically Eligible Basal Cell Carcinomas
TEAE	Treatment-emergent Adverse Event
VAS	Visual Analog Scale
WHO	World Health Organization

1. INTRODUCTION

This document outlines the statistical methods to be implemented during the analyses of data collected within the scope of PellePharm, Inc. (PellePharm) Phase 3 Protocol Pelle-926-301 [redacted] and associated country specific amendments. The purpose of this plan is to provide specific guidelines from which the statistical analyses will proceed. Any deviations from this plan will be documented in the clinical study report (CSR). [redacted].

1.1 Study Background

PellePharm is developing Patidegib Topical Gel to help reduce the disease burden of basal cell carcinomas (BCCs) in patients with basal cell nevus syndrome (BCNS or Gorlin Syndrome). Patidegib is a semi-synthetic small molecule which when given orally has good therapeutic efficacy in the treatment of locally advanced and metastatic BCCs but produces the same types of adverse events (AEs) as do other systemic hedgehog inhibitors (HHIs). Patidegib Topical Gel is manufactured with excipients generally accepted as safe, is stable in the developed gel formulation. Topical application of patidegib significantly reduces murine BCC tumor size in vivo and reduces GLI1 biomarker expression in vitro in human BCC tumor explants. The clinical results from the Phase 2 study in patients with Gorlin syndrome are discussed in the Sample Size Determination section.

1.2 Study Design

This is a global, multicenter, randomized, double-blind, stratified, vehicle-controlled study of the efficacy and safety of Patidegib Topical Gel, 2%, compared with Vehicle Gel, applied topically twice daily to the face of adult subjects with Gorlin Syndrome. Subjects are required to apply the study treatment for 12 months. The primary endpoint is a comparison between the 2 treatment arms of the number of new surgically eligible BCCs (nSEBs) that develop over the 12-month period. A nSEB is defined as a BCC that meets protocol-defined criteria for surgical eligibility and that was not eligible for surgery at baseline. These criteria are that a nSEB is a histologically verified BCC that (1) was surgically removed because of possible functional facial/health impairment as determined by the Investigator or (2) met objective size criteria (irrespective of whether or not it was surgically removed). These objective size criteria are that post Baseline, the BCC has a [redacted].

BCCs will be imaged and tracked consistently throughout the study in order to identify nSEBs.

Subjects will be stratified for enrollment based on the following criteria:

1. Age [redacted]
2. Gender
3. History of prior HHI therapy [redacted]

Digital imaging of the BCCs will be taken using Canfield 2-D camera. Photos will be taken to assess 5 views of the face in the same plane of an arc at a fixed distance from the face [redacted] The unmarked and marked digital facial images will be uploaded to the imaging vendor portal for further review. Full details for image capture can be found in the Imaging Review Work Instruction ([redacted]). Lesions suspicious for a BCC will be measured by the PI for their longest diameter.

[redacted].

1.3 Study Objectives

1.3.1 Primary Objectives

The primary objective of the study is to assess the number of nSEBs in the 2 arms (Patidegib Topical Gel, 2%, and Vehicle Gel) when applied twice daily to the face of subjects with Gorlin Syndrome.

1.3.2 Secondary Objectives

The secondary objective of the study is to assess the safety and tolerability of Patidegib Topical Gel, 2%, in subjects treated twice daily for 12 months.

2. STATISTICAL METHODOLOGY

2.1 General Principles

Continuous parameters will be summarized by number of non-missing observations (n), mean, standard deviation (SD), median, minimum, and maximum (with the exception of laboratory and vital sign parameters which will be summarized by mean, SD, and median). Categorical parameters will be summarized by count and percentage of the non-missing responses. All statistical analyses will be performed using SAS® Version 9.4 or later. AEs and medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 21.1 or later. Prior and concomitant medications and procedures/therapies will be coded using the World Health Organization (WHO) Drug Dictionary version Global B3 01SEP2018 or later.

2.2 Sample Size Determination

In the UK Phase 2 study of patients with Gorlin syndrome (Pelle-926-201), the mean number of nSEBs (erroneously stated in the protocol as mean number of new BCCs in the Section on Sample size determinations) at 6 months was [redacted] in the Patidegib Topical Gel, 2%, group with an SD of [redacted], and a mean of 1.4 in the Vehicle Gel group with a SD of 1.52. The effect size was 0.81. Because the current trial is longer, more conservative assumptions are made. It is assumed that at 6 months, the mean number of nSEBs in the Patidegib Topical Gel, 2%, group will be 0.6, with an SD of [redacted], while at 6 months the mean and SD in the Vehicle Gel group will be the same as in the Phase 2 study at 6 months for an effect size of 0.63. At 12 months, it is assumed that the means and SDs will be twice those of 6 months. With these assumptions, [redacted] subjects in the Patidegib Topical Gel, 2%, group and 63 in the Vehicle Gel group (for a total of [redacted] subjects) will give >90% power for a 2-sided 0.05 level statistical significance. To account for a 16% drop-out rate, approximately 150 subjects will be randomized.

The primary endpoint of number of qualifying nSEBs per subject by Month 12 may be affected by follow-up time. Based on the UK Phase 2 study (Pelle-926-201), it is projected that the power calculated above will be substantially similar unless marked differences are seen between the treatment arms in follow-up time. Sample size was calculated based on a Wilcoxon-Mann-Whitney test which is conservative if the underlying distribution is Negative Binomial.

2.3 Study Populations

The enrolled population, defined as all subjects who signed informed consent, will be used for study disposition summaries. The primary population for efficacy analyses will be the Intent-to-Treat (ITT) population, defined as all subjects who were randomized. Subjects will be assigned to treatment groups as randomized. The per-protocol (PP) population will be used for supportive efficacy analyses and is defined as all randomized subjects who completed 12 months of treatment with between [redacted] treatment compliance and without any major protocol deviations in regard to inclusion/exclusion criteria or prohibited concomitant medications. In the PP population, subjects will be assigned to treatment groups as treated (not as randomized). The Safety population will consist of all subjects who have applied treatment at least once and will be assigned to treatment groups as treated. Subjects who receive both Patidegib and Vehicle will be summarized in the Patidegib group.

Listings of all subjects who signed informed consent but did not fulfill all eligibility criteria, randomization assignments, and subject disposition will be created.

2.4 Subject Accounting and Baseline Characteristics

Subject disposition including population counts, treatment discontinuation reason, duration on study, completion of study (defined as being contacted for the 30 day follow-up), and reason for study discontinuation will be summarized for all subjects who signed an informed consent (enrolled population). A Kaplan-Meier plot of the time to treatment discontinuation will be presented in the Safety

Population. Demographic and baseline characteristics will be summarized for all subjects in the ITT Population. Demographics will be presented in a listing. A summary of baseline efficacy parameters including time since diagnosis of BCNS, number of SEBs at baseline, number of BCCs including and excluding SEBs, and history of prior HHI therapy use will be provided in the ITT Population. Medical history, prior and concomitant medications/products, prior and concomitant procedures/therapies, study completion status and reasons for discontinuation will also be summarized and listed for this population. Prior medications/products and procedures/therapies are those that started and ended prior to the date of initial dose of study drug and are not ongoing at baseline. Concomitant medications are defined as any medication ongoing at baseline (date of first dose) or which are taken at/after date of first study medication.

A summary and a listing of subjects with any blood relatives with BCNS will be provided. A table of Gorlin diagnostic criteria will be presented.

2.5 Efficacy Analyses

2.5.1 Primary Efficacy Analysis

The primary endpoint is the number of nSEBs per subject by Month 12. A nSEB is a histologically verified BCC that was surgically removed because of possible functional facial/health impairment as determined by the Investigator. Also, nSEBs are facial BCCs with a longest diameter of ≥5 mm that:

a. were not SEBs at Baseline,
b. have grown by [redacted] mm in longest diameter from Baseline, and
c. have been verified histologically.

If two or more lesions merge together and become a nSEB, then this will count as only 1 nSEB. If an investigator indicates that the lesion grew at least [redacted] mm from Baseline and the lesion is subsequently histologically confirmed, then it will count as a nSEB for the purposes of analysis.

Primary Estimand

The primary estimand – "Treatment policy" is defined as follows:

The difference in rates of nSEBs if subjects were to continue to be treated for the entire study unless forced to discontinue treatment due to an AE or lack of efficacy. The imputation of missing data in accordance with this estimand is discussed in detail in Section 3.2.

The treatment policy estimand assesses the expected effect on nSEBs in a patient population offered treatment with Patidegib Topical Gel, 2%, as compared to vehicle.

This endpoint will be analyzed using a Negative Binomial regression to compare treatment groups, with number of BCCs that are not SEBs at Baseline, prior HHI therapy, gender, age [redacted]),

and geographic region strata (North America, Europe), as covariates. PROC GENMOD of SAS will be used with distribution=NEGBIN.

A non-parametric rank analysis of covariance (ANCOVA, Stokes et al) will be provided as a sensitivity analysis with number of BCCs that are not SEBs at Baseline, prior HHI therapy, gender, age ([redacted] years old), and geographic region as covariates.

The Negative Binomial regression was selected as the primary method of analysis to account for the potential for overdispersion. Using the Negative Binomial model, the power will be somewhat higher than the one calculated using a non-parametric Wilcoxon-Mann-Whitney test. The proposed sensitivity analysis, Rank ANCOVA, was selected to corroborate the results from the Negative Binomial model. If the distribution is markedly different from Negative Binomial (and unknown to sufficiently allow a different parametric regression model), then the non-parametric ANCOVA gives type-1 error closer to the nominal level.

	ITT		**PP**	
Analysis	**Negative Binomial**	**Rank**	**Negative Binomial**	**Rank**
MI	x; s	x; s		
LOCF	x	x		
OC	x	x	x; s	x; s
Derived	Listing Only			

Note: 'x' denotes that the full population will be used, and 's' denotes that subgroup analyses will be provided in the given population and analysis type.

The primary method for dealing with missing data will be the multiple imputation (MI) method discussed in detail in Section 3.2. An analysis in the PP Population using the observed case (OC) method with the offset option specified to account for different follow-up times will be provided as a sensitivity analysis. A box plot of the mean number of nSEBs per subject over time will be presented. Additionally, a stacked histogram showing the proportion of subjects with 0, 1, 2, 3, 4, and [redacted] nSEBs will be presented over time in the ITT Population.

Last Observation Carried Forward (LOCF) and OC analyses using both the Negative Binomial and Rank ANCOVA models will be done as sensitivity analyses. Additional sensitivity analyses (conservative multiple imputations and tipping point analysis) are described in Section 3.2.

The derived number of nSEBs, determined using the actual measurements provided instead of relying on the answer to the question "Lesion has a longest diameter [redacted] mm and has grown by at least [redacted] mm in longest diameter from Baseline:", will be presented in the listings. If the post-baseline measurement is missing which makes the derivation impossible and the baseline measurement is present, then the question will be used instead to determine if the BCC qualifies as a nSEB. However, if there is no baseline measurement, then the lesion will not count as a nSEB.

Listings of derived efficacy endpoints, BCC measurements, facial BCC procedures, and digital imaging of BCCs will be created.

2.5.2 Secondary Efficacy Analysis

The secondary efficacy endpoints are:

1. The number of new BCCs that develop from Baseline to Month 12.
2. The proportion of subjects developing ≥2 facial nSEB(s) by Month 12.
3. The proportion of subjects developing ≥1 facial nSEB(s) by Month 12.
4. The number of qualifying nSEBs per subject at Month 9 (defined as the number of qualifying nSEBs per subject by Month 9).
5. The number of qualifying nSEBs per subject at Month 6 (defined as the number of qualifying nSEBs per subject by Month 6).
6. Advanced BCC Index (aBCCdex) change in lesion score from Baseline to Month 12.

The number of new BCCs that develop from Baseline to Month 12 will be analyzed using a negative binomial regression to compare treatment groups with number of BCCs at Baseline, Age ([redacted] years), gender, prior HHI therapy, and geographic region strata (North America, Europe) as covariates. Secondary endpoints that are based on number of nSEBs will be analyzed similar to the primary endpoint. The proportion of subjects with at least 2 nSEB lesions and the proportion of subjects with at least 1 nSEB lesion will be analyzed using a logistic regression to compare treatment groups with number of BCCs that are not SEBs at Baseline, Age ([redacted] years), gender, prior HHI therapy, and geographic region strata (North America, Europe) as covariates. Change in aBCCdex scores will be analyzed using an ANCOVA model with treatment group as a main effect, the corresponding baseline aBCCdex scale score, prior HHI therapy, gender, age [redacted] old), and geographic region strata as covariates. Please see Appendix A for the scoring instructions for the aBCCdex lesion scale score. A box plot of the number of new BCCS per subject over time will be presented. An analysis of the number of new BCCs by Month 12 in the PP Population using OC will be provided as a sensitivity analysis.

The secondary endpoints will be tested sequentially in the order above using a gatekeeping strategy with type-1 error of 0.05 at 12 months. If and only if the primary endpoint is statistically significant will the secondary efficacy endpoints be tested using the gate-keeping procedure.

All efficacy assessments (secondary and exploratory endpoints) will be summarized descriptively by treatment group and visit.

2.5.3 Exploratory Endpoints

The exploratory endpoints are:

1. Time to first nSEB.
2. The proportion of SEBs that undergo clinical resolution at Month 12.
3. The number of BCC surgeries from Baseline to Month 12.
4. Change in aBCCdex Worry About Future Lesions scale score from Baseline to Month 12.
5. Change in aBCCdex Mental Health scale score from Baseline to Month 12.
6. Change in aBCCdex Social/Relationships scale score from Baseline to Month 12.
7. Change in aBCCdex Life Impact scale score from Baseline to Month 12.
8. Change in Skindex-16 subscales (Symptoms, Emotions, and Functioning) from Baseline to Month 12
9. Change in Dermatology Life Quality Index (DLQI) total score from Baseline to Month 12
10. Shift in EuroQol (Quality of Life) using 5 levels (EQ-5D-5L) from Baseline to Month 12
11. Change in EQ-5D-5L Visual Analog Scale (VAS) from Baseline to Month 12

The time to first nSEB, defined as the date of first new SEB – date of randomization + 1, will be summarized using the Kaplan-Meier method. Subjects who do not develop an nSEB will be censored at the date of last tumor assessment, regardless of whether or not the subject was still on treatment. A Kaplan-Meier plot will also be provided.

The proportion of SEBs that undergo clinical resolution at Month 12 (or last assessment utilizing LOCF imputation) and the number of BCC surgeries will be summarized descriptively.

Change in Skindex-16 subscale scores and the DLQI total score will be summarized descriptively by visit Please see Appendix B for the Skindex-16, and Appendix C for the DLQI.

A separate analysis plan for the pharmacokinetic data will be provided.

A shift table will be presented for changes from baseline of the EQ-5D-5L dimension scores. Additionally, the VAS Scores will be summarized descriptively by visit.

2.5.4 Interim Analysis

An Interim analysis was initially planned per protocol when all subjects have completed 6 months of treatment (or terminated the study earlier) and if at least 126 nSEBs have been observed. However, based on advice received during 1H 2020 interactions with FDA and EMA, and given both agencies' recommendation to not conduct the interim analysis, the option for an interim analysis was cancelled and the DMC disbanded prior to the original IA time point and prior to any unblinding; only a final analysis will be performed.

2.6 Safety Analyses

Dermal Safety and Tolerability

Shift tables of the frequency of dermal safety and tolerability assessments including pain/burning, pruritus, erythema, edema, and crusting will be summarized using counts and percentages by treatment group and visits of the most severe post baseline assessment and end of treatment from baseline.

A listing of all dermal safety and tolerability assessments will be provided.

Adverse Events

Subjects will be assessed for the occurrence of new and ongoing AEs from the time of informed consent. Descriptions of AEs will include the dates of onset and resolution (if resolved) including study days, maximum severity, seriousness, action taken regarding the treatment, corrective treatment, outcome, and Investigator's assessment of causality. All AEs will be recorded and classified using terminology from the MedDRA. AEs prior to treatment application will be summarized by preferred term. All reported treatment-emergent AEs (TEAEs), defined as any AE with an onset on or after the date of first treatment application and on or prior to the date of last application plus 30 days, or worsening of a condition with an onset prior to first treatment application and on or prior to the date of last application plus 30 days, will be summarized by treatment group, the number of subjects reporting TEAEs, system organ class, preferred term, severity, and relationship to treatment. Additionally, a summary of TEAEs by preferred terms will be provided for all preferred terms with at least 1 TEAE in the Patidegib arm. When summarizing TEAEs by severity or relationship to treatment, each subject will be counted only once within a system organ class or a preferred term using the event with the greatest severity or causality, respectively, within each category. All reported treatment-emergent Serious AEs (SAEs) will be summarized by treatment group, the number of subjects reporting SAEs, system organ class and preferred term as well as by preferred term. All reported TEAEs leading to treatment discontinuation (those TEAEs with action taken with study treatment of Drug Withdrawn) will be summarized by treatment group, the number of subjects reporting treatment-related TEAEs/TEAEs leading to treatment discontinuation, system organ class and preferred term as well as by preferred term. Additionally, paralesional TEAEs and TEAEs on complete area of treatment application will be summarized by preferred term.

AEs with a missing relationship to treatment will be classified as related to treatment. AEs with missing severity will be classified as severe.

All information pertaining to AEs (separately for AEs prior to treatment application and TEAEs) noted during the study will be listed by subject and will include a verbatim description of the event as reported by the Investigator, as well as the preferred term, system organ class, start date, stop date (if stopped), seriousness, severity, action taken regarding the treatment, corrective treatment, outcome and relationship to the treatment. In addition, separate listings of subjects who prematurely discontinue from

the study due to TEAEs, subjects who reported an AE leading to death, and subjects who experienced treatment-emergent SAEs will be provided.

Safety Laboratory Tests

Safety laboratory tests were initially intended to be done at one central laboratory (Clinical Reference Laboratory; CRL). However, since CRL decided to shut down its operations, midway through the study, laboratory testing has been migrated to a second laboratory (Cerba). Since two central labs will be used during the trial, the data from CRL will be normalized to the data from Cerba data using the methodology outlined in Karvanen (2003). The lab data from CRL will be normalized to the Cerba data by applying the following location-scale family transformation for each parameter using the female range if the ranges are different for males and females:

$$normalized\ result = L_{Cerba} + (x_{CRL} - L_{CRL}) * \left(\frac{U_{Cerba} - L_{Cerba}}{U_{CRL} - L_{CRL}}\right)$$

Where L_{Cerba} and L_{CRL} are the lower limits of normal of the parameter from Cerba and CRL, respectively, x_{CRL} is the result of the parameter for the subject, and U_{Cerba} and U_{CRL} are the upper limits of normal of the parameter from Cerba and CRL, respectively.

Then, if any normalized result for a given parameter is negative, then the scale family transformation will be applied instead:

$$normalized\ result = (x_{CRL}) * \left(\frac{U_{Cerba}}{U_{CRL}}\right)$$

Changes from Baseline to last on treatment in safety laboratory values, including chemistry and hematology, will be summarized separately with descriptive statistics in SI units for each treatment group at all applicable study visits. Separate listings for each panel will be provided.

Shift tables will be presented for changes in safety laboratory values to summarize laboratory test results from baseline to last on treatment. Normal ranges established by each of the central laboratories will be used to determine the shifts. A listing of all out-of-range laboratory test results at any assessment time point will also be provided. Determination of clinical significance for all out-of-range laboratory values, derived as any AE with a system organ class of Investigations, will be made by each Investigator and included in the listing. In addition, a listing of all clinically significant laboratory test results will be provided.

Pregnancy and urinalysis results will be provided in listings. A shift table for urinalysis results (normal vs. abnormal) from baseline to last on treatment will be provided.

Physical Examinations

Changes from Baseline for last on treatment in vital sign measurements will be summarized with descriptive statistics for each treatment group. A listing of vital signs results will be provided.

3. DATA HANDLING

3.1 Baseline and Study Visits

Visits for all analyses will be as recorded on the electronic Case Report Forms. Baseline will be the latest of the non-missing values prior to receiving treatment. The following analysis windows will be used for all efficacy and safety data:

Visit	Target Day	Beginning of the Window	End of the Window
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

Note: All days are relative to the date of first treatment application.

If there is more than 1 assessment of the same data in the same window, then the latest assessment within the window will be used for analysis in the tables.

3.2 Missing Data

For the primary analysis, the MI method will be the primary method for dealing with missing data. LOCF and OC analyses will be done as sensitivity analyses. For the count secondary endpoints, MI will be used; for the dichotomous secondary endpoints and the aBCCdex endpoint, LOCF will be used.

For the MI model, 100 imputations will be generated using PROC MI of SAS. Fully conditional specification (FCS) model using the regression method will be used with the change from the last time-point in the number of SEBs at prior post-baseline visits, number of BCCs that are not SEBs at baseline, prior HHI therapy, gender, geographic region, and age [redacted]) as covariates to impute any missing data including missed visits and missing data due to dropouts. Missing data of subjects on Patidegib who discontinued due to intercurrent events of AEs, use of a prohibited concomitant medication or lack of efficacy (determined during blind data review of the reasons for study treatment stopped) will be imputed based on vehicle and imputed based on randomized treatment for remaining subjects to estimate the treatment effect if patients would continue to be treated for the entire study unless forced to discontinue treatment due to an AE or lack of efficacy or used a prohibited concomitant medication. If there is a problem with imputing based on the full model, then region will be dropped. The ROUND and MINIMUM options will be utilized to ensure imputed values are non-negative integers. The seed to be used is 20180330. The results from the MI will then be used to derive the number of nSEBs at each timepoint and then analyzed using the primary analysis model. The results of the 100 analyses will be transformed into a normal statistic and combined into a single analysis using

PROC MIANALYZE. The Wilson-Hilferty transformation (Wilson and Hilferty, 1931) will be utilized to combine the results of the 100 non-parametric Rank ANCOVA analyses.

For the primary and secondary efficacy endpoints related to number of nSEBs, a nSEB that was first reported at or prior to Month 12 will be counted in the endpoint. The lesion will be considered a nSEB at the time it was first reported clinically even if it is biopsied and confirmed later.

The following additional sensitivity analyses utilizing Negative Binomial regression will be provided for the primary endpoint to explore the space of plausible missing data assumptions:

- An MI analysis using the model for the primary imputation but by vehicle for all subjects who dropped out, irrespective of the cause, and MI by randomized treatment for the remaining subjects.
- A tipping point analysis which will impute the results at the months prior to the analysis using the same model for the primary imputation will be performed if and only if the result of the primary endpoint is significant. One hundred imputations will be generated using PROC MI of SAS. The imputation at Month 12 will be imputed using the MNAR ADJUST statement within SAS to shift the imputed values for the increase in nSEBs for subjects randomized to the Patidegib arm by different values until the p-value changes from significant to non-significant. The seed to be used is 20180330.

In order to classify a medication as prior or concomitant or classify an AE as treatment-emergent or not, missing or partial dates will be imputed. Please see Appendix D for full details of the imputation of missing/partial dates.

3.3 Pooling of Sites

Approximately [redacted] centers are planned in this study in 2 geographic regions: North America (USA and Canada), and Europe. ANCOVA analyses of primary and secondary endpoints will include geographic region, in addition to other factors, as a covariate. Sub-group analyses of primary efficacy endpoints will be done by geographic region in addition to other stratification factors age [redacted]), gender, and prior HHI therapy if at least 30 subjects are in a subgroup.

3.4 Protocol Deviations

Protocol deviations and violations will be monitored. Before unblinding the study, protocol deviations will be classified as minor, or those major deviations that may impact on the interpretation of the results. Major protocol deviations in regard to inclusion/exclusion criteria, prohibited concomitant medications, or treatment compliance not between 80% – 120% will cause the removal of subjects from the PP population. A summary of deviations by impact (major vs. minor) and COVID-19 related deviations will be summarized in the ITT Population. All deviations will be provided in a listing. Additionally, a listing of missed visits due to COVID-19 will be presented.

3.5 Compliance with Study Drug Application

At each visit when study drug is returned by the subject, returned tubes will be weighed. Compliance with study drug application will be calculated from the previous dispensation of study drug. Compliance will be calculated using the following formula:

[redacted]

Where DW represents the dispensed weights and RW represents the returned weights of each kit. If a tube is not returned, then it will be assumed that the subject used none of it, i.e. the returned weight equals the dispensed weight. An alternative compliance calculation based on the following formula:

[redacted]

will be presented.

Total dose received, duration of treatment, and compliance will be compared between the 2 treatment groups using summary statistics. Compliance will also be a factor in defining the PP population as noted in Section 2.3.

Listings of baseline study drug application, study drug accountability, compliance, and study drug interruption will be created.

Compliance will also be calculated based on the cumulative number of days that study drug was interrupted. If drug is interrupted for one day it will be assumed that two doses were missed which will be a conservative estimate of the number of doses missed.

4. CHANGES FROM THE PROTOCOL

This section highlights text in the SAP that overrides the cited text in the protocol.

The protocol (Section 8.2.1) refers to an Imaging Review Charter. In actuality this is an Imaging Review Work Instruction.

In the section on Sample Size Determination (Section 12.9) the protocol states ‘mean number of new **BCC** lesions at 6 months was 0.42’ instead of ‘mean number of new **SEB** lesions at….’

The secondary endpoints for the number of nSEBs at Months 6 and 9 were clarified to reflect that they are “by Month” 6 or 9 to bring these endpoints in line with the primary endpoint. Exploratory endpoints for the time to first nSEB, Skindex-16, DLQI, and EQ-5D-5L questionnaires were added.

The primary endpoint analysis distributional assumption was changed from Poisson to Negative Binomial based on FDA and EMA feedback, however the primary endpoint remains unchanged.

The interim analysis will no longer be performed based on EMA Scientific Advice, dated 26Mar2020, and FDA advice during a Type B Meeting held on 11May2020 recommending that the interim analysis should not be conducted; only a final analysis will be performed at an alpha level of 0.05 instead of 0.045 (both primary and secondary endpoints).

Laboratory test results will be summarized as change from baseline to last result on treatment instead of at each visit to provide a more concise, medically relevant summary.

5. REFERENCES

Chren, M.M., Lasek, R.F., Sahay, A.P, and Sands, L.P. Measurement Properties of Skindex-16: A Brief Quality-of-Life Measure for Patients with Skin Diseases. *Journal of Cutaneous Medicine and Surgery*. 2001; 105-110.

Finlay, A.Y. and Khan, G.K. 1992. Cardiff University Department of Dermatology. www.dermatology.org.uk. January 21, 2019.

Mathias, S.D., Chren, M.M., Crosby, R.D., et al. Reliability and validity of the Advanced Basal Cell Carcinoma Index (aBCCdex). *British Journal of Dermatology*. 2015;173:713-719.

Stokes, M.E., Davis, C.S., and Koch, G.G. (2012), ***Categorical Data Analysis Using the SAS System***, Third Edition, Cary, NC: SAS Institute Inc.

Wilson, E.B, and Hilferty, M.M. 1931. The distribution of chi-squared. *Proceedings of the National Academy of Sciences*. 1931; 16: 219-242.

APPENDIX A: ABCCDEX SCORING

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	

Items answered as not relevant will be scored as ‘0’ for the purposes of scoring.

APPENDIX B: SKINDEX-16 SCORING

During the past week, how often have you been bothered by:	Range	Scale (Range)
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	

A scale will be set to missing if >25% of the items in the scale were left blank. Individual missing items will not be imputed.

APPENDIX C: DLQI SCORING

The scoring of each question is as follows:

Answer	Score
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	3

The total score is calculated by summing each question's score with a score ranging from 0 to 30. The higher the score, the more impairment in quality of life.

APPENDIX D: PRESENTATION OF DATA AND PROGRAMMING SPECIFICATIONS

General

- Specialized text styles, such as bold, italics, borders, and shading will not be used in tables, figures, and data listings unless they add significant value to the table, figure, or data listing.
- Only standard keyboard characters are to be used in tables and data listings.
- Special characters, such as nonprintable control characters, printer-specific, or font-specific characters, will not be used on a table, figure, or data listing.
- Hexadecimal character representations are allowed (eg, μ, α, ß).
- All footnotes will be left justified and at the bottom of a page. Footnotes must be used sparingly and must add value to the table, figure, or data listing.
- Outputs will be numbered in accordance with Synteract's SAS Programming Standards (WI-SOP-0205-001) and will follow the listed examples: t14.3.4.x-vs-<yyyy-mm-dd>.rtf, l16.2.4.x-mh-<yyyy-mm-dd>.rtf, and f14.x.x-ttdis-<yyyy-mm-dd>.rtf.

Tables

- Formal organization of tabulations may be changed during programming, if appropriate, eg, tables for the different variables may be combined into a single table, or tables with more than 1 variable may be split into several tables.
- Means and medians will be presented to 1 more decimal place than the raw data. SDs will be presented to 2 more decimal places than the raw data. Minimums and maximums will be reported with the same number of decimal places as the raw data.
- Percentages will be presented to the tenths place.
- For frequency counts of categorical variables, categories whose counts are zero will be displayed for the sake of completeness. For example, if none of the subjects discontinue due to "lost to follow-up," this reason will be included in the table with a count of 0. Categories with zero counts will not have zero percentages displayed.
- Lower and upper CI values must be presented to 1 decimal place more than the raw/derived data (ie, to the same number of decimal places as the mean).
- Percentiles (eg, 25%, 75%) must be presented to 1 decimal place more than the raw/derived data.
- For all inferential analyses, *P* values will be rounded to 4 decimal places (or at the highest level of precision) with a leading zero (0.0001). *P* values less than 0.0001 will be presented as "<0.0001."
- The last footnotes will be
  - "Source: xxx", where xxx indicates the source **table number**(s) if applicable (in case aggregated results like mean or median are plotted) or the source listing(s) (in case individual responses are plotted) and/or source datatset(s) (eg, ADaM).
  - "PROGRAM SOURCE: ...\\xx.sas, DATA CUT OFF DATE: DDMMMYYYY, RUN DATE: DDMMYY hh:mm".
    where extract date is the datestamp of the data snapshot used.

Figures

- Legends will be used for all figures with more than 1 variable or item displayed. Treatment group sizes (n=xx) will be included, as appropriate.
- Figures will be in black and white but can be in color to add value to the clarity and readability of a figure. Lines must be wide enough to see the line after being copied.
- The last footnotes will be
  - "Source: xxx", where xxx indicates the source listing number(s) and/or source datatset(s) (eg, ADaM).
  - "PROGRAM SOURCE: ...\\xx.sas, DATA CUT OFF DATE: DDMMMYYYY, RUN DATE: DDMMYY hh:mm".
    where extract date is the datestamp of the data snapshot used.
- Line graphs over time of change from baseline results will include a horizontal dashed reference line at zero.
- For box plots, the horizontal line will represent the median, + represents the group mean, the length of the box represents the interquartile range (25th-75th percentiles), and the whiskers will represent the minimum and maximum.

Listings

- Formal organization of the listing may be changed during programming, if appropriate, eg, additional variables may be included, change in the column order, or the listing may be split into multiple parts due to space constraints.
- If not otherwise specified, all data listings will be sorted by treatment, center, subject number, visit, and date/time, as appropriate.
- All date values will be presented in a SAS date (eg, 29AUG2001) format showing actual dates entered and not imputed dates.
- All observed time values will be presented using a 24-hour clock HH:MM:SS format (eg, 01:35:45 or 11:26). Seconds will only be reported if they were measured as part of the study.
- The last footnote will be
  - "PROGRAM SOURCE: ...\\xx.sas, DATA CUT OFF DATE: DDMMMYYYY, RUN DATE: DDMMYY hh:mm".
    where extract date is the datestamp of the data snapshot used.

Missing or incomplete dates (ie, AEs (for determination of treatment-emergence only) and concomitant medications)
The most conservative approach will be systematically considered. If the AE onset date is missing/incomplete, it is assumed to have occurred during the study treatment phase (ie, considered a TEAE) except if the partial onset date or other data, such as the stop date, indicates differently. Similarly, a medication with partial start and stop dates could be considered as both a prior and concomitant treatment.

The following algorithms will be applied to missing and incomplete start and stop dates:

Start Dates

- If the day portion of the start date is missing, then the start date will be estimated to be equal to the date of first dose of study drug, provided the start month and year are the same as the first dose of study drug and the stop date is either after the first dose of study drug or completely missing. Otherwise, the missing day portion will be estimated as "01."
- If both the day and month portions of the start date are missing, then the start date will be estimated to be equal to the date of first dose of study drug, provided the start year is the same as the first dose of study drug and the stop date is either after the first dose of study drug or completely missing. Otherwise, the event will be assumed to start on the first day of the given year (eg, ??-???-2013 is estimated as 01-JAN-2013).
- If the start date is completely missing and the stop date is either after the dose of study drug or completely missing, the start date will be estimated to be the day of study drug dosing. Otherwise, the start date will be estimated to be the first day of the same year as the stop date. All other non-AE and non-concomitant medication day calculations where only partial dates are available will be handled as follows: the first day of the month will be used in the calculations if the day part of a start date is missing while January 1 will be employed if both the month and day parts of a start date are missing.

Stop Dates

- If only the day of resolution is unknown, the day will be assumed to be the last of the month (eg, ??-JAN-2013 will be treated as 31-JAN-2013).
- If both the day and month of resolution are unknown, the event will be assumed to have ceased on the last day of the year (eg, ??-???-2013 will be treated as 31-DEC-2013).
- if the stop date is completely missing or the event is continuing, the event will be assumed to be after first dose of study drug and will be imputed using the last known date on the study.

Standard Calculations

Variables requiring calculation will be derived using the following formulas:

Days – A duration expressed in days between 1 date (date1) and another later date (date2) is calculated using the formulas noted below:

duration in days = date2 – date1 + 1

Months – A duration expressed in months is calculated using the INTCK function of SAS as follows: months=intck('month','date1'd,date2'd, 'continuous').

Years – A duration expressed in years between 1 date (date1) and another later date (date2) is calculated as follows:
duration in years = intck('year,'date1'd,date2'd, 'continuous').

Age – Age is calculated as the number of years from the date of birth (DOB) to the date of informed consent (DOIC). If the month of DOIC < month of DOB or the month of DOIC=DOB and the day of DOIC < day of DOB, then the following formula is used:

age (years) = year of DOIC – year of DOB – 1.

Otherwise, the following formula is used:

age (years) = year of DOIC – year of DOB.

Height – Height entries made in inches (in) are converted to centimeters (cm) using the following formula:
height (cm) = height (in) × 2.54.

Weight – Weight entries made in pounds (lb) are converted to kilograms (kg) using the following formula:
weight (kg) = weight (lb)/2.2046.

Body Mass Index (BMI) – BMI is calculated using height (cm) and weight (kg) using the following formula:
BMI (kg/m^2) = weight (kg)/[[height (cm)/100]^2].

Temperature – Temperature entries in degrees Fahrenheit are converted to degrees centigrade using the following formula:
temp (degrees centigrade) = 5/9 × [temp (degrees Fahrenheit) – 32].

Change from baseline – Change from baseline will be calculated as:
Change = postbaseline value – baseline value.

APPENDIX E: SAS PROGRAMMING QC REQUIREMENTS

Derived datasets are independently reprogrammed by a second programmer. The separate datasets produced by the 2 programmers must match 100%. Detailed specifications for the derived datasets are documented in the study Analysis Dataset Specifications provided to the client at study conclusion.

Tables are independently reprogrammed by a second programmer for numeric results.

Listings are checked for consistency against corresponding tables, figures, and derived datasets.

Figures are checked for consistency against corresponding tables and listings, or independently reprogrammed if there are no corresponding tables or listings.

The entire set of TLFs is checked for completeness and consistency prior to its delivery to the client by the lead biostatistician and a seior level, or above, reviewer.