MEK-NF-201: A Phase 2b Trial of the MEK 1/2 Inhibitor (MEKi) PD-0325901 in Adult and Pediatric Patients with Neurofibromatosis Type 1 (NF1)-Associated Inoperable Plexiform Neurofibromas (PNs) that are Causing Significant Morbidity

MEK-NF-201 SAP, v1.0.0: 02 November 2023

NCT03962543

Integrated Quanticate

Document Approval Workflow

Approved

Document Name : **MEK-NF-201 SAP v1.0.0**

Document Type : **Customer/Study related document**

Quanticate Customer Q No. : **146**

Version : **v.1.0.0**

Quanticate Project Q No. : **03974**

DOCUMENT APPROVALS

Workflow Role	Name	Job Title	Document Role	Project Role	Date Completed
Initiator	[redacted]	Head of Statistics	Not Applicable	Not Applicable	03-Nov-2023 10:33:20 UTC
Author		Head of Statistics	Author of latest changes	N/A	03-Nov-2023 10:38:42 UTC
Approver Level 1		Senior Statistician I	QC statistician	Lead Statistician	03-Nov-2023 10:53:41 UTC
Approver Level 2		Senior Statistician II	Peer reviewer	N/A	03-Nov-2023 10:55:50 UTC

Quanticate
The Clinical Data Experts

Compound: Mirdametinib (PD-0325901)
Protocol Number: MEK-NF-201
Study Phase: 2b

A Phase 2b Trial of the MEK 1/2 Inhibitor (MEKi) PD-0325901 in Adult and Pediatric Patients with Neurofibromatosis Type 1 (NF1)-Associated Inoperable Plexiform Neurofibromas (PNs) that are Causing Significant Morbidity

Statistical Analysis Plan

Version: 1.0.0

Prepared for:
SpringWorks Therapeutics
100 Washington Blvd
Stamford, CT 06902
United States

Prepared by:
Quanticate International Ltd
9-11, Bancroft Court, Bevan House
Hitchin
Hertfordshire
SG5 1LH

QUANTICATE APPROVAL			
Name:	[redacted]		
Job Title:	Senior Statistician I		
Role in Project:	Lead Statistician		
Signature:	[redacted]	Date:	03-Nov-2023

SPRINGWORKS APPROVAL			
Name:	[redacted]		
Job Title:	Vice President, Head of Biometrics		
Role in Project:	Vice President, Head of Biometrics		
Signature:	[redacted]	Date:	03-Nov-2023
Name:	[redacted]		
Job Title:	Director, Clinical Development		
Role in Project:	Clinical Program Lead		
Signature:	[redacted]	Date:	03-Nov-2023

Table of Contents

List of Abbreviations

ADL	activities of daily living
AE	adverse event
AESI	adverse event of special interest
AR	adverse reaction
ATC	anatomical therapeutic chemical
BICR	blinded independent central review
BID	twice daily
BSA	body surface area
CI	confidence interval
CR	complete response
CRF	case report form
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DCR	disease control rate
DMC	data monitoring committee
DoR	duration of response
ECG	electrocardiogram
eCRF	electronic case report form
EOT	end of treatment
ePRO	electronic patient reported outcome
ET	early termination
HLT	high-level term
HLGT	high-level group term
HRQoL	Health-Related Quality of Life
IB	investigator brochure
ICF	informed consent form
LLT	lowest level term
LTFU	Long-Term Follow-up
MAR	missing at random
MCL	Medpace Core Labs

MedDRA	Medical Dictionary for Regulatory Activities
MNAR	missing not at random
MMRM	mixed model repeated measures
MRC	medical research council
MRI	magnetic resonance imaging
NF1	neurofibromatosis type 1
NRS-11	Numeric Rating Scale-11
ORR	objective response rate
PD	pharmacodynamic
PedsQL	Pediatric Quality of Life Inventory
PFS	progression free survival
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PII	Pain Interference Index
PK	pharmacokinetic
PN	plexiform neurofibroma
P-OMAQ	Pediatric Oral Medication Acceptability Questionnaire
P-OMAQ-C	Pediatric Oral Medication Acceptability Questionnaire – Caregiver Version
P-OMAQ-P	Pediatric Oral Medication Acceptability Questionnaire – Participant Version
PR	partial response
PRO	patient reported outcomes
PROMIS	Patient-Reported Outcome Measurement Information System
PT	preferred term
REiNS	Response Evaluation in Neurofibromatosis and Schwannomatosis
RSI	reference safety information
QoL	quality of life
SAE	serious adverse events
SAP	statistical analysis plan
SAR	serious adverse reaction
SCR	screening
SD	standard deviation

Sday	study day for individual participant
SFU	Safety Follow-Up
SoA	schedule of activities
SOC	standard of care
SUSAR	suspected unexpected serious adverse reaction
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TTP	time to progression
Who-DD	World Health Organization-Drug Dictionary

1. Introduction

This statistical analysis plan (SAP) is based on MEK-NF-201 protocol amendment 4, dated 17 May 2022. The SAP contains a complete and detailed specification of the statistical analyses.

1.1 Rationale

MEK-NF-201 is an open-label, multi-center, Phase 2b study being conducted to determine the efficacy and safety of mirdametinib (PD-0325901) in participants ≥ 2 years old with symptomatic, inoperable, neurofibromatosis type 1 (NF1)-associated plexiform neurofibromas (PNs).

A Phase 2 study with mirdametinib was previously conducted in participants with NF1-associated PNs (Weiss, 2021). This Phase 2, open-label study evaluated 19 participants (range 16 to 39 years) with symptomatic or progressing PNs. Quantitative radiographic response in a target PN after administration of mirdametinib at a dose 2 mg/m^2 twice daily (BID) (maximum dose of 4 mg BID on an intermittent (3 weeks on/1 week off) dosing schedule in 28-day cycles was assessed. Overall, mirdametinib was well tolerated; no participants discontinued treatment because of a dose-limiting toxicity (DLT). One participant developed two treatment-related Grade 3 adverse events (AEs) (back and abdominal pain) simultaneously during Cycle 1; the pain resolved upon holding the drug and did not recur at the protocol-mandated reduced dose. There were no Grade 4 or 5 AEs. Five participants (26.3%) required dose reductions while on study: for Grade 3 abdominal and/or back pain (as described above), Grade 1 rash (n = 2), Grade 2 nausea (n = 1), and Grade 2 fatigue (n = 1). This study demonstrated activity of mirdametinib in participants with a PN with eight participants (42.1%; 95% confidence interval [CI]: 20%, 67%) responding to mirdametinib as defined by a ≥ 20% reduction of the target PN after 1 year of therapy.

Analysis of pharmacokinetic (PK) and pharmacodynamic (PD) data from samples collected during Study MEK-NF-201 is described in a separate PK/PD Analysis Plan which documents the integrated PK/PD analyses for mirdametinib.

2. Study Overview

2.1 Study Objectives

Table 1: Study Objectives and Endpoints

Objectives	Endpoints
Primary	
To evaluate the confirmed complete and partial response (CR + PR) rate of mirdametinib using volumetric magnetic resonance imaging (MRI) analysis, as assessed by a blinded independent central review (BICR), in participants with an inoperable NF1 associated PN that is causing significant morbidity.	Objective response rate at the end of the Treatment Phase (defined as PN decrease ≥ 20% compared to baseline in consecutive scans 2-6 months apart) using centrally read MRI volumetric analysis.

Secondary	
To evaluate the safety and tolerability of mirdametinib in participants with NF1 and symptomatic inoperable PN as measured by the incidence of adverse events (AEs)	Safety endpoints will include incidence of treatment-emergent adverse events, changes in laboratory parameters, vital signs, physical examination findings, and electrocardiograms (ECGs); Incidence of treatment-emergent adverse events will also be assessed by dose formulation. Tolerability will be assessed according to toxicities graded by National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) Version (v) 5.0;
To determine the duration of response (DoR)	Duration of response for participants whose best overall confirmed response is complete response (CR) or partial response (PR) using centrally read MRI volumetric analysis
To evaluate the effect of mirdametinib on quality of life as measured by patient-reported outcomes (PROs)	Change from baseline in quality of life (QoL) assessed by the age-specific Pediatric Quality of Life Inventory The number of participants with a meaningful score change as derived from the anchor analysis.
To evaluate the effect of mirdametinib on pain as measured by PROs	Change from baseline in pain assessed using the Numeric Rating Scale-11 and Pain Interference Index (PII) The number of participants with a meaningful score change as derived from the anchor analysis.
Exploratory	
To evaluate the effect of mirdametinib on physical functioning as measured by PROs	Change from baseline in physical function status (Patient-Reported Outcomes Measurement Information System [PROMIS] – Physical Function/Mobility/Upper Extremity Scale) in the Treatment Phase
To evaluate baseline functional impairments secondary to PN, and the effect of mirdametinib on functional outcomes depending on PN location	Change in localized strength (dynamometer and Medical Research Council Muscle Scale), range of motion, and endurance (6- minute walk test) of PN associated functional impairment in the Treatment Phase
To evaluate time to progression (TTP)	Time to progressive disease using centrally read MRI volumetric analysis
To evaluate progression free survival (PFS)	Time to progressive disease or death using centrally read MRI volumetric analysis

To assess TTP and PFS of mirdametinib versus historical controls	TTP and PFS using MRI volumetric analysis compared against historical control data
To evaluate the effect of mirdametinib on PN associated disfigurement	Change from baseline in PN-associated disfigurement using standardized photography
To correlate MRI responses with pharmacodynamic (PD) biomarkers in PN biopsies obtained pre- and post-treatment with mirdametinib	Change in levels of pERK and other applicable biomarkers in tumor biopsies at baseline and Cycle 24 / Early Termination (ET) as detailed in a separate analysis plan
To assess self-reported overall severity of symptoms and status	PGIC and PGIS scores throughout the Treatment Phase
To evaluate the acceptability of the dispersible tablet formulation of mirdametinib	Pediatric Oral Medicines Acceptability Questionnaire scores recorded as self-report and by caregiver.
To evaluate volumetric change over time in the target PN	Change from baseline in target tumor volume by centrally read MRI volumetric analysis Best volumetric percentage change from baseline as well as the time to the best volumetric percentage change from baseline
To evaluate the disease control rate of mirdametinib using volumetric MRI analysis	Disease control rate using centrally read MRI volumetric analysis
To evaluate time to response	Time to response for participants whose best overall confirmed response is complete response (CR) or partial response (PR) using centrally read MRI volumetric analysis

2.2 Study Design

MEK-NF-201 is a multi-center, open-label, single-arm, longitudinal, uncontrolled, Phase 2b study to evaluate the efficacy, safety, and tolerability of mirdametinib in participants ≥ 2 years of age with an inoperable NF1-associated PN that is causing significant morbidity.

This study will consist of two phases: the Treatment Phase and the optional Long-Term Follow-Up (LTFU) Phase. Approximately 100 participants will be enrolled and stratified by age (approximately 50 participants 2 to 17 years of age and 50 participants ≥ 18 years of age). The first five participants enrolled in the pediatric cohort must be 9 to 17 years of age at the time of signing informed consent/assent. After the first 5 participants have enrolled in the pediatric cohort, enrollment into this cohort will be paused, and only after Data Monitoring Committee (DMC) evaluation of at least 2 cycles of cumulative safety data for these 5 participants will the pediatric cohort be re-opened and potentially expanded down to 2 years of age. Note the adult cohort will not be affected by this special procedure; participants ≥ 18 years of age will be enrolled continuously from the outset.

Participants will be screened for up to 45 days prior to the first dose of study treatment (mirdametinib) and eligibility will be based on the inclusion and exclusion criteria listed in Sections 5.1 and 5.2 of the protocol, respectively. Study treatment will then be administered orally at a dose of 2 mg/m^2 BID with a maximum dose of 4 mg BID. Dosing in the Treatment Phase will be on a 28-day cycle with a 3 week on/1 week off schedule for up to 24 Cycles. Eligible participants will take their first dose of study treatment following all pre-dose assessments on Cycle 1 Day 1. All participants will remain in the Treatment Phase of the study until radiographic disease progression (confirmed by central review), they discontinue study treatment for any other reason, the study is stopped by the sponsor for any reason, or the participant has completed the study through Cycle 24.

Participants who complete Cycle 24 of the Treatment Phase and did not meet withdrawal criteria detailed in the protocol are eligible to participate in the LTFU Phase. Eligible participants will continue taking the last dose assigned in the Treatment Phase (approximately 2 mg/m^2 BID [maximum dose of 4 mg BID]) of mirdametinib, on an intermittent (3 weeks on/1 week off) dosing schedule in 28-day cycles. Participants will remain in the LTFU Phase until a withdrawal criterion is met as described in Section 7 of the protocol or mirdametinib is commercially available.

A participant is considered to have completed the Treatment Phase if he/she has completed the Cycle 24 visit (End of Treatment Phase) or has entered the LTFU Phase.

The treatment group labels that will be used in all outputs are presented in Table 2.

Table 2: Study Treatments

Actual Treatment	Treatment Label
PD-0325901 (approximately 2 mg/m^2 BID)	Mirdametinib

Table 3 and Table 4 present the visit labels that will be used in all outputs for the Treatment and LTFU Phases, respectively.

Table 3: Study Visits – Treatment Phase

Visit	Visit Label
Screening	Screening
Cycle 1 – Week 1 (Baseline, study visit day 1)	C1D1
Cycle 1 - Week 3 (Cycle Day 15)	C1D15
Cycle 2 – Week 7 (Cycle Day 15)	C2D15
Cycle 3 – Week 11 (Cycle Day 15)	C3D15
Cycle 4 – Week 15 (Cycle Day 15)	C4D15
Cycle 5 – Week 19 (Cycle Day 15)	C5D15
Cycle 6 – Week 23 (Cycle Day 15)	C6D15
Cycle 7 – Week 27 (Cycle Day 15)	C7D15
Cycle 9 – Week 35 (Cycle Day 15)	C9D15
Cycle 11 – Week 43 (Cycle Day 15)	C11D15
Cycle 13 – Week 51 (Cycle Day 15)	C13D15
Cycle 15 – Week 59 (Cycle Day 15)	C15D15
Cycle 17 – Week 67 (Cycle Day 15)	C17D15
Cycle 19 – Week 75 (Cycle Day 15)	C19D15
Cycle 21 – Week 83 (Cycle Day 15)	C21D15
Cycle 23 – Week 91 (Cycle Day 15)	C23D15
Cycle 24 (End of Treatment Phase) – Week 96 (Cycle Day 21)	C24D21
Early Termination	ET
Safety Follow-up – 30 days after last dose	SFU

Table 4: Study Visits - Long-Term Follow-up Phase

Visit	Visit Label
Cycle 28 (Cycle Day 15)	C28D15
Cycle 32 (Cycle Day 15)	C32D15
Cycle 36 (Cycle Day 15)	C36D15
Cycle XX (Cycle Day 15)	CXXD15[1]
End of Treatment	EOT
Safety Follow-up – 30 days after last dose	LTSFU

[1]The Long-Term Follow-up phase lasts an indefinite amount of time; visits will occur on day 15 of every 4th cycle with the cycle number inserted into this visit label format.

2.3 Sample Size Determination

Approximately 150 participants will be screened in this study to assure approximately 100 participants are assigned to the open-label treatment with mirdametinib. These 100 participants will be stratified by age, with 50 participants ages 2 to 17 years for the pediatric cohort and 50 participants ≥ 18 years of age for the adult cohort. The sample size was calculated separately for each stratum to allow for different hypothesized magnitudes of response depending on the age group of the participants.

An observed objective response rate (ORR) of at least 42% (at least 21 objective responses) in 50 treated adult participants will result in a lower bound of the 95% confidence interval (CI) that excludes 28%. This will be considered evidence of clinically significant activity of mirdametinib. This is equivalent to testing a null ORR rate of 23% compared to a target rate of 43% with 80% power and 2-sided alpha of 0.05. Similarly, an observed ORR of 64% (at least 32 objective responses) in 50 treated pediatric participants will result in a lower bound of the 95% CI that excludes 49%.

For the analysis of safety, there is a 95% chance of observing at least 1 adverse event (AE) that occurs at an underlying rate of 3.0% in a sample size of 100 participants.

The assumptions for the response rate in the adult population used for the basis of sample size estimation are based on the study conducted by Weiss et al (Weiss 2017). As outlined in Section 2.1, the preliminary results from this study describe an ORR (as defined by a 20% reduction by of the target lesion from baseline) of 42% (95% CI: 20% to 67%) in response to one year of mirdametinib therapy. The results of tumor response post-therapy at > 1 year have not yet been released. This finding is consistent with the preliminary data of cabozantinib in patients with a morbid or progressing PN. In a Phase 2 study of cabozantinib in adults > 16 years of age with an evaluable PN, 8 of 19 patients achieved an objective response (42%) (Shih, 2018).

Sample size assumptions for the pediatric population are based on the interim results from the SPRINT study conducted by Gross et al (Gross, 2018). The SPRINT study enrolled 50 patients aged 3-18 years with inoperable morbid PNs. To date, this study has demonstrated a confirmed ORR of 68%. These findings are consistent with the preliminary data of another PN study with trametinib. Twenty-six patients aged 1 to 17 years of age (median 5.5 years) with a measurable PN were treated with trametinib for a median of 61 weeks (range 3 to 124 weeks). In this study, 12 patients (46%) achieved an objective response (Moertel, 2018). Section 5.2 discusses the hypotheses that will be evaluated for the primary efficacy analysis in each age cohort.

2.4 Schedule of Activities

The full schedule of activities (SoA) can be found in Sections 1.3 (Treatment Phase) and 1.4 (LTFU Phase) of the protocol.

3. Analysis Populations

The analysis populations of interest are:

- Full Analysis Set (FAS)
- Per-Protocol Set (PPS)
- Safety Analysis Set

In addition to these, the Enrolled Set is defined as all participants who signed the ICF and passed Screening.

Each of the analysis populations, as well as their associated criteria for inclusion, are described in the following sections.

3.1 Full Analysis Set

The FAS is defined as participants who received at least one dose of mirdametinib.

The efficacy endpoints will be based on the FAS.

3.2 Per-Protocol Set

The PPS is defined as the set of all participants in the FAS who have completed at least one post-baseline disease evaluation and who do not have any major CSR reportable deviations, including but not limited to efficacy assessments or compliance.

The PPS will be used in a supportive analysis of confirmed objective response rate.

3.3 Safety Analysis Set

The Safety Analysis Set is defined as participants who received at least one dose of mirdametinib. This is an equivalent definition to the FAS.

The Safety Analysis Set will be used for all safety analyses. Since it is equivalent to the FAS, it will not be presented in any baseline summaries such as demographics; instead, the FAS summary will be footnoted to highlight that it is equivalent to the Safety Analysis Set.

4. Study Measures

This section describes the measures that are collected and/or derived during the study at the time points specified in the SoA (Sections 1.3-1.4 of the protocol). This includes efficacy, safety, tolerability, and participant characteristics data.

4.1 Efficacy Measures

The efficacy measures summarized for this study are confirmed ORR, duration of response, quality of life and pain evaluations by PRO, physical function assessments, time to progression, progression free survival, and formulation acceptability by PRO. The endpoints described in this section will be analyzed according to the analysis methods described in Section 6.11.

Baseline efficacy measure values are defined as the most recent assessment taken on or before treatment start date, unless otherwise specified.

4.1.1 Primary Efficacy Measure

The primary efficacy endpoint of interest is the confirmed ORR by the end of Treatment Phase (i.e., Cycle 24). The confirmed ORR is defined as the proportion of participants who have a confirmed ≥ 20% reduction in target tumor volume as compared to baseline as assessed by a BICR, and the response needs to be confirmed by BICR in a consecutive tumor assessment within 2 – 6 months.

To determine the level of target tumor response (complete response [CR], partial response [PR], stable disease [SD], or progressive disease [PD]) which are described below, post-baseline measurement scans are compared to the target PN volume in the pre-treatment MRI scan using volumetric data analysis:

- CR: A complete resolution of the target PN
- PR: A ≥ 20% reduction in the volume of the target PN lesion compared to baseline
- SD: Between a < 20% increase and < 20% decrease in the volume of the target PN lesion compared to baseline
- PD: A ≥ 20% increase in the volume of the target PN compared to baseline. (Radiographic disease progression)

The best overall confirmed response (CR, PR, SD, PD, NE - where SD/PD/NE only require a singular response to be deemed confirmed at that level) will be summarized. For participants that remain on treatment after PD, only the best response up to PD will be considered. For participants who only have one response measurement, and this is a CR or PR (and therefore unconfirmed), their best overall response status will be set to "Missing".

All tumor volume measurements will be independently calculated by two BICR readers. In the case that the two measurements represent different tumor response statuses, as defined above, an independent adjudicator will make the final decision on which response category is recorded at that visit. The primary endpoint analysis will be performed using the adjudicator-selected reader for those visits where the two BICR readers did not agree.

Tumor volume (mL) will be assessed using MRI scans that will be taken at screening, then every 4 cycles at Cycles 5, 9, 13, 17, 21, 24, and ET for the Treatment Phase, and Cycle 28 and every 4 cycles after that, up to and including end of treatment (EOT), for the LTFU Phase, if applicable. Only tumor assessments during the Treatment Phase and prior to the first treatment start date in the long-term follow up will be considered for the determination of confirmed response. Unscheduled visits will be mapped to a visit window based on the criteria in Section 5.6.

The FAS will be used to calculate the confirmed ORR. Participants who do not have a post-baseline MRI assessment or do not achieve a confirmed response will be categorized as non-responders, and their best overall response will be deemed as "Missing".

4.1.2 Secondary Efficacy Measures

4.1.2.1 Duration of Response

A secondary efficacy objective is to determine the DoR for participants who attain a best confirmed response of PR or CR using BICR MRI volumetric analysis. Confirmed response is defined in Section 4.1.1.

DoR is defined as the time in months between the first instance of response that is subsequently confirmed, until the date of radiographic disease progression or death, whichever occurs first. The DoR will be summarized in months as follows: (date of event (i.e. radiographic disease progression or death) or censoring – date of first response + 1) / 365.25 *12.

For participants who enter the LTFU Phase, all MRI assessments in both the Treatment Phase and LTFU Phase will be used to determine DoR. Those who achieved a DoR of at least 12 months are considered to have experienced a durable response.

Participants without radiographic disease progression or death while on study will have their results censored to their most recent adequate (i.e. evaluable) tumor assessment date.

If participants receive any surgery impacting the target PN or start subsequent systemic therapy for the treatment of PNs prior to radiographic progression, then the participant is censored at the last adequate tumor assessment prior to target PN surgery or subsequent PN treatment. The rules for censoring and events for the duration of response are summarized in Table 5 below.

Table 5: Censoring and Event rules for DoR

Situation	Date of Censoring or Event	Outcome
No adequate tumor assessment post confirmed response	Date of tumor assessment that resulted in response confirmation	Censored
No radiographic disease progression or death post confirmed response	Date of last adequate tumor assessment	Censored
New surgery or subsequent anticancer PN therapy started prior to radiographic disease progression post confirmed response	Date of last adequate tumor assessment before the start of new therapy or surgery	Censored
Two or more consecutive missing or non-adequate tumor assessments post confirmed response	Date of last adequate tumor assessment before the first missing or non-adequate assessment	Censored
Radiographic disease progression that has been verified by BICR post confirmed response	Date of the earliest tumor assessment that results in a finding of progression	Event
Death before radiographic disease progression post confirmed response	Date of death	Event

4.1.2.1.1 Sensitivity Analyses of Duration of Response

According to the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) criteria (Dombi, 2013), progressive disease (PD) is defined as an "increase in the volume of the target lesion by 20% or more compared to baseline or the time of best response after documenting a PR." In order to assess the durability of the response using a similar criterion, the following sensitivity analyses will be performed using the PN volume from the two central readers, separately by each reader:

- DoR applying the REiNS criteria using all scans (I.e. considering scans from both the Treatment Phase and the LTFU phase); defined as the time from the first response which was subsequently confirmed to the derived disease progression scan using BICR PN volume data or death. That is, if a participant had an early confirmed response, a subsequent increase of at least 20% from the confirmed response would not necessarily be considered PD at that timepoint; for example, if the participant had MRI scans showing an even larger shrinkage in target PN volume at later scans, then the largest target PN shrinkage at the later timepoint would be considered the best response when determining PD according to the REiNS criteria.
- DoR applying REiNS using real-time review; defined as the time from first response which was subsequently confirmed to the derived disease progression scan using BICR PN volume data or death considering only the scans up to and including each visit, to assess PD in a real-time review. That is, if a participant had an early confirmed response, a subsequent increase of at least 20% from the confirmed response would be considered PD at that timepoint, regardless of if the participant had MRI scans showing an even larger shrinkage in target PN volume at later scans.

The censoring rules for the sensitivity analyses will be the same as for the main DoR analysis and will follow Table 5.

4.1.2.2 Patient-Reported Outcomes (PROs)

Additional secondary efficacy objectives focus on the impact of mirdametinib on participant quality of life (QoL) and pain using the following PROs:

- Pediatric Quality of Life Inventory (PedsQL)
- Numeric Rating Scale-11 (NRS-11)
- Pain Interference Index (PII).

4.1.2.2.1 Pediatric Quality of Life Inventory (PedsQL)

The PedsQL 4.0 Generic Core Scales are multidimensional self-report and parent proxy-report scales to assess health-related quality of life (HRQoL) in children, adolescents, young adults, and adults with a recall period of 7 days. It is a brief standardized HRQoL scale with good reliability and validity. It consists of a 23-item core measure of global QoL and can be completed in approximately 5 minutes. The questions are split into four dimensions, each of which can have an overall scale score calculated.

The secondary endpoint Change from baseline in quality of life (QoL) assessed by the age specific Pediatric Quality of Life Inventory will be evaluated using the following PedsQL scores:

- Physical functioning,
- Emotional functioning,

- Social functioning,
- School/work functioning,
- Total score.

PedsQL items are answered on a Likert scale with responses ranging from 0 to 4 (where 0 means it is never a problem and 4 means it is almost always a problem). These items are then reverse scored and linearly transformed to a 0-100 scale as follows: 0=100, 1=75, 2=50, 3=25, 4=0. Overall scale scores are calculated as the mean of the scores for the questions in said scale (or of all questions for the total score).

Participants complete PedsQL at the time points specified in the SoA (Sections 1.3-1.4 of the protocol).

4.1.2.2.2 Numerical Rating Scale–11 (NRS-11)

The NRS-11 is a self-report segmented 11-point numeric scale that assesses pain severity over the past 24 hours. The NRS-11 is administered daily over 7 consecutive days, inclusive of the study visit day, during the Treatment Phase as described in the SoA (Sections 1.3-1.4 of the protocol). For the ET visit, the assessment is only conducted once at the site and not for the preceding 6 days. Participants ≥ 8 years of age are administered the NRS-11 as a self-report.

Participants are asked to select the one number from 0 to 10 that best describes their worst pain over the past 24 hours for target tumor pain with 0 representing "no pain" and 10 representing "worst pain you can imagine". The NRS-11 score presented each visit will be taken as the average of the responses over the 7 consecutive days leading up to and including the visit day, with no transformation applied. The scale is presented in Appendix 1.3.

4.1.2.2.3 Pain Interference Index (PII)

The PII assesses self-reported consequences of pain on relevant aspects of one's life over the past 24 hours. The PII will be administered daily over 7 consecutive days, inclusive of the study visit day, as described in the SoA (Sections 1.3-1.4 of the protocol). For the ET visit, the assessment is only conducted once at the site and not for the preceding 6 days. Pain interference refers to the degree to which pain interferes with or limits an individual's daily activities. The PII also incorporates items probing sleep and mood. The PII is universal rather than disease specific. All participants ≥ 6 years of age complete the PII as a self-report during the Treatment Phase. Parents or legal guardians of children from 6 to 17 years of age will complete the parent proxy PII in parallel.

The PII consists of 6 questions, each asking the responder to select one number from 0 to 6 that best describes how their/their proxy's pain has impacted various items over the past 24 hours with 0 representing "Not at all" and 6 representing "Completely". The mean of the completed items is taken as the PII score for a single assessment. The PII score presented each visit will be taken as the average of the PII scores over the 7 consecutive days up to and including visit day, with no transformation applied. The full questionnaire is presented in Appendix 1.4.

4.1.3 Exploratory Efficacy Measures

4.1.3.1 Patient-Reported Outcomes Measurement Information System (PROMIS) – Physical Function/Mobility/Upper Extremity Scale

The PROMIS Physical Function instruments measure self-reported capability rather than actual performance of physical activities. This includes the functioning of one's upper extremities (dexterity), lower extremities (walking or mobility), and central regions (neck, back), as well as instrumental activities of daily living, such as running errands. A single Physical Function capability score is obtained from each short form. This PRO is only completed during the Treatment Phase.

All participants in the adult cohort with a target PN causing significant physical functioning impairment at baseline should complete the PROMIS Short Form v2.0 – Physical Function 8b. For all participants in the pediatric cohort, the questionnaire will be specific to the location of the target PN, either PROMIS Pediatric Short Form v2.0 – Upper Extremity 8a or PROMIS Pediatric Short Form v2.0 – Mobility 8a. Pediatric participants of age ≥ 5 will also have a parent proxy version of the relevant form completed by an adult caregiver. Further details of the assessment of PROMIS can be found in the Section 8.1.2.3 of the protocol.

All PROMIS items are answered on a Likert scale with responses ranging from 1 to 5, with higher scores representing higher reported physical capability. For each PROMIS Scale, a total raw score will be calculated as the sum of the individual item responses at a given visit (total raw scale range 8-40). The total raw scores for the measure will be converted to T-scores using the applicable score conversion table in the user manual and scoring instructions (PROMIS, 2023). All items must be answered to produce a valid score using the scoring tables. A higher T-score indicates better reported physical capability.

Participants complete the applicable PROMIS Physical Function instrument at the time points specified in the SoA (Section 1.3 of the protocol).

4.1.3.2 Physical Functioning Assessments

4.1.3.2.1 Range of Motion

All participants with a target PN that causes motor dysfunction or weakness at baseline will undergo a focused evaluation of range of motion of the affected joints during the Treatment Phase, in accordance with the SoA (Sections 1.3-1.4 of the protocol). For each participant, range of motion results will be summed and this will be taken as an overall score for each visit.

4.1.3.2.2 Strength Assessments

All participants with a target PN that causes motor dysfunction or weakness at baseline will undergo a focused evaluation of strength of the affected muscle groups in accordance with the SoA (Sections 1.3-1.4 of the protocol) and will be included for this analysis. Strength will be assessed utilizing a Sponsor provided dynamometer and Medical Research Council (MRC) Muscle Scale. The MRC grading assessment should be conducted prior to the dynamometer assessment. Dynamometer assessment and MRC grading (0-5/5) must be conducted in accordance with the study reference manual. For each participant, the sum of the scores for the muscle groups tested will be taken as an overall score for each visit.

4.1.3.2.3 6-Minute Walk Test

Participants with a target PN that causes airway or lower extremity motor dysfunction at baseline will undergo a 6-minute walk test in accordance with the SoA (Sections 1.3-1.4 of the protocol) that will be included for this analysis. Additionally, pulse rate, systolic blood pressure and diastolic blood pressure will be collected following a 5-minute rest after the 6-minute walk test is completed.

4.1.3.3 Patient Global Impression of Severity

The patient global impression of severity (PGIS) is a single item scale that evaluates the participant's perception of the overall severity of their NF1-related symptoms over the past week on a 4-point scale ranging from "none" to "severe." The PGIS has a 7-day recall period and is completed by participants ≥ 6 years of age. The scale is presented in Appendix 1.1.

The PGIS is assessed at the time points specified in the SoA (Section 1.3 of the protocol).

4.1.3.4 Patient Global Impression of Change

The patient global impression of change (PGIC) is a single item scale that evaluates the participant's perception of the overall change in their overall status since the start of the study treatment on a 7-point scale ranging from "very much better" to "very much worse." The PGIC will be completed by participants ≥ 6 years of age during the Treatment Phase. The scale is presented in Appendix 1.2.

The PGIC is assessed at the time points specified in the SoA (Section 1.3 of the protocol).

4.1.3.5 Target Tumor Biopsy

See Section 4.3.8.

4.1.3.6 Time to Progression

Time to progression (TTP) is defined as the time in months between the start of treatment to the first date of radiographic disease progression as measured by MRI. TTP will be summarized in months as follows: (date of event (i.e., radiographic disease progression) or censoring – treatment start date + 1) / 365.25 *12. Participants without radiographic disease progression will have their results censored at the last adequate tumor assessment date. Participants who lack any tumor response data post baseline will have their event time censored at treatment start date.

If participants receive any surgery impacting the target PN or start subsequent systemic therapy for the treatment of PNs prior to radiographic progression, then the participant's event time will be censored at the last adequate tumor assessment date prior to surgery or subsequent PN treatment. The rules for censoring and events for TTP are summarized in Table 6.

TTP will be analyzed for the FAS and also a subset of participants in the FAS who have progressive disease at baseline.

Table 6: Censoring and Event Rules for TTP

Situation	Date of Censoring or Event	Outcome
No adequate tumor assessment post baseline	Treatment start date	Censored

Situation	Date of Censoring or Event	Outcome
No radiographic disease progression	Date of last adequate tumor assessment	Censored
Two or more consecutive missing or non-adequate tumor assessments	Date of last adequate tumor assessment before the first missing or non-adequate assessment	Censored
Death before radiographic progression	Date of last adequate tumor assessment	Censored
New surgery or subsequent anticancer PN therapy started prior to radiographic disease progression	Date of last adequate tumor assessment before the start of new therapy or surgery	Censored
Radiographic disease progression that has been verified by BICR	Date of the earliest tumor assessment that results in a finding of progression	Event

4.1.3.6.1 Sensitivity Analysis of Time to Progression

The following sensitivity analyses using the REiNS criteria (Dombi, 2013) will be performed using the target PN volume from the two central readers, separately:

- TTP using all scans; defined as the time from start of treatment to the derived disease progression using BICR PN volume data considering all MRI scans.
- TTP using real-time review; defined as the time from start of treatment to the derived disease progression using BICR PN volume data considering only the scans up to and including each visit, in order to assess PD in a real-time review.

The censoring rules for the sensitivity analyses will be the same as for the main TTP analysis and will follow Table 6.

4.1.3.7 Progression Free Survival

Progression free survival (PFS) is defined as the time from treatment start date to the first occurrence of radiographic disease progression or death, whichever occurs first. PFS will be summarized in months as follows: (date of event (i.e. radiographic progression or death) or censoring – treatment start date + 1) / 365.25 *12. Any participant who does not experience radiographic disease progression or death while on study will have their results censored to their last adequate tumor assessment date. Participants who lack any tumor response data post-baseline will have their event time censored at treatment start date.

If participants receive any surgery impacting the target PN or start subsequent systemic therapy for the treatment of PNs prior to radiographic progression, then the participant is censored at the last adequate tumor assessment date prior to target PN surgery or subsequent systemic PN treatment. The rules for censoring and events for PFS are summarized in Table 7.

PFS will be analyzed for the FAS and also a subset of participants in the FAS who have progressive disease at baseline.

Table 7: Censoring and Event rules for PFS

Situation	Date of Censoring or Event	Outcome
No adequate tumor assessment post baseline	Treatment start date	Censored
No radiographic disease progression or death	Date of last adequate tumor assessment	Censored
New surgery or subsequent anticancer PN therapy started prior to radiographic disease progression	Date of last adequate tumor assessment before the start of new therapy or surgery	Censored
Two or more consecutive missing or non-adequate tumor assessments	Date of last adequate tumor assessment before the first missing or non-adequate assessment	Censored
Radiographic disease progression that has been verified by BICR	Date of the earliest tumor assessment that results in a finding of progression	Event
Death before radiographic disease progression	Date of death	Event

4.1.3.7.1 Sensitivity Analysis of Progression Free Survival

The following two types of sensitivity analyses using the REiNS criteria (Dombi, 2013) will be performed using the target PN volume from the two central readers, separately.

- PFS applying REiNS using all scans; defined as the time from start of treatment to the derived disease progression using BICR PN volume data considering all the MRI scans.
- PFS applying REiNS using real-time review; defined as the time from start of treatment to the derived disease progression using BICR PN volume data considering only the scans up to and including each visit, in order to assess PD in a real-time review.

The censoring rules for the sensitivity analyses will be the same as for the main PFS analysis and will follow Table 7.

4.1.3.8 Historical Controls

.

Full details on the analyses performed using historical control data (NCI 01-C-0222 and NCI 08-C-0079) will be summarized separately in an addendum to this SAP.

4.1.3.9 Pediatric Oral Medicine Acceptability Questionnaire (P-OMAQ)

The P-OMAQ is a clinical outcome assessment conducted as a self-report by pediatric participants (P-OMAQ-P), with an associated adult caregiver observer report (P-OMAQ-C) assessing the acceptability of the dispersible tablet (pediatric) dosage formulation. The P-OMAQ-P is a 12-item PRO for respondents of ≥ 8 years of age. The P-OMAQ-C is a 19-item PRO for adult caregivers (aged ≥18 years) overseeing study drug dosing of pediatric participants aged 2 to 17 years.

Each questionnaire is provided with a "past 7 days" recall with all items using a 5-point Numerical Rate Scale with higher item-scores reflecting greater oral treatment acceptability. The overall P–OMAQ-P and P-OMAQ-C scores for a participant in the study will be taken as the mean of all answered questions on the respective questionnaire.

The P-OMAQ will only be assessed once per participant in the study taking the tablet. This assessment can take place at any of the time points specified in the SoA (Sections 1.3-1.4 of the protocol).

4.1.3.10 Tumor Volume

Tumor volume (mL) will be measured independently by two blinded independent central reviewers, as described in Section 4.1.1. Unscheduled visits may be mapped to a visit window based on the criteria in Section 5.6.

The volume measured at the screening (SCR Volume) visit will be taken as the baseline for calculation of confirmed objective response rate, with percentage change from baseline at Cycle X (CX) calculated as:

((CX Volume – SCR Volume)/SCR Volume)*100.

The above calculation will be done for the average of the volumes recorded by each reviewer, unless there was a response adjudication in which the adjudicated reviewer's volume will be selected. In addition, this calculation will be done for each reviewer separately.

4.1.3.11 Disease Control Rate (DCR)

Disease control rate (DCR) is defined as having a confirmed BOR of SD or better post-baseline during the Treatment Phase. The number and percentage of participants achieving disease control along with the 2-sided 95% Clopper-Pearson exact CI for the DCR will be presented by age cohort. This will consider all subjects who have a baseline scan.

All tumor volume measurements will be independently calculated by two blinded reviewers. In the case that the two measurements represent different tumor response statuses, as defined in Section 4.1.1, an independent adjudicator will make the final decision on which response category is recorded at that visit.

As a sensitivity analysis, the DCR will also be summarized by each reader separately.

4.2 Safety Measures

The safety measures in this study include assessing treatment-emergent adverse events (TEAEs) (additionally by dose formulation), electrocardiograms (ECGs), Echocardiograms, clinical laboratory evaluations, vital signs assessments, and ophthalmic measurements. The endpoints described in this section will be analyzed according to the analysis methods described in Section

6.12.1. For all applicable safety measures, the baseline value will be defined as the most recent assessment taken on or before the treatment start date, unless otherwise specified.

4.2.1 Adverse Events

An adverse event (AE) is any untoward medical occurrence in a patient or clinical study participant, temporally associated with the use of study treatment, whether or not considered related to the study treatment.

An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of study treatment.

AE data is collected from the time of informed consent/assent to 30 days after the last dose of study treatment at the timepoints specified in the SoA (Sections 1.3-1.4 of the protocol). Missing AE data will be handled according to the rules specified in Section 5.5.2.1.

4.2.1.1 Adverse Event Definitions

Certain AEs are specified as adverse events of special interest (AESI). AESIs for this study were marked by the investigators and will be reported as such. AESIs include but are not limited to the following cases in Table 8.

Table 8: Adverse Events of Special Interest

Gastrointestinal (reported as AESI if Grade ≥3)		
Diarrhea	Nausea	Vomiting
Eye Disorders (reported as AESI if Grade ≥2)		
Retinal Vein Occlusion (RVO)	Uveitis	Optic neuropathy
Retinopathy	Retinal detachment	
Cardiac Disorders (reported as AESI if Grade ≥2)		
Ejection fraction decreased	Cardiac failure	Left ventricular dysfunction
Neurological Disorders (reported as AESI if Grade ≥2)		
Confusion	Hallucinations	Delirium
Infections and Infestations (reported as AESI if Grade ≥3)		
Viral Rash		
Skin Disorders (reported as AESI if Grade ≥3)		
Rash*	Acneiform Rash*	

*Rash and Acneiform Rash were considered an AESI for Grade ≥2 in earlier versions of the protocol. Due to this, there may be some Grade 2 rashes marked as AESIs by the investigators, which will be reported as such.

The AESI of "Rash" is a composite term, all PTs which will be considered as a "Rash" AESI are listed in Appendix 2.

AEs will be summarized combined across the Treatment and LTFU phases. A TEAE is defined as an AE that starts or worsens on or after the first dose of study treatment is taken, including those occurring or increasing in severity up to 30 days after the last dose of study treatment.

A treatment-emergent serious adverse event (SAE) is defined as a TEAE that, at any dose; results in death; is life-threatening; requires inpatient hospitalization or prolongation of existing hospitalization; results in persistent disability/incapacity; is a congenital anomaly/birth defect; or is assessed by the investigator as an otherwise important serious medical event.

An AE will be classified as a study treatment-related AE if it has an investigator determination of related to study treatment. Investigators will provide their assessment as related to study treatment, not related to study treatment, or not applicable. AEs that are missing a treatment related/unrelated classification will not be imputed and "missing" will be presented as a category in summary tables.

An AE will be classified as a study procedure-related adverse event if it has an investigator determination of related to study procedure. Investigators will provide their assessment as related to study procedure, not related to study procedure, or not applicable. AEs that are missing a study procedure related/unrelated classification will not be imputed and "missing" will be presented as a category in summary tables.

Additionally, the following terms are defined:

- Adverse reaction (AR) - any noxious and unintended response to a medical product or procedure, for which a causal relationship with the product or procedure is at least a reasonable possibility (i.e., the relationship cannot be ruled out).
- Serious adverse reaction (SAR) - an SAE for which a causal relationship with the product or procedure is at least a reasonable possibility (i.e., the relationship cannot be ruled out).
- Suspected unexpected serious adverse reaction (SUSAR) - an SAR that is judged as unexpected. An event is considered "unexpected" if it is not listed as expected in the reference safety information (RSI) section of the investigator brochure (IB) or summary of product characteristics.

The intensity of all AEs will be graded according to the NCI Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. For AEs not listed in the CTCAE, the following grading system should be used:

- CTCAE Grade 1 Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; treatment not indicated.
- CTCAE Grade 2 Moderate; minimal, local, or noninvasive treatment indicated; limiting age-appropriate instrumental activities of daily living (ADL)*.
- CTCAE Grade 3 Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL**.
- CTCAE Grade 4 Life-threatening consequences; urgent treatment indicated.
- CTCAE Grade 5 Death related to AE.

*Instrumental ADL refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.
**Self-care ADL refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden

An AE leading to study discontinuation will be defined as an AE where the primary reason for discontinuation of study (on the 'End of Study' CRF page) was recorded as being an AE.

An AE leading to treatment discontinuation will be defined as an AE where the action taken with study treatment was recorded as 'Study treatment withdrawn' on the 'Adverse Events' CRF page.

4.2.1.2 Coding of Adverse Event Terms

The AE term (Investigator term) will be assigned to the lowest level term (LLT), and a preferred term (PT) will be classified by a high-level term (HLT), a high-level group term (HLGT), and a system organ class (SOC) according to the Medical Dictionary for Regulatory Activities (MedDRA) thesaurus, Version 24.0 or higher, depending on the latest version available during the study.

Although there can be multiple SOCs for a PT, each PT will be linked with one SOC, namely the primary SOC that is automatically assigned by MedDRA via one HLT, HLGT route.

The following coding data will be presented:

- LLT (Investigator term).
- PT.
- Coding data per primary SOC:
  - HLT.
  - HLGT.

In the event that no coding information is available for a specific AE, the AE will be presented as an 'Uncodable Event' in all summary tables.

AEs will be reported on a per-participant basis and per-event. On a per-participant basis, this means that even if a participant reported the same event repeatedly (i.e., events mapped to the same PT) during the study period, the event will be counted only once. In the latter case, the event will be assigned the worst severity and the strongest relationship to the study treatment. The earliest date will be regarded as start date of the event and the latest date/time will be regarded as stop date of the event within the assigned study period.

4.2.2 Clinical Laboratory Evaluations

Blood and urine samples are collected and analyzed by a central laboratory. All laboratory results were standardized to Système International (S.I.) units. For all tables and figure summaries, only central laboratory data will be used. All data (including local laboratory data) will be included in listings.

In the case that multiple assessments map to the same visit, then the value at the scheduled assessment will be taken over any unscheduled assessments. If there are multiple unscheduled assessments, the assessment closest to the target study day will be used.

For the duration of the study, all safety laboratory assessments will be checked for abnormalities and measured against the NCI-CTCAE Version 5.0 for toxicities. The worst on-treatment assessment is determined by the value corresponding to the highest NCI-CTCAE Version 5.0 toxicity grade.

Hematology and biochemistry parameters with either a change of ≥ 2 in CTCAE Grade from baseline or a CTCAE Grade ≥ 4 will be flagged as markedly abnormal.

For the quantitative laboratory tests, the change from baseline value at each post-baseline visit will be calculated as the difference between the measurement obtained at the specific post-baseline visit and the baseline value.

In addition, quantitative test results at each visit will be categorized by the laboratory as 'Normal' (within the reference range) or 'Abnormal' (outside the reference range) according to the

reference ranges specified in the laboratory manual. Abnormal results will be further classified as being 'Low' or 'High', depending on whether the result is below or above the reference range limits. For all abnormal values, clinical significance (as determined by the investigator) should be indicated.

4.2.2.1 Hematology

Hematology laboratory evaluations are performed at all scheduled visits in accordance with the SoA (Sections 1.3-1.4 of the protocol). The evaluated parameters are hemoglobin (HGB); hematocrit (HCT); platelet count; red blood cell count; red blood cell indices including mean cell volume, mean cell hemoglobin, and %reticulocytes; and white blood cell count with differential including neutrophils, lymphocytes, monocytes, eosinophils, and basophils.

4.2.2.2 Serum Chemistry

Serum chemistry laboratory evaluations are performed at all scheduled visits in accordance with the SoA (Sections 1.3-1.4 of the protocol). The evaluated parameters are aspartate aminotransferase (AST)/serum glutamic-oxaloacetic transaminase (SGOT), alanine aminotransferase (ALT)/serum glutamic-pyruvic transaminase (SGPT), total and direct bilirubin, gamma-glutamyl transferase (GGT), sodium, chloride, potassium, calcium, bicarbonate, inorganic phosphorous, alkaline phosphatase, creatinine, creatine phosphokinase (CPK), blood urea nitrogen (BUN), non-fasting glucose, uric acid, albumin, total protein, triglycerides, and total cholesterol.

4.2.2.3 Urinalysis

Urinalysis evaluations are performed at the time points specified in the SoA (Sections 1.3-1.4 of the protocol). The evaluated parameters are specific gravity, bilirubin, glucose, leukocyte esterase, nitrite, protein, urobilinogen, blood, ketones, pH, and microscopy (only if blood or protein is abnormal).

4.2.2.4 Serology

Serology is performed at screening. Tests are be performed for hepatitis B virus (HBV), hepatitis B surface antigen (HBsAg), hepatitis C virus (HCV) and hepatitis C antibody.

4.2.3 Electrocardiogram (ECG) Evaluations

Triplicate 12-lead electrocardiogram (ECG) evaluations are performed at the time points specified in the SoA (Sections 1.3-1.4 of the protocol). At each visit, three ECG assessments are performed in succession, approximately 3 minutes apart, these triplicate recordings are averaged to give the result for a given visit and parameter. In the case that not all ECG assessments were recorded, the average of the recording(s) taken will be used. At the C1D1 visit measurements are taken both pre-dose and post-dose.

Table 9 presents the quantitative and qualitative ECG parameters that were collected.

Table 9: ECG Parameters

ECG Parameters (Unit)
Conduction Times
Heart rate (bpm)
RR Interval (msec)
PR Interval (msec)

QRS Interval (msec)
QT Interval (msec)
QTc Interval (msec)
QTcF Interval (msec)

For the quantitative ECG parameters, the change from baseline value at each post-baseline visit will be calculated as the difference between the measurement obtained at the specific post-baseline visit and the baseline value.

4.2.4 Echocardiograms

Echocardiograms are performed to assess left ventricular ejection fraction (LVEF) (%) in accordance with the SoA (Sections 1.3-1.4 of the protocol).

4.2.5 Vital Signs Evaluations

Height and weight are measured at screening, baseline visit (C1D1), and Cycles 3, 7, 11, 15, 19 and 23 for the Treatment Phase, and Cycle 28 and every 4 cycles after that during the study (I.e. LTFU phase). All other vital signs evaluations were performed at all visits described in accordance with the SoA (Sections 1.3-1.4 of the protocol).
The following variables were collected:

- Height (cm)
- Weight (kg)
- Systolic blood pressure (SBP) (mmHg)
- Diastolic blood pressure (DBP) (mmHg)
- Respiratory rate (breaths/min)
- Pulse rate (beats/min)
- Body temperature (degrees C)

In addition to the collected variables, the following variables will be derived at all visits with recorded height and weight data:

- BSA (m^2) will be calculated according to the Du Bois formula as:
  $BSA\ (m^2) = 0.007184 \times Weight(kg)^{0.425} \times Height(cm)^{0.725}$.

For each variable the change from baseline value at each post-baseline visit will be calculated as the difference between the measurement obtained at the specific post-baseline visit, and the baseline value.

4.2.6 Ophthalmic Measurements

Participants have a standard of care ocular evaluation in accordance with the SoA (Sections 1.3-1.4 of the protocol).

4.3 Other Measures

4.3.1 Participant Disposition

Participant disposition data is primarily collected on the 'End of Study' case report form (CRF) page when a participant completed or discontinued from the study. The following data will be summarized:

- Participants in each analysis set (I.e. Full Analysis Set, Safety Analysis Set, Per-Protocol Set)
- Participants who complete the Treatment Phase
- Participants who discontinued from study in the Treatment Phase and reason for discontinuation
- Participants who discontinued from treatment in the Treatment Phase and reason for discontinuation
- Participants withdrawing consent and reason for withdrawal
- Participants who complete Treatment Phase and do not enter LTFU Phase and reason for not entering LTFU Phase
- Participants entering the LTFU Phase
- Participants who discontinued from study in the LTFU Phase and reason for discontinuation
- Participants who discontinued from treatment in the LTFU Phase and reason for discontinuation
- Participants withdrawing consent in the LTFU Phase and reason for withdrawal

4.3.2 Protocol Deviations

Protocol deviations are reviewed in accordance with the Protocol Deviation Plan. All deviations will be reviewed prior to database lock to determine which deviations are deemed important to report in the CSR in accordance with ICH E3 guidelines. A CSR reportable protocol deviation is typically considered a significant deviation that should be addressed in the CSR as it may impact the accuracy, interpretation, and/or reliability of the study data, participant rights, safety, or well-being. Major protocol deviations are defined as reportable deviations that may impact the accuracy and or reliability of the efficacy data. As described in Section 3.2, any participants with major reportable protocol deviation leads to exclusion from the PPS.

4.3.3 Demographics

Demographic data are primarily collected on the 'Demographics' CRF page at screening. The following data will be summarized:

- Age at time of consent (years)
- Sex
- Childbearing potential at screening
- Ethnicity
- Race
- Karnofsky/Lanksy Performance Status at baseline (Cycle 1 Day 1)

Missing demography data will be handled according to the rules specified in Section 5.5.2.2.

4.3.4 Baseline Participant Characteristics

Baseline participant characteristics include characteristics that participants presented with prior to the first administration of study treatment.

4.3.4.1 Baseline Disease Characteristics

Tumor history and location is recorded at screening on the "Plexiform Neurofibroma History" CRF. History of tumor therapies, such as surgery and radiotherapy, and baseline tumor volume will also be assessed. The full list of disease characteristics and history to be presented in summaries is as follows:

- Target tumor location
- Target tumor type (nodular, solitary nodular, typical) by central reader (reader A)
- Time since diagnosis (in months)
- NF1 mutation status (CLIA/COA testing)
- Target tumor progression at baseline
- Reported morbidities
- Prior PN treatments
- Prior PN surgery
- Prior PN radiotherapy

Prior PN treatment information will be taken from the 'Prior Plexiform Neurofibroma Surgeries', 'Prior Plexiform Neurofibroma Radiotherapy' and 'Prior Plexiform Neurofibroma Targeted Therapies/ Medication' CRF pages.

4.3.5 Medical History

Medical history contains information about conditions that a participant might have suffered from prior to the first administration of study treatment or conditions that were ongoing at the time of the first administration of study treatment.

Medical history includes any history of clinically significant disease. Cancer history will include cancer type, date of diagnosis, and an assessment of prior surgery, prior radiotherapy, prior drug therapy including start and stop dates, best response and reason for discontinuation.

4.3.5.1 Coding of Medical History Terms

The medical history term (Investigator term) is assigned to the LLT, and a PT will be classified by a HLT, a HLGT and a SOC according to the MedDRA thesaurus, Version 24.0 or higher, depending on the latest version available during the study.

Although there can be multiple SOCs for a PT, each PT will be linked with one SOC, namely the primary SOC which is automatically assigned by MedDRA via one HLT, HLGT route.

The following coding data will be presented:

- LLT
- PT
- Coding data per primary SOC:
  - HLT
  - HLGT

Medical history will be reported on a per-participant basis. This means that even if a participant suffered the same clinical event repeatedly (i.e., events mapped to the same PT), the event will be counted only once and the earliest date will be regarded as start date of the event and the latest date will be regarded as stop date of the event.

The same rule for counting applies for PTs mapped to the same HLT and for HLTs mapped to the same HLGT and for HLGTs mapped to the same SOC.

4.3.6 Concomitant Medications and Procedures

Concomitant medications and procedures are defined as any medication or procedure that did not end prior to first dose of study treatment or did not start after the applicable safety follow-up period.

Medication and procedures will be considered prior if they stopped before the first dose of study treatment. Concomitant medications data is collected throughout the study on the 'Concomitant Medications' CRF page, Concomitant procedures data is collected throughout the study on the 'Concomitant Procedures' CRF page.

Missing concomitant medications dates will be handled according to the rules specified in Section 5.5.2.4.

4.3.6.1 Coding of Concomitant Medication Terms

Concomitant medications are classified according to active drug substance using the World Health Organization-Drug Dictionary (WHO-DD), Global B3 Version Mar 2021.

In this study, anatomical therapeutic chemical (ATC) codes are defined to the 4th level.

Although there can be multiple ATC classes for a drug, each drug will be linked with one ATC class which will be assigned manually during the coding process, based on information about the indication and route in relation to the study therapeutic area. This one ATC class will be indicated as the 'primary' ATC class, and only the primary class will be presented.

4.3.7 Exposure to Study Treatment

There is only a single dose level on this study. All participants receive mirdametinib at a dosage level of approximately 2 mg/m^2 BID (up to a maximum of 4 mg BID). All participants follow an intermittent dosing schedule of 3 weeks on/1 week off, which creates 28-day cycles. The Treatment Phase comprises 24 cycles.

Participants enrolled in the LTFU Phase of the study continue taking the last dose assigned in the Treatment Phase (approximately 2 mg/m^2 BID [maximum dose of 4 mg BID]) of mirdametinib, unless a dose modification is required (e.g., body surface area [BSA] change), on an intermittent (3 weeks on/1 week off) dosing schedule in 28-day cycles. This phase is of indefinite length.

At any point in the study, if a participant exhibits sufficiently worrisome AEs of significant toxicity, dosing can be interrupted and/or reduced. If dose is reduced, then it may only be increased in the Treatment Phase following a BSA change and with approval from the medical monitor and sponsor. The criteria for dose reduction, and the reduced dose plan are in Section 6.6 of the protocol.

The following exposure-related measures will be summarized:

- Duration of exposure (months)
- Total duration of exposure (participant years)
- Time to first dose reduction (months)
- Participants with dose modification along with reasons for dose modification
- Participants with dose reduction along with reasons for dose reduction
- Participants with dose interruption along with reasons for dose interruption
- Actual dose intensity
- Relative dose intensity

Duration of exposure in months is calculated as (Date of last dose – Date of first dose +1)/30.4375.

Total duration of exposure in participant years is calculated as (total sum of all participants duration of exposure in months)/12

Time to first dose reduction in months is calculated as (Date of first dose reduction – Date of first dose + 1)/30.4375

Actual and relative dose intensity will be calculated as follows:

- Actual dose intensity – cumulative dose received/number of prescribed dosing days.
- Relative dose intensity – 100*(cumulative dose received/planned dose).

Cumulative dose received will be calculated using dispensed and returned pill counts. Actual dose intensity will also be derived accounting for BSA.

4.3.8 Pharmacokinetic and Pharmacodynamic Analyses

A separate supplementary SAP will describe the PK parameters, PopPK/PD models, and analyses. Plasma PK collection dates, times, and concentration results will be displayed in a data listing.

5. Statistical Methodology

5.1 General Statistical Methods

5.1.1 General Information

All analysis data sets, and output will be produced by the Biostatistics Department of Quanticate Ltd using the SAS® system Version 9.4 or higher.

5.1.2 Default Descriptive Statistics and Data Rules

Unless otherwise stated, summary statistics including the number of participants, mean, standard deviation (SD), median, minimum, and maximum, will be presented by age cohort for all continuous variables. Minimum and maximum values will be presented to the same decimal

precision as the raw values, the mean and median values to one more, and the standard deviation, to two more decimal places than the raw values.

For categorical variables, per category, the absolute counts (n) and percentages (%) of participants in each with data, and if appropriate, the number of participants with missing data, will be presented for each age cohort. All percentages will be presented to one decimal place.

For AEs reported on a per-participant basis, medical history and concomitant medications, the denominator for the percentage calculation will be the number of participants in each age cohort. A participant will be considered at risk if the participant is in the Safety Analysis Set/FAS and entered the respective study period.

5.2 Hypotheses and Decision Rules

The primary efficacy endpoint will be assessed separately for each age cohort. The lower limit of the 95% confidence interval for the primary endpoint will be compared to a minimal clinically relevant response in the age groups.

The study was originally set up using minimally clinically relevant thresholds of 23% in adult participants and 42% in pediatric participants. As such, the null hypothesis for the participants ≥ 18 years of age at baseline (the adult population) is that the ORR is equal to 23%. The alternative hypothesis is that the ORR is not equal to 23%. The null hypothesis for participants 2 to 17 years of age at baseline (the pediatric population) is that the ORR is equal to 42%. The alternative hypothesis is that the ORR is not equal to 42%.

Due to emerging data and discussion with the FDA, as outlined in Section 7 (changes to Planned Analyses), the threshold for the pediatric cohort was updated. The null hypothesis for participants 2 to 17 years of age at baseline (the pediatric population) is that the ORR is equal to 20%. The alternative hypothesis is that the ORR is not equal to 20%.

5.3 Covariates

No covariates will be considered for the primary efficacy analysis. For all mixed models for repeated measures (MMRM) used in secondary and exploratory analyses, the only covariate included will be the respective baseline score.

5.4 Multi-Center Data

This is a multi-center study and all the data from all sites will be pooled. No consideration of center effect is planned for the listed analyses.

5.5 Handling of Missing Data

Missing data will only be imputed in the following cases:

5.5.1 Secondary Efficacy Endpoints

5.5.1.1 Change from Baseline in QoL assessed by the age specific PedsQL

If ≤ 50% of questions in a given scale are missing, then the scale score will be imputed as the mean of the non-missing questions. If > 50% of questions are missing, then the scale score should not be calculated for that scale. The main analysis for this endpoint will be a mixed model for repeated measures (MMRM), which treats all missing data as missing at random (MAR).

A sensitivity analysis will be performed to assess the potential impact of this assumption if some data is missing not at random (MNAR). For this purpose, monotone and non-monotone missing values will be treated differently. Non-monotone missing values are values missing intermittently, where a participant may miss some PRO assessments but has PRO assessments for the same score later in the study. Monotone missing values are such that once a value is missing for a given score, no subsequent values for this score are available. Any given participant may have a combination of non-monotone and monotone missing values.

Non-monotone missing values are assumed to be MAR so will not be directly imputed as the analysis by MMRM already makes this assumption. Monotone data, however, will be considered MNAR and imputed to be equal to the participant's baseline score. This choice of imputation represents a negative outcome of no improvement in QoL for each participant.

The details of the sensitivity analysis to be performed with this imputed data are given in Section 6.11.2.2.

5.5.1.2 Change from Baseline using the NRS-11

If ≥4 of the 7 NRS-11 readings associated with a single visit are missing, then the score for that visit will be treated as missing. If <4 readings are missing, the value associated with a visit will be the average of the non-missing readings.

A sensitivity analysis, similar to that described for PedsQL in Section 5.5.1.1, will be performed to assess the impact of MNAR data on the NRS-11 results. The imputation described in Section 5.5.1.1 will be performed for missing NRS-11 data, and the sensitivity analysis will be performed as detailed in Section 6.11.2.3.

5.5.1.3 Change from Baseline using the PII

If >3 of the 6 PII questions associated with a single assessment are missing, then the score for that assessment will be treated as missing. Otherwise, the score for a given assessment will be taken as the average of the non-missing assessment scores. If ≥4 of the 7 PII assessments corresponding to a single visit are missing, then the score for that visit will be treated as missing. Otherwise, the value associated with a visit will be the average of the non-missing readings.

A similar sensitivity analysis will also be performed to assess the impact of MNAR data on the PII results. The imputation described in Section 5.5.1.1 will be performed for missing PII data, and the sensitivity analysis will be performed as detailed in Section 6.11.2.3.

5.5.2 Other Endpoints

5.5.2.1 Adverse Events

Missing and/or incomplete dates/times for AEs are imputed in a manner resulting in the earliest onset or the longest duration during the treatment period, taking additionally into account that the

start date/time should not be after the stop date/time. Stop dates/times will not be imputed if the AE is ongoing. To do this, it is necessary to define the earliest and latest possible date for an event, which is done as follows:

The earliest possible date is derived as:

- The date itself if it is complete.
- The date of the first day of the month, if month and year are available but day is missing.
- The date of the first day of the year, if year is available but day and month are missing.
- A very early date, e.g., 01JAN1900 if the date is completely missing.

The latest possible date is derived as:

- The date itself if it is complete.
- The date of the last day of the month, if month and year are available but day is missing.
- The date of the last day of the year, if year is available but day and month are missing.
- A very late date, e.g., 31DEC2100 if the date is completely missing.

These earliest/latest possible dates are used to impute the missing/incomplete start/stop dates.

Missing/incomplete start dates will be imputed according to the following:

1. If the earliest possible start date is on or after the start of treatment administration:
    - The earliest possible start date.

2. If the earliest possible start date is before the start of treatment administration, the minimum of:
    - The date of first study treatment administration.
    - The latest possible start date.
    - The latest possible stop date.

Missing/incomplete stop dates will be imputed according to the following:

1. If the latest possible stop date is before the end of treatment administration:
    - The latest possible stop date.

2. If the latest possible stop date/time is after the end of treatment administration, the maximum of:
    - The date of final study treatment administration
    - The earliest possible stop date
    - The earliest possible start date

The imputation method will only be used to determine the time of the event relative to the first administration of study treatment.

In the occasion of missing severity data for an AE, no imputation will be performed, and severity will be presented as "Missing" in summary tables.

In the event that no coding information is available for a specific AE, the AE will be presented as an 'Uncodable Event' in summary tables.

5.5.2.2 Demographics

For determining age when the date of birth is not known completely, a missing day only will be imputed as the 15th, a missing day and month will be imputed as the 2nd of July which is day 183 in the year.

5.5.2.3 Medical History

Partially missing diagnosis dates will not be imputed.

5.5.2.4 Concomitant Medications

Missing concomitant medication dates will be handled in a similar fashion as described for Adverse Events in Section 5.5.2.1.

5.6 Windowing Conventions

For safety data which are not linked to nominal visits (e.g. AEs) as per the protocol schedule of assessments, these data will be mapped to the cycle of treatment in which they onset according to Table 10 and Table 11.

For safety data collected at nominal visits, the recorded nominal visit will be used and unscheduled visits will be mapped to the nearest planned visit where it would be expected to be recorded, by study visit day, as per the protocol. If a record is mapped to a visit which has a valid reading nominally assigned, the nominally assigned visit value will be used in all summaries and analysis.

Table 10: Visit Windows – Treatment Phase

Assessment Study Day	Visit	Visit Label	Study Visit Day
	Screening	Screening	-45
2≤Sday<8	Cycle 1 – Week 1 (Baseline, study visit day 1)	C1D1	1
8≤Sday<29	Cycle 1 - Week 3 (Cycle Day 15)	C1D15	15
29≤SDay<57	Cycle 2 – Week 7 (Cycle Day 15)	C2D15	43
57≤SDay<85	Cycle 3 – Week 11 (Cycle Day 15)	C3D15	71
85≤SDay<113	Cycle 4 – Week 15 (Cycle Day 15)	C4D15	99
113≤SDay<141	Cycle 5 – Week 19 (Cycle Day 15)	C5D15	127
141≤SDay<169	Cycle 6 – Week 23 (Cycle Day 15)	C6D15	155
169≤SDay<197	Cycle 7 – Week 27 (Cycle Day 15)	C7D15	183
197≤SDay<225	Cycle 8 – Week 31 (Cycle Day 15)	C8D15	211
225≤SDay<253	Cycle 9 – Week 35 (Cycle Day 15)	C9D15	239
253≤SDay<281	Cycle 10 – Week 39 (Cycle Day 15)	C10D15	267
281≤SDay<309	Cycle 11 – Week 43 (Cycle Day 15)	C11D15	295
309≤SDay<337	Cycle 12 – Week 47 (Cycle Day 15)	C12D15	323
337≤SDay<365	Cycle 13 – Week 51 (Cycle Day 15)	C13D15	351
365≤SDay<393	Cycle 14 – Week 55 (Cycle Day 15)	C14D15	379

393≤SDay<421	Cycle 15 – Week 59 (Cycle Day 15)	C15D15	407
421≤SDay<449	Cycle 16 – Week 63 (Cycle Day 15)	C16D15	435
449≤SDay<477	Cycle 17 – Week 67 (Cycle Day 15)	C17D15	463
477≤SDay<505	Cycle 18 – Week 71 (Cycle Day 15)	C18D15	491
505≤SDay<533	Cycle 19 – Week 75 (Cycle Day 15)	C19D15	519
533≤SDay<562	Cycle 20 – Week 79 (Cycle Day 15)	C20D15	547
561≤SDay<589	Cycle 21 – Week 83 (Cycle Day 15)	C21D15	575
589≤SDay<617	Cycle 22 – Week 87 (Cycle Day 15)	C22D15	603
617≤SDay<645	Cycle 23 – Week 91 (Cycle Day 15)	C23D15	631
645≤SDay<673	Cycle 24 – Week 95 (Cycle Day 15)	C24D15	659
	Safety Follow-up – 30 days after last dose	SFU	-

Abbreviations: SDay: study day for individual participant

Table 11: Visit Windows - LTFU Phase

Assessment Study Day	Visit	Visit Label	Study Visit Day
673≤SDay<701	Cycle 25 – Week 99 (Cycle Day 15)	C25D15	687
701≤SDay<729	Cycle 26 – Week 103 (Cycle Day 15)	C26D15	715
(X[a]-1)*28≤SDay<X[a]*28	Cycle X[a] – Week 4X[a]-1 (Cycle Day 15)	CX[a]D15	(X[a]-1)*28+15
	Safety Follow-up – 30 days after last dose	LTSFU	

[a]For X>26 and a whole number Abbreviations: SDay: study day for individual participant

For tumor volume/tumor response data, the nominal visit as entered in the CRF will be used unless a visit is recorded as "Unscheduled". If an assessment is unscheduled, it will be assigned to a cycle using the windowing laid out in Table 12. If a record is mapped to a visit which has a valid reading nominally assigned, the nominally assigned visit value will be used in all summaries and analysis.

Table 12: Tumor Visit Windows

Assessment Study Day	Visit	Visit Label	Study Visit Day
2≤Sday≤182	Cycle 5	C5D15	127
182<Sday≤294	Cycle 9	C9D15	239
294<Sday≤406	Cycle 13	C13D15	351
406<Sday≤518	Cycle 17	C17D15	463
518<Sday≤616	Cycle 21	C21D15	575
616<Sday≤714	Cycle 24	C24D21	665
714<Sday≤826	Cycle 28	C28D15	771
826<Sday≤938	Cycle 32	C32D15	883
((XX[a]-2)*28)-14<Sday≤((XX[a]+2)*28)-14	Cycle XX	CXXD15	(XX[a]*28)-13

[a]For XX>32 and divisible by 4. Abbreviations: SDay: study day for individual participant

For PRO assessments the nominal visit will be used if it is recorded as C1D1 or is recorded in the following table. For any other case, the visit will be mapped by study day to the closest visit using the table below. If a record is mapped to a visit which has a valid reading nominally assigned, the nominally assigned visit value will be used in all summaries and analysis.

Table 13: PRO Visit Windows

Assessment Study Day	Visit	Visit Label	Study Visit Day
2≤Sday≤99	Cycle 3	C3D15	71
99<Sday≤155	Cycle 5	C5D15	127
155<Sday≤211	Cycle 7	C7D15	183
211<Sday≤294	Cycle 9	C9D15	239
294<Sday≤406	Cycle 13	C13D15	351
406<Sday≤518	Cycle 17	C17D15	463
518<Sday≤616	Cycle 21	C21D15	575
616<Sday≤714	Cycle 24	C24D21	665
714<Sday≤826	Cycle 28	C28D15	771
826<Sday≤938	Cycle 32	C32D15	883
((XX[a]-2)*28)-14<Sday≤((XX[a]+2)*28)-14	Cycle XX	CXXD15	(XX*28)-13

[a]For XX>32 and divisible by 4. Abbreviations: SDay: study day for individual participant

For all other efficacy assessments, a similar algorithm to above will be used, with all unscheduled and ET assessments mapped to the nearest scheduled assessment visit by study day, with the earlier scheduled visit being used in the case of a tie. In the case that multiple records are recorded for a participant as the same nominal visit, the average of the results will be taken. If both unscheduled visits are equidistant from the visit being mapped to, then the earlier visit will be mapped.

The Study Visit Day of an event/assessment will be calculated relative to the first study treatment administration. The Study Visit Day of events/assessments occurring before the first administration will be calculated as follows:

Study Visit Day = (Date of assessment/event - Date of first study treatment administration).

For events/assessments occurring at baseline or post-baseline, Study Visit Day will be calculated as follows:

Study Visit Day = (Date of assessment/event - Date of first study treatment administration) + 1.

For DoR and other time-based endpoints, such as exposure, time periods may be presented in months rather than days. In these cases, the following formula will be used for conversion:

time in months = (time in days/365.25)*12.

5.7 Interim Analyses

No formal interim analysis is planned for this study.

An independent DMC will review safety and efficacy data at regular intervals during the study. Further details can be found in Protocol Section 9.5 and the DMC charter and will not be covered in this SAP.

5.8 PRO Analysis

All PROs analyzed in this study will be assessed using the mean change from baseline as well as by participants achieving of a meaningful score difference at the Cycle 13 visit.

Cycle 13 has been selected as the timepoint that will best allow the treatment effect to be captured in PRO scores. It is projected that a strong majority of participants who present a response will do so at or before Cycle 13, with the median time of response expected to be Cycle 5 to Cycle 9, so any associated pain or QoL improvements should be captured at the Cycle 13 visit. Choosing a later cycle would capture few additional responses and may be susceptible to increased amounts of missing data.

An addendum meaningful change threshold (MCT) SAP will provide detailed descriptions of the statistical methods, data derivations, and data displays for the meaningful change analyses of the clinical outcome assessment (COA) data collected in the MEK-NF-201 study. This addendum (PRO Addendum) for the meaningful change threshold analysis plan (MCTAP) is based on the MEK-NF-201 Protocol Amendment 4 dated 17 May 2022.

6. Statistical Analyses

6.1 Participant Disposition

Participant disposition information, as listed in Section 4.3.1, will be summarized separately for each age cohort by absolute counts (n) and percentages (%) for the Enrolled Set. Percentages will be based on the number of participants with data in the relevant age cohort.

All disposition data will also be listed.

6.2 Protocol Deviations

All protocol deviations that are categorized as defined in Section 4.3.2, will be summarized by the number of participants (n) and percentages (%) for the FAS. Percentages will be based on the number of participants with data in the relevant age cohort.

.

6.3 Demographics

Baseline demographics data as laid out in Section 4.3.3 include continuous and categorical variables, which will be summarized as described in Section 5.1.2. Percentages will be based on the number of participants with data. These summaries will be presented for the FAS and PPS.

All demographic data will be listed, with a separate listing for changes in childbearing potential during the study.

6.4 Baseline Disease Characteristics

Baseline disease characteristics, as listed in Section 4.3.4.1, comprise continuous and categorical variables, which will be summarized as described in Section 5.1.2. Percentages will be based on the number of participants with data. These summaries will be presented for the FAS and PPS.

All baseline disease characteristics data will be listed.

6.5 Medical History

Medical history will be summarized by absolute counts (n) and percentages (%) for the FAS. All medical history records will be presented by SOC and PT as laid out in Section 4.3.5.1.

Additional summaries will be presented for absolute counts (n) and percentages (%) of participants who have prior PN or cancer related radiotherapy, surgery, or procedures.

All medical history data will be listed.

6.6 Concomitant Medication

Number of participants with at least one recorded concomitant medication, 4th level ATC codes and preferred terms will be summarized by absolute counts (n) and percentages (%) for the FAS. Percentages will be calculated based on the number of participants at risk for the specific age cohort. This summary will be presented for the Treatment Phase and the LTFU Phase.

All concomitant medication data will be listed.

6.7 Concomitant Surgeries and Procedures

Surgeries and procedures will be summarized by absolute counts (n) and percentages (%) for the FAS. Unless specified otherwise, results will be sorted alphabetically by SOC, and by decreasing frequency of participants presenting PTs within a SOC. In the event of PTs having equal frequencies, the PTs will be sorted alphabetically within the SOC. This summary will be presented by Treatment Phase and for the LTFU Phase.

All concomitant surgeries and procedures data will be listed.

6.8 Treatment Compliance

All treatment compliance data including tablets dispensed/returned and recorded instances of non-compliant dosing will be listed.

6.9 Exposure to Study Treatment

Exposure to study treatment will be summarized by the parameters listed in Section 4.3.7, as described in Section 5.1.2. Percentages will be based on the number of participants with data.

6.10 PRO Compliance & Completion

Compliance and completion rates will be provided for the PRO instruments corresponding to the secondary endpoints of PedsQL, NRS-11, and PII.

Compliance rate is calculated as participants with minimum requirements for scoring divided by the number of participants with PRO assessments expected. For compliance rates, the summaries will include reasons for missing observations among patients who were expected to complete the questionnaire by assessment.

For completion rate, this will be derived in a similar manner as the compliance rate, however the subset of the FAS who are considered eligible to complete each questionnaire, as per Table 14, is used as a fixed denominator. For completion rates, the summaries will include reasons for missing observations among the eligible participants within the eligible FAS population by assessment. Reasons for missing observations include, but are not limited to, progression, adverse event, subject withdrawal, lost to follow-up, patient or physician decision, patient was too ill to complete, not completed due to site staff error, technical problems with eDiary, patient non-compliance, etc.

Table 14: PRO Age Range Eligibility

PROs	PedsQL		NRS-11	PII	
Age of Participant (years)	Self-report	Parent Proxy	Self-Report	Self-report	Parent Proxy
2		X			
3		X			
4		X			
5	X	X			
6	X	X		X	X
7	X	X		X	X
8-17	X	X	X	X	X
≥18	X	X[a]	X	X	

[a] If a participant is ≥ 18 years of age, but in the opinion of the investigator has cognitive deficiencies that would inhibit accurate collection of PRO data, a parent or legal guardian must also complete the appropriate parent-proxy PRO.

For the diary instruments where more than one day was expected (NRS-11 and PII), the number and percent of participants with ≥1 complete diary day, ≥2 complete diary days, ≥3 complete diary days, etc. for the diary periods associated with the study visits will also be provided.

All summaries of compliance/completion will be displayed with the relevant PROs.

6.11 Efficacy Analyses

The efficacy endpoints will include collected and derived continuous and categorical variables and will be summarized as described in Section 5.1.2, unless otherwise specified.

The FAS will be the primary analysis set of interest in all efficacy analyses, unless otherwise specified.

6.11.1 Primary Efficacy Analysis

To assess the primary efficacy endpoint, confirmed ORR by the end of the Treatment Phase as assessed by BICR will be presented as a percentage alongside the corresponding 2-sided 95% Clopper-Pearson CI and the p-value from a one-sample two-sided binomial test for the confirmed ORR compared to the minimum clinically relevant response rates defined in Section 5.2. The overall assessment of the endpoint will be based on the exclusion of the minimum clinically relevant response rate by the lower tail of the 95% CI.

All tumor data from MRI assessments will be listed across both the Treatment and LTFU Phases.

6.11.1.1 Sensitivity Analysis of Primary Efficacy Analysis

A sensitivity analysis will be performed that mirrors the primary efficacy analysis but uses the PPS rather than the FAS.

As additional sensitivity analyses, the confirmed ORR will be summarized by each central BICR reader separately for both the FAS and the PPS.

6.11.2 Secondary Efficacy Analyses

6.11.2.1 Duration of response for participants whose best response is CR or PR using centrally read MRI volumetric analysis

DoR will be analyzed using Kaplan-Meier (KM) methods, with the median, lower, and upper quartile estimates of DoR and the corresponding 95% CIs (using the Brookmeyer and Crowley method). The number of participants with events (i.e., radiographic disease progression or death) or who are censored, as well as reason for censoring, will be presented separately by age cohort. "NE" will be presented where estimates of such were not estimable. Additionally, KM plots will be presented separately for each age cohort and will contain the number of participants at risk and the event rate. In addition, KM estimates with 95% CIs (using the log-log transformation) will also be presented at key timepoints such as 8, 12, 16, 20, and 24 months.

Censoring rules and the definition of confirmed response are laid out in Section 4.1.2.1.

The sensitivity analyses using REiNS criteria as outlined in Section 4.1.2.1.1 will also be summarized using the same method as described above.

In addition, a swimmer plot will be used to present duration of treatment (in days) where radiographic progression, first onset and confirmation of confirmed response, whether still on study at cutoff and death are displayed, split by cohort.

6.11.2.2 Change from Baseline in quality of life assessed by the age specific Pediatric Quality of Life Inventory

All PedsQL records will be listed by participant, visit and questionnaire type (self-report/parent proxy).

Actual values and change from baseline for each score listed in Section 4.1.2.2.1 will be summarized for each age cohort by visit, for participants overall and by confirmed objective response status (as defined with confirmed response rate in Section 4.1.1 using adjudicated values where appropriate) with self-report and parent proxy scores presented separately. The summaries presented will be number and percentage of participants with a non-missing score, mean, SD, median, minimum, and maximum values.

For PedsQL, a within-patient meaningful score change, or meaningful score difference (FDA, 2023) , will be derived by anchor-based analysis of the study's data, as described in the meaningful change threshold analysis plan (Version 1.0, dated 12Sep2023). The primary focus of this will be on if a relevant difference exists at Cycle 13.

In addition to the anchor-based estimates described above, the number and percentage of participants with a within-patient meaningful score change estimates are defined below and will be considered as a supportive analysis. For PedsQL Scale Scores, the meaningful score difference will be defined separately for adult, pediatric self-report, and pediatric parent proxy data. In each case, it will be defined as 0.5*SD, where SD is calculated from the baseline Scale Score data. This is in-line with the difference used on the selumetinib SPRINT study (Gross, 2018) and serves as the lower bound for minimally detectable change (Norman, 2003).

The number and percentage of participants with a change representing a clinically meaningful improvement (alongside associated 95% Clopper-Pearson CIs) will be presented for both thresholds.

In addition, within each age cohort, the proportion of participants reaching the anchor-based and distribution threshold will be analyzed in the subgroup of participants who have a confirmed objective response status versus the subgroup of those who do not. This is to further characterize the relationship between PRO improvement and the clinical endpoint of tumor response. The proportion of participants reaching the anchor-based (and distribution) thresholds will be summarized with 95% Clopper-Pearson CIs by confirmed objective response status, and the differences in the proportions of participants reaching the anchor-based/distribution thresholds between responders and non-responders will be tested using a Chi-Square test at a 5% significance level.

An MMRM will be fitted separately by age cohort and questionnaire type for each PedsQL score listed in Section 4.1.2.2.1. Change from baseline in PedsQL score will be the dependent variable, visit will be fitted as a fixed effect, and baseline PedsQL score used as a covariate. Covariance structures will be trialed in the order of decreasing complexity, with the first structure that converges without error used. The structures tested are as follows: Unstructured, Heterogenous Toeplitz, Heterogenous Autoregressive, Autoregressive, Heterogenous Compound Symmetry, Compound Symmetry, Variance Components. A Kenward-Roger approximation will be used for the degrees of freedom. This model will be fitted twice, once using only data through Cycle 13, and once including data from the entire study (timepoints with < 5 subjects will not be presented). EOT and unscheduled visits will be mapped to the closest scheduled tumor visit following the rules outlined in Table 13, and will only be included in the model if the participant has no other record at that timepoint. For the MMRM change from baseline tables, the least square (LS) Means estimates will be presented with associated standard errors, 95% CIs and p-values by visit. Plots of the LS Means estimates over time with associated 95% CI bands will be presented separately by age cohort and questionnaire type for both models.

Box and whiskers plots of change from baseline in total score will be presented by visit and confirmed objective response status for the adult cohort, pediatric cohort self-report, and pediatric cohort parent proxy separately.

A waterfall plot of change from baseline will be presented by visit for adult cohort, pediatric cohort self-report and pediatric cohort parent proxy separately for each PedsQL domain at Cycles 5 and 13 with a line representing the respective within-patient meaningful change threshold derived by anchor-based analysis. This plot will be color-coded by confirmed objective response status.

.

Additionally, a sensitivity analysis will be performed for each PedsQL score using the imputed data described in Section 5.5.1.1. The MMRM analysis described above will be performed with data up to Cycle 13. LS Means estimates, associated standard errors, 95% CIs and p-values will be presented by visit.

6.11.2.3 Change from Baseline in pain (assessed using the NRS-11 and PII)

6.11.2.3.1 NRS-11

All NRS-11 records will be listed by participant and visit.

The number and percentage of participants in each of the result categories (Appendix 1.3) at baseline will be summarized. Absolute values and change from baseline will be summarized overall and by confirmed objective response status (as defined with confirmed response rate in Section 4.1.1 using adjudicated values where appropriate), by visit and age cohort, the summaries presented will be number and percentage of participants with a non-missing score, mean, SD, median, minimum, and maximum values.

For NRS-11, a within-patient meaningful score change, or meaningful score difference (FDA, 2023) , will be derived by anchor-based analysis of the study's data as described in the meaningful change threshold analysis plan (Version 1.0, dated 12Sep2023). The primary focus of this will be on if such a relevant difference exists at Cycle 13.

Additionally the number and percentage of participants with a change representing a clinically meaningful improvement (alongside associated Clopper-Pearson confidence intervals) will be presented using change from baseline of ≤ -1 (This cut point has previously been researched and assessed as a meaningful change in pediatric populations (Hirschfeld, 2014)), representing a clinically meaningful improvement, will be presented, as well as change from baseline greater than the within-patient meaningful change threshold (i.e., derived by anchor-based analysis of the study's data in accordance with the MCT SAP; see Section 5.8 above) will be presented.

In addition, within each age cohort, the proportion of participants reaching the anchor-based and distribution threshold will be analyzed in the subgroup of participants who have a confirmed objective response status versus the subgroup of those who do not. This is to further characterize the relationship between PRO improvement and the clinical endpoint of tumor response. The proportion of participants reaching the anchor-based (and distribution) thresholds will be summarized with 95% Clopper-Pearson CIs by confirmed objective response status, and the differences in the proportions of participants reaching the anchor-based/distribution thresholds between responders and non-responders will be tested using a Chi-Square test at a 5% significance level.

An MMRM will be fit separately by age cohort with change from baseline in NRS-11 score as the dependent variable, visit as a fixed effect and baseline NRS-11 score as a covariate. The covariance structure will be decided in the manner described in Section 6.11.2.2, and a Kenward-Roger approximation will be used for the degrees of freedom. This model will be fitted twice, once using only data through Cycle 13, and once through Cycle 24. EOT visits will be mapped to the closest scheduled tumor visit following the rules outlined in Table 13, and will only be included in the model if the participant has no other record at that timepoint. LS Means estimates will be presented with associated standard errors, 95% CIs and p-values by visit. Plots of the LS Means estimates over time with associated 95% CI bands will be presented separately by age cohort for both models.

Box and whiskers plots of change from baseline will be presented by visit and confirmed objective response status for each age cohort separately.

A waterfall plot of change from baseline in NRS-11 will be presented by visit for adult cohort, pediatric cohort self-report and pediatric cohort parent proxy separately at Cycles 5 and 13, with a line representing the respective within-patient meaningful change threshold derived by anchor-based analysis. This plot will be color-coded by confirmed objective response status.

Additionally, a sensitivity analysis will be performed on the NRS-11 results using the imputed data described in Section 5.5.1.2. The MMRM analysis described above will be performed with data up to Cycle 13. LS Means estimates, associated standard errors, 95% CIs and p-values will be presented by visit.

6.11.2.3.2 PII

All PII records will be listed by participant, visit, and questionnaire type (self-report/parent proxy).

Absolute values and change from baseline will be summarized, overall and by confirmed objective response status (as defined with confirmed response rate in Section 4.1.1 using adjudicated values where appropriate), by visit and age cohort with self-report and parent proxy scores presented separately. The summaries presented will be number and percentage of participants with a non-missing score, mean, SD, median, minimum, and maximum values.

For PII a within-patient meaningful score change, or meaningful score difference (FDA, 2023) , will be derived by anchor-based analysis of the trial's data, as described in the meaningful change threshold analysis plan (Version 1.0, dated 12Sep2023). The primary focus of this will be on if such a relevant difference exists at cycle 13.

Additionally, the number and percentage of participants with a change representing a clinically meaningful improvement (alongside associated 95% Clopper-Pearson confidence intervals) will be presented using the change from baseline of <-0.5*SD (using SD from baseline)(This is in-line with the difference used for analyzing PII on the selumetinib SPRINT study (Gross, 2018).) as well as change from baseline greater than the within-patient meaningful change threshold (i.e., derived by anchor-based analysis of the trial's data in accordance with the MCT SAP; see Section 5.8 above) in separate analyses.

In addition, within each age cohort, the proportion of participants reaching the anchor-based and distribution threshold will be analyzed in the subgroup of participants who have a confirmed objective response status versus the subgroup of those who do not. This is to further characterize the relationship between PRO improvement and the clinical endpoint of tumor response. The proportion of participants reaching the anchor-based (and distribution) thresholds will be summarized with 95% Clopper-Pearson CIs by confirmed objective response status, and the differences in the proportions of participants reaching the anchor-based/distribution thresholds between responders and non-responders will be tested using a Chi-Square test at a 5% significance level.

An MMRM will be fit separately by age cohort and questionnaire type, with change from baseline in PII score as the dependent variable, visit as a fixed effect and baseline PII score as a covariate. The covariance structure will be decided in the manner described in Section 6.11.2.2, and a Kenward-Roger approximation will be used for the degrees of freedom. This model will be fitted twice, once using only data through Cycle 13, and once through Cycle 24. Unscheduled and ET visits will be mapped to the closest scheduled tumor visit following the rules outlined in Table 13, and will only be included in the model if the participant has no other record at that timepoint. LS Means estimates will be presented with associated standard errors, 95% CIs and p-values by visit. Plots of the LS Means estimates over time with associated 95% CI bands will be presented separately by age cohort and questionnaire type for both models.

Box and whiskers plots of change from baseline will be presented by visit for adult cohort, pediatric cohort self-report and pediatric cohort parent proxy separately.

A waterfall plot of change from baseline in PII will be presented by visit for adult cohort, pediatric cohort self-report and pediatric cohort parent proxy separately at cycles 5 and 13, with a line representing the respective within-patient meaningful change threshold derived by anchor-based analysis. This plot will be color-coded by confirmed objective response status.

Additionally, a sensitivity analysis will be performed on the PII results using the imputed data described in Section 5.5.1.2. The MMRM analysis described above will be performed with data up to Cycle 13. LS Means estimates, associated standard errors, 95% CIs and p-values will be presented by visit.

6.11.3 Exploratory Efficacy Analyses

6.11.3.1 Change from Baseline in physical function status (PROMIS)

All PROMIS records will be listed by participant, visit, and questionnaire type (self-report/parent proxy).

Actual values and change from baseline will be summarized by visit and confirmed objective response status for each PROMIS Scale, with self-report and parent proxy records summarized separately. The summaries presented will be the mean, SD, median, minimum, and maximum values.

Box and whiskers plots of change from baseline will be presented by visit for each scale separately.

6.11.3.2 Change in Functional Impairments

Change from baseline in functional impairments, measured according to Section 4.1.3.2, will be summarized by visit and age cohort.

In participants who have tumors that affect functional status at baseline, the change from baseline in muscle strength and range of motion in the PN affected areas, as described in Section 4.1.3.2, will be presented by visit.

In addition, for all physical functioning assessments, plots will be created for scores at each visit. This will be presented individually for each participant, as well as all being included in a whole population plot.

All recorded functional impairment data of any form will be listed.

6.11.3.2.1 Strength

For each of the individual strength assessments, to assess the change from baseline at a visit, a participant will need assessments in which all the following categories, as assessed at baseline, match at that visit:

- Side affected
- Position during assessment (sitting/supine/lateral decubitus)
- Muscle group assessed

Assessment of strength will be conducted only in the area affected by the target tumor. Measurements will be obtained using Medical Research Council (MRC) grading followed by quantitative assessments using the sponsor provided MicroFET2 dynamometer.

The MRC grading will be presented by each affected area summarizing the number and percent of participants in each category below:

0 – No movement is observed.
1 – Only a trace or flicker of movement is seen or felt in the muscle or fasciculations are observed in the muscle.
2 – Muscle can move only if the resistance of gravity is removed. As an example, the elbow can be fully flexed only if the arm is maintained in a horizontal plane.
3 – Muscle strength is further reduced such that the joint can be moved only against gravity with the examiner's resistance completely removed. As an example, the elbow can be moved from full extension to full flexion starting with the arm hanging down at the side.
4 – Muscle strength is reduced but muscle contraction can still move joint against resistance.
5 – Muscle contracts normally against full resistance.

The strength measured by the dynamometer by each area location will be summarized and the percent change from baseline using standard summary statistics will be summarized by age cohort and by visit.

6.11.3.2.2 Range of Motion

For the range of motion assessments, to assess the change from baseline at a visit, a participant will need assessments in which all three categories, as assessed at baseline, match at that visit:

- Joint assessed
- Motion tested
- Type (passive/active)

The range of motion for that location will be summarized and the percent change using standard summary statistics will be summarized by age cohort and by visit.

6.11.3.2.3 6-Minute Walk Test

Participants with a target PN causing airway or lower extremity motor dysfunction at baseline will be assessed for change in 6-minute walk test distance. Vital signs are performed prior to the start of the walk test. The object of this test is to walk as far as possible over 6 minutes. After completing the test, participant will be instructed to rest for 5 minutes, after which time vital signs are repeated. The total meters walked will be captured. Standard summary statistics of length walked and the pre-/post vitals will be summarized by each visit.

6.11.3.3 Time to Progression and Progression Free Survival

TTP and PFS will be analyzed and presented separately for each age cohort, and censoring rules will be applied as laid out in Section 4.1.3.6 and Section 4.1.3.7 respectively.

TTP and PFS will each be analyzed using KM methods; with median, lower, and upper quartile estimate and associated 95% CIs (using the Brookmeyer-Crowley method), the number of participants with events or who are censored, as well as reason for censoring, will be presented separately by age cohort. "NE" will be presented where estimates of such were not estimable. Additionally, KM plots will be presented by age cohort. These presentations will be produced for the FAS as a whole and for the subgroup of participants in the FAS who have an investigator determination of progressive disease at baseline.

The sensitivity analyses of TTP and PFS using REiNS criteria as outlined in Section 4.1.3.6.1 and Section 4.1.3.7.1 will also be summarized using the same method as described above.

6.11.3.4 Patient Global Impression of Change/Severity

All PGIS and PGIC data will be listed by participant and visit.

The absolute count and percentage of participants in each of the PGIS and PGIC categories, as defined in Appendix 1.1 and Appendix 1.2 respectively, will be summarized separately by visit and by confirmed objective response status.

6.11.3.5 Acceptability of Tablet Formation (Pediatric Oral Medicines Acceptability Questionnaire)

All P-OMAQ-P and P-OMAQ-C data will be listed by participant.

P-OMAQ-P and P-OMAQ-C individual question scores and total scores will be summarized separately, the summaries presented will be the mean, SD, median, minimum, and maximum values. For these summaries data from all visits in either phase will be pooled.

6.11.3.6 Change in Tumor Volume

Target tumor volume, change from baseline in tumor volume, and percentage change from baseline in tumor volume calculated as per Section 5.1.2 will be summarized by the number of participants, mean, SD, median, minimum, and maximum values by visit for each age cohort separately. This will be presented by averaging the volume across the two readers (unless response adjudication was performed) and reported separately for each reader (A vs B) from the BICR data, since the volumetric measurement was not triggered for adjudication.

An MMRM will be fit separately by age cohort with change from baseline in tumor volume as the dependent variable, visit as a fixed effect and baseline tumor volume as a covariate. The covariance structure will be decided in the manner described in Section 6.11.2.2, and a Kenward-Roger approximation will be used for the degrees of freedom. This model will be fitted for each cohort for all scheduled visits up until the first point where there are fewer than 5 participants with valid tumor volume measurements. EOT visits will be mapped to the closest scheduled tumor visit following the rules outlined in Table 12 and will only be included in the model if the participant has no other record at that timepoint. LS Means estimates will be presented with associated standard errors, 95% CIs and p-values by visit. Plots of the LS Means estimates over time with associated 95% CI bands will be presented separately by age cohort.

Best volumetric percentage change from baseline as well as the time, in months, to the best volumetric percentage change from baseline will be summarized by the number of participants, mean, SD, median, minimum, and maximum. This will be presented by averaging the volume across the two readers (unless response adjudication was performed) and reported separately for each reader (A vs B) from the BICR data. In addition, a waterfall plot of the best volumetric % change from baseline will be presented using the average (unless adjudication of response occurred), and by the two readers separately.

6.11.3.7 Disease Control Rate

The number and percentage of participants achieving disease control and accompanying DCR by the end of the treatment phase will be presented by age cohort, alongside the corresponding 2-sided Clopper-Pearson 95% CI.

6.11.3.8 Time to Response

Time to response will also be summarized for confirmed responders by age cohort and is defined as the time in months from the start of treatment to the date of first response that was subsequently confirmed. Summary statistics presented will include mean, SD, median, minimum, and maximum. The time to response will be summarized for the primary endpoint and by each of the two readers separately for the FAS.

6.12 Other Analyses

6.12.1 Safety Analyses

All safety analyses will be performed on the Safety Analysis Set as defined in Section 3.3.

6.12.1.1 Adverse Events

Only TEAEs will be presented in summary tables, as defined in Section 4.2.2.1. An overview of these TEAEs will be presented by age cohort summarizing the number and percentages of participants with TEAEs; SAEs; TEAEs of at least Grade 3 severity; study-drug related AEs; and AEs leading to dose modification, interruption, reduction, and discontinuation. The number and percentage of participants with a maximum recorded toxicity Grade 1-5 will also be presented.

Separate summaries by SOC and PT will be provided for the incidence of TEAEs (also for TEAEs by PT only), incidence of TEAEs by highest severity, incidence of SAEs, TEAEs of at least Grade 3 severity, study drug-related TEAEs, study drug related SAEs, study drug-related TEAEs of at least Grade 3 severity, study procedure related TEAEs, TEAEs leading to discontinuation, most common TEAEs related to study drug (occurring in ≥5% of overall cohort), TEAEs leading to dose reduction, TEAEs leading to dose interruption and AESIs – all as defined in Section 4.2.2.1. In addition, a summary of TEAEs by PT only will also be presented. All summaries will present absolute participant counts (n) and TEAE event counts. Percentages (%) will be calculated based on the number of participants at risk. Additionally, TEAEs will be summarized by cycle of first onset (After Cycle 6, cycles will be grouped either as Cycles 7-12, or Cycle 13 and later).

6.12.1.2 Clinical Laboratory Parameters

All clinical laboratory parameters listed in Section 4.2.2 will be grouped by category and summarized by visit, with hematology, chemistry, urinalysis, and serology summarized separately. Absolute values and, where applicable, change from baseline for quantitative parameters will be summarized by the number of participants, mean, SD, median, minimum, and maximum values. Qualitative parameters will be summarized by absolute counts (n) and percentages (%). Percentages will be based on the number of participants with data.

Additionally, shift tables based on the NCI-CTCAE Version 5.0 toxicity grading scale will be presented for each laboratory parameter between pre-treatment and worst case on-treatment.

For parameters presenting both hyper- and hypo- NCI-CTCAE toxicities pre-treatment, both changes to lowest and highest on-treatment values will be presented as shift tables. Additionally, an evaluation of drug-induced serious hepatotoxicity (eDISH) plots of max ALT versus max Bilirubin and max AST versus max Bilirubin will be produced by age cohort and overall. ALT/AST and Bilirubin will be displayed as multiple of the upper limit of normal (ULN) with lines denoting 2 x ULN for Bilirubin and 3 x ULN for ALT/AST.

6.12.1.3 Electrocardiogram (ECG) Parameters

Absolute value and change from baseline for all ECG parameters as listed in Section 4.2.3 will be summarized by visit as per Section 5.1.2. Additionally, summaries will be provided by post-baseline visit for the following categories:

- Number (%) of participants with QTcF between 450 and 480
- Number (%) of participants with QTcF >480
- Number (%) of participants with QTcF >500
- Number (%) of participants with QTcF change from baseline <30
- Number (%) of participants with QTcF change from baseline between 30 and 60
- Number (%) of participants with QTcF change from baseline >60.

These summaries will be presented as per Section 5.1.2.

6.12.1.4 Echocardiogram

Absolute values and change from baseline in LVEF will be summarized by visit as per Section 5.1.2. Additionally, shift tables based on the NCI-CTCAE Version 5.0 toxicity grading scale will be presented for LVEF between pre-treatment and worst case on-treatment.

6.12.1.5 Vital Signs

Absolute value and change from baseline for all vital signs parameters as listed in Section 4.2.5 will be summarized by visit. Continuous variables will be summarized by the number of participants, mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by absolute counts (n) and percentages (%). Percentages will be based on the number of participants with data.

6.12.1.6 Ophthalmic Assessments

Intraocular pressure will be summarized descriptively, for left and right eye separately, by visit and age cohort using the standard statistics laid out in Section 5.1. Visual acuity will be summarized categorically, for left and right eye separately, by absolute counts (n) and percentages (%). Percentages will be based on the number of participants with data.

6.12.1.7 Evaluate Key Safety by Formulation

AEs by SOC and PT, and PT only will be assessed split by formulation. The following groupings will be used:

- Pediatrics receiving tablets
- Pediatrics receiving capsules
- Combined pediatrics and adults receiving capsules
- All participants

The summaries provided will mirror the summary of all TEAEs presented by SOC and PT, and PT only as described in Section 6.12.1.1.

7. Changes to the Planned Analyses

7.1 Changes to the Analyses Described in the Study Protocol and Protocol Amendments

The key assumption of the sample size and associated clinically relevant ORR was the claimed ORR of 66% from selumetinib SPRINT trial. During study conduct, it became known to SpringWorks, that the actual ORR based on independent central review (ICR) from the SPRINT trial was 44%, c.f., Page 20 of "Multi-disciplinary Review and Evaluation NDA 213756, Koselugo (selumetinib), Version July 24, 2019" by the agency.

As per the meeting minutes of the Type C meeting conducted on 16 November 2022, the confirmed ORR will be assessed primarily by the exclusion of a minimum clinically relevant rate using the lower bound of the 95% CI. Section 6.11.1 of this SAP was updated to specify the type of confidence interval as a 95% Clopper-Pearson CI.

In addition, as per the meeting minutes of the Type C meeting conducted on 16 November 2022, assessment of tumor volume according to MRI will be done independently by two readers and, in the case of the two readings suggesting different tumor response status, an independent adjudicator will make the final determination of response category as per Section 4.1.1. Thus, further clarification has been added to the SAP using adjudication, but also sensitivity analyses by each of the two independent readers.

As suggested by the agency, a specific timepoint has been fixed for the analysis of PROs in this study, Cycle 13 will be considered as the time of interest for all PRO analysis. This update is described in Section 5.8 and the PRO-specific subsections of Sections 6.11.2 and 6.11.3. This is accompanied by the specification of a meaningful score difference used in analysis of secondary PROs. Additionally, an anchor-based analysis of PROs will be performed, this is detailed fully in a separate meaningful change threshold analysis plan (Version 1.0, dated 12 September 2023).

Following the Type C meeting on 16 November 2022, the planned analysis of historical controls data was expanded. A separate SAP will be produced as an addendum to this SAP to provide full details of the expanded historical controls data analysis.

As per FDA correspondences on 23 May 2023, the use of a balancing model, (e.g.: propensity score matching) was considered for the comparisons for the TTP and PFS endpoints between the pediatric cohort, and the NF1 natural history study (Akshintala, 2020) and placebo arm from the prior clinical trial of tipifarnib (Widemann, 2014).

8. References

Dombi, E. A.-H.-V. (2013). Recommendations for imaging tumor response in neurofibromatosis clinical trials. *Neurology*, 81(21 suppl 1):S33–S40.

FDA. (2023). *Draft Guidance 4 Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making.* Retrieved from https://www.fda.gov/media/166830/download

Gross, A. M. (2018). SPRINT: Phase II Study of the MEK 1/2 inhibitor selumetinib (AZD6244, ARRY142886) in children with neurofibromatosis type 1 (NF1) and inoperable plexiform neurofibromas (PN). American Society of Clinical Oncology (ASCO) Annual Meeting 2018.

Hirschfeld, G. W. (2014). Minimally clinically significant differences for adolescents with chronic pain-variability of ROC-based cut points. *The journal of pain, 15(1), 32–39. https://doi.org/10.1016/j.jpain.2013.09.006*.

Moertel, C. L. (2018). Trametinib in Pediatric Patients with Neurofibromatosis Type 1-Associated Plexiform Neurofibroma: A Phase I/II Study. *Joint Global Neurofibromatosis Conference.*

Murphy, R. (2018). On the use of one-sided statistical tests in biomedical research. Clinical and experimental pharmacology & physiology, 45(1), 109–114. https://doi.org/10.1111/1440-1681.12754.

Norman, G. R. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*, 582-592.

PROMIS. (2023, February 20). *PROMIS PHYSICAL FUNCTION SCORING USER MANUAL AND SCORING INSTRUCTIONS*. Retrieved from https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Physical_Function_User_Manual_and_Scoring_Instructions_03Oct2023.pdf

Shih, C. S. (2018). NF105: A Neurofibromatosis Clinical Trials Consortium (NFCTC) Phase II study of Cabozantinib (XL184) for neurofibromatosis type 1 (NF1) associated plexiform neurofibromas. *Joint Global Neurofibromatosis Conference.*

Weiss, B. (2021). NF106: A Neurofibromatosis Clinical Trials Consortium Phase II Trial of the MEK Inhibitor Mirdametinib (PD-0325901) in Adolescents and Adults With NF1-Related Plexiform Neurofibromas. Journal of clinical oncology : official journal of the American Society. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 39(7), 797–806. https://doi.org/10.1200/JCO.20.02220*.

Widemann, B. C. (2014). Phase 2 randomized, flexible crossover, double-blinded, placebo-controlled trial of the farnesyltransferase inhibitor tipifarnib in children and young adults with neurofibromatosis type 1 and progressive plexiform neurofibromas. *Neuro-Oncology, 16(5), 707-718. doi:10.1093/neuonc/nou004*.

9. Attachments and Appendices

Appendix 1 PRO Scoring

Appendix 1.1 PGIS

Please choose the response below that best describes the severity of your tumor-related symptoms over the past week.

□ None (0)
□ Mild (1)
□ Moderate (2)
□ Severe (3)

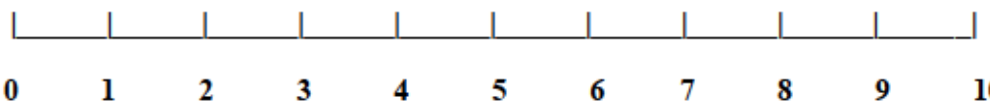
Appendix 1.2 PGIC

Please choose the response below that best describes the overall change in your general state of health since you started taking your study medication.

□ Very much Better (1)
□ Moderately Better (2)
□ A Little Better (3)
□ No Change (4)
□ A Little Worse (5)
□ Moderately Worse (6)
□ Very much Worse (7)

Appendix 1.3 NRS-11

Please select the one number that best describes your target tumor pain at its worst during the past 24 hours.



0 1 2 3 4 5 6 7 8 9 10

No pain **Worst pain you can imagine**

Appendix 1.4 PII

Below you will find a list of questions about you and your pain. Please answer each question by selecting a number between 0 and 6.

Please note that we are asking about your pain during the past 24 hours.

In the past 24 hours:

		Not at all			Some			Completely
1.	Has your pain made it difficult for you to pay attention (focus at work or school, complete tasks)?	0	1	2	3	4	5	6
2.	Has your pain made it difficult for you to do activities outside of work (leisure activities)?	0	1	2	3	4	5	6
3.	Has your pain made it difficult for you to spend time with friends and family members?	0	1	2	3	4	5	6
4.	Has your pain affected your mood?	0	1	2	3	4	5	6
5.	Has your pain affected your ability to do physical activities (like run, walk upstairs, play sports, do chores)?	0	1	2	3	4	5	6
6.	Has your pain affected your sleep?	0	1	2	3	4	5	6

Appendix 1.5 PROMIS score mapping

Adult v2.0 – Physical Function 8b		
Short Form conversion Table		
Raw summed score	T-score	SE*
8	20.3	3.7
9	23.9	2.5
10	26.0	2.2
11	27.5	2.1
12	28.8	2.0
13	29.8	1.9
14	30.8	1.8
15	31.7	1.8
16	32.5	1.7
17	33.2	1.7
18	34.0	1.7
19	34.7	1.7
20	35.4	1.6
21	36.1	1.6
22	36.7	1.6
23	37.4	1.6
24	38.1	1.6
25	38.8	1.6
26	39.5	1.6
27	40.1	1.6
28	40.8	1.6
29	41.6	1.7
30	42.3	1.7
31	43.1	1.7
32	43.9	1.7
33	44.7	1.8
34	45.7	1.8
35	46.7	1.9
36	47.8	2.1
37	49.2	2.3
38	50.8	2.6
39	53.0	3.0
40	60.1	5.9
*SE – Standard Error on T-score metric		

Pediatric v2.0 – Upper Extremity 8a – Self-Report		
Short Form Conversion Table		
Raw summed score	T-score	SE*
8	10	4
9	12	4
10	14	3
11	15	3
12	17	3
13	18	3
14	19	3
15	20	3
16	21	3
17	22	3
18	23	3
19	24	3
20	24	3
21	25	3
22	26	3
23	27	3
24	28	3
25	29	3
26	30	3
27	31	3
28	32	3
29	33	3
30	34	3
31	35	4
32	37	4
33	39	4
34	40	4
35	42	5
36	45	5
37	49	5
38	57	7
*SE – Standard Error on T-score metric		

Pediatric v2.0 – Mobility 8a – Self-Report		
Short Form Conversion Table		
Raw summed score	T-score	SE*
8	14	4
9	17	3
10	19	3
11	20	3
12	21	3
13	22	3
14	23	3
15	24	3
16	25	3
17	26	3
18	27	3
19	28	3
20	28	3
21	29	3
22	30	3
23	31	3
24	32	3
25	33	3
26	33	3
27	34	3
28	35	3
29	36	3
30	37	3
31	38	3
32	39	3
33	40	3
34	41	3
35	43	4
36	45	4
37	46	4
38	48	4
39	52	5
40	59	7
**SE – Standard Error on T-score metric*		

Pediatric v2.0 – Upper Extremities 8a - Parent Proxy		
Short Form Conversion Table		
Raw summed score	T-score	SE*
8	13	3
9	16	3
10	17	3
11	18	2
12	19	2
13	20	2
14	21	2
15	22	2
16	22	2
17	23	2
18	24	2
19	24	2
20	25	2
21	25	2
22	26	2
23	26	2
24	27	2
25	28	2
26	28	2
27	29	2
28	30	2
29	30	2
30	31	2
31	32	2
32	33	2
33	34	3
34	35	3
35	37	3
36	38	4
37	40	4
38	42	4
39	45	5
40	55	8
*SE – Standard Error on T-score metric		

Pediatric v2.0 – Mobility 8a – Parent Proxy		
Short Form Conversion Table		
Raw summed score	T-score	SE*
8	14	4
9	17	3
10	20	3
11	21	3
12	22	3
13	23	2
14	24	2
15	25	2
16	26	2
17	27	2
18	27	2
19	28	2
20	29	2
21	29	2
22	30	2
23	31	2
24	31	2
25	32	2
26	33	2
27	33	2
28	34	2
29	35	2
30	35	2
31	36	2
32	37	2
33	38	3
34	39	3
35	40	3
36	42	4
37	43	4
38	45	4
39	48	4
40	56	7
**SE – Standard Error on T-score metric*		

Appendix 2 Rash Composite Term PT List

The following PTs will be considered as a Rash AESI event:

- Dermatitis
- Dermatitis acneiform
- Dermatitis exfoliative
- Dermatitis exfoliative generalized
- Rash
- Rash maculopapular
- Rash erythematous
- Rash macular
- Rash popular
- Rash pustular
- Rash pruritic
- Exfoliative rash
- Papule
- Skin exfoliation
- Rash maculovesicular
- Rash morbilliform
- Rash papulosquamous
- Rash vesicular
- Eyelid rash
- Nodular rash

MEK-NF-201 SAP v1.0.0

Final Audit Report 2023-11-03

Created:	2023-11-03 (Greenwich Mean Time)
By:	[redacted]
Status:	Signed
Transaction ID:	CBJCHBCAABAA7hJCHnYK4rze2T3XvqmYUoRwSSfE133n
Number of Documents:	1
Document page count:	61
Number of supporting files:	0
Supporting files page count:	0

[redacted]