STATISTICAL ANALYSIS PLAN FOR PATIENT-REPORTED OUTCOMES

An Addendum to NIR-DT-301 Statistical Analysis Plan (PRO Addendum)

Protocol NIR-DT-301

Protocol Number:	NIR-DT-301
Protocol Version and Date:	Amendment 5: 09 February 2021 Amendment 4: 07 July 2020 (not released to sites) Amendment 3: 27 January 2020 Amendment 2: 14 October 2019 Amendment 1: 28 November 2018 Original: 03 August 2018
Name of Test Drug:	PF-03084014 (nirogacestat)
Phase:	Phase 3
Methodology:	Randomized, Double-Blind, Placebo-Controlled
Sponsor:	SpringWorks Therapeutics 100 Washington Blvd. Stamford, CT 06902 United States
Sponsor Representative:	[redacted] Sponsor Biostatistician
Analysis Plan Date:	05 Apr 2022
Analysis Plan Version:	Final Version 2

APPROVAL SIGNATURE PAGE

Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Trial of Nirogacestat Versus Placebo in Adult Patients with Progressing Desmoid Tumors/Aggressive Fibromatosis (DT/AF)
Sponsor:	SpringWorks Therapeutics 100 Washington Blvd Stamford, CT 06902 United States
Protocol Number:	NIR-DT-301
Document Date / Version:	05 Apr 2022 / Final Version 2

IQVIA Author:

[redacted]
IQVIA

Signature:_____

Date:_____

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol, clinical development plan, and all applicable regulatory guidances and guidelines.

I have discussed any questions I have regarding the contents of this document with the biostatistical author.

I also understand that any subsequent changes to the planned statistical analyses, as described herein, may have a regulatory impact and/or result in timeline adjustments. All changes to the planned analyses will be described in the clinical study report.

Sponsor Signatory:

[redacted]
Sponsor Biostatistician
SpringWorks Therapeutics

Signature:_____

Date:_____

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	14Oct2021	[redacted]	Not Applicable – First Version
2.0	12Jan2022	[redacted]	Clarifications added in response to FDA comments and to enhance readability. Specifically, the following key modifications were performed: • statistical test was added for the proportion of patients with improvement at Cycle 10 in section 5.6 • clarification on the seed used in the multiple imputation was added in section 5.5.2 • clarified that the formal statistical testing on PRO-related endpoints is conducted at the 1-sided, 0.025 level of significance. • Supplementary analysis added where GODDESS Total Symptom Score is derived excluding item 2 (dull pain) and item 3 (shooting pain) • Proportion of participants with improvement in BPI-SF API score at Cycle 10 has been removed from the hierarchical testing of secondary endpoints and added as an exploratory endpoint.

TABLE OF CONTENTS

Section **Page**

TABLES INCLUDED IN THE TEXT

Page

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
AP	Appetite loss
API	Average Pain Intensity
BPI-SF	Brief Pain Inventory Short Form
CDF	Cumulative Distribution Function
CF	Cognitive functioning
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
CO	Constipation
CSR	Clinical Study Report
DI	Diarrhea
DT/AF	Desmoid Tumors/Aggressive Fibromatosis
DTIS	Desmoid Tumor Impact Scale
DTRF	Desmoid Tumor Research Foundation
DTSS	Desmoid Tumor Symptom Scale
DY	Dyspnea
EF	Emotional functioning
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
FA	Fatigue
GHS/QoL	Global health status/Quality of life
GODDESS	GOunder/DTRF DEsmoid Symptom/Impact Scale
HR	Hazard Ratio
ITT	Intention-to-treat
LS	Least squares
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MMRM	Mixed model repeated measures
MNAR	Missing not at random
NDA	New Drug Application
NRS	Numeric rating scale
NV	Nausea and vomiting
OLE	Open-label extension
OR	Odds ratio
PA	Pain
PF	Physical functioning
PGIC	Patient Global Impression of Change

Abbreviation	Definition
PGIS	Patient Global Impression of Severity
PMM	Pattern mixture model
PRO	Patient-reported outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
PROMIS PF	Patient-Reported Outcomes Measurement Information System Physical Function
QoL	Quality of life
REML	restricted maximum likelihood
RF	Role functioning
SAP	Statistical analysis plan
SF	Social functioning
SL	Insomnia

1. INTRODUCTION

This document is a patient-reported outcome (PRO) data analysis addendum to the statistical analysis plan (SAP) of NIR-DT-301, a Phase 3, randomized, double-blind, placebo-controlled study that compares the efficacy, safety, and tolerability of nirogacestat and placebo in adult participants with progressing Desmoid Tumors/Aggressive Fibromatosis (DT/AF).

To evaluate desmoid tumor symptoms and impacts in patients with progressing DT/AF, the following PROs were collected in the study NIR-DT-301:

1. GOunder/Desmoid Tumor Research Foundation (DTRF) DEsmoid Symptom/Impact Scale (GODDESS)

2. Brief Pain Inventory Short form (BPI-SF)

3. European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30)

4. Patient-Reported Outcomes Measurement Information System Physical Function (PROMIS PF) short form 10a + 3 additional items from PROMIS item banks

5. Patient Global Impression of Severity (PGIS)

6. Patient Global Impression of Change (PGIC)

This addendum (PRO Addendum) is designed to outline the methods used in the analysis of the PRO data. Populations for analysis, data handling rules, statistical methods, and formats for data representation follow those specified in the NIR-DT-301 SAP, unless otherwise noted. The statistical analyses and summary tabulations described in this addendum will provide the basis for the results sections of the PRO analysis of the Clinical Study Report (CSR) for this trial.

2. INFORMATION FROM THE STUDY PROTOCOL

2.1. Study Design and Objectives

This is a multi-center, randomized, double-blind, placebo-controlled, parallel group, event-driven, Phase 3 study to compare the efficacy, safety, and tolerability of nirogacestat and placebo in adult participants with progressing DT/AF.

The primary objective of this study is:

- To determine the efficacy (as defined by progression-free survival) of nirogacestat in adult participants with progressing DT/AF

One of the secondary objectives of this study, relating to the PRO data, is:

- To evaluate desmoid tumor symptoms and impacts using the following PROs:
    - GODDESS
    - BPI-SF
    - EORTC QLQ-C30

Exploratory objectives of this study, relating to the PRO data, are:

- To evaluate desmoid tumor symptoms and impacts using the following PROs:
    - PROMIS PF short form 10a and 3 additional items from PROMIS item banks
    - PGIS
    - PGIC

Additional analyses where PRO based outcomes are correlated with clinical data are documented in the Main SAP (Section 4.3.3.9 Change in Symptoms by Exposure and Change in Tumor Size/Volume).

2.1.1. Synopsis of PRO Data Collection

This study will consist of two phases: the double-blind phase and the optional, open-label extension (OLE) phase.

The following PRO assessments will be conducted during the double-blind phase:

- Screening PRO assessment:
    - On Day 1 of the screening visit, participants will receive training by the site staff on how to use the home ePRO device and will include a practice questionnaire to be completed by the participant prior to leaving the site.

- Participants will then begin the screening PROs assessments that same day (more details outlined in Table 1).
- The PGIC is intentionally omitted from the screening PRO assessments.

- Baseline PRO assessment:
  - The baseline PRO assessments will begin 7 days prior to the Cycle 1 Day 1 visit (more details outlined in Table 1).
  - The PGIC is intentionally omitted from the baseline PRO assessments.
- Monthly PRO assessments are required throughout the study (Cycle 2, 3, 4 and on).

The following PRO assessments will be conducted during the OLE phase:

- Monthly PRO assessments are required for the first year (Cycle 2-12).
- Quarterly PRO assessments are required after the first year (Cycle 13, 16, 19 and on).

2.1.2. PRO Data Collection Procedures

The schedules of PRO assessments, as outlined in the study protocol, are provided in Table 1 (double-blind phase) and Table 2 (OLE phase).

Table 1: Double-Blind Phase: PRO Assessment Administration Schedule (Reproduced from Table 1-1 of main SAP)

PROs	Screening PRO Assessments First 7 days of screening period							Baseline PRO Assessments 7 day prior to baseline visit							Base-line Visit	Monthly PRO Assessments 7 days prior to Cycle X/FU (X = Cycles 2, 3, 4 & on)							Cycle X/FU Visit
	d 1	d 2	d 3	d 4	d 5	d 6	d 7	d -7	d -6	d -5	d -4	d -3	d -2	d -1		d -7	d -6	d -5	d -4	d -3	d -2	d -1	
GODDESS (symptom scale)	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	
BPI-SF	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	
PROMIS							X							X								X	
GODDESS (impact scale)							X							X								X	
EORTC QLQ-C30							X							X								X	
PGIS							X							X								X	
PGIC																						X	

BPI = brief pain inventory; d = day; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30; FU = follow-up; GODDESS = GOunder/DTRF DEsmoid Symptom/Impact Scale; PGIC = patient global impression of change; PGIS = patient global impression of severity; PRO = patient-reported outcome; PROMIS PF= Patient-Reported Outcomes Measurement Information System Physical Function

* PROMIS PF short form 10a plus 3 additional items from the PROMIS item bank.

Table 2: OLE Phase: PRO Assessment Administration Schedule (Reproduced from Table 1-2 of main SAP)

PROs	Monthly PRO Assessments *7 days prior to Cycle X (X = Cycles 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12)*							Cycle X	Quarterly PRO Assessments *7 day prior to Cycle X (X = Cycles 13, 16, 19, 22 & on)*							Cycle X	Follow-Up PRO Assessments *7 days prior to the FU visit*							FU Visit
	d -7	d -6	d -5	d -4	d -3	d -2	d -1		d -7	d -6	d -5	d -4	d -3	d -2	d -1		d -7	d -6	d -5	d -4	d -3	d -2	d -1	
GODDESS (symptom scale)	X	X	X	X	X	X	X		X	X	X	X	X	X	X		X	X	X	X	X	X	X	
BPI-SF	X	X	X	X	X	X	X		X	X	X	X	X	X	X		X	X	X	X	X	X	X	
PROMIS*							X								X									X	
GODDESS (impact scale)							X								X									X	
EORTC QLQ-C30							X								X									X	
PGIS							X								X									X	
PGIC							X								X									X	

BPI = brief pain inventory; d = day; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30; FU = follow-up; GODDESS = GOunder/DTRF DEsmoid Symptom/Impact Scale; PGIS = patient global impression of severity; PRO = patient-reported outcome; PROMIS PF = Patient-Reported Outcomes Measurement Information System Physical Function

* PROMIS PF short form 10a plus 3 additional items from the PROMIS item bank.

2.1.3. Efficacy Endpoints

The primary efficacy endpoint of the study is progression free survival. It is documented in the Main SAP for DeFi clinical data analyses.

Secondary efficacy endpoints related to the PRO are as follows:

- Mean change from baseline at Cycle 10 in BPI-SF Average Pain Intensity (API) score
- Mean change from baseline at Cycle 10 in Desmoid Tumor Symptom Scale (DTSS) Total Symptom Score
- Mean change from baseline at Cycle 10 in Desmoid Tumor Impact Scale (DTIS) Physical Functioning Domain Score
- Mean change from baseline at Cycle 10 in EORTC QLQ-C30 Global health status/Quality of life (GHS/QoL)
- Mean change from baseline at Cycle 10 during the double-phase period in EORTC QLQ-C30 Physical Functioning
- Mean change from baseline at Cycle 10 during the double-phase period in EORTC QLQ-C30 Role Functioning

Exploratory endpoints related to PRO are as follows:

- Mean change from baseline at Cycle 10 in DTSS Pain Domain Score
- Mean change from baseline at Cycle 10 in EORTC QLQ-C30 pain
- Mean change from baseline at Cycle 10 in PROMIS PF10a (sum score) and PF13 (sum score)
- Proportion of participants with improvement at Cycle 10 in DTSS Total Symptom Score and DTSS Pain Domain Score
- Proportion of participants with improvement at Cycle 10 in DTIS Physical Functioning Domain Score
- Proportion of participants with improvement at Cycle 10 in BPI-SF API score
- Time to first DT symptom improvement (using DTSS Total Symptom)
- Time to pain response (using DTSS Pain Domain Score)
- Time to first control of pain symptoms (using BPI-SF API)
- Time to pain response (using BPI-SF API)

2.2. PRO Measures

2.2.1. The GOunder/DTRF DEsmoid Symptom/Impact Scale (GODDESS)

The GODDESS tool was developed to measures signs and symptoms of desmoid tumors and their impact on patients' lives, using two separate scales – the DTSS and the DTIS. The DTSS consists of 11 items assessing the severity of key signs and symptoms, including pain, fatigue, swelling, muscle weakness, difficulty moving, and tumor location-specific signs/symptoms. The DTIS includes 17 items that measure the impact of the symptoms of DTSS on daily life. The items from the GODDESS tool are presented in Appendix 8.1.

2.2.1.1. Desmoid Tumor Symptom Scale (DTSS)

Items of DTSS are evaluated on an 11-point, numeric rating scale (NRS) from 0 to 10 to measure severity from "none" to "as bad as you can imagine," with a 24-hour recall period. The DTSS will also have daily total scores and weekly average scores computed for items 1-7 and 9-11. Item 8 refers to the location of the tumor and will not be includied in the scoring. Weekly average scores will be computed for the 7-day periods at Screening, Baseline, and the 7 days preceding each Cycle using the mean of the period's total daily scores but only if there are four or more days of assessments within that period.

The following daily scores will be derived for the DTSS for each day of completion:

- Total Symptom Score:
  - Mean of Pain items (Items 1-3) as a single score, then a mean of this with items 4-7)
    - $\frac{((Item\,1+Item\,2+Item\,3)/3)+Item\,4+Item\,5+Item\,6+Item\,7}{5}$
- Total Symptom Score-5 Items:
  - Mean of items 1, 4, 5, 6 and 7
- Total Symptom Score -Average
  - Mean of Items 1-7
- Pain Domain Score
  - Mean of Items 1-3
- Extra-abdominal Symptoms Domain Score
  - Mean of Items 5-7
- Abdominal Symptoms Domain Score
  - Mean of Items 9-11

In the case of missing item-level data (not expected, as all items were administered electronically, and mandatory) participants will have a missing daily value.

Weekly summary scores will be created by averaging the daily scores over the 7 days period prior to each visit. A weekly score will be derived only if 4 or more out of 7 days period have non-missing scores. The weekly summary score will be used in analyses. If no weekly summary score is calculable, the participant will have data considered as missing at that visit.

Higher scores represent worse symptom severity. There is no total score for the DTSS.

2.2.1.2. Desmoid Tumor Impact Scale (DTIS)

Items of DTIS are evaluated either on an 11-point NRS to measure severity, or a 5-point Likert scale ranging from "none of the time" to "all of the time," to measure frequency with a 7-day recall period.

The following scores will be derived for the DTIS:

- Physical Function Domain Score
    - Mean of: Item 01 Moving, Item 02 Reaching (Freq), Item 06 Vigorous Activity, Item 07 Moderate Activity, Item 08 Accomplished Less
- Sleep Domain Score
    - Mean of: Item 03 Falling Asleep, Item 04 Comfortable in Bed, Item 05 Staying Asleep
- Emotion Domain Score
    - Mean of: Item 12 Fear Tests, Item 13 Fear Growth/Reoccurrence, Item 14 Hopelessness, Item 15 Anger, Item 16 Anxiety, Item 17 Frustration

In the case of missing item-level data (not expected as all items were electronically administered, and mandatory) participants will have a missing domain value.

Higher scores represent worse impact severity.

2.2.2. BPI Short Form (BPI-SF)

The BPI-SF is a measurement tool for assessing clinical pain and allows patients to rate the severity of their pain and the degree to which their pain interferes with common dimensions of feeling and function. The BPI-SF consists of 9 questions and utilizes an 11-point NRS from 0-10 (0 being "no pain" and 10 being "highest pain level") with a 24-hour recall period. BPI-SF is presented in Appendix 8.2.

The BPI-SF assesses pain at its "worst," "least," "average," and "now" (current pain) through four **pain severity** items (items 3, 4, 5 and 6) rated on 0–10 scale, with 0 = "no pain" and 10 = "pain as bad as you can imagine". To measure the average pain severity, a Pain Severity

Subscale will be derived both daily and weekly. Daily score will be calculated as the average of the 4 items responses for each day, and the weekly score will be calculated as the average of the daily scores. All 4 questions must be answered for the daily average pain severity score to be computed. A weekly score will be derived only if 4 or more days out of 7 days period have non-missing scores.

The BPI-SF also measures how much pain has interfered with seven daily activities, including general activity, walking, work, mood, enjoyment of life, relations with others, and sleep through seven **pain interference** items (items 9A to 9G of question 9) rated on 0–10 scales (with 0 = "no interference" and 10 = "interferes completely"). To measure average pain interference, a Pain Interference Subscale will be derived both daily and weekly. Daily score will be calculated as the average of the 7 items responses for each day provided that 4 out of the 7 items have non-missing responses. The weekly score will be the average of the daily scores over the 7 days period prior to each visit. A weekly score will be derived only if 4 or more days out of 7 days period have non-missing scores.

In addition, Average Pain Intensity (API) will be calculated as the average of the daily BPI-SF Item 3 "Worst Pain in Past 24 hours" over the 7 day period prior to each visit. API will be derived only if 4 to 7 days have non-missing scores.

2.2.3. EORTC QLQ-C30

The EORTC QLQ-C30 is a quality of life (QoL) questionnaire used for assessing the health-related quality of life of cancer patients participating in international clinical trials. The items from the questionnaire are presented in Appendix 8.3.

EORTC QLQ-C30 version 3.0 will be used in this study, with a 7-day recall period. It consists of 30 questions, with all items scored 1 ("not at all") to 4 ("very much") except for the 2 items contributing to the global health status/QoL, which are scored 1 ("very poor") to 7 ("excellent"). The recall period for each question is "during the past week". The instrument yields the following scales:

- GHS/QoL scale with 2 items
- 5 Functional scales :
    - Physical functioning (PF) with 5 items
    - Role functioning (RF) with 2 items
    - Emotional functioning (EF) with 4 items
    - Cognitive functioning (CF) with 2 items
    - Social functioning (SF) with 2 items
- 3 Symptom scales / single items :
    - Fatigue (FA) with 3 items

- Nausea and vomiting (NV) with 2 items
- Pain (PA) with 2 items
- Dyspnea (DY) , insomnia (SL), appetite loss (AP), constipation (CO), diarrhea (DI) with 1 item each

- Financial impact of disease with 1 item .

The EORTC QLQ-C30 scale scores will be calculated using the EORTC QLQ-C30 Scoring Manual (Fayers, et al., 2001). Each of the scales will have a raw mean score computed as long as at least 50% of the scales' questions have responses. The raw score will be used in a scoring formula described below that will transform the raw score linearly onto a score of 0 to 100. Details on computing raw scores and transforming them are found in Appendix 8.3.

A higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms, i.e., a high score for a functional scale (PF, RF, EF, CF, SF) represents a high/healthy level of functioning, a high score for the GHS/QoL and FI represents a high QoL or less impact, but a high score for a symptom scale (FA, NV, PA, DY, SL, AP, CO, DI) represents a high level of symptomatology.

The financial impact domain will not be included in this PRO Addendum to the DeFi Main SAP analysis as it is not related to symptoms, functioning or health-related quality of life, and therefore is not relevant for the context of the clinical trial.

2.2.4. PROMIS PF Short Form 10a + 3 Additional Items from PROMIS Item Bank

The PROMIS PF 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.

The PROMIS PF short form 10a version 2.0 will be used in this study with a 7-day recall period. The items from the questionnaire are presented in Appendix 8.4. This PRO assessment consists of 10 questions and was constructed with a focus on representing the range of the trait and the content of the item bank, as well as mapping the questions in the instrument to qualitative evidence of the physical function concepts important to patients. A total sum score will be computed using all 10 items. A T-score and standard error of the T-score associated with the total score will be assigned based on the short form conversion table accompanying the scale and found in Appendix 8.4. A higher PROMIS T-score represents more of the concept being measured. For positively-worded concepts like physical function, a T-score of 60 is one SD better than average, and a participant with a T-score of 40 is one SD worse than the average (PROMIS, s.f.).

To supplement the PROMIS PF short form 10a, 3 additional questions representing other elements of physical function found to be important to patients were selected from the PROMIS Physical Function, Upper Extremity, and Ability to Participate item banks.

- Additional item 1: "Are you able to bend or twist your back",

- Additional item 2: "Are you able to reach into a high cupboard",
- Additional item 3: "I have trouble doing my regular daily work around the house".

A total score for the PROMIS-PF 10a with these three additional items has been supported by further psychometric evaluation. Therefore, a PROMIS PF13 score will be derived as a sum score of all 13 items, provided that there are no missing items.

Due to limitations in available translations in some countries, the additional 3 questions were not available at all sites. As such, this analysis only applies to those participants with available data at baseline including PROMIS PF short for 10a plus the 3 additional questions.

2.2.5. Patient Global Impression of Severity (PGIS)

The PGIS is a single-item scale that evaluates the participant's perception of the overall severity of their desmoid related symptoms over the past week on a 4-point scale ranging from "none" to "severe.". The PGIS has a 7-day recall period. The scale is presented in Appendix 8.5.

2.2.6. Patient Global Impression of Change (PGIC)

The PGIC is a single-item scale that evaluates the participant's perception of the 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 has a 7-day recall period. The scale is presented in Appendix 8.6.

3. SUBJECT POPULATION

3.1. Population Definitions

The Intent-to-Treat (ITT) Population will be evaluated and used for presentation and analysis of the PRO data. The ITT Population will consist of all participants who are enrolled and randomized to study treatment (nirogacestat or placebo). Participants will be analyzed according to the treatment to which they were randomized and the strata to which they have been assigned. Participants who were randomized but did not subsequently go on to receive study treatment are included in the ITT population.

4. GENERAL STATISTICAL METHODS

4.1. General Methods

Continuous variables will be described by the number of participants, mean, median, standard deviation, minimum, and maximum. Categorical variables will be described by the number and percentage of participants within each category (with a category for missing data). Time-to-event data will be summarized using Kaplan-Meier methodology using 25th, 50th (median), and 75th percentiles with associated 2-sided 95% confidence intervals (CIs), as well as percentage of censored observations. Additional conventions for presentation of study data are laid out in the main body of the NIR-DT-301 SAP and will be applied for the PRO analyses as appropriate.

Formal statistical hypothesis testing on the PRO-related secondary endpoints listed in section 2.1.3, for the purpose of New Drug Application (NDA), will be conducted at the 1-sided, 0.025 level of significance.

Specific details on analysis of selected PRO measures are laid out in Section 5. All additional PRO measures, as well as all data collected after crossover to nirogacestat in the OLE phase of the study will be analyzed as described in an exploratory PRO statistical analysis plan and reported separately.

4.2. Baseline Definitions

For all PRO analyses unless specified, baseline for the double-blind phase will be defined as the most recent measurement prior to the first administration of study drug.

For weekly summary scores of the GODDESS and the BPI-SF, the baseline score will be set as the weekly average of the Baseline period (e.g., study days -7 through -1) if there are 4 or more days of assessments during that Baseline period. If there are less than 4 days of assessments during Baseline period, the baseline weekly score will be set to weekly average of the Screening visits' assessments, provided that there are 4 or more days of assessments during that period. Otherwise, it will be set as missing.

Baseline assessments of PGIC do not exist, as PGIC assessments are post-baseline only.

4.3. Adjustments for Covariates

In general, the stratification factor (as reported in randomization) will be included in the analysis of all PRO endpoints. In addition, longitudinal change from baseline models will account for the baseline scores.

4.4. Multiple Comparisons/Multiplicity

Multiplicity will be controlled via hierarchical testing method for the primary and secondary endpoints as defined in the main study SAP.

4.5. Subgroups

The longitudinal analysis of change from baseline in PRO endpoints and the time to pain response will be examined in selected subgroups as listed in Table 3.

Table 3 List of subgroups

Subgroup Name	**Subgroup Levels**
Primary tumor location as reported in randomization	Intra-abdominal
	Extra-abdominal
Baseline worst pain score	Uncontrolled (API > 4)
	Controlled (API ≤ 4)
Tumor focality	Multi-Focal Disease
	Single Tumor
Desmoid tumor treatment status	Treatment naïve, measurably progressing DT/AF that is deemed not amenable to surgery
	Recurrent, measurably progressing DT/AF following at least one line of therapy
	Refractory, measurably progressing DT/AF following at least one line of therapy
Sex	Male
	Female
Race	White
	Non-White
Geographic region	North America Sites
	Rest of World Sites
Age group	First age quartile
	Second age quartile
	Third age quartile
	Fourth age quartile
Ethnicity	Hispanic or Latino
	Not Hispanic or Latino

4.6. Withdrawals, Dropouts, Loss to Follow-up

Subjects who are withdrawn or discontinue from the study will not be replaced.

4.7. Missing, Unused, and Spurious Data

In case of missing items, the scale and total scores will be calculated as indicated in the scoring manual for the PRO instrument for the existing instruments. For GODDESS all items for a given total or domain score should be present in order to create the total or domain score.

Weekly scores (GODDESS and BPI-SF) will be calculated only if there are 4 or more days of assessments with non-missing data within the respective week.

It is anticipated that the great majority of missing data in this study will have a monotone pattern, meaning that once a participant has missing data at one visit, data will be missing at all subsequent visits. There may be some small amount of intermittent (non-monotone) missing data (when participant skips intermediate visits but returns for evaluations at subsequent visits).

Missing PRO data will be assumed to be missing at random (MAR) in general and will be analyzed using mixed model repeated measures (MMRM) analysis when appropriate. To address the possibility that missing data may not be MAR, sensitivity analysis with pattern mixture model (PMM) will be conducted.

4.8. Visit Windows

In line with the clinical SAP, no windowing conventions will be applied for the analysis of the PRO data. All data will be tabulated per the evaluation visit as recorded on the eCRF. In the case of multiple observations at a specific visit, the first observation will be used.

5. STUDY ANALYSES

5.1. Participant Disposition

The participant disposition by treatment arm for all PRO assessment timepoints in the double-blind phase will be provided:

- The number of participants with PRO assessment expected
- The number and % of participants with PRO assessment not expected due to progression
- The number and % of participants with PRO assessment not expected due to death
- The number and % of participants with PRO assessment not expected due to other reasons

The participant disposition by treatment arm per timepoint will also be provided graphically using bar charts.

A participant is expected to complete the PRO assessment as long as he/she is still alive and have not discontinued the study.

5.2. PRO Completion

PRO completion for all instruments will be examined at each timepoint in the double blind phase. Specifically, the following will be examined:

- Unadjusted completion rate at each timepoint will be calculated as the number of participants meeting at least the minimum requirements for scoring of each instrument divided by the number of participants in the ITT population.
- Adjusted completion rate at each timepoint will be calculated among participants who are expected to have PRO assessments.

For the adjusted completion rate described above, the following will be provided:

- The number and % of participants with all questions completed
- The number and % of participants meeting at least the minimum requirements for scoring of the instrument (see section 2.2
- The number and % of participants with at least one question completed (for multi-item instruments)

The completion rates (both adjusted and unadjusted) by treatment arm per timepoint will also be provided graphically: This will be displayed using grouped bar charts with visit on the x-axis and percent completion on the y-axis. Groups will be defined by treatment arm.

5.3. Exploration of Missing Data

Tabular summaries for the percentage of participants by the reason for discontinuation of study treatment, as well as for withdrawal from the study, are provided in the clinical SAP and will not be repeated herein.

The number of participants with missing PRO data will be summarized by treatment arm and timepoint (all scheduled timepoints in the double blind and OLE phases). For this analysis, the frequency and percentage of:

- Informative missingness, versus
- Non-informative missingness

The denominator will be the number of participants in the ITT.

This will be tabulated over time by treatment arm for each PRO endpoint. Informative missingness will be based on discontinuation due to an adverse event or due to lack of efficacy.

A comprehensive evaluation of missing data may warrant a modification to the planned analysis Any such modification will be detailed in the CSR.

5.4. Descriptive Analyses

Descriptive statistics for the observed scores as well as change from baseline scores for all the scores resulting from the instruments described in Section 2.2 except PGIS and PGIC will be provided for the ITT population by treatment arm at each timepoint in the double blind. Graphical representation will be also provided for the observed scores, e.g., box and whisker plots will be presented by treatment arm at each timepoint. Number and percentage of participants with each response value for PGIS and PGIC will be prestend by treatment arm at each timepoint.

A cumulative distribution function (CDF) plot showing a continuous plot of the change from baseline during the study, with change scores presented on the x-axis and the cumulative percent of participants experiencing that change on the y-axis, will be presented by treatment arm at all post-baseline timepoints up to and including Cycle 10. The cumulative distribution plot will be produced for the following PRO scores:

- DTSS scores: Total Symptom Score (weekly summary), Total Symptom Score – 5 Items (weekly summary), Pain Domain Score (weekly summary)
- DTIS scores: Physical Functioning Domain Score
- BPI-SF: API
- EORTC QLQ-C30: GHS/QoL, PF, RF and PA
- PROMIS: PF10a (sum score) and PF13 (sum score).

In addition, the number and proportion of participants by pain severity at baseline will be reported by treatment arms. Pain severity at baseline will be defined using the API score at baseline and classified as follows:

- Controlled pain at baseline: is API ≤ 4
- Uncontrolled pain at baseline: if API >4.

5.5. Longitudinal Analysis of Change from Baseline

Change from baseline during the double-blind phase in PRO scores will be analyzed using a restricted maximum likelihood (REML) based repeated measures approach (Brown & Prescott, 2006) (MMRM – Mixed Model Repeated Measures) in a primary analysis and a PMM (O'Kelly & Ratitch, 2014) in a sensitivity analysis. The PMM will be used to assess the robustness of the MMRM estimate with regard to missing data, when the MAR assumption is replaced by assumptions that are likely to be relatively less favorable to the experimental treatment.

The longitudinal analysis of change from baseline will include only the on-treatment assessments (thus excluding the unscheduled, EOT and follow-up visits).

5.5.1. Mixed Model Repeated Measures – Main Analysis

The primary objective of this analysis is to examine the treatment difference at Cycle 10. Data from all PRO assessments at all scheduled timepoints will be reported, although analysis will be limited to timepoints at which at least 10 participants have non-missing data in both treatment arms through Cycle 10 (i.e, Cycle 11 on will not be included in the model) .

The response variable will be the change from baseline to each PRO assessment. The model will include the treatment arm and timepoint as fixed-effect categorical factors, the baseline PRO score and stratification factor (primary tumor location: intra-abdominal vs extra-abdominal) as fixed effects covariates, and the baseline x time and treatment x time interactions. Both main effects and the interaction terms will remain in the model, regardless of significance. The model will present least squares (LS) mean estimates, standard errors, 2-sided 95% CIs, and 1-sided p-values for mean changes from baseline to each visit.

A plot of the LS means accompanied by the 2-sided 95% CI will be produced for each PRO measure by treatment arm.

The analysis will be conducted using PROC MIXED in SAS. The model will assume unstructured covariance among the within-participant repeated measurements. If the algorithm does not converge, a heterogeneous Toeplitz (the TOEPH option in SAS PROC MIXED) will be tried first and then heterogenious autoregressive (ARH)(1) as a covariance structure to achieve convergence. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

The analysis will be performed on the ITT population. Separate models will be considered for each of the following PRO score:

- DTSS scores: Total Symptom Score, Total Symptom Score – 5 Items, Pain Domain Score

- DTIS scores: Physical Functioning Domain Score
- BPI-SF: API
- EORTC QLQ-C30: GHS/QoL, PF, RF and PA
- PROMIS: PF10a (sum score) and PF13 (sum score).

For the subgroup examinations, a separate model will be considered for each subgroup listed in Section 4.5. The model will include the treatment arm, and timepoint as fixed-effect categorical factors, the baseline PRO score, stratification factor (primary tumor location: intra-abdominal vs extra-abdominal, except for the stratification factor subgroup) and the subgroup as fixed effect covariates, and the baseline x time, treatment x time and treatment x subgroup interactions. The same hierarchy will be used to chose a covariance structure. Subgroup analyses will be run only for variables resulting in subgroups of at least 10 participants in each treatment arm. The following will be presented: least squares (LS) mean estimates for the overall treatment effect (across all timepoints), standard errors, 2-sided 95% CIs, and 1-sided p-values for mean changes from baseline to each visit, as well as the p-value for the treatment by subgroup interaction.

5.5.2. Pattern Mixture Model – Sensitivity Analysis

The MMRM assumes that the missing observations are MAR. To address the possibility of the data being missing not at random (MNAR) (e.g., non-ignorable missing data), a sensitivity analysis using a PMM with sequential modelling with multiple imputation and delta-adjustment will be used. Change from baseline to Cycle 10 in PRO scale will be analyzed using a pattern-mixture model using control-based approach by means of sequential modelling with multiple imputation as described by O'Kelly (O'Kelly & Ratitch, 2014). The results from this analysis will be used to judge the validity of the MAR assumption. Similar conclusions from MMRM and PMM would suggest that the results are not overly dependent on the assumptions of the primary analysis with regard to the missing data.

We distinguish between monotone and non-monotone missing values. 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 on. 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 and will be multiply imputed using a Markov Chain Monte Carlo (MCMC) method of Proc MI in SAS. The imputation model will include the following:

- Treatment arms (nirogacestat or placebo)
- Stratification factors (primary tumor location:intra-abdominal vs extra-abdominal)
- Age (continuous)
- Gender (male or female)

- Geographic region (North America vs the rest of world)
- Desmoid Tumor Treatment Status (treatment naïve, recurrent or refractory)
- Any prior treatment (Yes vs No).

Non-monotone missing data will be imputed first, followed by the imputation of monotone missing data.

To impute monotone missing values, we define patterns depending on reason and timing of missingness as follows:

- **Pattern 1**: missing values before or at Cycle 10 due to participant's death;
- **Pattern 2**: missing values before or at Cycle 10 due to adverse events (AEs) or due to progression (clinical or radiographic);
- **Pattern 3**: missing values before or at Cycle 10 with missingness that does not satisfy conditions of patterns 1 to 2.

The patterns are mutually disjointed, i.e., each participant with monotone missing data belongs only to one pattern.

The following assumptions will be made for the missing data in each pattern:

For **pattern 1**, the worst score (e.g., 10 for DTSS Total Symptom Score) will be assigned as a penalty for unobservable values up Cycle 10 after participant's death. This will be applied to both treatment arms.

For **pattern 2**, control-based approach will be used for the nirogacestat arm. For the placebo arm, multiple imputation under MAR assumption will be used.

For **pattern 3**, data will be assumed to be MAR in both treatment arms.

The imputation of monotone missing data will be done sequentially for each scheduled PRO assessment visit, k=k1,…,K (where K corresponds to Cycle 10) as follows:

a) Impute monotone missing data in pattern 1 at visit k as the worst possible score for the imputed PRO score.

b) For the pattern 3 and for pattern 2 placebo participants only, impute the monotone missing values at visit k using an MAR-based multiple imputation regression model (PROC MI option MONOTONE REG) including the effects for baseline covariates as listed for the imputation model of non-monotone missing data above and PRO values at each schedule assessment time point up to (k-1).

c) For the pattern 2 nirogacestat participants only, impute the monotone missing values at visit k using multiple imputation regression model including the effects for baseline covariates as listed for the imputation model of non-monotone missing data above and PRO

values at each schedule assessment time point up to (k-1); at this step we will include all participants from the nirogacestat arm with missing at visit k, plus participants from the placebo arm with visit k observed. We omit participants from the nirogacestat arm with outcomes observed at vist k. Multiple imputation will now estimate regression parameters for visit k using data from the placebo arm only. The imputed data for vist k for a participant from the nirogacestat arm will look similar to the imputed data for a similar participant from the placebo arm.

The above steps (a)-(c) are performed for each visit k, before proceeding with the imputations of the next visit (k+1).

A total of 50 multiply-imputed datasets will be created for each PRO. The random number generator seed for the imputation of non-monotone missing values using MCMC will be 5414, and the random seed for imputation of monotone missing values will be 5414+k, for k=1, 2, … for each sequential visit with monotone missing data. For monotone missing imputation, the specified seed will be used for the first dataset. Data will be sorted so that the 1st dataset for each instrument is imputed, followed by the second dataset for each instrument, etc. Instruments will be ordered in the following manner:

- BPI-SF: API
- DTIS scores: Physical Functioning Domain Score
- DTSS scores: Total Symptom Score, Total Symptom Score – 5 Items, Pain Domain Score
- EORTC QLQ-C30: GHS/QoL, PF, RF and PA
- PROMIS: PF10a (sum score) and PF13 (sum score).

The MMRM modeling with the identical setup as described above in Section 5.5.1 will be performed, i.e., at each timepoint and also overall across all timepoints giving each visit equal weight, for each imputed dataset.

The SAS MIAnalyze procedure will be used to combine the results of these analyses for the imputations. For a more detailed description of the implementation MNAR imputation, see Ratitch B and O'Kelly M (O'Kelly & Ratitch, 2014).

The treatment differences will be estimated from the final model with LS-means differences and using the REML method. The degrees of freedom will be estimated with the Kenward-Roger approximation. The LS mean treatment difference, 2-sided 95% CI, and 1-sided p-value will be presented. A plot of the LS means accompanied by the 2-sided 95% CI will be produced by treatment and at each timepoint.

The analysis will be performed on the ITT population. Separate models will be considered for each of the following PRO scores:

- DTSS scores: Total Symptom Score, Total Symptom Score – 5 Items
- DTIS scores: Physical Functioning Domain Score

- BPI-SF: API
- EORTC QLQ-C30: GHS/QoL, PF and RF.

5.6. Responder Analysis

To understand meaningful changes experienced by participants at Cycle 10 (during the double-blind phase only), the frequency of responders vs. non-responders on the selected PRO measures will be reported and compared between the two arms. These change groups will be defined using either established (e.g.- literature based) thresholds where they exist, or empirically-derived thresholds for the newly developed instrument or when no literature values were observed (see Table 4). The responder analysis will be performed for the following measures:

- DTSS scores: Total Symptom Score, Pain Domain Score
- DTIS scores: Physical Functioning Domain Score
- BPI-SF: API.

For the GODDESS, as part of the psychometric analysis performed by IQVIA, clinically meaningful thresholds for improvement were also derived and will be used to define the responders and non-responders as described in Table 4 below.

Cut-off values of 30% or greater, or 2-point or greater change in numerical rating BPI-SF scores, have been proposed in the literature to detect clinically important improvements in cancer-related breakthrough pain and chronic pain states [(Farrar, et al., 2000) (Dworkin , et al., 2008)]. For this study, the value of 2 points will be used for API.

Table 4 Visit Responses for PRO Measures Based on Clinically Meaningful Thresholds

Instrument / Score	Primary threshold	Sensitivity 1 threshold	Sensitivity 2 threshold	Visit Response
GODDESS				
DTSS scores				
Total Symptom Score	CFB ≤ -1.4	CFB ≤ -1.0	CFB ≤ -1.7	Responder
	CFB > -1.4	CFB > -1.0	CFB > -1.7	Non-responder
Pain Domain Score	CFB ≤ -1.9	CFB ≤ -1.7	CFB ≤ -2.4	Responder
	CFB > -1.9	CFB > -1.7	CFB > -2.4	Non-responder
DTIS scores				
Physical Functioning Domain Score	CFB ≤ -0.8	CFB ≤ -0.6	CFB ≤ -1.0	Responder
	CFB > -0.8	CFB > -0.6	CFB > -1.0	Non-responder

CFB=change from baseline

The proportion of participants with improvement at Cycle 10 in PRO scores (e.g., responders) will be compared between the nirogacestat and placebo arms using using Cochran–Mantel–Haenszel test (CMH) test stratified by primary tumor location (intra-abdominal or extra-abdominal). Missing data at Cycle 10 will be imputed under the missing not at random

assumption. Specifically, multiple imputation under the MNAR assumption as described in Section 5.5.2 will be performed first to impute the continuous PRO scores. Next, a visit response (responder vs non-responder) at Cycle 10 will be derived as described above for each PRO score for each of the 50 datasets generated in the imputation procedure. The analysis will include only those patients with a baseline score ≥ threshold value used to define improvement (see Table 4 for GODDESS, for BPI-SF a value of 2 will be used).

Each of the multiply-imputed data sets with the response visit status at Cycle 10 will be analysed using the CMH test adjusted by primary tumor location (intra-abdominal or extra-abdominal) used. Statistical inference obtained from all imputed data will be combined using Rubin's rule (Ratitch, et al., 2013). The odds ratio will be log-transformed and the Wilson-Hilferty transformation will be applied to the CMH statistitic prior to combining all results with PROC MIANALYZE. The visit response by treatment arm (average, minimum and maximum percent responder), odds ratio (OR) as well as the corresponding 2-sided 95% CI will be provided along with the 1-sided p-values.

A descriptive analysis will also be performed. The number and proportion of participants who are responders vs. non-responders (as defined in the Table 4 above) will be summarized at each post-baseline timepoint (the double-blind phase only) by treatment arm. The proportions will be derived in two ways. In the first analysis, the denominator will be the number of participants with non-missing data at specific cycle. In a second analysis, the denominator will be the number of participants in the ITT population for whom a PRO is expected (i.e., including participants with missing data) at specific cycle. Participants with missing data at specific cycle will be considered non-responders in this second analysis. The proportion of participants who are responders vs non-responders in the second analysis described above will be presented graphically using bar charts.

In addition, for the DTSS Total Symptom Score, Pain Domain Score, and BPI-SF API, a symptom improvement rate will be defined as the number and proportion of participants who are responders at two or more consecutive timepoints during the double-blind period (see Table 4). This analysis will be performed only among those particpants with uncontrolled pain at baseline (e.g., with a baseline API ≥4) and will use all the available datapoints during double-blind phase. The improvement rate will be compared using a logistic regression stratified by primary tumor location (intra-abdominal or extra-abdominal). Additional covariates, such as age and gender, will also be examined and included as appropriate. The results of the analysis will be presented in terms of an OR together with its associated 2-sided 95% CI and 1-sided p-value.

5.7. Time to Event Analysis

Time to event analyses will include all PRO assessments during the double-blind phase (thus any unscheduled and EOT visits) and will be performed separately for the following PRO scores:

- DTSS scores: Total Symptom Score, Pain Domain Score
- DTIS scores: Physical Functioning Domain Score
- BPI-SF: API.

The following time to event endpoints will be defined:

- Time to first DT symptom improvement (using DTSS Total Symptom)
- Time to first control of pain symptoms (using BPI-SF API)
- Time to pain response (using BPI-SF API, DTSS Pain Domain Score)
- Time to first improvement in functioning (using DTIS: Physical Functioning Domain Score)

Time to first DT symptom improvement will be defined as the duration of time from the date of randomization to the date of the first time reduction of at least a X points in DT symptoms using the DTSS Total Symptom Score (X is the threshold value which is provided in Table 4) as compared to the baseline score. The primary threshold as well as both sensitivity thresholds will be used. Participants without observed symptom improvement at the time of analysis will be censored at the date of last available PRO assessment (i.e., date of the last non-missing value) on or before the analysis data cutoff date. Participants who were randomized but with no baseline or whose baseline scores do not allow for further improvement will be censored on the date of randomization. Participants with a baseline score but no post-baseline assessments will be censored at the baseline assessment date.

Time to first control of pain symptoms will be defined as the time from randomization to first time the BPI-SF API score is ≤ 4. The analysis will include only those participants whose BPI-SF API baseline scores are > 4. Participants without observed control of pain symptoms at the time of analysis will be censored at the date of last pain assessment on or before the analysis data cutoff date. Participants who were randomized but with no baseline score will be censored on the date of randomization. Participants with a baseline score but no post-baseline assessments will be censored at the baseline assessment date.

Time to pain response will be defined using the BPI-SF API and DTSS Pain Domain Score as follows: time to pain response is defined as the time from randomization to first occurence of pain response (using 2 points for BPI-SF API and the values in Table 4 for DTSS Pain Domain Score). Participants without observed pain response at the time of analysis will be censored at the date of last pain assessment on or before the analysis data cutoff date. Participants who were randomized but with no baseline or whose baseline scores do not allow for further reduction in pain will be censored on the date of randomization. Participants with a baseline score but no post-baseline assessments will be censored at the baseline assessment date.

Time to first improvement in functioning will be defined using the DTIS Physical Functioning Domain Score as follows: time to clinically meaningful improvement is defined as the time from randomization to first occurrence of improvement (e.g., response, see Table 4). Participants without observed improvement at the time of analysis will be censored at the date of last PRO assessment on or before the analysis data cutoff date. Participants who were randomized but with no baseline or whose baseline scores do not allow for further improvement will be censored on the date of randomization. Participants with a baseline score but no post-baseline assessments will be censored at the baseline assessment date.

For all time to event analyses, the time to event will be analyzed in months and will be presented

by treatment arm. Kaplan-Meier curves will be presented. The hazard ratio (HR) and the 95% CI will be estimated using a Cox proportional hazards model controlling for stratification factor (primary tumor location [intra-abdominal or extra-abdominal]). Additional covariates, such as age or gender, will also be examined and included as appropriate.

A 1-sided stratified log-rank test on the time to event will be performed using SAS PROC LIFETEST with method=PL option (Kaplan-Meier estimates, also known as the product-limit estimates). The HR with 2-sided 95% CI will be estimated from the stratified Cox proportional hazards model using SAS PHREG procedure with ties=EXACT option in the model. In this analysis, the baseline hazard function will be allowed to vary across strata; i.e., the MODEL statement will include the treatment arm variable as the only covariate and the STRATA statement will include the prespecified variable. The assumption of proportionality will be tested by producing plots of complementary log-log (event times) versus log(time).

Kaplan-Meier plots of the survival distribution function will be presented and include the number of participants at risk over time by treatment arm.

The time to pain response will also be examined in selected subgroups (see Section 4.5). For the subgroup examinations, separate unstratified Cox models will be employed for each subgroup listed in Section 4.5. The model will include the treatment arm, the subgroup and treatment x subgroup interaction. Subgroup analyses will be run only if at least 10 events occurred in one of the subgroups across treatment groups. The following will be presented for each subgroup: the number of participants, events and censored cases, the median time to event and the corresponding 2-sided 95% CI , HR with 2-sided 95% CI and the 1-sided p-value corresponding to the stratified log-rank test, as well as the 2-sided p-value obtained from the unstratified Cox for the treatment by subgroup interaction.

6. CHANGES TO PLANNED ANALYSES

As of this date, there have been one notable change between the protocol-defined statistical analyses of the PRO data and those presented in this statistical analysis plan:

"Patient-Reported Outcomes Measurement Information System Physical Function (PROMIS PF) short form 10a plus 3 additional items from PROMIS item banks" has been moved from secondary to exploratory endpoint, due to duplications to other PROs instruments.

7. REFERENCES

Brown, H. & Prescott, R., 2006. *Applied Mixed Models in Medicine.* s.l.:John Wiley and Sons Ltd.

Dworkin , R., Turk, D. & Wyrwich, K., 2008. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Pain,* Volume 9, pp. 105-21.

Farrar, J. et al., 2000. Defining the clinically important difference in pain outcome measures. *Pain,* Volume 88, pp. 287-94.

Fayers, P. et al., 2001. EORTC QLQ-C30 Scoring Manual (3rd ed.). *EORTC.*

O'Kelly, M. & Ratitch, B., 2014. *Clinical Trials with Missing Data: A Guide for Practitioners..* s.l.:John Wiley & Sons Ltd.

PROMIS, n.d. Physical functioning: A brief guide to the PROMIS Physical Function instruments. *Patient-Reported Outcomes Measurement Information System (PROMIS).*

Ratitch, B., Lipkovich, I. & O'Kelly, M., 2013. *Combining Analysis Results from Multiply Imputed Categorical Data.* [Online]
Available at: https://www.lexjansen.com/pharmasug/2013/SP/PharmaSUG-2013-SP03.pdf
[Accessed 17 December 2021].

8. APPENDICES

8.1. The GOunder/DTRF DEsmoid Symptom/Impact Scale (GODDESS)

8.1.1. Items of Desmoid Tumor Impact Scale (DTIS)

Item #	Question	Scale	Range
1	During the past 7 days How often have you had difficulty moving (for example twisting or bending) near your tumor?	5	None - All the time
2	How often have you had difficulty with reaching up, such as reaching shelves that were above your head?	5	None - All the time
3	How often have you had trouble falling asleep?	5	None - All the time
4	How often have you had difficulty getting comfortable in bed?	5	None - All the time
5	How often have you had trouble staying asleep at night?	5	None - All the time
6	How often have you had difficulty doing vigorous activities (such as running, lifting heavy objects, or participating in strenuous sports)?	5	None - All the time
7	How often have you had difficulty doing moderate activities (such as pushing a vacuum cleaner, playing with children, or taking a long walk)?	5	None - All the time
8	How often have you accomplished less than you would like when doing work or other regular daily activities?	5	None - All the time
9	How often have you avoided people because of the way you feel about your appearance?	5	None - All the time
10	What was your worst difficulty with reaching up, such as reaching shelves that were above your head?	11	0 - 10
11	Have you been dissatisfied about your appearance?	11	0 - 10
12	How much fear of future diagnostic tests did you have?	11	0 - 10
13	How much fear of recurrence / growth of your desmoid tumor(s) did you have?	11	0 - 10
14	How much hopelessness did you have?	11	0 - 10
15	How much anger did you have?	11	0 - 10
16	How much anxiety did you have?	11	0 - 10
17	How much frustration did you have?	11	0 - 10

8.1.2. Items of Desmoid Tumor Symptom Scale (DTSS)

Item #	Question	Scale	Range	Note
	During the past 24 hours			
1	How bad was your worst feeling of pain?	11	0 - 10	
2	How bad was your worst feeling of dull pain?	11	0 - 10	
3	How bad was your worst feeling of shooting pain?	11	0 - 10	
4	How bad was your worst feeling of fatigue?	11	0 - 10	
5	What was your worst swelling around your tumor(s)?	11	0 - 10	
6	What was your worst muscle weakness around your tumor(s)?	11	0 - 10	
7	At its worst, how difficult was moving (for example twisting or bending) near your tumor(s)?	11	0 - 10	
8	Please indicate the location(s) of your desmoid tumor(s). Select all that apply.			Gate Q for Q9 - Q11
9	How bad was your worst feeling of abdominal pain?	11	0 - 10	If Abdominal Wall in Q8
10	How bad was your worst feeling of nausea?	11	0 - 10	If Abdominal Wall in Q8
11	How bad was your worst feeling of fullness after beginning to eat?	11	0 - 10	If Abdominal Wall in Q8

8.2. BPI Short Form

STUDY ID #:_____ DO NOT WRITE ABOVE THIS LINE HOSPITAL #:_____

Brief Pain Inventory (Short Form)

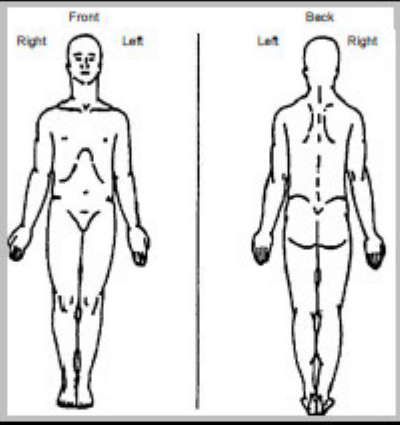
Date: ____/____/____ Time: _____

Name: _____ _____ _____
Last First Middle Initial

1. Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches). Have you had pain other than these every-day kinds of pain today?

1. Yes 2. No

2. On the diagram, shade in the areas where you feel pain. Put an X on the area that hurts the most.



3. Please rate your pain by circling the one number that best describes your pain at its worst in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

4. Please rate your pain by circling the one number that best describes your pain at its least in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

5. Please rate your pain by circling the one number that best describes your pain on the average.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

6. Please rate your pain by circling the one number that tells how much pain you have right now.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

Page 1 of 2

STUDY ID #:_____ DO NOT WRITE ABOVE THIS LINE HOSPITAL #:_____

Date: ___/___/____ Time: _____
Name: _____ _____ _____
Last First Middle Initial

7. What treatments or medications are you receiving for your pain?

8. In the last 24 hours, how much relief have pain treatments or medications provided? Please circle the one percentage that most shows how much relief you have received.

0% No Relief 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% Complete Relief

9. Circle the one number that describes how, during the past 24 hours, pain has interfered with your:

A. General Activity

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

B. Mood

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

C. Walking Ability

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

D. Normal Work (includes both work outside the home and housework)

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

E. Relations with other people

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

F. Sleep

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

G. Enjoyment of life

0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

Page 2 of 2

8.3. EORTC QLQ-C30

ENGLISH

EORTC QLQ-C30 (version 3)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials: ⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋
Your birthdate (Day, Month, Year): ⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋
Today's date (Day, Month, Year): 31 ⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋⌊ ⌋

	Not at All	A Little	Quite a Bit	Very Much
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	1	2	3	4
2. Do you have any trouble taking a long walk?	1	2	3	4
3. Do you have any trouble taking a short walk outside of the house?	1	2	3	4
4. Do you need to stay in bed or a chair during the day?	1	2	3	4
5. Do you need help with eating, dressing, washing yourself or using the toilet?	1	2	3	4

During the past week:	Not at All	A Little	Quite a Bit	Very Much
6. Were you limited in doing either your work or other daily activities?	1	2	3	4
7. Were you limited in pursuing your hobbies or other leisure time activities?	1	2	3	4
8. Were you short of breath?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Did you need to rest?	1	2	3	4
11. Have you had trouble sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseated?	1	2	3	4
15. Have you vomited?	1	2	3	4
16. Have you been constipated?	1	2	3	4

Please go on to the next page

ENGLISH

During the past week:	**Not at All**	**A Little**	**Quite a Bit**	**Very Much**
17. Have you had diarrhea?	1	2	3	4
18. Were you tired?	1	2	3	4
19. Did pain interfere with your daily activities?	1	2	3	4
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	1	2	3	4
21. Did you feel tense?	1	2	3	4
22. Did you worry?	1	2	3	4
23. Did you feel irritable?	1	2	3	4
24. Did you feel depressed?	1	2	3	4
25. Have you had difficulty remembering things?	1	2	3	4
26. Has your physical condition or medical treatment interfered with your family life?	1	2	3	4
27. Has your physical condition or medical treatment interfered with your social activities?	1	2	3	4
28. Has your physical condition or medical treatment caused you financial difficulties?	1	2	3	4

For the following questions please circle the number between 1 and 7 that best applies to you

29. How would you rate your overall health during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

30. How would you rate your overall quality of life during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

Scoring the EORTC QLQ-C30 version 3.0

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	Version 3.0 Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL:**

$$Score = \{(RS-1)/range\} \times 100$$

Examples:

Emotional functioning

$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

8.4. PROMIS

8.4.1. PROMIS PF Short Form 10a

PROMIS® Item Bank v2.0 – Physical Function – Short Form 10a

Physical Function – Short Form 10a

Please respond to each question or statement by marking one box per row.

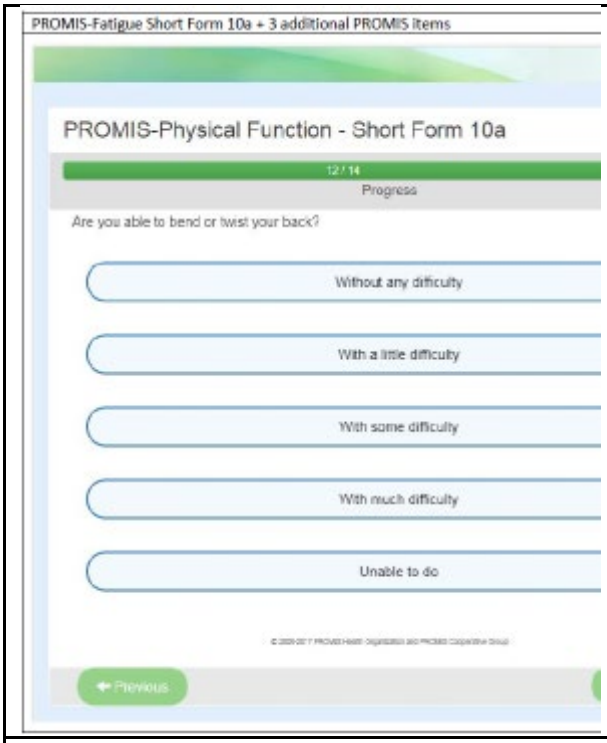
		Not at all	Very little	Somewhat	Quite a lot	Cannot do
PFA1	Does your health now limit you in doing vigorous activities, such as running, lifting heavy objects, participating in strenuous sports?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFC36r1	Does your health now limit you in walking more than a mile (1.6 km)?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFC37	Does your health now limit you in climbing one flight of stairs?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA5	Does your health now limit you in lifting or carrying groceries?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA3	Does your health now limit you in bending, kneeling, or stooping?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
		Without any difficulty	**With a little difficulty**	**With some difficulty**	**With much difficulty**	**Cannot do**
PFA11	Are you able to do chores such as vacuuming or yard work?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA10r1	Are you able to dress yourself, including tying shoelaces and buttoning your clothes?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFB26	Are you able to shampoo your hair?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA56	Are you able to wash and dry your body?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFC45r1	Are you able to sit on and get up from the toilet?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1

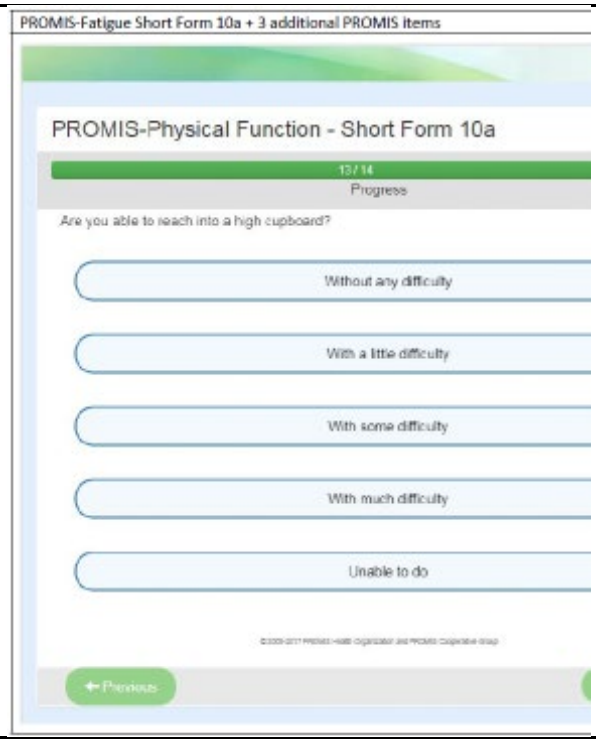
28 March 2017
 Page 1 of 1

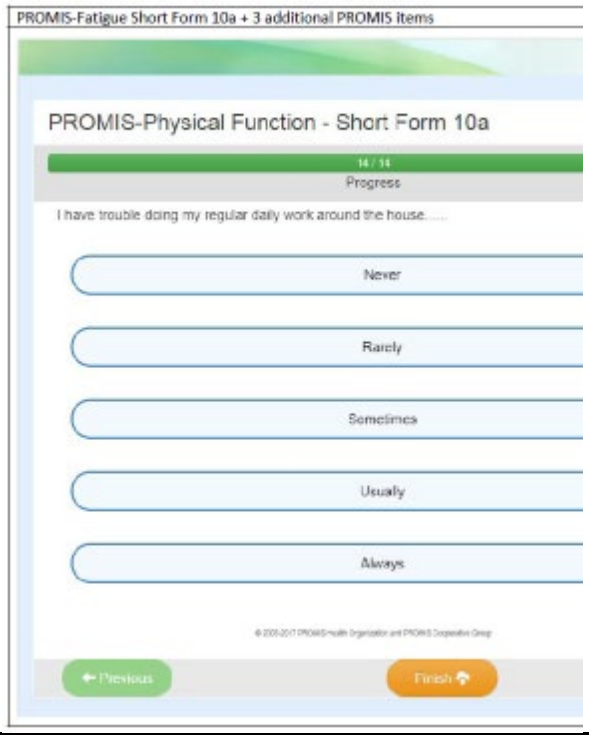
8.4.2. PROMIS PF Short Form 10a Look-Up Table

Adult v2.0 – Physical Function 10a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
10	13.5	3.6
11	16.6	2.8
12	18.3	2.7
13	19.7	2.5
14	20.9	2.4
15	22.1	2.3
16	23.1	2.2
17	24.1	2.2
18	25.0	2.1
19	26.0	2.0
20	26.9	2.0
21	27.7	1.9
22	28.6	1.9
23	29.4	1.9
24	30.2	1.8
25	31.0	1.8
26	31.8	1.8
27	32.5	1.8
28	33.3	1.7
29	34.0	1.7
30	34.8	1.7
31	35.5	1.7
32	36.3	1.7
33	37.0	1.7
34	37.8	1.7
35	38.5	1.8
36	39.3	1.8
37	40.1	1.8
38	40.9	1.9
39	41.7	1.9
40	42.6	1.9
41	43.5	2.0
42	44.4	2.1
43	45.5	2.1
44	46.6	2.3
45	47.9	2.5
46	49.4	2.8
47	51.2	3.2
48	53.4	3.6
49	55.8	3.9
50	61.9	5.9
*SE = Standard Error on T-score metric		

8.4.3. 3 Additional Items from PROMIS Item Banks







8.5. PGIS

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

□ None

□ Mild

□ Moderate

□ Severe

8.6. 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

□ Moderately Better

□ A Litter Better

□ No Change

□ A Little Worse

□ Moderately Worse

□ Very much Worse