Protocol B7981032

A Phase 3 Open-Label, Multi-Center, Long-Term Study Investigating the Safety and Efficacy of PF-06651600 in Adult and Adolescent Participants with Alopecia Areata

Statistical Analysis Plan
(SAP)

Version: 6.0

Date: 12 Apr 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Major Changes in SAP Amendments

Version/Date	Associated Protocol Amendment	Change	Rationale
1.0 06 JUN 2019	Original 03 APR 2019	Not Applicable	Not Applicable
2.0 19 SEP 2019	Amendment 1 15 OCT 2019	1. Section 9.4.1 Analyses Sets: Updated the Safety and Full Analysis Sets to be defined as all participants who take at least 1 dose of study intervention. 2. Section 5.2 General Methods: The section is updated to clarify that we will summarize by de-novo participants (i.e., those who did not receive study intervention in either Study B7931005 or B7981015) and participants originating from Study B7931005 or B7981015 (i.e., those who received study intervention in either Study B7931005 or B7981015). 3. Section 6 Analyses and Summaries: This section is updated to clarify that all reporting will be done by the groups specified in section 5.2. 4. Section 6.6 Baseline and Other Summaries: All the subsections in this section is updated to clarify that all reporting will be done by the groups specified in specified in section 5.2.	All these changes to the SAP were made to be consistent with the protocol amendment 1 and in alignment with a request from the Voluntary Harmonisation Procedure (VHP).

		5. Section 6.7.1 Adverse Events: A phrase has been added "but are not limited to" to specify that all safety events will be summarized, though we are listing some of them in this section.	
3.0 01 SEP 2020	Amendment 4 31 AUG 2020	1. Section 2: Added a sentence mentioning a separate SAP for the vaccine sub-study. 2. Section 2.1 Study Objectives: Updates were made on the Objectives and Endpoints table 3. Section 2.2 Study Design: The number of de novo participants was updated from approximately 350 to approximately 450, with a corresponding update to the total sample size. In addition, the number of sites was updated from approximately 120 to approximately 170. 4. Section 3.2 Secondary Endpoints; Section 3.3 Tertiary/Exploratory Endpoints: updated to align with the endpoints specified in Section 2.1 with details of endpoint definitions. 5. Section 6.2 Secondary Endpoints; Section 6.3 Tertiary/Exploratory Endpoints: These sections were updated to align with the endpoints specified in Section 2.1.	These changes to the SAP were made to be consistent with the protocol amendments 2-4, and to provide additional details and clarification.

		6. Section 8 References: additional references were added. 7. Appendix 3 Details on Selected PRO Endpoints was added. 8. Appendix 4 and Appendix 5 were added.	
4.0 26 APR 2021	Amendment 5 23 APR 2021	1. Section 2.1 Study Objectives: Updates were made on the Objectives and Endpoints table to be consistent with the PA5. 2. Section 2.2 Study Design: The maximum duration of study was updated from approximately 26 months to approximately 38 months. The duration of open-label treatment period was updated from 24 months to 36 months. Deleted "from approximately 170 sites globally". Figure 1 was updated to align with that in PA5. The treatment duration was updated from up to 24 months to 36 months. 3. Section 3 Endpoints and Baseline Variables: Definitions and Conventions: Updated to align with the endpoints specified in Section 2.1 with details of endpoint definitions. 4. Section 4.2 Full Analysis Set: The definition of FAS was updated from "all participants who take at least 1 dose of study intervention" to "all participants regardless of	These changes to the SAP were made to be consistent with the protocol amendments 2-5, and to provide additional details and clarification.

		whether they received study intervention", in order to keep consistency with that in Study B7981015. 5. Section 5.2 General Methods: Removed "standard error of the mean" for continuous variables. Added the presentation (number and percentage) for categorical variables. 6. Sections 6.1-6.4 Primary Endpoints-PK Endpoints: Updated to align with the endpoints specified in Section 3 with details of endpoint definitions. Removed confidence intervals in "Statistical Method" for continuous and categorical variables to align with Section 5.2. Added a summary ("A descriptive summary of the distribution of PGI-C will also be provided with the breakdown of all PGI-C categories.") for PGI-C to align with Study B7981015. 7. Section 6.6.1 Baseline Summaries: Added "alopecia areata history" for summary. 8. Section 6.7.4 Electrocardiogram: Updated from "baseline and End of Treatment visits" to "baseline and all post-baseline visits as per schedule of activities". 9. Section 8 REFERENCES: Added a reference to BRIEF®2.	

		10. Appendix 1 Definition and Use of Visit Windows in Reporting: Added Months 28, 32 and 36 to align with the study design, and updated the associated target days and analysis visit windows. 11. Appendix 2 Severity of Alopecia Tool (SALT): Removed content under copyright restrictions, and the formula used for derive the SALT score. Updated assessment time point from "Month 24" to "Month 36". 12. Appendix 3 Details on Selected PRO Endpoints: Added introduction to BRIEF®2.	
5.0 11 DEC 2023	Amendment 6 28 MAR 2022	General	In general, changes to the SAP were made to be consistent with the protocol amendment 6 (PA6), and to provide additional details and clarification.
		1. Table 2 in Section 2.1: adjusted the objectives and endpoints for consistency with PA6.	PA6 was implemented to increase treatment duration and to incorporate an Observation Period where patients would remain in the study after discontinuation of study drug.
		2. Section 2.2: updates were made on the study design for consistency with PA6	
		3. Sections 3.1-3.3: updates were made on the endpoints for consistent with PA6.	

		4. Sections 3.6.1-3.6.3: added the definitions of risk periods for AEs, lab data and vital signs.	The risk periods were added for reporting safety data in the on-treatment period and the Observation Period separately.
		5. Section 5.2: added "Graphical presentations may be provided" for clarification.	This sentence was added for flexibility of data summarization.
		6. Section 6.1: specified the analysis methods using risk periods for the primary endpoints.	Given that the risk periods are needed for the safety endpoints, updated the analysis methods for the primary endpoints.
		7. Sections 6.2-6.3: updates were made on the analyses for the secondary and exploratory endpoints for consistency with PA6.	The study was planned to extend, and Observation Period was added to the study duration. To keep consistency with PA6.
		8. Section 6.7: added summaries and listings related to COVID-19; added a listing for all AEs; clarification was made for consistency with PA6.	Per team's alignment, such a listing for all AEs is needed. To keep consistency with PA6.
		9. Section 6.7.1: updates were made on the analysis of adverse events, to incorporate the concept of risk periods.	Given that the risk periods are needed for the safety endpoints, updated the analysis methods for AEs.
		10. Sections 6.7.3-6.7.5: added the reporting scopes for vital signs, ECG and PE.	Added the scopes for accuracy of analysis methods.
		11. Section 8: this section was newly added to clarify the	In addition to the interim CSR for the purpose of

		reporting scopes for the PCD CSR and the supplemental CSR.	submissions, 2 CSRs (PCD CSR and supplemental CSR) are planned, so it is necessary to specify what content should be in these 2 CSRs, respectively.
		12. Appendix 1 in Section 10: changed the analysis visit window for Month 36, added the analysis visit windows for Months 40-60, removed the analysis visit window for Follow-Up; deleted the paragraph describing the imputation approach for unscheduled visits.	The study was planned to extend by up to 2 years with TP2 and Observation Period, so the analysis windows for the subsequent visits after Month 36 should be defined. The visit windows will be defined using study day as in the table of analysis visit windows rather than the nominal protocol defined visit labels. The visit labels conflict with the window definitions in the table.
		13. Appendix 2 in Section 10: updates were made on the definition of SALT AGA for consistency with PA6	The study was planned to extend by up to 2 years with TP2 and Observation Period. To keep consistency with PA6.
6.0 12 APR 2024	Amendment 6 28 MAR 2022	General	In general, changes to the SAP were made to be consistent with the protocol

			amendment 6 (PA6), and to provide additional details and clarification.
		1. Table 2 in Section 2.1: added clarifications for BRIEF®2.	BRIEF®2 is decided to be both an ObsRO and safety endpoint.
		2. Sections 3.3.1 and 3.3.2: added clarifications for BRIEF®2.	
		3. Section 3.6.1: updated the definitions of RPs from "for TP1 and TP2 + for the Observation Period" to "for TP1 and TP2 + for the entire period".	All available safety data to the data cutoff date will be included in the PCD CSR to be more aligned with the intent of the study protocol.
		4. Sections 3.6.2-3.6.4: updated to include clarifications for RPs for these endpoints.	To be consistent with the RPs for safety endpoints.
		5. Section 3.6.5: added a new section to include BRIEF®2.	The endpoint of BRIEF®2 is considered as both an ObsRO and safety endpoint.
		6. Section 5.3: updated to include clarifications for the derivation of SALT AA scores.	Updated the algorithm for calculating SALT AA scores for clarity.
		7. Section 6.1: added a clarification for the primary safety estimand and removed the summaries for the Observation Period.	Use the estimand concept and exclude the RP for the Observation Period. Per SoA, no data of laboratory and VS will be collected during the Observation Period.
		8. Section 6.2.1: added clarifications for the	Use the estimand concept.

		secondary/supportive safety estimand.	
		9. Sections 6.3.9 and 6.3.14: updated the summaries for BRIEF®2.	To be consistent with the RPs with the estimand concept.
		10. Section 6.7.1: added a new flag (identifying an AE is in the active treatment period or the Observation Period) for the AE listing.	For identifying an AE whether it is in the active treatment period or the Observation Period
		11. Section 8: updated to clarify that the PK data will be reported once throughout the whole study period.	The PK results were already reported in an interim CSR previously.
		12. Sections 8.1 and 8.2: updated to include clarifications for the reporting scopes.	Clarifications for the reporting scope in the PCD CSR and sCSR.
		13. Appendix 2 in Section 10: updated the algorithm for the derivation of SALT AA scores.	Updated the algorithm for calculating SALT AA scores for clarity.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7981032. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment. The statistical analyses of data collected in the vaccine sub-study will be provided in a separate SAP for this sub-study.

2.1. Study Objectives

The study objectives and corresponding endpoints are listed below in Table 2.

Table 2. Study Objectives and Endpoints

Primary Objective:	**Primary Endpoints:**

To evaluate the long-term safety and tolerability of PF-06651600 in adult and adolescent participants with Alopecia Areata (AA).	Through the time the last participant completes the Follow-up visit or 28 days after the Month 36 visit • Incidence of treatment-emergent adverse events (TEAEs); • Incidence of serious adverse events (SAEs) and adverse events (AEs) leading to discontinuation; • Incidence of clinically significant abnormalities in vital signs; • Incidence of clinically significant abnormalities in clinical laboratory values.
Secondary Objectives:	**Secondary Endpoints:**
To evaluate the long-term safety and tolerability of PF-06651600 in adult and adolescent participants with AA.	Through the time of the last participant visit: • Incidence of TEAEs; • Incidence of SAEs and AEs leading to discontinuation; • Incidence of clinically significant abnormalities in vital signs; • Incidence of clinically significant abnormalities in clinical laboratory values.
To evaluate the long-term efficacy of PF-06651600 in adult and adolescent participants with AA.	• Response based on achieving absolute Severity of Alopecia Tool (SALT) score ≤10 through Month 36, for overall and AA SALT score*; • Response based on achieving absolute SALT score ≤20 through Month 36, for overall and AA SALT score; • Change from baseline in SALT scores through Month 36, for overall and AA SALT score; • Response based on achieving at least 75% improvement in SALT (SALT75) from baseline through Month 36, for overall and AA SALT score; • Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in Eyebrow Assessment (EBA) score through Month 36;

	• Response based on achieving at a least 2-grade improvement from baseline or a score of 3 in Eyelash Assessment (ELA) score through Month 36.
To evaluate the effect of PF-06651600 on patient-centered outcomes and payer relevant measures to assess treatment benefit from the patient perspective and to demonstrate value.	• Patient's Global Impression of Change (PGI-C) response defined as PGI-C score of "moderately improved" or "greatly improved" through Month 36; • Change from baseline in Alopecia Areata Patient Priority Outcomes (AAPPO) scales through Month 36; • Change from baseline in the depression subscale scores of the Hospital Anxiety and Depression Scale (HADS) through Month 36; • Change from baseline in the anxiety subscale score of the HADS through Month 36; • Improvement on HADS among participants with a baseline subscale score indicative of depression and who achieved a "normal" subscale score indicative of an absence of depression through Month 36; • Improvement on HADS among participants with a baseline subscale score indicative of anxiety and who achieved a "normal" subscale score indicative of an absence of anxiety through Month 36.
Tertiary/Exploratory Objectives:	**Tertiary/Exploratory Endpoints:**
To evaluate the long-term efficacy of PF-06651600 in adult and adolescent participants with AA.	• Response based on achieving at least 50% improvement in SALT (SALT50) from baseline through Month 36, for overall and AA SALT score; • Absolute SALT scores through Month 36, for overall and AA SALT score.

	For all scheduled timepoints after Month 36 through Month 60: • Response based on achieving absolute SALT score ≤10, for overall and AA SALT score; • Response based on achieving absolute SALT score ≤20, for overall and AA SALT score; • Change from baseline in SALT score, for overall and AA SALT score; • Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in EBA score; • Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in ELA score.
To evaluate the effect of PF-06651600 on patient centered outcomes and payer relevant measures to assess treatment benefit from the patient perspective and to demonstrate value.	• Change from baseline in EuroQoL 5 dimensions (EQ-5D-5L) or EuroQoL 5 dimensions-Youth (EQ-5D-Y) through Month 24; • Improvement on PGI-C defined as "slightly improved", "moderately improved", or "greatly improved" through Month 36; • Improvement on Patient's Satisfaction with Hair Growth (P-Sat) items defined as slightly, moderately, or very satisfied through Month 36; • Change from baseline in Alopecia Areata Resource Utilization (AARU) through Month 24; • Change from baseline in Work Productivity and Activity Impairment items: Alopecia Areata (WPAI: AA) through Month 24; • Change from baseline in 36-Item Short Form Health Survey version 2 Acute (SF36v2 Acute) through Month 24; • Behavior Rating Inventory of Executive Function (BRIEF®2) index scores through Month 36. BRIEF®2 is considered both an Observer Reported Outcome (ObsRO) and safety endpoint;

	• PGI-C response defined as PGI-C score of "moderately improved" or "greatly improved" after Month 36 through Month 60; • Improvement on P-Sat items defined as slightly, moderately, or very satisfied after Month 36 through Month 60; • BRIEF®2 index scores after Month 36 through Month 60. BRIEF®2 is considered both an ObsRO and safety endpoint.
To evaluate the long-term efficacy of PF-06651600 in AA nail disease over time.	• Change from baseline in fingernails affected by AA through Month 36.
To evaluate the long-term effect of PF-06651600 on the clinician global impression of severity of scalp hair loss.	• Change from baseline in the Clinician Global Impression - Alopecia Areata (CGI-AA) through Month 36.
To evaluate pharmacodynamic and disease-related biomarkers over time.	• Change from baseline in lymphocyte subsets (T-cell, B-cell, and natural killer [NK] cells) through Month 36; • Change from baseline in immunoglobulins (IgA, IgG, IgM) through Month 36. • Change from baseline in lymphocyte subsets (T-cell, B-cell, and natural killer [NK] cells) after Month 36 through Month 60; • Change from baseline in immunoglobulins (IgA, IgG, IgM) after Month 36 through Month 60.

PK Objectives:	PK Endpoints:
To characterize the pharmacokinetics of PF-06651600.	• Plasma concentrations of PF-06651600 at Month 1 and Month 3.

* See Appendix 2 for details on overall SALT score and AA SALT score calculation. In this table, and throughout this document, AA SALT and SALT AA are used interchangeably as well as for overall SALT and SALT overall.

2.2. Study Design

This is a Phase 3, open-label, multi-center, long-term study designed to evaluate the safety and efficacy of PF-06651600 in adults and adolescents aged 12 years and older with AA. It is estimated that a total of approximately 960 participants will be enrolled. This will include eligible prior participants from the index studies B7931005 and B7981015, as well as approximately 450 de novo participants (ie, those who have not previously received study intervention in Study B7931005 or B7981015). The study will have a maximum duration of approximately 62 months. The study consists of two treatment periods (treatment period 1 [TP1] and TP2) and one Observation Period.

TP1:

TP1 includes up to a 5-week screening period, a 36-month open-label treatment period, and a 4-week follow-up period after completion of study intervention at Month 36 (for participants not continuing to the next treatment period [TP2]) or discontinuation of study intervention (see Schema in Figure 1).

Screening will occur within 35 days prior to the first dose of study intervention to confirm that selection criteria for the study are met for de novo participants and participants originating from Study B7931005 or B7981015 with >30 days between the last dose in Study B7931005 or B7981015 and their first visit in Study B7981032.

Participants enrolling from B7931005 and B7981015 will receive open-label 50 mg PF-06651600 once a day (QD) for 36 months and de novo participants will receive open-label 200 mg PF-06651600 QD for 4 weeks followed by open-label 50 mg PF-06651600 QD for 35 months.

Following the last dose of study intervention, both discontinued and completed participants not continuing to TP2 will enter into a 4-week follow-up period for safety monitoring. Participants in countries where PF-06651600 is not commercially available at the time of their Month 36 visit will have the opportunity to enter TP2, as described below. In this case, the Follow-up visit will be completed at the end of their participation in TP2.

TP2:

TP2 will be of variable length for individual participants; a participant may continue to receive PF-06651600 in TP2 for a maximum of 24 months or until availability of commercial

product in their country, or until the sponsor terminates the study in that country, whichever occurs first.

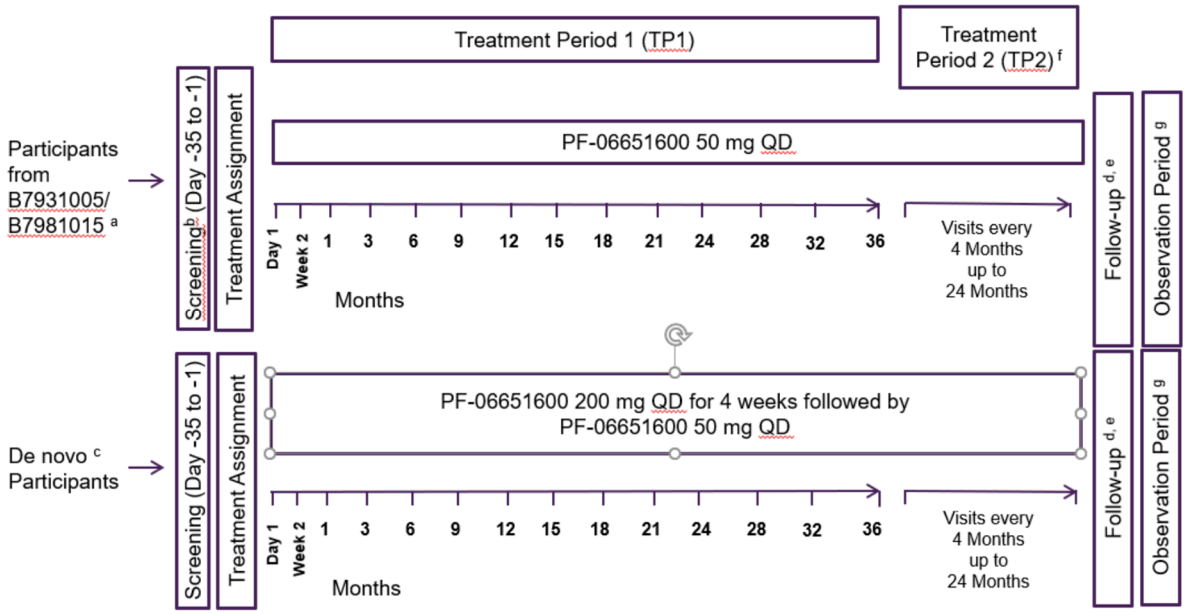
In TP2, participants will receive 50 mg QD PF-06651600. The total duration of participation in TP2 is approximately 25 months, including up to 24 months of study intervention, and a Follow-up period of 4 weeks after completion or discontinuation of study intervention (See Schema in Figure 1).

Observation Period:

If study intervention is permanently discontinued, the participant will be asked to remain in the study after the Follow-up visit without study intervention and continue to comply with study visit schedules for approximately 2 years after the last dose of study intervention or until study end, whichever occurs first. Refer to Section 7.1.1.1 of the protocol amendment 6 for further details.

At visits during the Observation Period, only the SALT, ELA, EBA, PGI-C, and concomitant medications/treatments will be collected at the visits specified in the applicable Schedule of Activities; AE and SAE reporting will follow the guidelines in Section 8.3 of the protocol. Specifically for AEs, the time period for actively eliciting and collecting AEs and SAEs for each participant begins from the time the participant provides informed consent through and including a minimum of 28 calendar days after the last administration of the study intervention (exceptions are described in Section 8.3.1 of the protocol amendment 6).

Figure 1. Study Design Schematic



Abbreviations: QD = once daily.

a. Participants originating from Study B7931005 or B7981015 are defined as those who received study intervention in one of these studies.

b. Participants with ≤30 days between the first study visit in B7981032 and the last dose in Study B7981015 will not require a screening period.
c. De novo participants are defined as those who did not previously receive study intervention in Study B7931005 or B7981015; this includes, but is not limited to, those consented and screened for Study B7931005 or B7981015 but who did not receive study intervention in one of these studies.
d. After completion of TP1 (for participants not continuing to TP2) or discontinuation of study intervention, a Follow-up period of 4 weeks will occur. Participants in countries where PF-06651600 is not commercially available at the time of their Month 36 visit will have the opportunity to enter TP2.
e. In TP2, after completion or discontinuation of study intervention, a Follow-up period of 4 weeks will occur.
f. TP2 will be of variable length for individual participants for a maximum of 24 months or until availability of commercial product in their country, or until the sponsor terminates the study in that country, whichever occurs first.
g. If study intervention is permanently discontinued, the participant will be asked to remain in the study after the Follow-up visit for the Observation Period without study intervention and continue to comply with study visit schedules for approximately 2 years after the last dose of study intervention or until study end, whichever occurs first.

To be eligible to enroll in this study, participants enrolling from Study B7931005 or B7981015 must not have had any events meeting the B7981032 discontinuation criteria or discontinued for safety-related events. In addition, participants enrolling from Study B7931005 must have taken their last dose of PF-06700841 (a TYK2/JAK1 inhibitor) in Study B7931005 >12 weeks prior to the B7981032 Day 1 visit. There is no necessary washout period for participants who took PF-06651600 in Study B7931005 or B7981015. Participants enrolling from B7981015 must have completed ≥34 weeks of study intervention. De novo participants ≥12 to <18 years of age must have a clinical diagnosis of AA with no other etiology of hair loss other than androgenetic alopecia with ≥50% terminal hair loss of the scalp due to AA at both the screening and Day 1 visits which, in the opinion of the investigator, is appropriate for systemic therapy. De novo participants ≥18 years of age and participants originating from Study B7931005 or B7981015 with >30 days between the last dose in Study B7931005 or B7981015 and their first visit in Study B7981032 must have a clinical diagnosis of AA with no other etiology of hair loss other than androgenetic alopecia with ≥25% terminal hair loss of the scalp due to AA at both the screening and Day 1 visits which, in the opinion of the investigator, is appropriate for systemic therapy. The full list of eligibility criteria for the study is included in Section 5 of the protocol amendment 6.

A participant is considered to have completed the study:

- If participant has completed all phases of TP1, including follow-up visit and does not continue to TP2 due to local commercial availability of PF-06651600 for AA; OR
- If participant has completed all phases of TP2, including Follow-up visit; OR
- If participant discontinued from the study during TP2 due to local commercial availability of PF-06651600 for AA.

The primary completion date (PCD) is defined as the date when the last participant completes the Follow-up visit or 28 days after the Month 36 visit. The end of the study is defined as the date of the last visit of the last participant in the study.

Study Treatments

- Doses of PF-06651600 will be 200 mg taken QD orally for 1 month, followed by 50 mg taken QD orally thereafter for the de novo participants.
- Doses of PF-06651600 will be 50 mg taken QD orally for participants enrolling from B7931005 and B7981015.
- Treatment duration will be up to 60 months.

Sample Size Determination

Sample size is not based on any formal hypothesis testing but instead driven by the regulatory requirement for the safety database. The sample size is determined by the number of participants who enroll from the Phase 2a study (B7931005) and participants who enroll from the Phase 2b/3 study (B7981015) as well as the number of de novo participants enrolled. It is estimated that a total of approximately 960 patients will be enrolled in Study B7981032, including approximately 450 de novo participants. The number of de novo participants enrolled may be modified based on the actual number of participants who continue from study B7931005 and study B7981015 into study B7981032.

Analyses performed, including interim analyses, if any, will primarily be for safety though some measures of efficacy may be analyzed.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoints

The primary endpoints, through the PCD, are:

- Incidence of treatment-emergent adverse events (TEAEs);
- Incidence of serious adverse events (SAEs) and adverse events (AEs) leading to discontinuation;
- Incidence of clinically significant abnormalities in vital signs;
- Incidence of clinically significant abnormalities in clinical laboratory values.

3.2. Secondary Endpoints

The secondary endpoints are for safety and efficacy:

3.2.1. Safety Endpoints

Secondary safety endpoints, through the time of the last participant visit are:

- Incidence of treatment emergent adverse events (TEAEs);

- Incidence of serious adverse events (SAEs) and adverse events (AEs) leading to discontinuation;
- Incidence of clinically significant abnormalities in vital signs;
- Incidence of clinically significant abnormalities in clinical laboratory values.

3.2.2. Efficacy Endpoints

- Response based on achieving absolute Severity of Alopecia Tool (SALT) score ≤10 through Month 36, for SALT overall and AA score;
- Response based on achieving absolute SALT score ≤20 through Month 36, for SALT overall and AA score;
- Change from baseline in SALT overall score and AA score through Month 36;
    - Change from baseline in SALT overall score and AA score are the same numerically;
- Response based on achieving at least 75%, improvement in SALT (SALT75) from baseline through Month 36, for SALT overall and AA score;
- Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in Eyebrow Assessment (EBA) score through Month 36;
- Response based on achieving at a least 2-grade improvement from baseline or a score of 3 in Eyelash Assessment (ELA) score through Month 36;
- PGI-C response defined as PGI-C score of "moderately improved" or "greatly improved" through Month 36;
- Change from baseline in Alopecia Areata Patient Priority Outcomes (AAPPO) scales through Month 36;
    - Current hair loss on scalp, eyebrows, eyelash and body hair (Items 1-4) for each individual item scored as 0= 'no hair loss', 1=' little hair loss' and 2-4=' moderate-complete hair loss';
    - Improvement on AAPPO items (1-4) among participants with a baseline score 2-4 indicating moderate-complete hair loss who achieved a score of 0= 'no hair loss' or 1=' little hair loss';
    - Change from baseline in AAPPO Emotional Symptoms: Emotional Symptoms is defined as mean of items 5-8;
    - Change from baseline in AAPPO Activity Limitations: Activity Limitations is defined as mean of items 9-11.

- Change from baseline in the depression subscale score of the Hospital Anxiety and Depression Scale (HADS) through Month 36;
- Change from baseline in the anxiety subscale score of the HADS through Month 36;
- Improvement on HADS among participants with a baseline subscale score indicative of depression who achieved a "normal" subscale score indicative of an absence of depression through Month 36;
    - Among adults, a score 0-7 is considered normal, a score of >7 is indicative of depression. Among adolescents (using baseline age), a score of 0-6 is considered normal, a score of >6 is indicative of depression;[1,2]
- Improvement on HADS among participants with a baseline subscale score indicative of anxiety who achieved a "normal" subscale score indicative of an absence of anxiety through Month 36;
    - Among adults, a score 0-7 is considered normal, a score of >7 is indicative of anxiety. Among adolescents (using baseline age), a score of 0-8 is considered normal, a score of >8 is indicative of anxiety.[1,2]

3.3. Tertiary/Exploratory Endpoints

3.3.1. Tertiary Endpoints through Month 36

The following are the tertiary/exploratory endpoints through Month 36, unless specified otherwise:

- Response based on achieving at least 50% improvement in SALT (SALT50) from baseline, for SALT overall and AA score;
- Absolute SALT scores, for SALT overall and AA score;
- Change from baseline in EuroQoL 5 dimensions (EQ-5D-5L) or EuroQoL 5 dimensions-Youth (EQ-5D-Y) through Month 24;
    - Will be summarized for adults and adolescents separately. UK specific utility weights[3] will be applied for the scoring of these endpoints;
- Improvement on PGI-C defined as slightly, moderately or greatly improved;
- Improvement on Patient's Satisfaction with Hair Growth (P-Sat) items defined as slightly, moderately or very satisfied;
- Change from baseline in Alopecia Areata Resource Utilization (AARU) questions through Month 24;
    - Proportion of participants with any healthcare professional (HCP) visits;

- Among those with HCP visits, mean total number of visits for any reason, mean change from baseline in total number of visits for any reason, mean total number of visits related to AA and mean change from baseline in AA-related visits.

- Change from baseline in Work Productivity and Activity Impairment domains: Alopecia Areata (WPAI: AA) through Month 24;
  - Change from baseline in absenteeism, presenteeism, work productivity loss (overall) and activity impairment in adults;

- Change from baseline in 36-Item Short Form Health Survey version 2 Acute (Short SF36v2 Acute) through Month 24;
  - Eight domain scales: physical function (PF), bodily pain (BP), role physical (RP), role emotional (RE), Vitality (VT), general health (GH), social function (SF) and mental health (MH);
  - Two summary scales: Mental Component Summary (MCS) Score and Physical Component Summary (PCS) Score.

- Behavior Rating Inventory of Executive Function (BRIEF®2) index scores through Month 60 for all the data that is available; BRIEF®2 is considered as both an ObsRO and safety endpoint.
  - T-scores for three Index scores: Behavior Regulation Index (BRI), Emotional Regulation Index (ERI) and Cognitive Regulation Index (CRI);

- Change from baseline in fingernails affected by AA;
- Change from baseline in the Clinician Global Impression - Alopecia Areata (CGI-AA);
  - CGI-AA scored as 0= 'None (no hair loss)', 1='Minimal hair loss', 2-4='Moderate-very severe or complete hair loss';
  - Improvement among participants with a baseline score 2-4 indicating Moderate-very severe or complete hair loss who achieved a score of 0= 'None (no hair loss)' or 1='Minimal hair loss';

- Change from baseline in lymphocyte subsets (T-cell, B-cell, and natural killer [NK] cells);
- Change from baseline in immunoglobulins (IgA, IgG, IgM).

3.3.2. Tertiary Endpoints after Month 36

The following are the tertiary/exploratory endpoints for all scheduled timepoints after the PCD (ie, after Month 36 through Month 60), unless specified otherwise:

- Response based on achieving absolute SALT score ≤10, for overall and AA SALT score;
- Response based on achieving absolute SALT score ≤20, for overall and AA SALT score;
- Change from baseline in SALT score, for overall and AA SALT score;
- Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in EBA score;
- Response based on achieving at least a 2-grade improvement from baseline or a score of 3 in ELA score.
- PGI-C response defined as PGI-C score of "moderately improved" or greatly improved";
- Improvement on P-Sat items defined as slightly, moderately, or very satisfied;
- Behavior Rating Inventory of Executive Function (BRIEF®2) index scores after Month 36 through Month 60; BRIEF®2 is considered as both an ObsRO and safety endpoint.
  - T-scores for three Index scores: BRI, ERI and CRI;
- Change from baseline in lymphocyte subsets (T-cell, B cell, and natural killer [NK] cells);
- Change from baseline in immunoglobulins (IgA, IgG, IgM).

3.4. PK Endpoints

- Plasma concentrations of PF-06651600 at Month 1 and Month 3.

3.5. Baseline Variables

For safety measurements relative to baseline, baseline values will be defined as follows:

- For de novo participants and participants originating from study B7931005 or B7981015 with >30 days between the first visit of study B7981032 and the last dose in study B7931005 or B7981015:
  - Baseline values will be defined as the values from Day 1 of study B7981032 (or from screening of study B7981032 for evaluations not performed at Day 1).
- For participants enrolled from study B7981015 with ≤30 days between the first visit of study B7981032 and the last dose in study B7981015:

- Baseline values will be defined as the values from Day 1 of Study B7981015 (or from screening of study B7981015 for evaluations not performed at Day 1).

For efficacy measurements, baseline values will be defined as follows:

- For participants originating from study B7931005 or B7981015:
  - Baseline values are defined as the Day 1 values from Study B7981015 or B7931005. For assessments which were not collected in the respective index study, the baseline value is defined as the value collected on Day 1 of Study B7981032 (or from screening of study B7981032 for evaluations not performed at Day 1).
- For the de novo participants:
  - Baseline values will be defined as the values from Day 1 of study B7981032 (or from screening of study B7981032 for evaluations not performed at Day 1).

For either safety or efficacy measurements, if the baseline of a participant is from the index study (if applicable), then the screening or Day 1 assessment of Study B7981032 will be regarded as a post-baseline assessment.

3.6. Safety Endpoints

Primary safety endpoints are defined above in Section 3.1.

3.6.1. Adverse Events

An adverse event (AE) is considered a Treatment-Emergent Adverse Event (TEAE) if the event has start date on or after the first dosing date of this study.

If any AEs are ongoing at the time of the End of Treatment visit within the index study and the participant is enrolling into this study, such AEs will be transcribed and followed within this study. AE updates following the index study End of Treatment visit date will only be made within this study for subjects who enroll into this study.

The AE summaries in the PCD Clinical Study Report (CSR) (Section 8.1) and in the supplemental CSR (Section 8.2) will be reported for the risk periods (RP) that are for the active treatment periods (TP1 and TP2, which will be referred to as On-Treatment Plus 35-Day Estimand) and for the entire period (which will be referred to as Treatment Policy Estimand) defined as follows:

For the active treatment periods (TP1 and TP2)

- For all participants, the RP is defined as that from the first dosing date of TP1 to min{last contact date, last dosing date + 35}, where last contact date is max{date of discontinuation from study, date of participant's last visit, date of PCD data cut-off if ongoing}, where max{*a*, *b, c*} denotes the maximum value among *a*, *b* and *c* (if any value is invalid, then calculate the maximum value using the valid ones). If a

participant dies, the death date will be the last contact date. This RP includes events that start at any time during the active treatment periods up to 35 days after the last dose.

For the entire period

- For all participants, the RP is defined as that from the first dosing date of TP1 to the last contact date, where the last contact date is the same as that defined above.

3.6.2. Laboratory Data

See protocol Appendix 2 for the list of clinical laboratory tests to be performed.

RPs for the purpose of laboratory data summaries (if applicable) will follow the same principles as those in Section 3.6.1, except that the laboratory data collected on the treatment period switch date should be counted in the pre-switch period rather than in the post-switch period, ie, a laboratory test result collected on the switch date should be counted in the pre-switch period.

3.6.3. Vital Signs, including Height and Weight

Vital sign measurements are pulse rate, blood pressure.

Height and weight will be collected pre- and post-dose. After the screening visit, height will not be collected from participants who are ≥18 years of age at the time of the visit.

For adolescent participants (ie, <18 years of age at the time of the assessment), in addition to the above summary, height, weight, and body mass index (BMI) will also be summarized using Z-scores. Z-score is calculated using the following formulation:

$$Z = \frac{(X/M)^L - 1}{LS}, L \neq 0$$

or

$$Z = \ln(X/M)/S, L = 0$$

Where X is the physical measurement (eg, height, weight, calculated BMI) and L, M, and S are the values from standardized growth charts (Centers for Disease Control and Prevention [CDC] web page[4]) for different age (12 years to <18 years) and sex.

RPs for the purpose of vital signs data summaries (if applicable) will follow the same principles as those in Section 3.6.2.

3.6.4. Physical Examinations and Tanner Stage Assessments

Complete physical examinations consist of assessments of general appearance; skin; head, eyes, ears, nose and throat (HEENT); mouth, heart; lungs; abdomen; extremities; neurologic function, back, and lymph nodes. In addition, dermatological full body exam must be

performed by the investigator, sub investigator or a qualified health professional per local guidelines. Dermatological examinations should also include visual inspection of the breasts and external genitalia.

Targeted physical examinations should include skin, heart, lung, and abdomen, neurologic function, and examination of body systems where there are symptom complaints by the participant.

Determination of physical and sexual maturation will be performed using the Tanner stages of development for participants who are <18 years of age at baseline. For these participants, Tanner stages will be collected at visit(s) only until the participant has reached a score of 5 on all applicable domains. For participants with AU (ie, total loss of hair on the scalp, face and body) at the Day 1 visit for whom the stage for pubic hair cannot be reliably assessed, the Tanner stages will be collected at visit(s) after the Day 1 visit only until the participant has reached a score of 5 on the remaining applicable domain (ie, breasts for females and genitalia for males).

RPs for the purpose of the physical examination and Tanner stage data summaries (if applicable) will follow the same principles as those in Section 3.6.2.

3.6.5. Behavior Rating Inventory of Executive Function®, Second Edition (BRIEF®2)

The BRIEF®2 consists of 63 items and yields scores on nine scales, with the Inhibit and Self-Monitor scales comprising the Behavior Regulation Index (BRI), the Shift and Emotional Control scales comprising the Emotional Regulation Index (ERI), and the Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials scales comprising the Cognitive Regulation Index (CRI).

RPs for the purpose of BRIEF®2 data summaries (if applicable) will follow the same principles as those in Section 3.6.2.

4. ANALYSIS SETS

4.1. Safety Analysis Set

The Safety Analysis Set (SAS) is defined as all participants who take at least 1 dose of study intervention. SAS is the primary analysis set.

4.2. Full Analysis Set

The Full Analysis Set (FAS) is for all the efficacy analyses and is defined as all participants regardless of whether they received study intervention.

4.3. Pharmacokinetic (PK) Analysis Set

The PK Analysis Set (PKAS) is defined as all participants who take at least 1 dose of study intervention and have at least one measurable plasma concentration of PF-06651600.

4.4. Per Protocol Analysis Set

No per protocol analysis set will be defined for this study.

5. GENERAL METHODOLOGY AND CONVENTIONS

The final analysis and reporting of results will be performed after the completion of the study and the database is locked. Interim data cut and reporting may be performed from time to time, see Section 7.

5.1. Hypotheses and Decision Rules

B7981032 is an open-label and non-comparative study, hence no statistical hypotheses will be tested.

5.2. General Methods

In general, descriptive summary statistics such as number, percentage and 95% confidence interval based on normal approximation will be presented for binary variables, and number, mean, standard deviation (and/or standard error), median, minimum, and maximum will be presented for continuous variables. For categorical variables, descriptive summary statistics such as number and percentage will be presented. Graphical presentations may be provided.

Displays will be presented by all participants, as well as by de novo participants (ie, those who did not receive study intervention in either Study B7931005 or B7981015) and participants originating from Study B7931005 or B7981015 (ie, those who received study intervention in either Study B7931005 or B7981015). By study design, de novo participants are expected to receive PF-06651600 200 mg QD for 4 weeks followed by 50 mg QD, and participants originating from Study B7931005 or B7981015 are expected to receive PF-06651600 50 mg QD for the duration of the treatment period.

5.3. Methods to Manage Missing Data

Data will be summarized based on observed data. Missing data will not be imputed.

For the PRO variables such as AAPPO, CGI-AA, HADS, EQ-5D-5L/EQ-5D-Y, SF-36v2, PGI-C, P-Sat, and WPAI:AA, BRIEF®2, missing data will be handled in accordance with the developer's or Pfizer's (for AAPPO and P-Sat) established rules. If these rules are not enough for imputing a value, then it will be treated as missing, unless noted otherwise.

For handling of missing values and special situations in calculating the SALT AA scores, see Appendix 2.

6. ANALYSES AND SUMMARIES

The primary endpoints of incidence of adverse events, serious adverse events, clinically significant vital sign abnormalities and clinically significant laboratory abnormalities will be summarized descriptively through appropriate data tabulations, descriptive statistics, and graphical presentations. All summaries will be displayed by the groups mentioned earlier (Section 5.2). Efficacy analyses will be descriptive in nature; there will be no hypothesis testing, though 95% two-sided confidence intervals may be reported.

6.1. Primary Endpoints

Safety through the PCD using the On-Treatment Plus 35-Day Estimand:

- Summary: Incidence of treatment-emergent adverse events (TEAEs) in the Treatment Period (see Section 3.6.1 for the definition of RP)
- Summary: Incidence of treatment-emergent adverse events (TEAEs) in the Observation Period (see Section 3.6.1 for the definition of RP)
- Summary: Incidence of serious adverse events (SAEs) and adverse events (AEs) leading to discontinuation in the Treatment Period (see Section 3.6.1 for the definition of RP)
- Summary: Incidence of serious adverse events (SAEs) and adverse events (AEs) leading to discontinuation in the Observation Period (see Section 3.6.1 for the definition of RP).
- Summary: Incidence of clinically significant abnormalities in clinical laboratory values in the Treatment Period (see Section 3.6.2 for the definition of RP).
- Summary: Incidence of clinically significant abnormalities in vital signs in the Treatment Period (see Section 3.6.3 for the definition of RP).
- Population: SAS.
- Statistical Method: Descriptive statistics: number and percentage of participants with the events.

6.2. Secondary Endpoints

6.2.1. Safety Endpoints (ie, Safety in the Entire Study, Treatment Policy Estimand)

The analysis methods for the secondary safety endpoints are the same as those for the primary safety endpoints, but through the time of the last participant visit. This will be a supportive analysis to that using the On-Treatment Plus 35-Day Estimand.

6.2.2. Efficacy Endpoints

6.2.2.1. SALT ≤ 10 Response

- Summary: Proportion of participants achieving SALT score ≤ 10 through Month 36 for SALT overall and AA scores.
- Population: FAS.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.2. SALT ≤ 20 Response

- Summary: Proportion of participants achieving SALT score ≤ 20 through Month 36 for SALT overall and AA scores.
- Population: FAS.

- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.3. Change from Baseline in SALT Score

- Summary: Change from baseline in SALT overall score and SALT AA score through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.2.2.4. SALT75 Response

- Summary: Proportion of participants achieving at least 75% improvement in SALT scores from baseline through Month 36 for SALT overall and AA scores.
- Population: FAS.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.5. EBA Response

- Summary: Proportion of participants achieving at least 2-grade improvement from baseline or absolute score of 3 in the EBA through Month 36.
- Population: FAS on the subset of subjects who do not have normal (score of 3 on the EBA) eyebrow at baseline.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.6. ELA Response

- Summary: Proportion of participants achieving at least 2-grade improvement from baseline or absolute score of 3 in the ELA through Month 36.
- Population: FAS on the subset of subjects who do not have normal (score of 3 on the ELA) eyebrow at baseline.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.7. PGI-C Response

- Summary: Proportion of participants achieving a score of moderately improved or greatly improved on the PGI-C through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals. A descriptive summary of the distribution of PGI-C will also be provided with the breakdown of all PGI-C categories.

6.2.2.8. Change from Baseline on AAPPO

- Summary: Current hair loss on scalp, eyebrows, eyelash and body hair (Items 1-4) for each individual item scored as 0= 'no hair loss', 1=' little hair loss' and 2-4=' moderate-complete hair loss' through Month 36.
- Summary: Improvement on AAPPO items (1-4) among participants with a baseline score 2-4 indicating moderate-complete hair loss who achieved a score of 0= 'no hair loss' or 1=' little hair loss' through Month 36.
- Summary: Change from baseline in AAPPO Emotional Symptoms through Month 36.
- Summary: Change from baseline in AAPPO Activity Limitations through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.2.2.9. Change from Baseline in HADS depression subscale score

- Summary: Change from baseline in HADS depression subscale score through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.2.2.10. Change from Baseline in HADS anxiety subscale score

- Summary: Change from baseline in HADS anxiety subscale score through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.2.2.11. Improvement on HADS Depression Score

- Summary: Proportion of participants with improvement (participants with a baseline sub-scale score indicative of depression who achieved a "normal" score indicative of an absence of depression) through Month 36.
- Population: FAS on the subset of subjects with depression at baseline.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.2.2.12. Improvement on HADS Anxiety Score

- Summary: Proportion of participants with improvement (participants with a baseline sub-scale score indicative of anxiety who achieved a "normal" score indicative of an absence of anxiety) through Month 36.

- Population: FAS on the subset of subjects with anxiety at baseline.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.3. Tertiary/Exploratory Endpoints

The following are the tertiary/exploratory endpoints through Month 36, unless specified otherwise:

6.3.1. SALT-50 Response

- Summary: Proportion of participants achieving at least 50% improvement in SALT scores from baseline through Month 36 for SALT overall and AA scores.
- Population: FAS.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.3.2. Absolute SALT Score

- Summary: Absolute SALT score through Month 36 for SALT overall and AA scores.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.3.3. Change from Baseline in EQ-5D-5L/EQ-5D-Y Score

- Summary: Change from baseline in EuroQoL 5 dimensions (EQ-5D-5L) or EuroQoL 5 dimensions-Youth (EQ-5D-Y) through Month 24.
- Population: FAS.
- Statistical Method: Descriptive statistics for adults and adolescents separately.

6.3.4. Improvement on PGI-C

- Summary: Proportion of participants achieving PGI-C improvement defined as slightly, moderately or greatly improved through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.3.5. Improvement on P-Sat Score

- Summary: Proportion of participants achieving improvement on Patient's Satisfaction with Hair Growth items defined as slightly, moderately or very satisfied through Month 36.
- Population: FAS.

- Statistical Method: Descriptive statistics and 95% two-sided confidence intervals.

6.3.6. Change from Baseline in AARU

- Summary: Proportion of participants with any HCP visits through Month 24.
- Summary: Among those with HCP visits, mean total number of visits for any reason, mean change from baseline in total number of visits for any reason, mean total number of visits related to AA and mean change from baseline in AA-related visits through Month 24.
- Population: FAS.
- Statistical Method: Descriptive statistics for all variables, and 95% two-sided confidence intervals only for proportions.

6.3.7. Change from Baseline in WPAI:AA

- Summary: Change from baseline in absenteeism, presenteeism, work productivity loss (overall) and activity impairment through Month 24.
- Population: Adults in FAS.
- Statistical Method: Descriptive statistics.

6.3.8. Change from baseline in SF36v2

- Summary: Change from baseline in SF36v2 eight domain scales and two summary scales through Month 24.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.3.9. BRIEF®2 index Scores

- Summary: T-scores of BRIEF®2 three index scores BRI, ERI and CRI through the date of PCD data cut-off inside the RP for active treatment period (On-Treatment Plus 35-Day Estimand) defined in Section 3.6.1. Derivation of T-scores for each index can be found in Appendix 3. A summary using Treatment Policy Estimand will serve as a supportive analysis.
- Population: Adolescents in SAS.
- Statistical Method: Descriptive statistics.

6.3.10. Change from Baseline in Fingernails Affected by AA

- Summary: Change from baseline in number of fingernails affected by AA through Month 36, among those who had affected fingernails at baseline;

- Population: FAS;
- Statistical Method: Descriptive statistics.

6.3.11. Change from Baseline in CGI-AA Score

- Summary: CGI-AA scored as 0= 'None (no hair loss)', 1='Minimal hair loss', 2-4='Moderate-very severe or complete hair loss' through Month 36.
- Summary: Improvement among participants with a baseline score 2-4 indicating Moderate-very severe or complete hair loss who achieved a score of 0= 'None (no hair loss)' or 1='Minimal hair loss' through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.3.12. Change from Baseline in Lymphocyte Subsets

- Summary: Absolute and Change from baseline in T-cell, B-cell, and natural killer (NK) cells through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.3.13. Change from Baseline in Immunoglobulins

- Summary: Change from baseline in IgA, IgG and IgM through Month 36.
- Population: FAS.
- Statistical Method: Descriptive statistics.

6.3.14. Tertiary Endpoints after PCD

For all endpoints in Section 3.3.2 with scheduled timepoints after the PCD (ie, after Month 36 through Month 60), the following endpoints will be analyzed in the same way as their corresponding endpoints through PCD (ie, the endpoints in Section 3.3.1), unless specified otherwise.

- SALT ≤ 10 Response post-PCD
- SALT ≤ 20 Response post-PCD
- Change from Baseline in SALT Score post-PCD
- EBA Response post-PCD
- ELA Response post-PCD

- PGI-C Response post-PCD
- Improvement on P-Sat Score post-PCD
- BRIEF®2 index Scores (this is covered in Section 6.3.9)
- Change from Baseline in Lymphocyte Subsets post-PCD
- Change from Baseline in Immunoglobulins post-PCD

6.4. PK Endpoints

6.4.1. Plasma Concentration

- Summary: Mean plasma concentrations of PF-06651600 at Month 1 and Month 3.
- Population: PKAS.
- Statistical Method: Descriptive statistics.

6.5. Subset Analyses

No subgroup analyses will be performed.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

Demographics, medical and alopecia areata history will be summarized for all participants, as well as by de novo participants (ie, those who did not receive study intervention in either Study B7931005 or B7981015) and participants originating from Study B7931005 or B7981015 (ie, those who received study intervention in either Study B7931005 or B7981015). Baseline disease characteristics will also be summarized.

Targeted medical history using the list in Appendix 5 will also be summarized.

6.6.2. Study Conduct and Subject Disposition

Subjects evaluation, disposition, and discontinuation will be summarized for all participants, as well as by de novo participants and participants originating from Study B7931005 or B7981015.

6.6.3. Study Treatment Exposure

A summary of compliance and the number of doses received as well as the median total dose by groups (as described in Section 5.2) will be provided.

The exposure to study drug will be summarized by the total number of days of dosing.

6.6.4. Concomitant Medications and Non-Drug Treatments

Prior drug and non-drug treatments, concomitant drug and non-drug treatments will be summarized for all participants, as well as by de novo participants and participants originating from Study B7931005 or B7981015.

6.7. Safety Summaries and Analyses

Safety analyses will be based on SAS.

Safety data will be presented in tabular and/or graphical format and summarized descriptively, where appropriate. All safety endpoints will be listed and summarized in accordance with Pfizer Standards. Categorical outcomes (eg, AEs) will be summarized by subject counts and percentage. Continuous outcome (eg, blood pressure, pulse rate, etc) will be summarized using N, mean, median, standard deviation, etc. Change from baseline in laboratory data, ECGs and vital signs may also be summarized. Subject listings may be produced for these safety endpoints accordingly.

In order to report the impact of COVID-19 on clinical trial populations and study data, the following additional listings and summaries will be produced:

- Listing of subjects affected by COVID-19 related study disruption;
- Protocol deviations related to COVID-19;
- Adverse events related to COVID-19;
- Summary of drug interruptions due to COVID-19;
- Discontinuations from study intervention or study due to COVID-19 related AEs;
- Summary of discontinuations due to COVID-19 related reasons.

6.7.1. Adverse Events

The safety data will be summarized in accordance with CDISC and Pfizer Standards (CaPS). Adverse events of special interest will be summarized based on a list of preferred terms that will be provided by safety risk lead to the programming team prior to database lock.

All safety data will be summarized descriptively through appropriate data tabulations, descriptive statistics, categorical summaries, and graphical presentations. All AEs will be included in a single listing with flags for (including but not limited to) SAEs, causality, treatment-emergence, and being in the active treatment period or in the Observation Period. This listing includes all AEs, regardless of seriousness, causality, treatment-emergence, and period allocation.

6.7.2. Laboratory Data

Laboratory data will be listed and summarized in accordance with the Pfizer reporting standards. Summaries of subjects meeting pre-specified monitoring and discontinuation

criteria will be created using methods for categorical data. Laboratory data criteria for discontinuation and abnormalities can be found in Appendix 4.

6.7.3. Vital Signs, including Height and Weight

Vital signs will be summarized at baseline and all available post-baseline visits, according to the scope defined for each CSR (Section 8).

Height and weight will be collected at pre and post dose. After the screening visit, height will not be collected from participants who are ≥18 years of age at the time of the visit.

6.7.4. Electrocardiogram

ECG parameters will be summarized at baseline and at all post-baseline visits as per schedule of activities, according to the scope defined for each CSR (Section 8).

6.7.5. Physical Examination and Tanner Stage Assessment

Physical examinations will be summarized at baseline and all available post-baseline visits, according to the scope defined for each CSR (Section 8).

Tables or spaghetti plots may be produced for Tanner stage assessments by gender separately for participants <18 years old at baseline.

7. INTERIM ANALYSES

Interim data cuts may be performed to support regulatory submissions, for regulatory requests, or for study monitoring for internal decision making. Other periodical data cuts may be conducted to support regulatory queries. As the study is open-label and no hypotheses are being tested, there are no concerns regarding protecting the Type I error rate. This SAP will be used for any interim data cut.

This study uses an External Data Monitoring Committee (E-DMC). The E-DMC will be responsible for ongoing monitoring of safety of participants in the study according to the charter.

8. TIMING OF ANALYSES

There will be 2 CSRs for this study. One is the PCD CSR and the other is the supplemental CSR. If the PK data are collected and analyzed prior to the PCD CSR, they will be only reported once in an interim CSR, and not in the PCD CSR, nor in the supplemental CSR.

8.1. PCD CSR

The PCD CSR will include the primary analysis for this study, which will be conducted following the PCD (Section 2.2). The PCD CSR will include analyses of the following:

- All safety data collected through the PCD.
- Efficacy and pharmacodynamic data collected for the time points as specified in the applicable SoAs through the Month 36 visit for each participant.

For participants who permanently discontinue from study intervention but remain in the study, efficacy data collected during the Observation Period will be included in efficacy listings, but not in the efficacy summaries.

8.2. Supplemental CSR

The supplemental CSR will include analyses of the following:

- All safety data collected in the entire study.
- Efficacy and pharmacodynamic data collected for the time points post Month 36 visit, as specified in the applicable SoA of TP2. The efficacy summaries will not repeat the efficacy summaries contained in the PCD CSR

For participants who permanently discontinue from study intervention but remain in the study, efficacy data collected during the Observation Period will be included in efficacy listings (including the efficacy data for the Observation Period from the PCD CSR), but not in the efficacy summaries.

9. REFERENCES

1. Snaith RP, Zigmond AS. The Hospital Anxiety and Depression Scale with the Irritability-Depression-Anxiety Scale and the Leeds Situational Anxiety Scale Manual. GL Assessment, 1994.

2. White et al. Validation of the Hospital Anxiety and Depression Scale for use with adolescents. Br J Psy, 1999; 175:452-54.

3. EQ-5D country specific weights. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/.

4. CDC website for normalized height vitals calculation. https://www.cdc.gov/nchs/data/series/sr_11/sr11_246.pdf.

5. Gioia GA, Isquith PK, Guy SC, et al. BRIEF2® Behavior Rating Inventory of Executive Function®, Second Edition. PAR 2015.

10. APPENDICES

Appendix 1 Definition and Use of Visit Windows in Reporting

Visit windows will be used for efficacy variables and for any safety data that are displayed or summarized by study visit.

Visit Label	Target Day	Analysis Visit Window
Screening/Baseline	1	Up to Day 1
Week 2	15	Days 2 to 22
Month 1	31	Days 23 to 61
Month 3	91	Days 62 to 136
Month 6	181	Days 137 to 226
Month 9	271	Days 227 to 316
Month 12	361	Days 317 to 406
Month 15	451	Days 407 to 496
Month 18	541	Days 497 to 586
Month 21	631	Days 587 to 676
Month 24	721	Days 677 to 781
Month 28	841	Days 782 to 901
Month 32	961	Days 902 to 1021
Month 36	1081	Day 1022 to 1141
Month 40	1201	Day 1142 to 1261
Month 44	1321	Day 1262 to 1381
Month 48	1441	Day 1382 to 1501
Month 52	1561	Day 1502 to 1621
Month 56	1681	Day 1622 to 1741
Month 60	1801	Day 1742 to (last dose day + 28)

If more than one observation falls within a visit window, the observation that is the closest to the target day will be used for the summary tables. If two visits are equally distant from the target day, the data from the later visit will be used.

Appendix 2 Severity of Alopecia Tool (SALT)

Severity of alopecia tool (SALT) is a quantitative assessment of alopecia severity based on scalp terminal hair loss. The overall SALT score does not distinguish the reason for hair loss.

The secondary endpoints for SALT include both the SALT overall and SALT AA scores. The SALT overall score includes scalp hair loss regardless of etiology (eg, including scalp hair loss due to both androgenetic alopecia [AGA] and AA). The SALT androgenetic alopecia (SALT AGA) score only takes into account scalp hair loss due to AGA and is required to be assessed at the final on-therapy visit for each treatment period (ie, in TP1 at Month 36 or Early Termination, and in TP2 at Month 60, the last on-therapy visit, or Early Termination). SALT AA scores will not be calculated for visits within the Observation Period.

Imputation of the SALT AGA Scores: within each of the treatment periods (ie, TP1 and TP2), use the observed SALT AGA score at the latest visit within the applicable treatment period to impute the SALT AGA scores at the earlier visits. If there are multiple observed SALT AGA scores available within the treatment period, use the observed SALT AGA score at the latest visit within the treatment period to overwrite the SALT AGA scores at the earlier visits where the observed SALT AGA scores are available.

Calculation of the SALT AA Scores: the SALT AA score at each visit will be calculated as follows using the observed SALT overall score and the observed or imputed SALT AGA score,

SALT AA score = SALT overall score – SALT AGA score.

If either the SALT overall score or SALT AGA score is missing (ie, it is neither observed nor imputed), the resulting SALT AA score will be missing. In addition, if the calculated SALT AA score <0, reset the SALT AA score as missing.

Hypothetical Examples: the following hypothetical examples illustrate implementation of the algorithm as described above.

Table 3. Derivation of the SALT AA score - Hypothetical Case 1

Visit	Observed SALT Overall Score	Observed SALT AGA Score	Imputed SALT AGA Score	Calculated SALT AA Score	Final SALT AA Score
Month 6	11.36	Not planned to be collected as per SoA	8	3.36	3.36
Month 9	3.72	Not planned to be collected as per SoA	8	-4.28	Missing
Month 12	Missing	Not planned to be collected as per SoA			Missing
Month 15	2	Not planned to be collected as per SoA	8	-6	Missing
Month 18	2	Not planned to be collected as per SoA	8	-6	Missing
Month 21	0	Not planned to be collected as per SoA	8	-8	Missing
Month 24	0	Not planned to be collected as per SoA	8	-8	Missing
Month 28	0	Not planned to be collected as per SoA	8	-8	Missing
Month 32	0	Not planned to be collected as per SoA	8	-8	Missing

Month 36	8	8	0	0

Table 4. Derivation of the SALT AA score - Hypothetical Case 2

Visit	Observed SALT Overall Score	Observed SALT AGA Score	Imputed SALT AGA Score	Calculated SALT AA Score	Final SALT AA Score
Month 6	11.36	Not planned to be collected as per SoA			Missing
Month 9	3.72	Not planned to be collected as per SoA			Missing
Month 12	Missing	Not planned to be collected as per SoA			Missing
Month 15	2	Not planned to be collected as per SoA			Missing
Month 18	2	Not planned to be collected as per SoA			Missing
Month 21	0	Not planned to be collected as per SoA			Missing
Month 24	0	Not planned to be collected as per SoA			Missing
Month 28	0	Not planned to be collected as per SoA			Missing
Month 32	0	Not planned to be collected as per SoA			Missing
Month 36	8	Missing			Missing

Table 5. Derivation of the SALT AA score - Hypothetical Case 3

Visit	Observed SALT Overall Score	Observed SALT AGA Score	Imputed SALT AGA Score	Calculated SALT AA Score	Final SALT AA Score
Screening	27.5	4.1	3.4	24.1	24.1
Baseline	29.7	3.0	3.4	26.3	26.3
Month 1	10.5	3.4		7.1	7.1
Month 6	11.36	Not planned to be collected as per SoA			Missing
Month 9	3.72	Not planned to be collected as per SoA			Missing
Month 12	Missing	Not planned to be collected as per SoA			Missing
Month 15	2	Not planned to be collected as per SoA			Missing
Month 18	2	Not planned to be collected as per SoA			Missing
Month 21	0	Not planned to be collected as per SoA			Missing
Month 24	0	Not planned to be collected as per SoA			Missing
Month 28	0	Not planned to be collected as per SoA			Missing
Month 32	0	Not planned to be collected as per SoA			Missing
Month 36	8	Missing			Missing

Appendix 3 Details on Selected PRO Endpoints

AAPPO

- The AAPPO is a 11-item scale. Items 1-4 are an assessment of current hair loss, eyebrow loss, eyelash loss and body hair loss and will be as such analyzed separately on a scale of 0-4 with 0 = 'no hair loss' and 4= 'complete hair loss'. Items 5-8 are an assessment of emotional symptoms. Response choices on these items are scored from 0 ="never" to 4="always". Items 9-11 are an assessment of activity limitations. Response choices on these items are scored from 0='not at all' to 4= 'completely'.

    - AAPPO Items 1 through 4 are scored separately as 0= 'no hair loss', 1=' little hair loss' and 2-4=' moderate-complete hair loss';

    - Improvement on AAPPO items (1-4) among participants with a baseline score 2-4 indicating moderate-complete hair loss who achieved a score of 0= 'no hair loss' or 1=' little hair loss';

    - Emotional Symptoms: Mean of Items 5, 6, 7, 8 (missing rule: requires at least 2 nonmissing responses; otherwise missing);

    - Activity Limitations: Mean of Items 9, 10, 11 (missing rule: requires at least 2 nonmissing responses; otherwise missing).

AARU

- Assessment of total in-person and remote healthcare visits

- Assessment of AA-related in-person and remote healthcare visits Assessment of utilization of hair prostheses or camouflaging agents.

- Assessment of # of days hair prostheses or camouflaging agents were utilized.

- Percentage of patients not currently employed and not seeking employment due to alopecia areata.

- Percentage of patients who have not sought employment due to alopecia areata.

- Number of opportunities for employment not sought due to alopecia areata.

- Scoring:

  Among those with any HCP visits (Q1), a subject's total number of visits for any reason will be computed by adding total number of in-person visits for any reason and total number of remote visits for any reason. Total number of visits for a subject related to alopecia areata will be computed by adding the total number of in-person visits related to AA and total number of remote visits related to AA.

WPAI –AA

- WPAI-AA is an adaptation of the WPAI: SHP (Specific Health Problem) version. Outcomes from this scale are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity. The questionnaire is scored as follows:

  - Questions:

  1 = currently employed

  2 = hours missed due to specified problem

  3 = hours missed other reasons

  4 = hours actually worked

  5 = degree problem affected productivity while working

  6 = degree problem affected regular activities

  - Scoring

  Multiply scores by 100 to express in percentages.

  Percent work time missed due to problem (Absenteeism): Q2/(Q2+Q4)

  Percent impairment while working due to problem (Presenteeism): Q5/10

  Percent overall work impairment due to problem (Overall Work Impairment): Q2/(Q2+Q4)+[(1-(Q2/(Q2+Q4)))x(Q5/10)]

  Percent activity impairment due to problem (Activity Impairment): Q6/10

BRIEF®2 index Scores – Derivation of T-scores

- The BRIEF®2 consists of 63 items and yields scores on nine scales, with the Inhibit and Self-Monitor scales comprising the Behavior Regulation Index (BRI), the Shift and Emotional Control scales comprising the Emotional Regulation Index (ERI), and the Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials scales comprising the Cognitive Regulation Index (CRI).

- For participants who are <18 years of age at the time of the Study B7981032 Day 1 visit, the parent/caregiver will complete the BRIEF®2 questionnaire. The same parent/caregiver should complete the questionnaire at each visit throughout the study whenever possible. Once a participant reaches ≥18 years of age, this assessment will no longer be required. At the Early Termination visit, if the BRIEF®2 was administered within 2 months of the Early Termination visit, it does not need to be collected.

- No more than one missing item response is acceptable to generate a valid raw score for each scale. If there are two more missing responses, the raw score for that scale will not be calculated. Raw scores of each index scores will be converted to T-scores using the BRIEF®2 manual[5]. Further details of this derivations can be found in the Data Set Specifications document.

Appendix 4 Laboratory Data Criteria for Discontinuation and Abnormalities

- Labs Meeting Discontinuation Criteria

Hematology		Criterion #1	Criterion #2
	Hemoglobin	<90 g/L or a decrease of >30% from baseline	n/a
	Platelets	<75 x 10^9/L	n/a
	Lymphocytes (ABSOLUTE)	<0.5 x 10^9/L	n/a
	Neutrophils (ABSOLUTE)	<0.75 x 10^9/L	n/a
Chemistry	Aspartate Aminotransferase	>3 ULN W/ Bilirubin >2 ULN	Two Sequential Elevations >5 ULN
	Alanine Aminotransferase	>3 ULN W/ Bilirubin >2 ULN	Two Sequential Elevations >5 ULN
	Creatine Kinase	>10 x ULN	

- Labs Test Abnormalities

Hematology		Lower Limit	Higher Limit
	Hemoglobin	<0.8 x LLN	n/a
	Hematocrit	<0.8 x LLN	n/a
	Red Blood Cell Count	<0.8 x LLN	n/a
	MCV	<0.9 x LLN	>1.1 x ULN
	MCH	<0.9 x LLN	>1.1 x ULN
	MCHC	<0.9 x LLN	>1.1 x ULN
	Platelets	<0.5 x LLN	>1.75 x ULN
	White Blood Cell Count	<0.6 x LLN	>1.5 x ULN
	Reticulocytes	<0.5 x LLN	>1.5 x ULN
	Reticulocytes/Erythrocytes	<0.5 x LLN	>1.5 x ULN
	Leukocytes	<0.6 x LLN	>1.5 x ULN
	Lymphocytes (ABSOLUTE)	<0.8 x LLN	>1.2 x ULN
	Lymphocytes/Leukocytes	<0.8 x LLN	>1.2 x ULN
	Neutrophils (ABSOLUTE)	<0.8 x LLN	>1.2 x ULN
	Neutrophils/Leukocytes	<0.8 x LLN	>1.2 x ULN
	Basophils/Leukocytes	n/a	>1.2 x ULN
	Eosinophils	n/a	>1.2 x ULN
	Eosinophils/Leukocytes	n/a	>1.2 x ULN
	Monocytes	n/a	>1.2 x ULN
	Monocytes/Leukocytes	n/a	>1.2 x ULN
Chemistry			
	Blood Urea Nitrogen	n/a	>1.3 x ULN
	Urea	n/a	>1.2 x ULN
	Creatinine	n/a	>1.3 x ULN
	Glucose	<0.6 x LLN	>1.5 x ULN
	Calcium	<0.9 x LLN	>1.1 x ULN

	Sodium	<0.95 x LLN	>1.05 x ULN
	Potassium	<0.9 x LLN	>1.1 x ULN
	Chloride	<0.9 x LLN	>1.1 x ULN
	Bicarbonate	<0.9 x LLN	>1.1 x ULN
	Aspartate Aminotransferase	n/a	>3.0 x ULN
	Alanine Aminotransferase	n/a	>3.0 x ULN
	Bilirubin	n/a	>1.5 x ULN
	Direct Bilirubin	n/a	>1.5 x ULN
	Indirect Bilirubin	n/a	>1.5 x ULN
	Alkaline Phosphatase	n/a	>3.0 x ULN
	Uric Acid	n/a	>1.2 x ULN
	Albumin	<0.8 x LLN	>1.2 x ULN
	Creatine Kinase	n/a	>2.0 x ULN
	Cholesterol	n/a	>1.3 x ULN
	HDL Cholesterol	<0.8 x LLN	n/a
	LDL Cholesterol	n/a	>1.2 x ULN
	Triglycerides	n/a	>1.3 x ULN
Urinalysis			
	pH (Scalar)	<4.5	>8
	Glucose (Scalar)	n/a	>=1
	Ketones (Scalar)	n/a	>=1
	Protein (Scalar)	n/a	>=1
	Hemoglobin (No Unit)	n/a	>=1
	Nitrite (No Unit)	n/a	>=1
	Leukocyte Esterase (No Unit)	n/a	>=1
	Urobilinogen	n/a	>=1
	Erythrocytes (Scalar)	n/a	>=20
	Leukocytes (Scalar)	n/a	>=20
	Epithelial Cells (Scalar)	n/a	>=6
	Granular Casts (Scalar)	n/a	>1
	Hyaline Casts (Scalar)	n/a	>1
	Bacteria (No Unit)	n/a	>20

Appendix 5 List of Targeted Medical History Terms

Term	PT
Atopic dermatitis	DERMATITIS ATOPIC
Diabetes mellitus Type I	TYPE 1 DIABETES MELLITUS
Crohn’s disease	CROHN'S DISEASE
Ulcerative colitis	COLITIS ULCERATIVE
Systemic lupus erythematosus	SYSTEMIC LUPUS ERYTHEMATOSUS
Sjogren's syndrome	SJOGREN'S SYNDROME
Rheumatoid arthritis	RHEUMATOID ARTHRITIS
Psoriatic arthritis	PSORIATIC ARTHROPATHY
Psoriasis	PSORIASIS
Vitiligo	VITILIGO
Hashimoto’s thyroiditis	AUTOIMMUNE THYROIDITIS
Graves' disease	BASEDOW'S DISEASE