Protocol B7931028

A PHASE 2B, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED, MULTICENTER, DOSE-RANGING STUDY TO EVALUATE THE EFFICACY AND SAFETY PROFILE OF PF-06700841 IN PARTICIPANTS WITH ACTIVE SYSTEMIC LUPUS ERYTHEMATOSUS (SLE)

Statistical Analysis Plan (SAP)

[redacted]

Version: **5**

Date: **October 25, 2023**

TABLE OF CONTENTS

[redacted]

[redacted]

LIST OF TABLES

LIST OF FIGURES

APPENDICES

CCI

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
5 25 Oct 2023	Amendment 8 25 Jun 2023	Revised to align the SAP with the blinded data review #2 comments and Priovant comments.	• CCI [redacted] • Updated the SAS code in Appendix 2 for CMH and for Fisher's exact test • Removed continuity correction for CMH and removed normal approximation method. • Added MH risk difference and Sato variance estimate for CMH wording. • Removed Region as a stratification factor in CMH analysis and described a pooling strategy for undersized strata • Changed CMH to Fisher's Exact Test for subgroup analyses • Added wording to collapse strata <20 participants. • Changed Missing Data mechanism from LOCF48 full to LOCF48 mixed components. • NRI+LOCF48 is applied to treatment policy estimand for SRI-4 and BICLA • Added specific wording for exclusion of data quality events and background on issues. • Added Bar plots for some analyses • Added AST > 2*ULN • Updated risk period definition for capturing first severe flare. • CCI [redacted] [redacted]

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			• Removed the 3-tier approach for Adverse event analyses. • Updated Interim analysis wording to match protocol • Added table of abbreviations in Appendix 6.
4 7 Jul 2023	Amendment 8 25 Jun 2023	Revised to align SAP language with Protocol Amendment 8, blinded data review #1 comments, and Priovant comments.	• Changed Key Secondary Endpoint from mSSFI to BICLA • Added sub-group analyses for BICLA • Included 6.3.3.6, 6.3.3.7, and 6.3.4.2.1 • Removed anchor-based approach for FACIT-F • Updated wording for Russia/Ukraine and COVID-19 • Changed any binary endpoints using Chan and Zhang to CMH • Changed CMH CL from Wilson Corrected to Wald Corrected • Added SLEDAI-2K, BILAG 2004, and PhGA definitions in Section 3 • Changed any two-sided p-value output to both one-sided and two-sided p-values for formal hypothesis tests • Changed PRO endpoints to be descriptive [redacted] • Added SAS Modeling Code and Tipping Point Code to Appendix

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			• Revised Tipping Point Analysis Methodology to be consistent with A3921120 [redacted]
3 23 Aug 2022	Amendment 7 22 Jun 2022	Revised to align with the protocol amendment 7	• Refer to PA7 amendment changes • Updated 6.3.16.2.1, 6.3.16.2.2 and 6.3.16.5.1 to include use of anchor-based approach [redacted]
2 06 Nov 2019	Amendment 3 31 Oct 2019	Revised to align with the protocol amendment 3	• Section 2.1 Study Objectives, Endpoints, and Estimands. Revised to be consistent with the protocol amendment 3. • Section 2.1.1 Primary Estimand. Revised to be consistent with protocol amendment 3. • Section 3.5.4 Electrocardiograms. Baseline ECG is revised to be consistent with the protocol amendment 3. • Section 5.1. Hypotheses and Decision Rules. Revised the multiple testing procedure to be consistent with the protocol amendment 3. • Section 6.1.1.1. Main Analysis. Clarifications were added in the estimand and analysis details. • Section 6.1.1.2. Sensitivity/Supplementary Analyses. Revised supplementary analysis to be consistent with the protocol amendment 3.

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			• The Lupus QoL endpoint revised to individual domain scores to be consistent with protocol amendment 3 in multiple relevant sections. • Section 6.3.16.2.2. Responder analysis for both SRI-4 and FACIT-F (All Post-baseline Visits). CCI [redacted] • Section 6.3.16.9. Suicidal Behavior or Ideation. Revised suicidal behavior or ideation endpoint and new mapping algorithm from C-SSRS to C-CASA in Appendix 1.2 to be consistent with new Pfizer Standard and FDA recommendation. • Restricted the longitudinal GLMMRM analyses to selected binary endpoints; Simplifies the language to reduce duplication. • Updated visit windows for reporting in Appendix 1.1 • Revised the word "subjects" to "participants".
1 19 Jun 2019	Amendment 1 04 Feb 2019	N/A	N/A

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7931028. 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.

2.1. Study Objectives, Endpoints, and Estimands

Objectives	Endpoints	Estimands
Primary	**Primary**	**Primary**
• *To evaluate the efficacy of 3 QD dose levels of PF-06700841 compared to placebo in participants with active SLE*	• *Proportion of participants achieving the Systemic Lupus Erythematosus Responder Index (SRI) change of 4 (SRI-4) at Week 52*	*E1: This composite estimand is defined as a population average treatment difference between three PF-06700841 dose groups versus placebo in the proportions of SRI-4 responders at Week 52.* *Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *Population: Participants with active SLE as defined by the full analyses set (FAS)* *Variable: SRI-4 response at Week 52* *Intercurrent Events:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing, and such participants will be imputed as non-responders for the primary analysis.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the primary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the primary analysis.* *Population Level Summary: Treatment differences between three PF-06700841 dose groups and placebo in the proportions of SRI-4 responders at Week 52*
Key Secondary	**Key Secondary**	**Key Secondary**
• *To evaluate the efficacy of 3 QD dose levels of PF-06700841 compared to placebo in participants with active SLE*	• *Proportion of participants achieving the British Isles Lupus Assessment Group-Based Composite Lupus*	*E2: This composite estimand is defined as a population average treatment difference between three PF-06700841 dose groups versus placebo in the proportions of BICLA responders at Week 52, among participants with active SLE who are not*

Objectives	Endpoints	Estimands
	Assessment (BICLA) at Week 52	*impacted by the COVID-19 or Russia-Ukraine crises at Week 52.* *<u>Treatment</u>: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *<u>Population</u>: Participants with active SLE as defined by the FAS.* *<u>Variable</u>: BICLA response at Week 52* *<u>Intercurrent Events</u>:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing, and such participants will be imputed as non-responders for the key secondary analysis.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the key secondary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the key secondary analysis.* *<u>Population Level Summary</u>: Treatment differences between three PF-06700841 dose groups and placebo in the proportions of BICLA responders at Week 52*
• **Other Secondary**	• **Other Secondary**	**Other Secondary**
• *To assess attainment of low disease activity state in 3 QD dose levels of PF-06700841 compared to placebo in participants with active SLE*	• *Proportion of participants achieving the Lupus Low Disease Activity State (LLDAS) at Week 52*	*<u>E3</u>: This composite estimand is defined as a population average treatment difference between three PF-06700841 dose groups versus placebo in the proportions of LLDAS responders at Week 52, among participants with active SLE who are not impacted by the COVID-19 or Russia-Ukraine crises at Week 52.* *<u>Treatment</u>: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group*

Objectives	Endpoints	Estimands
		<u>Population</u>: Participants with active SLE as defined by the FAS *<u>Variable</u>: LLDAS response at Week 52* *<u>Intercurrent Events</u>:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing, and such participants will be imputed as non-responders for the other secondary analysis.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the other secondary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the other secondary analysis.* • *<u>Population Level Summary</u>: Treatment differences between three PF-06700841 dose groups and placebo in the proportions of LLDAS responders at Week 52*
• *To compare the corticosteroid use (prednisone or equivalent) in PF-06700841 treated participants relative to placebo*	• *Proportion of participants achieving a reduction in prednisone (or equivalent) dose to ≤7.5 mg/day at Week 52 and sustained for 12 weeks prior to Week 52, in the subset of participants on prednisone >7.5 mg/day (or equivalent) at baseline*	*There is no defined estimand for this endpoint.*
• *To evaluate the efficacy of PF-06700841 compared to placebo in participants with active SLE and with sustained reduction of oral corticosteroids*	• *Proportion of participants achieving SRI-4 response with dose of prednisone (or equivalent) reduced to ≤7.5 mg/day and sustained for 12 weeks at Week 52, in the subset of participants with prednisone dose*	*<u>E4</u>: This composite estimand is defined as a population average treatment difference between three PF-06700841 dose groups versus placebo in the proportions of participants with active SLE who achieved the binary SRI-4 endpoint with sustained prednisone dose reduction to ≤7.5 mg/day (or equivalent) for 12 weeks at Week 52 starting at Week 40, among participants*

Objectives	Endpoints	Estimands
	>7.5 mg/day (or equivalent) at baseline	*with active SLE who are not impacted by the COVID-19 or Russia-Ukraine crises at Week 52.* *<u>Treatment</u>: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *<u>Population</u>: Participants with active SLE as defined by the FAS and with prednisone dose >7.5 mg/day (or equivalent) at baseline* *<u>Variable</u>: SRI-4 at Week 52 with sustained prednisone dose reduction ≤7.5 mg/day (or equivalent) for 12 weeks at Week 52 starting at Week 40* *<u>Intercurrent Events</u>:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing, and such participants will be imputed as non-responders for the other secondary analysis.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the other secondary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the other secondary analysis.* *<u>Population Level Summary</u>: Treatment differences between three PF-06700841 dose groups and placebo in proportions of dual SRI-4 and sustained prednisone reduction responders at Week 52*
• *To evaluate the Cutaneous Lupus Erythematosus Disease Area and Severity Index activity (CLASI-A) score in the subset of participants with baseline CLASI-A score ≥10 in PF-06700841 treated*	• *Proportion of Participants with Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI-A) Total Activity Score ≥10 at Baseline with ≥50% Reduction in*	*There is no defined estimand for this endpoint.*

Objectives	Endpoints	Estimands
participants relative to placebo	*CLASI-A Total Activity Score at Week 52*	
• *To evaluate the effect on fatigue of PF-06700841 treated participants relative to placebo*	• *Change from baseline in the total scores of Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT-F) at Week 52*	*E5: This hypothetical estimand is defined as a population average treatment difference between three PF-06700841 dose groups and placebo in the FACIT-F score at Week 52 among participants with active SLE who are not impacted by the COVID-19 pandemic or Russia-Ukraine conflict at Week 52, as if none of the specified intercurrent events have occurred.* *Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *Population: Participants with active SLE as defined by the FAS* *Variable: Change from baseline in FACIT-F total scores at Week 52* *Intercurrent Events:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing and assumed missing at random by the analysis model.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the other secondary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the other secondary analysis.* • *Population Level Summary: Mean differences between three PF-06700841 dose groups and placebo in the change from baseline in FACIT-F score at Week 52*

Objectives	Endpoints	Estimands
• *To evaluate the effect on health-related quality of life of PF-06700841 treated participants relative to placebo*	• *Change from baseline in the individual domain scores of the Lupus Quality of Life (LupusQol) at Week 52*	*E6: This hypothetical estimand is defined as a population average treatment difference between three PF-06700841 dose groups and placebo in the LupusQol individual domain scores at Week 52 among participants with active SLE who are not impacted by the COVID-19 pandemic or Russia-Ukraine conflict at Week 52, as if none of the specified intercurrent events have occurred.* *Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *Population: Participants with active SLE as defined by the FAS* *Variable: Change from baseline in the individual domain scores of LupusQoL at Week 52* *Intercurrent Events:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent event will be set to missing and assumed missing at random by the analysis model.* *Participants who miss the Week 52 visit due to the COVID-19 pandemic will be excluded from the other secondary analysis.* *Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the other secondary analysis.* • *Population Level Summary: Mean differences between three PF-06700841 dose groups and placebo in the change from baseline in LupusQoL individual domain scores at Week 52*
• *To assess time to first severe flare in PF-06700841 treated participants relative to placebo*	• *Time-to-first severe flare as measured by the modified Safety of Estrogen in Lupus: National Assessment*	• *E7: This estimand is defined as a population average treatment hazard ratio between three PF-06700841 dose groups and placebo, based on the Cox PH model, for the occurrence of the first severe flare*

Objectives	Endpoints	Estimands
	(SELENA)-Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) Flare Index (m-SSFI)	*event, among participants with active SLE who are not impacted by the COVID-19 pandemic or Russia-Ukraine conflict at Week 52.* • *Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group* *Population: Participants with active SLE as defined by the FAS.* *Variable: Time-to-first severe flare measured by the modified SELENA-SLEDAI Flare Index* *Intercurrent Events:* *Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected more than 28 days after the intercurrent event will be set to missing, and such participants will be censored for the other secondary analysis at the time the intercurrent event occurred. If participants discontinue from study treatment due to COVID-19 or Russia/Ukraine, they will be censored at the time of discontinuation.* *Population Level Summary: Hazard ratio of first severe flare between three PF-06700841 dose groups and placebo*
• *To evaluate the safety and tolerability of PF-06700841 dose levels versus placebo*	• *Incidence of treatment-emergent adverse events (AEs)* • *Incidence of serious AEs (SAEs) and AEs leading to discontinuation* • *The incidence of clinically significant abnormalities in vital signs and ECGs* • *The incidence of clinically significant abnormalities in clinical laboratory values*	• *There is no defined estimand for these endpoints.*

Objectives	Endpoints	Estimands
CCI	[redacted]	[redacted]

Objectives	Endpoints	Estimands
CCI		

Objectives	Endpoints	Estimands
CCI		

Objectives	Endpoints	Estimands
CCI	[redacted]	[redacted]

2.1.1. Primary Estimand

The primary endpoint of the study is the proportion of participants achieving Systemic Lupus Erythematosus Responder Index (SRI) change of 4 (SRI-4) at Week 52, which is a composite endpoint where a responder must meet criteria in the assessment of Systemic Lupus

Erythematosus Disease Activity Index-2000 (SLEDAI-2K), British Isles Lupus Assessment Group (BILAG) and the Physicians Global Assessment (PhGA).

The primary estimand (E1) for the primary endpoint is defined according to the primary objective and is in alignment with the primary endpoint of SRI-4 response at Week 52. The primary estimand of this study will use the composite estimand strategy and estimate the population average treatment difference versus placebo on a binary endpoint. It includes the following 5 attributes:

- Population: Participants with active Systemic Lupus Erythematosus (SLE) as defined by the full analyses set (FAS) to reflect the targeted population
- Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group
- Variable: SRI-4 response at Week 52
- Intercurrent Events:

  Treatment discontinuation: Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent events will be set to missing, and such participants will be imputed as non-responders for the primary analysis. Refer to Section 5.3.1 for more information on missing data handling.

  COVID-19 pandemic: Participants who miss the Week 52 visit due to COVID-19 will be excluded from the primary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment, withdraw from the study, or dies (Meyer 2020) due to COVID-19 will be excluded from the primary analysis after the treatment discontinuation visit or study withdrawal visit.

  Russia-Ukraine Conflict: Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the primary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment or withdraws from the study due to the Russia-Ukraine Conflict will be excluded from the primary analysis after the treatment discontinuation visit or study withdrawal visit.

- Population Level Summary: Treatment difference between three PF-06700841 dose groups versus placebo in proportions of SRI-4 responders at Week 52

2.1.2. Treatment Policy Estimand

Supportive to the primary estimand, a treatment policy estimand analysis will be conducted to analyze the proportion of participants achieving SRI-4 response at Week 52. The treatment

policy estimand estimates the effect, regardless of treatment adherence. Refer to Section 5.3.1 for more information on missing data handling. It possesses the following 5 attributes:

- Population: Participants with active Systemic Lupus Erythematosus (SLE) as defined by the FAS to reflect the targeted population
- Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group
- Variable: SRI-4 response at Week 52
- Intercurrent Events:

  COVID-19 pandemic: Participants who miss the Week 52 visit due to COVID-19 will be excluded from the primary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment, withdraw from the study, or dies (Meyer 2020) due to COVID-19 will be excluded from the treatment policy analysis after the treatment discontinuation visit or study withdrawal visit.

  Russia-Ukraine Conflict: Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the primary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment or withdraws from the study due to the Russia-Ukraine Conflict will be excluded from the treatment policy analysis after the treatment discontinuation visit or study withdrawal visit.

- Population Level Summary: Treatment difference between three PF-06700841 dose groups versus placebo in proportions of SRI-4 responders at Week 52

2.1.3. Key Secondary Estimands

The key secondary endpoint of the study is the proportion of participants achieving the British Isles Lupus Assessment Group-Based Composite Lupus Assessment (BICLA) response at Week 52. BICLA response is a composite endpoint where a responder must meet criteria defined in the assessment of BILAG, SLEDAI-2K, and PhGA.

The key secondary estimand of this study, E2, will use the composite estimand strategy. E2 possesses the following 5 attributes:

- Population: Participants with active Systemic Lupus Erythematosus (SLE) as defined by the FAS
- Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group
- Variable: BICLA response at Week 52

- Intercurrent Events:

  Treatment discontinuation: Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected after the intercurrent events will be set to missing, and such participants will be imputed as non-responders for the key secondary analysis. Refer to Section 5.3.1 for more information on missing data handling.

  COVID-19 pandemic: Participants who miss the Week 52 visit due to COVID-19 will be excluded from the key secondary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment or withdraw from the study due to COVID-19 will be excluded from the key secondary analysis after the treatment discontinuation visit or study withdrawal visit.

  Russia-Ukraine Conflict: Participants who miss the Week 52 visit due to the Russia-Ukraine Conflict will be excluded from the key secondary analysis (i.e., excluded from the numerator and denominator). Participants who discontinue treatment or withdraw from the study due to the Russia-Ukraine Conflict will be excluded from the analysis after the treatment discontinuation visit or study withdrawal visit.

- Population Level Summary: Treatment differences between three PF-06700841 dose groups versus placebo in proportions of BICLA responders at Week 52

2.1.4. General Comments on Other Secondary Estimands

Estimands are also defined corresponding to other secondary objectives. In general, population will consist of participants with active SLE as defined by their respective analyses sets to reflect the target population, except those estimands for subgroups of participants.

For binary endpoints, a composite estimand similar to that for the primary endpoint will be used, with treatment discontinuation as the intercurrent event and the data post the intercurrent events excluded from analysis. A study participant will be considered as a non-responder after the occurrence of the intercurrent event. This will be the main estimand. As a supportive analysis, a treatment policy estimand may be used and defined similarly as in Section 2.1.2: all the data collected during the study will be used to derive the response (i.e., no intercurrent event will be defined for early treatment discontinuation or study withdrawal before the timepoint of interest).

For continuous endpoints, a hypothetical estimand (main estimand) will be used, with treatment discontinuation as the intercurrent event and the data collected post the intercurrent event censored in the analysis. As a supportive analysis, a treatment policy estimand may be used to assess continuous endpoints; data from participants impacted by the COVID-19 pandemic or the Russia-Ukraine Conflict will be removed, otherwise all data collected,

including from participants who discontinued treatment for any reason other than the COVID-19 pandemic or the Russia-Ukraine Conflict, will be used in the analysis.

Estimands E3 and E4 will be used to evaluate the binary endpoints of LLDAS response at Week 52 and proportion of participants achieving a reduction in prednisone (or equivalent) dose to ≤7.5 mg/day at Week 52 and sustained for 12 weeks prior to Week 52, respectively. Both estimands are defined as a composite estimand like that for the primary endpoint of SRI-4 response at Week 52. Intercurrent events will be handled in the same manner as for the primary estimand, with participants being removed from the given analysis when they have missed their Week 52 visit due to the COVID-19 pandemic or the Russia-Ukraine Conflict; and, with any data collected after treatment discontinuation being set to missing. Participants will be imputed as a non-responder after the occurrence of the intercurrent event of treatment discontinuation.

Estimands E5 and E6 will be used to evaluate the continuous endpoints of change from baseline in the FACIT-F total score at Week 52 and change from baseline in the individual domain scores of the LupusQoL at Week 52, respectively. Both estimands are defined as a hypothetical estimand. Intercurrent events will be handled in a similar manner as for the primary estimand, with participants being removed from the given analysis when they have discontinued treatment due to either the COVID-19 pandemic or the Russia-Ukraine Conflict; following a treatment discontinuation for any other reason, any data collected after the intercurrent event will be set to missing.

Estimand E7 is a hypothetical estimand that will be used to evaluate the time-to-first severe flare endpoint. E7 possesses the following 5 attributes:

- Population: Participants with active Systemic Lupus Erythematosus (SLE) as defined by the FAS.
- Treatment: Three dose groups of PF-06700841 (45 mg QD, 30 mg QD, 15 mg QD), one placebo group
- Variable: Time to first severe flare measured by the modified SELENA-SLEDAI Flare Index
- Intercurrent Events:

  Treatment discontinuation: Discontinuation of the study treatment prior to the Week 52 visit or study withdrawal due to any reason other than COVID-19 or the Russia-Ukraine Conflict will be considered an intercurrent event. Data collected more than 28 days after the intercurrent event will be set to missing, and such participants will be censored for the other secondary analysis at the time the intercurrent event occurred. Participants who discontinue study treatment due to COVID-19 or the Russia-Ukraine Conflict will be censored at the time of discontinuation.

- Population Level Summary: Hazard ratios of first severe flare rates between three PF-06700841 dose groups versus placebo

2.2. Study Design

This is a Phase 2b, double-blind, randomized, placebo-controlled, parallel design, multicenter, dose--ranging- study to assess the efficacy and safety of PF-06700841 in participants with active, moderate to severe generalized SLE. After an up to 7--week screening period, eligible participants will be randomly assigned to one of the four treatment groups (A-D in Table 3) in a 1:2:2:2 ratio such that participants will receive either 1 of three PF06700841 QD dose levels (15 mg, 30 mg- and 45 mg) or placebo every day for 52 weeks. All participants who are randomized should be encouraged to remain in the study to at least the end of the double-blind period to complete safety and efficacy assessments in order to reduce missing data as much as possible. It is expected that the only reasons for which a participant will withdraw from the study will be for withdrawal of consent or lost to follow--up.

Approximately 350 randomized participants (100 participants per treatment group except 50 participants for the PF-06700841 15 mg QD dose group) will participate in this study at approximately 185 investigative sites worldwide. Participants will participate in this study for approximately up to 63 weeks. This includes up to a 7-week screening period, a 52-week treatment period, and a 4-week follow-up period.

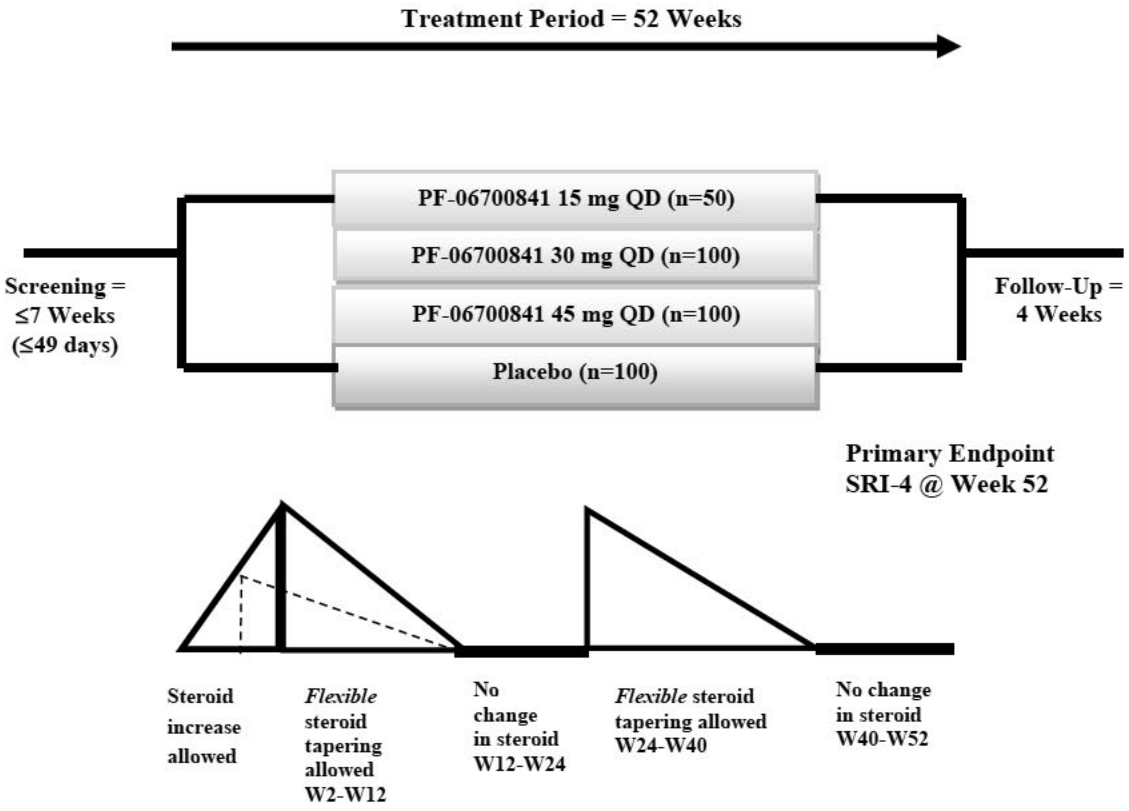
2.3. Changes from the Protocol

This SAP is based on the most recent version of the PF-06700841 Clinical Protocol Amendment 8, dated 15 June 2023. Contained within this section is a summary of planned analyses, conventions, or definitions which are inconsistent with the protocol.

- Anchor-based analyses to derive a responder endpoint for the FACIT-F will not be performed, therefore there will be no analyses for the endpoint of FACIT-F response, nor for the endpoint of SRI-4 and FACIT-F dual response.
- [redacted]

- CCI

Figure 1. Study Design



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

Study endpoints are listed in this section. See Section 3.4 for baseline definition.

3.1. Primary Endpoint

The primary endpoint is proportion of participants who achieve the SRI change of 4 (SRI-4) at Week 52. To be classified as a responder, participants must meet all of the following criteria compared with baseline:

- ≥4-point reduction in the SLEDAI-2K score, and
- No new BILAG A organ domain score or 2 new BILAG B organ domain scores, and
- No worsening (<0.3-point increase) in PhGA score

3.1.1. SLEDAI-2K

The SLEDAI-2K contains 24 individual manifestations, with each scored during a visit based on the presence or absence of the manifestation at that visit. Certain manifestations are

scored higher than others; refer to Section 10.11 of the protocol for the manifestations and their point totals. The SLEDAI-2K score is calculated as the sum of individual scores.

3.1.2. BILAG 2004

The BILAG is designed to capture disease activity related to SLE, based on considering 97 different clinical manifestations organized within 9 organ domains. These manifestations are scored based on whether the symptoms observed during a visit are Not present (0), Improving (1), Same (2), Worse (3), or New (4) compared to 4 weeks earlier. Refer to Section 10.12 of the protocol for the manifestations and organ domains. An organ domain is assigned a score from A to E depending on how the manifestations within that domain have been scored. An "A" score represents an increase in disease activity sufficient to normally require intensification of therapy with steroids or immunosuppressants. A "B" score represents moderate, reversible manifestations requiring antimalarials, NSAIDs, or low dose steroids. A "C" score reflects mild, stable disease. A "D" score indicates no current disease activity in a previously affected organ system. An "E" score indicates the organ system has never been involved. Numerical scoring is derived from the alphabetical score as: A = 12, B= 8, C= 1 and D/E = 0.

3.1.3. PhGA

Physician Global Assessment (PhGA) is a measure of worsening in the participants general health status. The investigator will mark a visual analog scale answering the question "Evaluate the participants lupus activity since the last visit using the visual analog scale" indicating the participants overall disease activity at a particular visit. The investigator's response will be recorded using a visual analog scale (VAS) by placing a mark on the scale between 0 (none), 1 (mild), 2 (moderate), and 3 (severe).

The VAS will need to be rescaled prior to any calculation and analysis. The VAS is recorded in terms of length at mark (X in mm) and overall length of line (Y in mm). The default value of Y is 100 mm or 10 cm if it is not captured in the eCRF. The rescaled VAS for use in analysis will be Z=X/Y × 100 mm or X/Y × 10 cm with the exception of PhGA. For PhGA, the re-scaled VAS for use in the analysis will be Z=X/Y × 3.

3.2. Secondary Endpoints

3.2.1. Key Secondary Endpoint

British Isles Lupus Assessment Group Based Composite Lupus Assessment (BICLA) is a composite endpoint that includes the BILAG2004, SLEDAI-2K, and PhGA of disease activity.

To be classified as a BICLA responder, participants must meet all of the following criteria compared with baseline (see study protocol Section 8.1.6):

- BILAG-2004 improvement (all A scores at baseline improved to B/C/D and all B scores improved to C or D), and

- No worsening in disease activity (no new BILAG-2004 A scores or ≤1 new B score), and

- No worsening of total SLEDAI-2K score, and

- No significant deterioration (<10% worsening) in analogue PhGA

3.2.2. Other Secondary Endpoints – Efficacy

- *Proportion of participants achieving the Lupus Low Disease Activity State (LLDAS) at Week 52*

  LLDAS is defined as (1) SLE disease activity index (SLEDAI-2K ≤4, with no activity in major organ systems (renal, central nervous system, cardiopulmonary, vasculitis, fever) and no haemolytic anaemia or gastrointestinal activity; (2) no new lupus disease activity compared with the previous assessment; (3) a Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA)-SLEDAI PhGA (scale 0-3) ≤1; (4) a current prednisolone (or equivalent) dose ≤7.5 mg/daily; and (5) well tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents.

- *Proportion of Participants with Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI-A) Total Activity Score ≥10 at Baseline with ≥50% Reduction in CLASI-A Total Activity Score*

  CLASI has separate scores for activity and damage, which are based upon degree of erythema, scale, mucous membrane lesions, and non-scarring alopecia. Increased weight is assigned to the face and neck, relative to other less frequently involved parts of the body. Total activity score is computed as the sum of the component score according to the assigned weights.

- *Time-to-first severe flare as measured by the mSSFI*

  This index categorizes SLE flares as either “mild or moderate”, based on 5 variables, or “severe”, based on 6 variables as defined in the study protocol Appendix 19. All severe SLE flares will be reviewed by a blinded central adjudication committee. The decisions on severe SLE flares made by the central adjudication committee will be databased and used for the analyses.

  Corresponding to the main estimand for this endpoint (see Section 2.1.4) the risk period will be the time from Day 1 to the minimum of (end of treatment +28 days, death, participant’s last study visit date), except for patients who discontinue due to COVID or due to the Russia/Ukraine conflict, whereby their risk period will be from Day 1 to the day they discontinue from the study.

For participants who experience the event within the risk period, the first event will be used in the analysis. For participants who did not experience the event or experienced the event but outside the risk period, the participants will be censored at the end of the risk period.

3.2.3. Other Secondary Endpoints – Steroid Use Related

- *Proportion of participants achieving a reduction in prednisone (or equivalent) dose to ≤7.5 mg/day at Week 52 and sustained for 12 weeks prior to Week 52, in the subset of participants on prednisone >7.5 mg/day (or equivalent) at baseline*

- *Proportion of participants achieving SRI-4 response with dose of prednisone (or equivalent) reduced to ≤7.5 mg/day and sustained for 12 weeks at Week 52, in the subset of participants with prednisone dose >7.5 mg/day (or equivalent) at baseline*

3.2.4. Other Secondary Endpoints – Patient Reported Outcome

- *Change from baseline in the total scores of Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT---F) at Week 52*

  The Functional Assessment of Chronic Illness Therapy - -Fatigue (FACIT--F) version 4 scale is a patient completed questionnaire consisting of 13 items that assess fatigue each measured on a 4-point Likert scale over the past 7 days. Instrument scoring yields a range from 0 to 52, with higher scores representing less fatigue.

- *Change from baseline in the individual domain scores of the Lupus Quality of Life (LupusQoL) at Week 52*

  The LupusQoL questionnaire is a validated questionnaire used to evaluate SLE -specific concepts of SLE as reported by patients. It consists of 34 items in 8 different domains: physical health, emotional health, body image, pain, planning, fatigue, intimate relationship, and burden to others measured over a 5-point Likert response (0-4) over the past 4 weeks. To facilitate analysis and result interpretation, the individual domain scores are transformed to a 0 to 100 scale wherein higher scores indicate better quality of life.

CCI [redacted]

[redacted]

[redacted]

CCI

CCI

[redacted]

3.4. Baseline Variables

Demographics will be collected at screening and medical history will be collected at screening and baseline. Prior SLE and non-SLE medications will be collected at screening.

Baseline is defined as pre-dose on Day 1. Data from the screening period may be used if Day 1 data are missing. If multiple data points are available, we will use the last observation before Day 1 dosing. Exceptions to this general baseline definition will be noted.

3.4.1. Stratification Factor

Participants will be stratified by screening disease severity (total SLEDAI-2K score <10 versus ≥10) and screening anti-dsDNA status (anti-dsDNA antibody level > ULN at the central laboratory versus other) to form a total 4 strata:

- Adjudicated screening total SLEDAI-2K score <10 and screening anti--dsDNA antibody level ≤ ULN

- Adjudicated screening total SLEDAI-2K score <10 and screening anti--dsDNA antibody level > ULN
- Adjudicated screening total SLEDAI-2K score ≥10 and screening anti--dsDNA antibody level ≤ ULN
- Adjudicated screening total SLEDAI-2K score ≥10 and screening anti--dsDNA antibody level > ULN

The strata derived from the data in the clinical database will be used in all the analyses where applicable. For the primary endpoint of SRI-4 at Week 52 only, a sensitivity analysis may be performed using the strata from the randomization system.

Prior to unblinding, any of the randomization substrata with counts below 20 will be pooled across either of the adjacent substrata in a manner which ensures balance across the resulting substrata.

3.5. Safety Endpoints

Safety endpoints include:

- Incidence of treatment-emergent adverse events (AEs)
- Incidence of serious AEs (SAEs) and AEs leading to treatment discontinuation
- The incidence of clinically significant abnormalities in vital signs
- The incidence of clinically significant/abnormal ECGs
- The incidence of clinically significant abnormalities in clinical laboratory values

3.5.1. Adverse Events

Adverse events will be classified as treatment emergent if the event start date occurred on or after the treatment start date, and before the treatment end date + 28 days. Events which are missing their start date will be classified as treatment emergent regardless of the imputed start date.

3.5.2. Laboratory Data

The laboratory tests will be performed at time points identified in the Schedule of Activities. Unscheduled clinical labs may be obtained at any time during the study to assess any perceived safety concerns at the investigator's discretion. Laboratory data will be summarized in participants while they are on treatment. If a participant discontinues treatment before Week 52, any laboratory data collected up to 28 days after the early treatment discontinuation will be summarized, and any data collected after will be listed only.

3.5.3. Vital Signs

Vital signs (blood pressure, pulse rate, respiratory rate, and temperature) will be measured after 5 minutes of rest as indicated in the Schedule of Activities. If a participant discontinues treatment before Week 52, any vital sign data collected up to 28 days after the early treatment discontinuation will be summarized, and any data collected after will be listed only.

3.5.4. Electrocardiograms

Standard 12-lead ECGs should be collected at times specified in the Schedule of Activities.

The Day 1 ECG values (average of the triplicate measures) will serve as each participant's baseline values. If Day 1 ECG is missing, then screening ECG will serve as baseline value. Data from central ECG readings will be used for statistical analyses.

Baseline and change from baseline in ECG will be summarized descriptively. Categorical summaries of the ECG data will also be provided.

If a participant discontinues treatment before Week 52, any ECG data collected up to 28 days after the early treatment discontinuation will be summarized, and any data collected after will be listed only.

3.6. COVID-19 Impact

In order to describe the impact of COVID-19, participants who discontinue from the study due to COVID-19 as the primary reason will be summarized in the disposition table. The final set of participants that are deemed as having discontinued due to COVID-19 will be determined by the study clinicians and agreed upon by the study team using an Analysis Population Exclusion list prior to unblinding the data.

The specific reasons for participants who discontinue from the study due to COVID-19, if collected in detail, such as participant decision, participant self-isolation, non-compliance with protocol, transportation issue, site restrictions, physician decision, site closed, lack of drug at site, sponsor decision, and severity of COVID-19 infection will be presented in the disposition listing.

The number and percentage of participants with at least one important PD and with at least one PD associated with COVID-19 will be summarized by treatment group, deviation category, and deviation sub-category using the FAS. Important PDs and PDs related to COVID-19 will be summarized similarly overall using the FAS.

A listing of all PDs will be provided. Important PDs and COVID-19 related PDs will be flagged in this listing.

3.7. Russia-Ukraine Conflict Impact

In order to describe the impact of the Russia-Ukraine Conflict, participants who discontinue from the study due to Russia-Ukraine Conflict as the primary reason will be summarized in the disposition table. The final set of participants that are deemed as discontinuation due to the Russia-Ukraine Conflict will be determined by the study clinicians and agreed upon by the study team using an Analysis Population Exclusion list prior to unblinding the data.

The specific reasons for participants who discontinue from the study due to the Russia-Ukraine Conflict, if collected in detail, such as participant decision, participant self-isolation, non-compliance with protocol, transportation issue, site restrictions, physician decision, site closed, lack of drug at site, and sponsor decision will be presented in the disposition listing.

The number and percentage of participants with at least one important PD and with at least one PD associated with the Russia-Ukraine Conflict will be summarized treatment group, deviation category, and deviation sub-category using the FAS. Important PDs and PDs related to the Russia-Ukraine Conflict will be summarized similarly overall using the FAS.

A listing of all PDs will be provided. Important PDs and Russia-Ukraine Conflict related PDs will be flagged in this listing.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database. The population sets will be tabulated and listed.

Prior to database lock, the following events occurred. Two sites were closed due to GCP violations. These two sites included 10 participants screened, 5 of whom were randomized into the study. At another site, 8 participants were screened, of which 5 were randomized into the study before the PI moved to another practice which could not accommodate clinical research. Prior to leaving the site, the PI discontinued all the participants prematurely from study intervention and the study. Data from these participants will be excluded from the Full, PK Concentration, and PD Analysis Sets. An additional row will be provided in the Participant Evaluation Groups table describing “Excluded due to data quality concerns” in which these participants will be tabulated. Additional details related to these events will be described in the CSR.

Population	Description
Full Analysis Set	All participants randomly assigned to study intervention and have received at least one dose of study intervention. The primary efficacy population for this study is defined by the FAS participants. Participants will be analyzed according to the product to which they were randomized. A narrative will be provided in the clinical study report (CSR) for any participant who is treated but not randomized.

Population	Description
CCI [redacted]	[redacted]
[redacted]	[redacted]
Safety Analysis Set	All participants randomly assigned to study intervention and who take at least one dose of study intervention. Participants will be analyzed according to the product they actually received on Day 1.

5. GENERAL METHODOLOGY AND CONVENTIONS

The primary analysis will be performed when the last participant has completed the last visit.

5.1. Hypotheses and Decision Rules

As specified in the protocol (Section 9.1.2), a sequentially rejective multiple testing procedure (Figure 2) will be performed to compare each of the PF-06700841 dose groups with the placebo group for both primary endpoint of SRI-4 response at Week 52 and the key secondary endpoint of BICLA response at Week 52 in six hypotheses (Table 2). This procedure will preserve an overall one-sided Type I error rate of no more than 0.025.

Table 2. Families of Hypotheses

Notation	Hypotheses	Family of Hypotheses
HvP1	H0: SRI-4 response rates at Week 52 are the same between the PF-06700841 45 mg dose group and the placebo group H1: The SRI-4 response rate at Week 52 for the PF-06700841 45 mg dose group is greater than the SRI-4 response rate for the placebo group	FA
HvP2	H0: BICLA response rates at Week 52 are the same between the PF-06700841 45 mg dose group and the placebo group H1: The BICLA response rate at Week 52 for the PF-06700841 45 mg dose group is greater than the BICLA response rate for the placebo group	FA
MvP1	H0: SRI-4 response rates at Week 52 are the same between the PF-06700841 30 mg dose group and the placebo group H1: The SRI-4 response rate at Week 52 for the PF-06700841 30 mg dose group is greater than the SRI-4 response rate for the placebo group	FA
MvP2	H0: BICLA response rates at Week 52 are the same between the PF-06700841 30 mg dose group and the placebo group H1: The BICLA response rate at Week 52 for the PF-06700841 30 mg dose group is greater than the BICLA response rate for the placebo group	FA

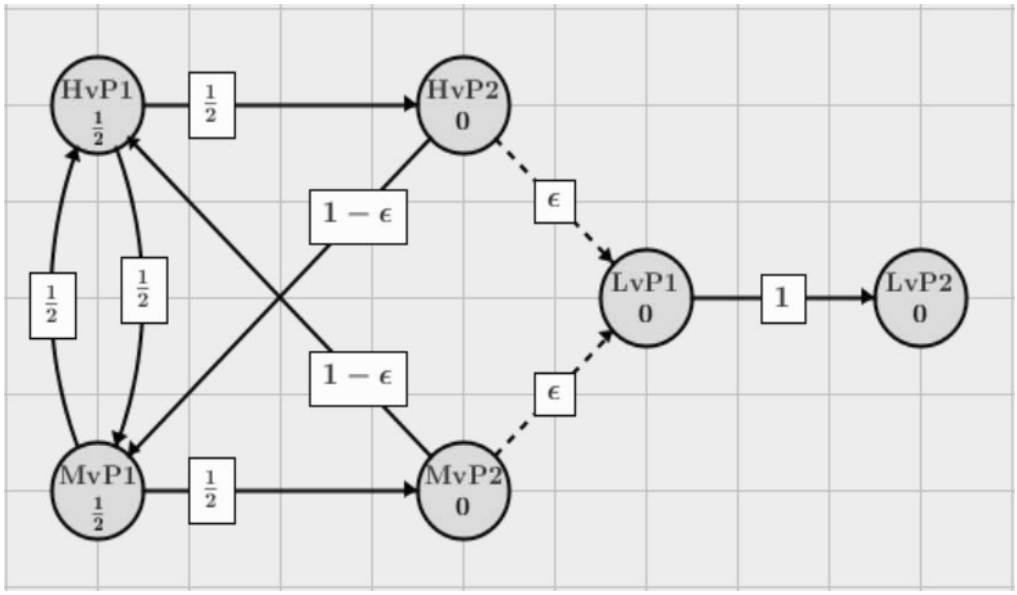
Table 2. Families of Hypotheses

Notation	Hypotheses	Family of Hypotheses
LvP1	H0: SRI-4 response rates at Week 52 are the same between the PF-06700841 15 mg dose group and the placebo group H1: The SRI-4 response rate at Week 52 for the PF-06700841 15 mg dose group is greater than the SRI-4 response rate for the placebo group	FB
LvP2	H0: BICLA response rates at Week 52 are the same between the PF-06700841 15 mg dose group and the placebo group H1: The BICLA response rate at Week 52 for the PF-06700841 15 mg dose group is greater than the BICLA response rate for the placebo group	FB

The family of hypotheses FA will serve as the gatekeeper for the testing of hypotheses in the family FB. The hypotheses in the family FB won't be formally tested unless all null hypotheses in the family FA are rejected. The ε in the Figure 2 is an infinitesimally small weight for the gatekeeping strategy. This testing procedure is a closed testing procedure and controls the family-wise type-I error rate for the multiple testing of both primary and the key secondary endpoints (Bretz, et al 2009).[1]

The testing procedure will start with the primary efficacy analysis of SRI-4 in either the high dose (45 mg) or medium dose (30 mg) arms, depending on which arm produces the smaller observed p-value. Note that the directed edge indicates the direction and associated weight that corresponds to the passing of α. Hypothesis nodes that are rejected are removed from the diagram and the weights along the directed edges get normalized to fulfill the regularity conditions of the graphical testing procedure. The value ε is set as an infinitesimally small number close to 0 but gets normalized to 1 once all other directed edges that link to the node besides the edge with ε are removed.

Figure 2. Testing Procedure



5.2. General Methods

5.2.1. Analyses for Binary Endpoints

In general, the data for all responder-type binary endpoints will be summarized by treatment group and visit in tabular and/or graphic format with descriptive statistics, including N (number of participants evaluable for the endpoint in the corresponding treatment group/visit), n (number of responders), response rate (%), standard error of the response rate, and 95% confidence interval (CI) based on Wald-type large sample approximation with correction for discrete distribution. The treatment comparisons between each dose group of PF-06700841 and the placebo group will also be summarized with treatment differences and associated 95% CIs based on the Cochran-Mantel-Haenszel (CMH) method using the Mantel-Haenszel estimate of the binomial difference along with the Sato variance estimator (Sato, 1989).).

When response rates of 0% or 100% are observed in both treatments in comparison and in all strata, no formal comparison will be performed. Estimated response rate of 0% or 100% will be reported as observed. Standard error will be reported as 0.

The primary endpoint of the study is a binary endpoint (i.e., SRI-4 response at Week 52). Both this endpoint and the key secondary endpoint for BICLA response at Week 52 will be formally tested as follows:

- The superiority testing of each PF-06700841 dose group over placebo in the primary endpoint of SRI-4 (or BICLA) responder rate at Week 52 is based on the CMH approach adjusting for the stratification factors(see Section 3.4.1).

- The SRI-4 (or BICLA) responder rate will be summarized by treatment group, and treatment effect estimation will be presented as the treatment difference in the responder rate between each PF-06700841 dose group versus placebo. These estimates will be reported with 95% confidence intervals, and the corresponding one-sided and two-sided p-values.

- The multiplicity adjustment will be made for multiple superiority testing using the sequentially rejective testing method (Section 5.1).

Binary secondary[CCI] endpoints will be analyzed similarly to the primary endpoint, with the exception that no formal testing will be made for treatment comparisons thus no multiplicity adjustment is applied.

For selected binary endpoints, treatment difference will also be summarized with 95% confidence interval using generalized linear marginal model for repeated measures (GLMMRM) model (see details in Section 6.1.1.2). More details will be described for each endpoint in Section 6.

5.2.2. Analyses for Continuous Endpoints

In general, the data for all continuous endpoints will be summarized by treatment group and visit in tables containing descriptive statistics (N, mean, standard deviation, minimum, median, and maximum) for actual and change from Baseline (or percent change from Baseline) values for those endpoints measured with respect to Baseline.

Repeated measures, continuous or ordered-categorical (analyzed as continuous) endpoints will be analyzed as change from baseline using a mixed model for repeated measures (MMRM) that includes fixed effects of treatment group, visit, treatment-group by visit interaction, stratification factor at randomization, stratification-factor by visit interaction, baseline value, and baseline-value by visit interaction. A common unstructured variance-covariance matrix will be used, provided the model converges, otherwise an alternative covariance structure (e.g., heterogeneous compound symmetry [CSH]) will be used and noted within the given output. The Kenward-Roger degrees of freedom approximation will be used. Comparison of the 45 mg, 30 mg, and 15 mg arms of PF06700841 with the placebo arm will be summarized (using least squares means [LSM] of the each of the treatments, LSM of the treatment difference, 2-sided p-value and 95% CI) at each visit. If the Baseline is missing or if there are no post-baseline measurements, the participant will be excluded from this analysis.

The secondary endpoints [CCI], shown below, will be analyzed using a MMRM:

- *Change from baseline in CLASI-A score in participants with baseline CLASI-A score ≥10 at all scheduled post-baseline visits*

- *Change from baseline in the total scores of Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT-F) at all scheduled post-baseline visits (including Week 52)*
- *Change from baseline in the individual domain scores of the Lupus Quality of Life (LupusQoL) at all scheduled post-baseline visits (including Week 52)*

CCI

5.2.3. Analyses for Time-to-Event Endpoints

Time to first severe flare as measured by the modified SELENA-SLEDAI Flare Index (mSSFI) will be tested to compare treatment groups with placebo based on a logrank test with estimates from the Kaplan-Meier product limit method. The corresponding Kaplan-Meier plot will also be used to visualize differences in the incidence of flares over time. Cox regression model will be used to estimate the hazard ratio and its 95% Cis for the treatment effect using treatment group and one stratification variable (based on the strata levels and

pooling strategy discussed in Section 3.4.1) as covariates. The number of participants, percentage of participants experienced the event and the number and percentage of participants censored, the cumulative event/censor time, and the incidence rate will be summarized in a table by treatment group.

5.3. Methods to Manage Missing Data

5.3.1. Efficacy Data

Participants with missing data due to treatment discontinuation prior to Week 52 (i.e., participants who have an intercurrent event) for the primary endpoint of SRI-4 response at Week 52 or for the key secondary endpoint of BICLA response at Week 52 will be considered non-responders within those respective primary analyses.

Participants who do not have an intercurrent event but have missing data at Week 52 not resultant from the COVID-19 pandemic or the Russia/Ukraine Conflict will also be considered a non-responder for the primary and key secondary analyses. This method of handling missing response is known as missing response as non-response (MR=NR). The only exception to this rule is for participants who have data available at Week 48 for components used to derive an SRI-4 response or a BICLA response. In this case, LOCF imputation will be used to carry forward any missing components for SRI-4 or BICLA at Week 52 from Week 48, where available. Thus, non-missing components from Week 48 can be combined with partial data at Week 52 to derive the endpoint at Week 52. If all components are missing at Week 52, then it will be imputed using all of Week 48's components, if available. Unless otherwise stated, this LOCF methodology applies only to the primary and key secondary analyses and their corresponding supplementary analyses using the treatment policy estimand but does not apply to analyses involving any other binary endpoints or supplementary analyses that is not specified.

If a participant has an SRI-4 (or BICLA) assessment prior to or on the same date as the discontinuation of the investigational product within the Week 52 visit window, this SRI-4 (BICLA) assessment will be used to evaluate the SRI-4 (BICLA) response as an on-drug observation for Week 52. This same rule will apply for other study visits and other binary endpoints.

In the supplementary analysis of longitudinal binary endpoints using a GLMMRM model (Section 6.1.1.2), missing data will be assumed missing at random.

For secondary endpoints which are binary endpoints, non-responder imputation will be applied for missing data caused by intercurrent events; for continuous events, missing data caused by intercurrent events will be assumed missing at random under a hypothetical estimand. Additionally, any missing values for safety endpoints will not be imputed.

For the time to first severe flare endpoint, the risk period will be the time from Day 1 to the minimum of (end of treatment + 28 days, death, participant's last study visit date), except for patients who discontinue due to COVID or due to the Russia/Ukraine conflict, whereby their risk period will be from Day 1 to the day they discontinue from the study. For participants who experience the event within the risk period, the first event will be used in the analysis. For participants who did not experience the event or experienced the event but outside the risk period, the participants' data will be set to missing and censored at the end of the risk period. In general, any participant who had data missing due to COVID-19 or Russia--Ukraine conflict at any visit of interest will be removed from analysis for that visit (e.g., not included into numerator and denominator in the calculation of proportion of responders at the visit where the data are missing due to COVID-19 or Russia-Ukraine conflict).

CCI

[redacted]

5.3.4. Missing Dates

If the day of the month is missing for any date used in a calculation, the 1st of the month will be used to replace the missing date unless the calculation results in a negative time duration (e.g., date of resolution cannot be prior to date of onset; if replacing resolution date with the 1st of the month results in a negative duration, the resolution date will be set to the onset date).

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint

6.1.1. Proportion of Participants Achieving SRI-4 Response at Week 52

6.1.1.1. Main Analysis

The primary efficacy analysis evaluating the proportion of participants who achieve SRI-4 response at Week 52 will be conducted using the composite estimand strategy, E1, described in Section 2.1.1. This analysis will be performed on the Full Analysis Set. The statistical model to be used is the CMH method, adjusting for stratification factors (Section 5.2.1). Missing data will be handled following the conventions described in Section 5.3.1. The conventions described below will also be followed:

- The number and proportion of participants meeting SRI-4 response at Week 52 will be presented for each treatment group.
- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% confidence interval for risk difference adjusting by CMH weights will be presented.
- Raw p-values (using the sequentially rejective testing approach described in Section 5.1) will be presented for the comparison of each PF-0670081 dose group and the placebo group.
- A bar plot of proportion of participants achieving SRI-4 response at Week 52 with corresponding 95% CIs by treatment arm will be presented.
- A plot of proportion of participants meeting SRI-4 with the corresponding SEs over time by treatment will be presented.

6.1.1.2. Supportive Analyses

A similar analysis as for the primary analysis but using the Treatment Policy estimand will be performed (Section 2.1.2).

A supplementary analysis using generalized linear marginal model for repeated measures (GLMMRM) will be performed for the longitudinal SRI-4 endpoint data. This model will include treatment, visit (discrete), treatment-by-visit interaction as fixed effects; the dependent response variable will be the logit of the probability of SRI-4 response. Note that the model includes fixed effects of treatment, visit, and treatment-by-visit interaction only to ensure the estimated response rates and the treatment difference have a population average or marginal interpretation. A common AR(1) variance-covariance matrix for all treatment groups will be used to model the variability among longitudinal observations within a participant. If there are any concerns with model convergence, covariance structures of compound symmetry will be attempted.

Unlike the non-responder imputation for missing data in the primary analysis, missing at random is assumed for missing data in the GLMMRM analysis. SRI-4 response rates and odds of response for each treatment group by visit will be estimated from the model. Odds ratios and associated confidence intervals for each PF-06700841 dose group compared to placebo by visit will be generated. The delta method will be applied to generate the confidence intervals for the treatment difference in SRI-4 response rates by visit for each dose group of PF-06700841 compared to placebo. Odds ratios and treatment differences (compared to placebo) with 2-sided 95% Cis will be plotted by treatment over time.

Tipping Point Analyses, using the Primary Estimand, and using the Treatment Policy Estimand

The impact of dropouts on the primary analysis will be evaluated based on a tipping point analysis, using model-based multiple imputation. Tipping point analysis is a supportive methodology to analyze longitudinal data of a binary endpoint and imputes missing values to assess the robustness of the missing at random (MAR) assumption. In this tipping point analysis, a single saturated generalized linear mixed effect model is used as the imputation model. Estimation of the model parameters is performed under the Bayesian framework using Markov Chain Monte Carlo (MCMC) methods.

Separately, imputation of the primary endpoint at Week 52 will also be done for patients who discontinued from study after randomization prior to the Week 52 visit. A series of imputation models will be fit for patients assigned to each arm of PF-06700841 and the placebo. The primary endpoint data will be imputed based on the predictive distribution of the generalized linear mixed effect model, in which the probabilities of achieving SRI-4 at Week 52 will be sampled from independent distributions based on study arm and over a grid of rates of mean missingness for each comparison of treatment arm of PF-06700841 with the placebo. The completed imputed data sets will be analyzed using the same approach as the primary analysis in Section 5.2.1. This will be repeated for 100 data sets. The multiple imputation analysis will be applied to each setting in the two-way tipping point analysis. More detailed descriptions are provided in the Appendix 1. Participants who are impacted at Week 52 by the COVID-19 pandemic or the Russia-Ukraine Conflict will be removed from the analysis.

Subgroup Analyses

Subgroup analyses for different races, geographic regions, disease durations, and oral corticosteroid dose at baseline will be performed. With each subgroup, the number and proportion of participants meeting the endpoint specified will be presented for each treatment group. Fisher's exact test without adjustment for stratification factors will be used to compute the 2-sided 95% CI for the risk difference in the endpoint specified for each treatment arm and presented.

6.2. Secondary Endpoints

6.2.1. Key Secondary Endpoint - Proportion of Participants Achieving BICLA Response at Week 52

6.2.1.1. Main Analysis

The key secondary efficacy analysis evaluating the proportion of participants who achieve BICLA response at Week 52 will be conducted using the composite estimand strategy, E2, described in Section 2.1.3. This analysis will be performed on the Full Analysis Set. The statistical model to be used is the CMH method, adjusting for stratification factors (Section 5.2.1). Missing data will be handled following the conventions described in Section 5.3.1. The conventions described below will also be followed:

- The number and proportion of participants meeting BICLA response at Week 52 will be presented for each treatment group.
- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% confidence interval for risk difference adjusting by CMH weights will be presented.
- Raw p-values (using the sequentially rejective testing approach described in Section 5.1) will be presented for the comparison of each PF-0670081 dose group and the placebo group.
- A plot of proportion of participants meeting BICLA response with the corresponding SEs over time by treatment will be presented.
- A bar plot of proportion of participants achieving BICLA response at Week 52 with corresponding 95% CIs by treatment arm will be presented.

6.2.1.2. Supportive Analyses

Supportive analyses will be performed in a manner consistent with the analyses described in Section 6.1.1.2.

6.2.2. Other Secondary Endpoint – Proportion of Participants Achieving the Lupus Low Disease Activity State (LLDAS) at Week 52

6.2.2.1. Main Analysis

- Estimand strategy: E3 (Section 2.1.4)
- Analysis set: Full Analysis Set (Section 4)
- Analysis methodology: The proportion of participants achieving the LLDAS at Week 52 will be analyzed using the CMH method adjusting for stratification factors (Section 5.2.1)

- Intercurrent events and missing data: Refer to the hypothetical estimand strategy, E3, described in Section 2.1.4 and the missing data strategy described in Section 5.3.1.

- The number and proportion of participants achieving the LLDAS will be presented at Week 52.

- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% confidence interval for risk difference will be presented. This will be analyzed using CMH as described in Section 5.2.1.

- A plot of proportion of participants achieving LLDAS with the corresponding SEs over time by treatment will be presented.

6.2.2.2. Supplementary Analyses

A supplementary analysis using GLMMRM model will be performed for the longitudinal LLDAS endpoint data. This model will include treatment, visit, and treatment by visit interaction as fixed effects. The same methodology as for the primary endpoint will be applied for this endpoint (Section 6.1.1.2).

6.2.3. Other Secondary Endpoint – Proportion of Participants Achieving a Reduction in Prednisone (or Equivalent) Dose to ≤7.5 mg/day at Week 52 and Sustained for 12 Weeks Prior to Week 52, in the Subset of Participants on Prednisone >7.5 mg/day (or Equivalent) at Baseline

6.2.3.1. Main Analysis

- Estimand strategy: No estimand has been defined for this endpoint, the analysis should be performed in a manner consistent with E4 (Section 2.1.4)

- Analysis set: Full analysis set, in the subset of participants on prednisone >7.5 mg/day (or equivalent) at baseline

- Analysis methodology: Proportion of participants achieving a reduction in prednisone (or equivalent) dose to ≤7.5 mg/day at Week 52 and sustained for 12 weeks prior to Week 52 will be analyzed using CMH method adjusting for stratification factors (Section 5.2.1).

- Missing data: Refer to the missing data strategy described in Section 5.3.1.

- The number and proportion of participants achieving the criteria will be presented by treatment group.

- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% for risk difference will be presented.

6.2.3.2. Supplementary Analyses

A supplementary analysis using GLMMRM model will be performed for the longitudinal proportion of participants achieving a reduction in prednisone (or equivalent) dose to ≤7.5 mg/day at Week 52 and sustained for 12 weeks prior to Week 52 endpoint data. This model will include treatment, visit, and treatment by visit interaction as fixed effects. The same methodology as for the primary endpoint will be applied for this endpoint (Section 6.1.1.2).

6.2.4. Other Secondary Endpoint – Proportion of Participants Achieving SRI-4 Response with Dose of Prednisone (or equivalent) Reduced to ≤7.5 mg/day and Sustained for 12 Weeks at Week 52

6.2.4.1. Main Analysis

- Estimand strategy: E4 (Section 2.1.4)
- Analysis set: Full analysis set, in the subset of participants on prednisone >7.5 mg/day (or equivalent) at baseline
- Analysis methodology: Proportion of participants achieving SRI-4 response with dose of prednisone (or equivalent) reduced to ≤7.5 mg/day and sustained for 12 weeks at Week 52 will be analyzed using CMH method adjusting for stratification factors (Section 5.2.1). The number and proportion of participants achieving the criteria will be presented by treatment group.
- Intercurrent events and missing data: Refer to the composite estimand strategy, E4, described in Section 2.1.4 and the missing data strategy described in Section 5.3.1.
- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% confidence interval for risk difference, the standard error will be presented.

6.2.4.2. Supplementary Analyses

A supplementary analysis using MMRM model may be performed for the longitudinal prednisone sustained reduction endpoint data. This model will include treatment, visit, and treatment by visit interaction as fixed effects. The same methodology as for the primary endpoint will be applied for this endpoint (Section 6.1.1.2).

6.2.5. Other Secondary Endpoint – Proportion of Participants with Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI-A) Total Activity Score ≥10 at Baseline with ≥50% Reduction in CLASI-A Total Activity Score at Week 52

6.2.5.1. Main Analysis

- Estimand strategy: No estimand has been defined for this endpoint, the analysis should be performed in a manner consistent with E4 (Section 2.1.4)

- Analysis set: Full analysis set, in the subset of participants with CLASI-A Total Activity Score ≥10 at baseline
- Analysis methodology: Proportion of participants achieving a reduction in CLASI-A of ≥50% at Week 52will be analyzed using CMH method adjusting for stratification factors (Section 5.2.1).
- Missing data: Refer to the missing data strategy described in Section 5.3.1.
- The number and proportion of participants achieving the criteria will be presented by treatment group.
- The risk differences between each PF-06700841 dose group and the placebo group and the 2-sided 95% confidence interval for risk difference will be presented.

6.2.5.2. Supplementary Analyses

Subgroup analyses for different races, geographic regions, disease durations, and oral corticosteroid dose at baseline will be performed in a manner consistent with the subgroup analyses described in Section 6.1.1.2.

6.2.6. Other Secondary Endpoint – Change from Baseline in the Total Scores of Functional Assessment of Chronic Illness Therapy Fatigue (FACITF) at Week 52

6.2.6.1. Main Analysis

- Estimand strategy: E5 (Section 2.1.4)
- Analysis set: Full Analysis Set (Section 4)
- Analysis methodology: The change from baseline in FACIT-F total score will be analyzed using a linear mixed effects model which utilizes the longitudinal measurements (Section 5.2.2).
- Intercurrent events and missing data: Refer to the hypothetical estimand strategy, E5, described in Section 2.1.4 and the missing data strategy described in Section 5.3.1.
- Descriptive statistics of change from baseline in FACIT-F total score will be presented for each treatment group at each visit.
- The LS means, 2-sided 95% confidence interval for the LS means, difference between the LS means for PF-06700841 dose groups and placebo and corresponding 2-sided 95% confidence interval will be presented.
- Plot of the LS means and 2-sided 95% confidence interval will be presented by visit for each treatment group.

6.2.6.2. Supplementary Analyses

A supplementary treatment policy analysis using MMRM model may be performed for the longitudinal FACIT-F CFB data. This model will include treatment, visit, and treatment by visit interaction as fixed effects. Data collected after early treatment discontinuation will be analyzed as collected. Participants who are impacted by the COVID-19 pandemic or the Russia-Ukraine Conflict will be removed from the analysis.

6.2.7. Other Secondary Endpoint – Change from Baseline in the Individual Domain Scores of the Lupus Quality of Life (LupusQoL) at Week 52

6.2.7.1. Main Analysis

- Estimand strategy: E6 (Section 2.1.4)
- Analysis set: Full Analysis Set (Section 4)
- Analysis methodology: Each individual domain score of the LupusQoL will be analyzed separately in a linear mixed effects model which utilizes the longitudinal measurements (Section 5.2.2). Subgroup analyses for different races, geographic regions, disease durations, and oral corticosteroid dose at baseline may be performed.
- Intercurrent events and missing data: Refer to the hypothetical estimand strategy, E6, described in Section 2.1.4 and the missing data strategy described in Section 5.3.1.
- Descriptive statistics of change from baseline in individual domain scores of the LupusQoL will be presented by treatment group at each visit.
- The LS means, 2-sided 95% confidence interval for the LS means, difference between the LS means for PF-06700841 and placebo and corresponding 2-sided 95% confidence interval will be presented.
- Plot of the LS means and 2-sided 95% confidence interval will be presented by visit for each treatment group.
- A spider plot of the observed change from baseline means will be generated to compare different domain scores between treatment.

6.2.7.2. Supplementary Analyses

A supplementary treatment policy analysis may be performed for the longitudinal LupusQoL CFB data in the same manner as Section 6.2.6.2

6.2.8. Other Secondary Endpoint – Time to First Severe Flare

6.2.8.1. Main Analysis

- Estimand strategy: E7 (Section 2.1.4)

- Analysis set: Full Analysis Set (Section 4)
- Analysis methodology: Time to first severe flare will be analyzed using a log-rank test adjusting for stratification factors (Section 5.2.3).
- Intercurrent events and missing data: Refer to the estimand strategy, E7, described in Section 2.1.4 and the missing data strategy described in Section 5.3.1.
- The number of participants, percentage of participants experienced the event and the number and percentage of participants censored, the cumulative event/censor time, and the incidence rate will be summarized in a table by treatment group.
- Kaplan-Meier survival curves will be plotted.

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

6.4. Subgroup Analyses

Subgroup analyses will be performed using the derivations shown below.

- Race: American Indian or Alaska Native, Asian, White, Black or African American, Multiracial, and Not Reported
- Region: US/Canada, Latin America, Western Europe, Central/Eastern Europe, and Asia Pacific
- Disease duration: <2 years vs. ≥2 years but <4 years vs. ≥4 years
- Oral corticosteroid dose at baseline: ≥10 mg/day (prednisone equivalent) vs. <10 mg/day (prednisone equivalent)

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographics and SLE medical history variables as defined in Section 3.4 will be summarized by treatment group.

6.5.2. Study Conduct and Participant Disposition

Participant evaluation groups will be summarized and listed to show the participants belonging to each of the analysis sets described in Section 4. A summary of participant disposition, along with reasons for discontinuation, will be provided for the blinded treatment phase and the follow-up phase of the study by treatment using the Safety Analysis Set. A listing will be provided which summarizes participants who screen failed. Participants who discontinued from study due to the COVID-19 pandemic will be summarized. Participants who discontinued from study due to the Russia-Ukraine Conflict will be summarized. Note: A participant is considered to have completed the blinded treatment phase if they have completed the Week 52 visit. A participant is considered to have completed the follow-up phase if they have completed the Week 56 visit/end of study visit. Participants who prematurely discontinue from the blinded treatment phase may continue in the study and therefore be counted as follow-up phase completers.

6.5.3. Study Treatment Exposure

A summary of overall exposure to study drug will be provided by treatment group and visit.

6.5.4. Concomitant Medications and Nondrug Treatments

All concomitant medication(s) as well as non-drug treatment(s) will be provided in the listings.

6.5.5. Protocol Deviations

Important protocol deviations, important protocol deviations related to COVID-19, and important protocol deviations related to the Russia-Ukraine Conflict will be summarized by treatment group. A listing will also be provided for all protocol deviations, with columns denoting importance and relatedness to COVID-19.

6.6. Safety Summaries and Analyses

6.6.1. Adverse Events

The Treatment emergent adverse events (TEAEs) by system organ class and preferred term will be generated, for both all causality and treatment-related AEs. The Serious TEAEs (both all causality and treatment-related) by system organ class and preferred term may also be generated.

For selected adverse events specified in the study protocol Section 9.5.2, separate adjudication committees will conduct formal adjudication procedures for the purpose of event classifications. The results of the adjudication committees' decisions will be used in safety analyses.CCI [redacted]. Treatment-Emergent Adverse Events Occurring in $\geqslant$ 4 Participants in any Treatment Group by Preferred Term will be summarized.

6.6.2. Laboratory Data

Laboratory data will be listed and summarized by treatment in accordance with CTCAE v4.03. Baseline is defined in Section 3.5.2. Additionally, the plots specified below will be generated using the safety analysis set (SAS):

- Box plot of absolute values in hemoglobin, reticulocytes, platelets, lymphocytes, neutrophils, creatine kinase, eGFRs, LDL cholesterol, HDL cholesterol and total cholesterol
- Plot of eDISH analysis
- Plot of proportion of participants with alanine aminotransferase (ALT) > 2*ULN over time with corresponding SEs.
- Plot of proportion of participants with aspartate aminotransferase (AST) > 2*ULN over time with corresponding SEs.

6.6.3. Vital Signs

Absolute values and changes from baseline in systolic and diastolic blood pressure, respiratory rate, pulse rate and temperature will be summarized by treatment and time post-dose, according to sponsor reporting standards. Baseline is defined in Section 3.5.2.

Table 3. Vital Sign Ranges

Vital Sign	Minimum	Maximum
Systolic BP (mm Hg)	min. <90	
Systolic BP (mm Hg) change from baseline	max. decrease ≥30	max. increase ≥30
Diastolic BP (mm Hg)	min. <50	
Diastolic BP (mm Hg) change from baseline	max. decrease ≥20	max. increase ≥20
Supine pulse rate (bpm)	min. <40	max. >120

6.6.4. Electrocardiograms

Baseline and change from baseline in ECG parameters (QT, heart rate, QTcF, PR and QRS) will be summarized descriptively by treatment, visit and nominal time, according to sponsor reporting standard.

Categorical analysis for QT/QTc will be summarized. The number (%) of participants with maximum post-dose QTc (ie, QTcF) values and maximum increase from baseline in the following categories will be tabulated by treatment:

Table 4. Safety QTc Assessment

Degree of Prolongation	Mild (ms)	Moderate (ms)	Severe (ms)
Absolute value of QTcF	>450-480	>480-500	>500
Increase from baseline		30-60	>60

Table 5. Abnormal Criteria for PR and QRS

Measure (units)	Abnormal Criteria
PR (ms)	Maximum ≥ 300
PR increase from baseline (ms)	1. Baseline > 200 and maximum ≥ 25% increase OR 2. Baseline ≤ 200 and maximum ≥ 50% increase
QRS (ms)	Maximum ≥ 140
QRS increase from baseline (ms)	≥ 50% increase

In addition, the number of participants with uncorrected QT values >500 ms will be summarized.

6.6.5. Physical Examination

All physical exam data will be provided in the listings.

7. INTERIM ANALYSES

An interim analysis may be conducted when sufficient number of participants complete week 52 visits and before the interim analysis is instigated, the details of the objectives, decision criteria, dissemination plan and method of maintaining the study blind per Pfizer's SOPs will be documented and approved in the E-DMC charter. In addition, the analysis details must be documented and approved in an interim analysis plan.

8. REFERENCES

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedure. Statistics in Medicine 2009; 28:586-604.

2. Fleiss JL, Levin B, and Paik MC, Statistical Methods for Rates and Proportions, Third Edition, 2003, New York: John Wiley & Sons, Inc.

3. Lai JS, Beaumont JL, Ogale S, Brunetta P, Cella D. Validation of the functional assessment of chronic illness therapy-fatigue scale in patients with moderately to severely active systemic lupus erythematosus, participating in a clinical trial. J Rheumatol. 2011 Apr;38(4):672-9.

4. McElhone K, Abbott J, Sutton C, et al. Sensitivity to Change and Minimal Important Differences of the LupusQoL in Patients With Systemic Lupus Erythematosus. Arthritis Care Res. 2016 Oct;68(10):1505-13.

5. Miettinen O and Nurminen M (1985), Comparative Analysis of Two Rates. Statistics in Medicine 1985; 4: 213-226.

6. Thanou A, Chakravarty E, James JA, Merrill JT. How should lupus flares be measured? Deconstruction of the safety of estrogen in lupus erythematosus national assessment-systemic lupus erythematosus disease activity index flare index. Rheumatology 2014; 53:2175-81.

7. Ware JE, Kosinski M. SF36® Physical and Mental Health Summary Scales: A manual for users version1. Second edition, Lincoln, RI, QualityMetric Inc, 2001.

8. Chan IS, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics. 1999 Dec;55(4):1202-9.

9. Sato, T. (1989). "On the Variance Estimator of the Mantel-Haenszel Risk Difference." Biometrics 45:1323–1324. Letter to the editor.

10. Rubin, D.B. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons Inc., New York. 1987 June.

11. R. Daniel Meyer, Bohdana Ratitch, Marcel Wolbers, Olga Marchenko, Hui Quan, Daniel Li, Christine Fletcher, Xin Li, David Wright, Yue Shentu, Stefan Englert, Wei Shen, Jyotirmoy Dey, Thomas Liu, Ming Zhou, Norman Bohidar, Peng-Liang Zhao & Michael Hale (2020) Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic, Statistics in Biopharmaceutical Research, 12:4, 399-411, DOI: 10.1080/19466315.2020.1779122

9. APPENDICES

Appendix 1. Data Derivation Details

Appendix 1.1. Definition and Use of Visit Windows in Reporting

For reporting purposes, the following visit windows will be used for efficacy and patient -reported outcomes display that show by scheduled visits. If two or more observations fall into the same visit window, the observation closest to the target day will be used in the analyses. If there is a tie, the later observation will be used.

Visit Label	Target Day	Start Day	End Day
Baseline	Day 1	Last non-missing assessment on or before Day 1 and prior to first dose of study treatment (i.e., PF-06700841 or placebo)	
Week 2	Day 15	Day 2	Day 21
Week 4	Day 29	Day 22	Day 35
Week 6	Day 43	Day 36	Day 49
Week 8	Day 57	Day 50	Day 70
Week 12	Day 85	Day 71	Day 98
Week 16	Day 113	Day 99	Day 126
Week 20	Day 141	Day 127	Day 154
Week 24	Day 169	Day 155	Day 182
Week 28	Day 197	Day 183	Day 210
Week 32	Day 225	Day 211	Day 238
Week 36	Day 253	Day 239	Day 266
Week 40	Day 281	Day 267	Day 294
Week 44	Day 309	Day 295	Day 322
Week 48	Day 337	Day 323	Day 350
Week 52	Day 365	Day 351	Day 378
Week 56	Day 393	≥ Day 379	

Appendix 1.2. C-SSRS Mapped to C-CASA – Suicidal Ideation and Behavior Events and Codes

Event Code	C-CASA Event	C-SSRS Response
Suicidal Ideation		
1	Passive	"Yes" on "Wish to be dead"
2	Active: Nonspecific (no method, intent, or plan)	"Yes" on "Non-Specific Active Suicidal Thoughts"
3	Active: Method, but no intent or plan	"Yes" on "Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act"
4	Active: Method and intent, but no plan	"Yes" on "Active Suicidal Ideation with Some Intent to Act, without Specific Plan"
5	Active: Method, intent, and plan	"Yes" on "Active Suicidal Ideation with Specific Plan and Intent"
Suicidal Behavior		
1	Completed suicide	"Yes" on "Completed Suicide"
2	Suicide attempt	"Yes" on "Actual Attempt"
3	Interrupted attempt	"Yes" on "Interrupted attempt"
4	Aborted attempt	"Yes" on "Aborted attempt"
5	Preparatory actions toward imminent suicidal behaviors	"Yes" on "Preparatory Acts or Behavior"
Self-injurious behavior, no suicidal intent		
	Self-injurious behavior, no suicidal intent	"Yes" on "Has subject engaged in Non-suicidal Self-Injurious Behavior?"
Self-injurious behavior, intent unknown*		
	Self-injurious behavior, intent unknown	"Yes" on "Has subject engaged in Self-Injurious Behavior, intent unknown?"

*This event is only captured on the pediatric version of the C-SSRS.

Appendix 2. SAS Modeling Code

CMH - Primary Analysis

```
/*Proportion and ASE for a single treatment*/
ods output BinomialCLs=bin
binomial=ase(where=(lowcase(strip(label1))='ase'));
proc freq data=_data;
  by visit treatment;
  tables response / binomial(CL=WALD(CORRECT) level="1");
  weight weight/zeros;
run;

/*CMH*/
ods output CommonPdiff=diff CommonPdiffTests=diff_p; *output the CMH diff,
SE, 95% CL, Z and pvalues;
proc freq data=indata;
  by avisitn; *this is for the by-visit tables, for just Week 52, can
remove this line;
    where trt01an in (1,2); *comparing treatment vs Placebo one treatment
at a time, can also do this in the data step;
  weight weight/zeros;
  tables stratum2*trt01an*resp/binomial commonriskdiff(TEST=MH)
ALPHA=0.05;
run;

data diff;
  set diff diff_p;
  if method='Mantel-Haenszel';
run;
/*sub-group analyses*/

ods output riskdiffcol1=rd1(where=(row="Difference"))
riskdiffcol2=rd2(where=(row="Difference"));
proc freq data=&dat.(where=(treatment in (1, &trt.)));
      by visit;
      tables treatment*response / binomial fisher exact riskdiff
alpha=0.05;
run;
```

MH

MMRM

```
proc mixed data=_data order=data;
  by parameter;
  class TRT Visit STRATA subjid;
  model chg = TRT Visit STRATA TRT*Visit STRATA*Visit base base*Visit
/solution ddfm=kr;
  repeated Visit /subject=subjid type=un; *csh as alternative covariance
structure;
  lsmeans TRT*Visit/cl alpha=0.05 pdiff;
  ods output lsmeans=means diffs=diffs;
run;
```

GLMMRM

```
proc glimmix data=_data method=rmpl;
  class trt01pn avisitn usubjid;
  model response (event='1')= trt01pn avisitn trt01pn*avisitn / alpha=0.05
dist=binary link=logit solution;
  random avisitn / subject=usubjid type=ar(1) residual; /* type=cs will be
attempted if there are any concerns with model convergence */
  lsmeans trt01pn*avisitn / pdiff slicediff=avisitn ilink oddsratio cl;
  ods output lsmeans=lsmeans SliceDiffs=diffs;
run;
```

Cox Regression

```
proc phreg data=_data;
  class trt stratum;
  model aval*cnsr(1)= trt stratum;
  hazardratio trt /diff=pairwise;
run;
```

Appendix 3. Statistical Methodology Details of Tipping Point Analysis - Generalized Linear Mixed-Effect Model-Based Imputation

Tipping point analysis is a method to analyze the longitudinal data of a binary endpoint measured during the placebo-controlled period under varying MNAR assumptions. This analysis is used as a supportive analysis for the primary endpoint of SRI-4 at Week 52. It assesses the robustness of the binary data to potential deviations from the MAR assumptions for the three dose groups of PF-06700841 and one placebo group and is based on MCMC and multiple imputation.

A saturated generalized linear mixed effect model (GLMM) is used as the imputation model in this tipping point analysis. The Mantel-Haenszel estimate of the difference in binomial proportions and the Sato variance estimator using the CMH method adjusting for the stratification factors (see Section 3.4.1) derived from the clinical database is used as the analysis model. The tipping point analysis will be implemented in the following steps:

Step 1: Model fitting using GLMM

The GLMM will include the fixed effects of treatment (trt: High (H), Medium (M), and Low (L) doses), visit (time1-timeN, discrete values), treatment by visit interaction, and the stratification factor (strata1-3: 3 levels) and include subject-level random effects (ϕ). If the binary response is denoted by Yi for subject i, then the saturated generalized linear mixed effect model for four treatment groups and the N post-baseline visits is parameterized as:

$$\begin{aligned} logit(\pi_i) = logit(P(Y_i = 1) \\ = \beta_0 + \beta_H trt_H + \beta_M trt_M + \beta_L trt_L + \beta_{t_1} time_1 + \cdots + \beta_{t_N} time_N \\ + \beta_{H1} trt_H time_1 + \cdots + \beta_{HN} trt_H time_N + \beta_{M1} trt_M time_1 + \cdots \\ + \beta_{MN} trt_M time_N + \beta_{L1} trt_L time_1 + \cdots + \beta_{LN} trt_L time_N + \beta_{strata_1} strata_1 \\ + \beta_{strata_2} strata_2 + \phi_i \end{aligned}$$

Reference coding is used for the variables in this model, where each variable is a dummy variable (0 or 1).

The GLMM is fit to the observed, on-drug data in the initial step to obtain initial estimates for the model parameters.

Step 2: Parameter estimation using Bayesian framework

Final model parameter estimates are obtained under the Bayesian framework and using MCMC sampling. All of the βs are assigned the same non-informative prior Normal distribution: $N(\mu_\beta = 0, \sigma_\beta^2 = 9)$, where the variance of 9 on the logit scale ensures this prior is non-informative on the probability scale over its support of [0, 1]. The subject-level random effects, $\phi_i's$, are assigned a common prior distribution, $N(0, \sigma^2)$, where σ^2 , the common variance, further assigned a weakly informative Inverse-Gamma prior distribution with shape=1 and scale=1, where the 90th percentile of this distribution is approximately 9.

Step 3: Imputation of missing responses at Week 52

A single imputation of missing SRI-4 response at Week 52 will be performed based on the predictive distribution of the final GLMM obtained from Step 2, with the final model parameters sampled from the posterior distributions using MCMC. The inference of earlier time points are not of interest. Steps 2 and 3 can be repeatedly performed to generate multiply-imputed datasets. The number of multiple imputations in this tipping point analysis, *R*, is specified as 100.

Step 4: MNAR parameter (δ) and predictive distribution

The probability of response at Week 52, π_i^*, is obtained from exponentiating the inverse of the logit based upon the sampled parameter estimates. A parameter, δ (-1< δ <1), is defined as an MNAR adjustment to π_i^* to obtain $\pi_i^P = \pi_i^* - \delta, \pi_i^P > 0$. A series of δ values (δ = -0.9, -0.7, -0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5, 0.7, 0.9) is used to obtain a range of MNAR assumptions (where δ = 0 represents MAR). Positive δ values represent unfavorable MNAR scenarios (ie, penalizing the subject's probability of response) while negative δ values represent favorable MNAR scenarios. Once all participants with missing SRI-4 values have π_i^P calculated a single δ value and their SRI-4 responses at Week 52 imputed, this step results in a single complete imputed dataset for Week 52.

Step 5: Analysis of complete, imputed data

Analysis of an imputed data set will produce an estimate as well as standard error of the treatment difference applying the normal approximation of the difference in binomial proportions using the CMH approach adjusting for stratification. For a given value of MNAR quantity, this is repeated for *R*=100 times to generate *R*=100 complete imputed data sets and these *R*=100 sets of estimates are combined using the Rubin's rules (Rubin, 1987).[9] The analysis method is the same as that used for the primary analysis.

Step 6: Assessment of tipping point

Steps 4-5 can be repeated for participants with missing SRI-4 values at Week 52, across the series of δ values, which represent different MNAR scenarios in order to evaluate the impact of MNAR upon treatment arm differences. The specific implementation (ie, seed for random number generation, burn-in and thinning of the MCMC samples) of this tipping point analysis will be pre-specified in the statistical programming plan. Note that there is no need to repeat Steps 1-3 for different values of δ.

CCI [redacted]

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

CCI	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

CCI

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

CCI

[redacted]

[redacted]

[redacted]

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

CCI	[redacted]
[redacted]	[redacted]

CCI	

CCI	

CCI	

CCI	

Appendix 6. Abbreviations

Abbreviation	Definition
ACR	American College of Rheumatology
AE	Adverse Event
ALT	alanine aminotransferase
CCI	[redacted]
ANCOVA	analysis of covariance
CCI	[redacted]
AR (1)	Autoregressive lag 1
AST	aspartate aminotransferase
BICLA	British Isles Lupus Assessment Group-Based Composite Lupus Assessment
BILAG	British Isles Lupus Assessment Group
CCI	[redacted]
C-CASA	Columbia-Classification Algorithm of Suicide Assessment
CI	Confidence Interval
CLASI-A	Cutaneous Lupus Erythematosus Disease Area and Severity Index - Activity
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus disease 2019
CS	Corticosteroids
CSH	heterogeneous compound symmetry
ECG	Electrocardiogram
CCI	[redacted]
EuroQol	European Quality of Life
FACIT-F	Functional Assessment of Chronic Illness Therapy – Fatigue
GLMMRM	Generalized Linear Mixed Model for Repeated Measures
CCI	[redacted]
CCI	[redacted]
LLDAS	Lupus Low Disease Activity State
LLOQ	lower limit of quantification
LOCF	Last Observation Carried Forward
LS	Least-square
LSM	least squares means
LupusQoL	Lupus Quality of Life
MH	Mantel-Haenszel
MCS	Mental Component Summary
MedRA	A Medical Dictionary for Regulatory Activities

MMRM	Mixed Model for Repeated Measures
MR=NR	missing response as non-response
NRI	Non-response Imputation
PCS	Physical Component Summary
CCI	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
PhGA	Physicians Global Assessment
PR Interval	Period of time on the ECG, in milliseconds, that extends from beginning of P wave until beginning of QRS complex
PRO	Patient Reported Outcome
PT	Preferred Term
CCI	[redacted]
QT/QTc	Correct QT Interval
QTcF	Frederica's corrected QT Interval
QT Interval	ECG measure between Q wave and T wave
SAE	Serious Adverse Event
SE	Standard Error
SELENA	Safety of Estrogens in Lupus Erythematosus National Assessment
SLE	Systemic Lupus Erythematosus
SLEDAI-2K	Systemic Lupus Erythematosus Disease Activity Index-2000
CCI	[redacted]
SOC	System Organ Class
SRI-4	Systemic Lupus Erythematosus Responder Index change of 4
TEAE	Treatment-Emergent Adverse Event
ULN	Upper Limit of Normal