Protocol B7921023

A 24-WEEK RANDOMIZED, DOUBLE-BLIND, PARALLEL GROUP, ACTIVE COMPARATOR, MULTICENTER STUDY TO ASSESS THE EFFICACY AND SAFETY OF PF-06650833, PF-06651600, AND TOFACITINIB ALONE AND IN COMBINATION IN PARTICIPANTS WITH MODERATELY-SEVERELY ACTIVE RHEUMATOID ARTHRITIS WITH AN INADEQUATE RESPONSE TO METHOTREXATE

Statistical Analysis Plan
(SAP)

Version: 2.0

Date: March 1, 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. VERSION HISTORY

This statistical analysis plan (SAP) for Study B7921023 is based on the protocol Amendment 3 dated by February 10, 2022.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
2	The updated evaluations of treatment efficacy use the "treatment policy" approach. The approach compares treatment efficacy across the treatment groups irrespective of the participant's compliance with the study treatment and use of concomitant medications. This change affected sections 2.1, 5.2, 5.3, 6.1, 6.5, 6.6, 6.7.	The update makes the analyses identical to the studies that are expected in the forthcoming Phase 3 study and other concurrent studies in the literature.
2	The details of notation and definitions of estimands are updated. This change affected sections 2.1,5.2,5.3,6.1, 6.5,6.6,6.7.	The update improves the match between the statistical terminology and the terminology in the recent statistical literature.
2	Added supportive estimands to evaluate sensitivity of the results of primary and key secondary analyses to the adjustment of efficacy comparisons for poor compliance and use of concomitant medications. This change affected sections 2.1, 5.2, 5.3, 6.1, CCI	The analysis explores the robustness of the primary and key secondary analyses to the confounding effects of poor compliance and use of concomitant medications.
2	Incorporated exact (rather than asymptotic) statistical inference methods for a comparison of binary outcomes. This change affected sections 5.3, 6.1, CCI	The update ensures a validity of the analyses at the visits where proportion of responders is low.
CCI [redacted]		

2	For each of the visits and treatment arms we will estimate the proportions of subjects in the mITT data set that were removed from the sensitivity data set because of the potential influence on poor compliance, use of prohibited concomitant medication or any of these reasons. This change affected section 6.7.	Confirm the expected balance of proportions of potentially confounding intercurrent events across the treatment arms.
2	Additional comparisons for the E1 analysis of DAS28-CRP at week 12, analyses E2 and E3 for the ratio and difference of DAS28-CRP remission rates at week 24 and E3 analyses of ACR 20, ACR 50, ACR 70 and ACR 90 responder rates at week 24. • A comparison of the treatment effect between the combination treatment arm PF-06650833 +TOFACITINIB and monotherapy arm PF-06650833 and the comparison of the combination arm PF-06650833 + PF06651600 with each of the mono-therapy treatment arms (PF-06650833 and PF06651600). • An estimation of the treatment effects of primary interest (each of the combination treatment arms versus tofacitinib) for the subset of anti-TNF naïve patients. This change affected section 6.7.	The results of these analyses may be included into the CSR.
2	Changed label of longitudinal analysis of covariance from LANCOVA to MMRM (The abbreviation for Mixed Model for Repeated Measures). This change affected sections 5.2, CCI	Retain consistency with the Protocol.

CCI [redacted]

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study B7921023.

2.1. Study Objectives, Endpoints and Estimands

A definition of an estimand includes four components (ICH, 2019, Mallinckrodt et al, 2020).[1,2]

1. The population targeted by the scientific question.
2. The endpoint (variable) to be obtained for each patient.
3. The population-level summary for the variable that provides a basis for a comparison between treatment conditions and metric for a comparison between the treatment groups.
4. How to account for the intercurrent events (ICEs). The intercurrent events are defined as the post-randomization events that alter the course of the randomized treatment.

We will use 5 different estimands presented in the Table 2. These estimands are characterized by the different definition of the key features above. We will use the estimand from the group E1-E3 corresponding to a particular study objective as the main estimand and the estimands E1*, E2* as supportive. The goal of the estimands E1* and E2* is to confirm a qualitative similarity of the results of primary and key secondary analyses based on the different approaches to the adjustment of the analyses for the presence of the intercurrent events.

The estimands E1-E3 rely on the “treatment policy” approach to the efficacy comparison across the treatment groups. The approach compares treatment efficacy across the treatment groups irrespective of the participant's compliance with study treatment and use of concomitant medications. No intercurrent events are included into the definition of these estimands and therefore no accounting for the intercurrent events is required. The estimands E1*, E2* classify the events of poor compliance to the randomized treatment or use of concomitant medications as the intercurrent events and use “hypothetical” strategy for accounting for these events. This strategy aims at the mitigation of the confounding of the comparison of the efficacy of randomized treatment by the intercurrent events. The estimands E1* and E2* will be used for the support of the primary and key secondary objective of the study.

We will also report the additional supportive analyses of continuous and binary outcomes based on the cross-sectional, visit specific summaries. These summaries are based on the data observed at a particular visit (also baseline observations for change from baseline). The analyses may be viewed as an estimation of the treatment effect in the population of subjects who stay in the study up to a given study visit.

In our presentation we will use labels E1, E1*, E2, E2*, E3 for the analyses which have as their target the estimands presented in the Table 2 and labels D1, D2, D3 for the descriptive analyses.

The qualitative similarity of the results of these different analyses is expected to support the conclusion based on the results of primary and key secondary analyses.

Table 2. Characterization of the Estimands

Notation for the Estimand	Population targeted by the Scientific Question	Type of Endpoint	Population-level Summary for a given Treatment	Metric for a Comparison between the Treatment Groups	Account for the ICE
E1	All treated study participants	Continuous	Mean change from baseline	Difference of mean values	None
E2	All treated study participants	Binary	Mean probability of response	Ratio of probabilities of response.	None
E3	All treated study participants	Binary	Probability of response	Difference of probabilities of response	None
E1*	All treated study participants	Continuous	Mean change from baseline	Difference of mean values	Censoring of observations (visits within a subject) potentially affected by the ICE
E2*	All treated study participants	Binary	Mean probability of response	Ratio of probabilities of response.	Censoring of observations (visits within a subject) potentially affected by the ICE

The correspondence between the study objectives, endpoints and estimands is shown in Table 3.

Table 3. List of Study Objectives, Endpoints and Estimands

Objectives	Endpoints	Key Estimand [Supportive estimand]
Primary	**Primary**	**Primary**
• To compare the efficacy of each of 2 treatment combination arms to tofacitinib alone at Week 12	• Change from baseline DAS28-CRP at week 12.	*E1 [E1*]*
Key Secondary	**Key Secondary**	Key Secondary
• To compare the remission rates for each of 2 treatment combinations arms to tofacitinib alone at Week 24	• DAS28-CRP remission at Week 24.	*E2 [E2*]*
Secondary	**Secondary**	**Secondary**
• To assess the safety of PF-06650833, PF-06651600, and tofacitinib alone and of the combinations of PF-06650833 with PF-06651600 and tofacitinib	Most of the safety outcomes are defined by the Sponsor's standards. These outcomes include • Incidence and severity of adverse events (AEs), serious adverse events (SAEs), and withdrawals due to adverse events. • Change from baseline in clinical laboratory values (chemistry, hematology parameters). • Change from baseline in vital signs (blood pressure, pulse rate and temperature measurements),ECG-based parameters. • Incidences of severe and opportunistic infection AEs; herpes virus infection AEs; clinically significant categorical increases in hepatic enzymes Aspartate aminotransferase (AST), and Alanine aminotransferase (ALT) and total bilirubin, and potential cases meeting Hy's Law criteria for increased risk of drug-induced liver injury (DILI); major adverse cardiovascular events (MACE), including pulmonary embolism (PE) and deep vein thrombosis (DVT), Cerebrovascular accident (CVA); Adverse events (AEs) for decreased renal function, acute kidney injury, clinically significant increases in serum creatinine (Scr) and decreases in	• There is no defined estimand for these endpoints. They will be analyzed using Pfizer data standards as applicable.

	estimated glomerular filtration rate (eGFR). The additions include the parameters of the audiometry test	
Secondary	**Secondary**	**Secondary**
• To explore other signs of clinical efficacy of all treatment arms at week 12 and 24	• Change from baseline in DAS28-CRP at week 24. • Change from baseline in Tender/Painful and Swollen Joint Count at Week 12 and Week 24. • Change from baseline in Physician's Global Assessment (PhGA) of Arthritis at Week 12 and Week 24. • American College of Rheumatology (ACR 20, ACR 50, ACR 70, and ACR 90) responder rates at Week 12 and Week 24.	• E1 for continuous endpoints. • E3 for binary endpoints.

CCI [redacted]

CCI

[redacted]

[redacted]

	CCI [redacted]	
• To explore treatment-related pharmacodynamic (PD) activity in all treatment arms	• Change from baseline in concentration of anti-citrullinated protein antibodies (ACPA). • Change from baseline of rheumatoid factor (RF). • Change from baseline of Interferon Gamma-Induced Protein 10 (IP-10).	
CCI [redacted]		
• *To evaluate the effects of PF-06650833, PF-06651600, and tofacitinib alone and in combination on joint inflammation assessed by magnetic resonance imaging (MRI), in a subset of participants*	• *Change from baseline for each of multiple MRI-based endpoints collected for the study participant.*	

2.2. Study Design

This is a Phase 2, 24-week, multicenter, randomized, double-blind, 5-arm, parallel group, active comparator study to evaluate the efficacy and safety profiles of 2 combinations of investigational products (IPs): PF-06650833 co-administered with PF-06651600 (PF-06650833 + PF-06651600) and PF-06650833 co-administered with tofacitinib (PF-06650833 + tofacitinib), as well as the efficacy and safety of each IP individually in the absence of background MTX in participants with moderately - severely active RA who have had an inadequate response to MTX (MTX-IR). Both PF-06650833 and tofacitinib will be administered as modified release (MR) tablets to allow for once daily administration. The primary objective is to demonstrate superiority of at least 1 combination to tofacitinib monotherapy as measured by CFB in DAS28-CRP at Week 12. A key secondary endpoint is to assess the remission rates of the combinations and individual IPs at Week 24, as defined by DAS28-CRP <2.6. The study is powered to show statistical superiority of the combinations to tofacitinib monotherapy at Week 12.

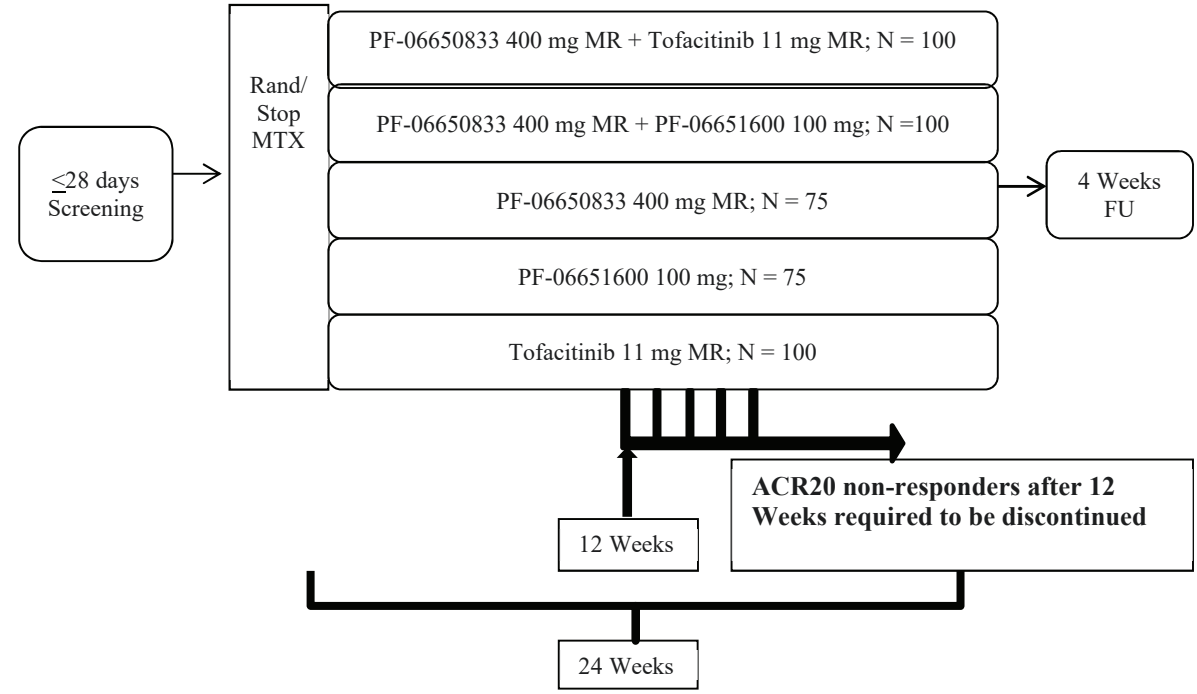
After an up to 28-day screening period, eligible participants will enter a 24-week active treatment period after being randomly assigned to one of the 5 treatment groups shown in Figure 1 below.

All doses will be administered orally once daily in a fasted state (about 4 hours after the last and 1.5 hours before the next meal). In order to maintain the blind and minimize bias, all participants will receive the same number and types of tablets each day as a mix of active and placebo tablets. MTX (and folate/folinic acid taken with MTX) will be discontinued at the time of randomization and washed out during the active treatment phase. The MTX washout is intended to reduce the potential for excessive immune suppression and additive safety liability.

No changes in the background concomitant steroid dose is allowed during the study. During the active dosing period, no "rescue" treatment is pre-specified. However, limited - duration increases in permitted analgesic medication (acetaminophen/paracetamol, nonsteroidal anti-inflammatory drugs [NSAIDs], opiates) will be permitted for acute worsening of arthritis pain (see Permitted Concomitant Medications, Protocol Sections 6.5.1 and 6.5.2).

Up to approximately 450 participants are planned to be randomized globally into the study to ensure at least up to approximately 370 participants complete at least 12 weeks of active dosing (assuming a dropout rate of approximately 18%). Participants will be in this study for a total of approximately 32 weeks which includes an up to 28-day screening period, a 24-week active treatment period, and a 4-week drug-free follow-up period.

Figure 1. Study Design Schematic



All doses are orally administered once daily under fasted conditions.
Non-responders (defined as not meeting ACR20 response criteria) by Week 12 (or later) are to be discontinued from study.
MR=modified release; MTX=methotrexate; Rand=randomization; FU=follow-up; ACR=American College of Rheumatology.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Efficacy Endpoints

The characterization of the treatment efficacy at a given visit relies on the four classes of outcomes

- Clinician Reported Outcomes (ClinRO);
- Patient-Reported Outcomes (PRO);
- Laboratory Measurements (LAB);
- Imaging Measurements (IMG).

The endpoints are designated as primary, secondary, CCI [redacted] in Table 2 in Section 2.1. This section further defines the components of the composite endpoints.

3.1.1. Clinician Reported Outcomes (ClinRO)

The outcomes include:

- Counts of Tender/Painful Joints (***TJC28, TJC68***) among the pre-defined sets of 28 (TJC28) and 68 (TJC68) joints. The 28-joints set is the subset of 68 joints set. Higher scores indicate higher level of disability.

- Counts of Swollen Joints (**SJC28, SJC66**) among the pre-defined sets of 28 (SJC28) and 66 (SJC66) joints. The 28-joints set is the subset of 66 joints set. Higher scores indicate higher level of disability.

- Physician Global Assessment Score (**PhGA**) is defined as a value within (0-100 mm or 0-10 cm) visual analog scale. Higher scores indicate higher level of disability.

3.1.2. Patient-Reported Outcomes (PRO)

The outcomes include:

CCI

[redacted]

[redacted]

3.1.3. Selected Laboratory Measurements (LAB)

The following laboratory measurements used in the computation of the efficacy endpoints are included below:

- CRP- concentration of high sensitivity C-reactive protein (mg/L). The normal range is 0.08-3.1 mg/L. Higher values of CRP indicate higher levels of inflammation.

- ESR- Erythrocyte Sedimentation Rate (mm/hr). The normal range is 0 to 22 mm/hr for men and 0 to 29 mm/hr for women. Higher values of ESR indicate higher level of inflammation.

3.1.4. Imaging Measurements (IMG)

The MRI imaging endpoints come from the dynamic contrast-enhanced MRI (DCE-MRI), RA MRI scoring system (RAMRIS) and automated quantitative RA MRI system (RAMRIQ)

- The DCE-MRI method provides 5 measurements of interest: Volume of enhancing synovium, Mean K^{trans}, Median K^{trans}, Median AUCBN (90), and Median V_e. Each of these values will be reported for 4 different subregions (*metacarpophalangeal joints*, *wrist joint*, *hand*, *wrist and metacarpophalangeal joints*) resulting in 20 endpoints for a study participant at each visit.

- The RAMRIS method provides 5 measurements/scores: bone erosion, synovitis, osteitis/bone marrow edema, tenosynovitis, and joint space narrowing. Bone erosions, Synovitis and Joint space narrowing will be summarized as *combined wrist and MCP locations*. Osteitis/bone marrow edema scores will be shown for the *combined wrist and MCP locations* in addition to *MCP joint locations* and *wrist locations*. Tenosynovitis scores will be shown for the *combined wrist and finger joint locations* in addition to *finger joint locations* and *wrist locations*. This results in 9 endpoints for a study participant at each visit.

- The RAMRIQ method provides 4 measurements/scores: bone erosion, synovitis, osteitis/bone marrow edema, and tenosynovitis. Bone erosions and synovitis will be summarized as *combined wrist and MCP locations*. Osteitis/bone marrow edema scores will be shown for the *combined wrist and MCP locations* in addition to *MCP joint locations* and *wrist locations*. Tenosynovitis scores will be shown for the *combined wrist and finger joint locations* in addition to *finger joint locations* and *wrist locations*. This results in 8 endpoints for a study participant at each visit.

Table 4 lists ClinRO, PRO and LAB measurements used in the evaluation of the treatment efficacy.

Table 4. ClinRO, PRO and LAB Outcomes

Outcome	Desirable Score (indicates lower level of disability/inflammation)	Group
TJC28, TJC68	Low	ClinRO
SJC28, SJC66	Low	ClinRO
CCI [redacted]		
CRP	Low	LAB
CCI [redacted]	[redacted]	[redacted]

Some of the ClinRO and PRO outcomes and LAB measurements listed in the Table 4 are used for forming composite variables. If any of the component of the score is missing, then the composite score is assigned to a missing value.

3.1.5. Continuous Composite Measures of Efficacy: DAS28-CRP, DAS28-ESR, CDAI, SDAI

Table 5. Continuous Composite Outcomes

Variable	Calculation
DAS28-CRP	$0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.36ln(CRP, mg/l + 1) + 0.014PtGA(mm) + 0.96$
DAS28-ESR	$0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.70ln(ESR, \frac{mm}{hr}) + 0.014PtGA(mm)$
CDAI	$TJC28 + SJC28 + PtGA(cm) + PhGA(cm)$
SDAI	$TJC28 + SJC28 + PtGA(cm) + PhGA(cm) + CRP(mg/dL)$

The variables CDAI and SDAI denote Clinical Disease Activity Index and Simplified Disease Activity Index, respectively. For all composite outcomes, higher scores correspond to higher disease severity.

3.1.6. Binary Composite Measures of Efficacy

The following binary measures of efficacy will be evaluated.

3.1.6.1. ACR20, ACR50, ACR70, ACR90 outcomes

For a given level of response of x% (x=20,50,70,90) the study participant will be declared a responder if the following three criteria are satisfied:

1. A decrease by x% or more from the baseline in the number of tender joints (TJC68).
2. A decrease by x% or more from the baseline in the number of swollen joints (SJC66).
3. A decrease by x% or more from the baseline in three of the following five variables (see Table 3)
    a. PhGA
    b. PAAP
    c. PtGA
    d. HAQ-DI
    e. CRP or ESR

Also, the decrease in either of the lab measures CRP or ESR is sufficient for satisfying requirement 3e.

CCI

3.2. Baseline Variables

Unless specifically stated otherwise, the baseline value is defined as the last non-missing measurement collected prior to the first administration of study drug at Day 1.

Baseline value will be used as a covariate in statistical models for change from baseline in efficacy endpoints to adjust for baseline disease severity.

3.3. Safety Endpoints

Most of the safety outcomes are defined by the Sponsor's standards. These outcomes include

- Incidence and severity of adverse events (AEs), serious adverse events (SAEs), and withdrawals due to adverse events.
- Clinical laboratory values (chemistry, hematology parameters).
- Vital signs (blood pressure, pulse rate and temperature measurements),ECG-based parameters.
- Incidences of severe and opportunistic infection AEs; herpes virus infection AEs; clinically significant categorical increases in hepatic enzymes Aspartate aminotransferase (AST), and Alanine aminotransferase (ALT) and total bilirubin, and potential cases meeting Hy's Law criteria for increased risk of drug-induced liver injury (DILI); major adverse cardiovascular events (MACE), including pulmonary embolism (PE) and deep vein thrombosis (DVT), Cerebrovascular accident (CVA); Adverse events (AEs) for decreased renal function, acute kidney injury, clinically significant increases in serum creatinine (Scr) and decreases in estimated glomerular filtration rate (eGFR).

The additions include the parameters of the audiometry test.

An adverse event (AE) is considered a Treatment-Emergent Adverse Event (TEAE) if the event started during the effective duration of treatment. All events that start on or after the first dosing day for B7921023 will be flagged as TEAEs. The algorithm will not consider any events that started prior to the first dose date. If an AE starts on the same day as the first dose date, it will be considered treatment emergent unless the CRF data indicates otherwise via explicitly recording time for AE onset which was occurred before the first treatment dosing.

3.4. CCI [redacted] PD Endpoints

The CCI [redacted]/PD endpoints include:

- CCI [redacted]
- Concentrations of anti-citrullinated protein antibodies (ACPA), rheumatoid factor (RF) and IP-10;

- CCI [redacted]

4. ANALYSIS SETS

For purposes of analysis, the following analysis sets are defined.

Table 7. Analysis Sets

Data Set	Description
Modified Intent to Treat (mITT)	All participants randomly assigned to investigational product who take at least 1 dose of investigational product. Participants will be analysed according to the randomized intervention.
Sensitivity Set (SNS)	All participants randomly assigned to investigational product who take at least 1 dose of investigational product. It is a subset of mITT data set where the observations that may be affected by the poor compliance or use of concomitant medications are excluded. Participants will be analysed according to the randomized intervention.
Safety Set	All participants randomly assigned to investigational product who take at least 1 dose of investigational product. Participants will be analysed according to the intervention they actually received.
CCI [redacted]	
PD Data Set (PD)	All participants randomly assigned to investigational product and who take at least 1 dose of investigational product and have at least 1 of PD parameters of interest.

Appendix 1 shows how to assign observations to the discrete set of study visits based on the observation time. CCI [redacted]

5. GENERAL METHODOLOGY AND CONVENTIONS

Final analyses will occur after database lock after Last Participant Last Visit (LPLV).

5.1. Hypotheses and Decision Rules

The primary clinical hypothesis is that mean decrease at week 12 in DAS28-CRP score in one or both combo arms CCI [redacted] exceeds the mean decrease in the reference (tofacitinib) treatment arm. The null hypothesis is that the mean decrease in DAS28-CRP score at week 12 is identical in the control (Tofacitinib arm) and combination arms. There are no multiplicity adjustments planned for this study.

5.2. Analyses of Continuous Data

The estimands E1 and E1* are described in the Section 2.1. The analysis for the estimand E1 is based on the mITT data set while the analysis for the estimand E1* is based on the sensitivity data set. We denote the analyses of interest by the names of the corresponding estimands and use the name D1 for the descriptive analysis. The analyses names, data sets and analyses methods are listed in the Table 8.

Table 8. Analyses of Continuous Data

Analysis Name	Data Set	Analytic Approach
E1	mITT	MMRM
E1*	Sensitivity data set (SNS)	MMRM
D1	mITT	Descriptive statistics

For the E1 and E1* analyses we will use the MMRM (Mixed Model for Repeated Measures) estimator while the usual descriptive estimator will be used for estimating E1*. The MMRM analysis will use the change from baseline value of the continuous variable as an outcome and treatment, scheduled study visit, baseline value of the variable of interest, treatment by visit interaction and baseline by visit interaction as fixed effects. The model will use the unstructured covariance matrix. The MMRM model is described in multiple textbooks including the books by Fitzmaurice, Laird and Ware (2011)[4] and by Mallinckrodt and Lipkovich (2017).[3] CCI [redacted]

The p-values and 90% confidence intervals will be reported for the E1 analysis of primary and secondary efficacy outcomes for the comparisons of combination arms with tofacitinib arm. Only the confidence intervals with 90% coverage will be used for all other analyses of continuous outcomes.

5.3. Analyses of Binary Data

The estimands E2, E2*, and E3 are described in the Section 2.1. The analyses for the estimands E2, E3, are based on the mITT data set while the analysis for the estimand E2* is based on the sensitivity data set. We denote the analyses of interest by the names of the corresponding estimands and denote the analyses based on descriptive statistics by D2, D3. The analyses names, data sets and analyses methods are listed in the Table 9.

Table 9. Analyses of Binary Data

Analysis Name	Data Set	Analytic Approach
E2	mITT	BSC + CZ (ratio of probabilities) + NRI
E2*	Sensitivity data set (SNS)	BSC + CZ (ratio of probabilities) + NRI
E3	mITT	BSC + CZ (difference of probabilities) + NRI
D2	mITT	BSC + CZ (ratio of probabilities) + NOIMP
D3	mITT	BSC + CZ (difference of probabilities) + NOIMP

*BSC+CZ= Blyth-Still-Casella + Chan and Zhang, NRI- non-responder imputation, NOIMP-no imputation.

We will use the Blyth-Still-Casella (Blyth and Still (1983)[5] and Casella (1986)[6]) method for the estimation of the probability of response at a given visit and a given treatment arm and Chan and Zhang (Chan and Zhang, 1999)[7] method for comparison of the probability of response across the treatment arms. For the E2, E2* and E3 analyses we will use the non-responder imputation treating missing value of the binary outcome as non-response with the exception of the two special cases of the reason for missing data. These special reasons for missing data include missingness related to the COVID infection (either patient illness or COVID-related problem of attending the visit) and missingness when subject attended the visit but some of the components for the calculation of the outcome (eg, ACR20) are missing. The observations that are missing for these special reasons will be excluded from the calculation of the probabilities of response at a given visit.

The p-values and 90% confidence intervals will be reported for the E2 and E3 analysis of key secondary and secondary efficacy outcomes for the comparisons of combination arms with tofacitinib arm. Only the confidence intervals with 90% coverage will be used for all other analyses of binary outcomes.

5.4. Analysis of Safety Data

Safety data will be summarized based on observed-case data by treatment arm.

6. ANALYSES AND SUMMARIES

6.1. Efficacy Analyses

The tables below summarize analyses of continuous and binary efficacy endpoints.

Table 10. Analyses of Continuous Efficacy Endpoints

N	Observations	*Group*	*E1*	*D1*	*E1**
1	DAS28-CRP at week 12		X	X	X
2	TJC28,	ClinRo	X	X	
3	TJC68	ClinRo	X	X	
4	SJC 28	ClinRo	X	X	
5	SJC 68	ClinRo	X	X	
6	PhGA	ClinRo	X	X	
CCI [redacted]					
15	DAS28-CRP	COMP-CONTI	X	X	
CCI [redacted]					

The analysis #1 for primary outcome will include data collected up to and including week 12 and will report the summaries for week 12 visit. All other analyses will include all collected data and will report summaries at all study visits.

Table 11. Analyses of Binary Efficacy Endpoints

N	Observations	*Group*	*E2*	*D2*	*E2**	*E3*	*D3*
1	DAS28-CRP remission rate at week 24		X	X	X		
2	ACR 20	ACR				X	X
3	ACR 50	ACR				X	X
4	ACR 70	ACR				X	X
5	ACR 90	ACR				X	X

CCI

[redacted]

6.2. Baseline and Other Summaries and Analyses

6.2.1. Baseline Summaries

Demographics and baseline characteristics will be summarized for each arm and total according to Pfizer standards.

6.2.2. Study Conduct and Participant Disposition

Participant's evaluation, disposition, discontinuation will be summarized according to Pfizer standards.

6.2.3. Study Intervention Compliance

A participant will be considered compliant with the dosing regimen if they receive at least 80% of the expected number of doses in accordance with the protocol. The number and percentage of participants who are compliant with the dosing regimen, number of dosing days and number of applications, and amount of intervention used will be summarized.

6.2.4. Concomitant Medications and Nondrug Treatments

Prior drug and non-drug treatment, concomitant drug and non-drug treatment will be summarized according to Pfizer standards.

6.3. Safety Summaries and Analyses

All safety analyses will be performed on the safety analysis set population (Section 4).

The safety data will be summarized in accordance with Pfizer Data Standards. All safety data will be summarized descriptively through appropriate data tabulations, descriptive statistics, categorical summaries, and graphical presentations.

6.3.1. Adverse Events

Overall standard AE endpoints for this study are:

- Incidence of TEAEs;
- Incidence of SAEs;
- Incidence of AEs leading to discontinuation;
- Incidence of special interest AEs: severe and opportunistic infection, herpes virus infection, MACE including PE and DVT, CVA, and decreased renal function, and acute kidney injury.

6.3.2. Laboratory Data

To determine if there are any clinically significant laboratory abnormalities, the safety laboratory tests will be assessed against the criteria specified in the protocol. The assessment will consider whether each participant's baseline test result is within or outside the laboratory reference range for the laboratory parameter. Absolute value, changes from baseline and percent changes from baseline in laboratory data will be summarized. The number (%) of subjects with abnormal AST, ALT, and alkaline phosphatase (>2 × ULN), and bilirubin (>1.5 × ULN) and the number (%) of subjects with abnormal AST, ALT, and alkaline phosphatase (>2 × baseline) will also be presented. Baseline of safety laboratory tests will be the last pre-dose measurement.

6.3.3. Vital Signs

Vital signs will be summarized for treatment arm in accordance with the Pfizer reporting standards.

6.3.4. Physical Examination

Physical examination data (including incidence of clinically significant changes in physical examination from baseline) will be summarized for treatment arm in accordance with the Pfizer reporting standards.

6.3.5. Electrocardiogram Analyses

Parameters of electrocardiogram analyses (including PR, QRS, QT, QT_C) will be summarized according to Pfizer's reporting standards.

CCI

[redacted]

6.5. Analysis of PD and Biomarker Endpoints

Concentrations of ACPA, RF, IP-10, hsCRP, and ESR will be analyzed using E1 and D1 methods. The log-transformation of the endpoints will be explored if needed for achievement symmetry of the distributions

6.6. Analysis of MRI Endpoints

Each of the 37 endpoints listed in Section 3.1.4 will be analyzed by the E1 and D1 methods. The log-transformation of the endpoints will be explored if needed for achievement symmetry of the distributions.

6.7. Additional Analyses

For the E1 analysis of DAS28-CRP at week 12, analyses E2 and E3 for the ratio and difference of DAS28-CRP remission rates at week 24 and E3 analyses of ACR 20, ACR 50, ACR 70 and ACR 90 responder rates at week 24 we may implement additional comparisons.

- A comparison of the treatment effect between the combination treatment arm PF-06650833 +TOFACITINIB and monotherapy arm PF-06650833 and the comparison of the combination arm PF-06650833 + PF06651600 with each of the mono-therapy treatment arms (PF-06650833 and PF06651600).

- An estimation of the treatment effects of primary interest (each of the combination treatment arms versus tofacitinib) for the subset of anti-TNF naïve patients.

Also, for each of the visits and treatment arms we will estimate the proportions of subjects in the mITT data set that were removed from the sensitivity data set because of the potential influence on these observations of poor compliance , use of prohibited concomitant medication or any of these reasons. The Blyth-Still-Casella method will be used for the estimation.

7. INTERIM ANALYSES

One or more interim analyses (IA) may be performed for internal business decision-making or sample size re-estimation (there is no intention a priori to terminate the study or a treatment arm based on the results of the IA). The final number and timing of any IA, if performed, will be defined by the sponsor.

If an IA is performed, an Interim Analysis Committee/Executive Committee will be constituted that will include members of senior Pfizer leadership. Before any interim analysis is initiated, the details of the objectives, decision criteria, dissemination plan, and method of maintaining the study blind as per Pfizer's SOPs will be documented and approved in an Interim Analysis Committee/Executive Committee charter. In addition, the analysis details must be documented and approved in an interim analysis SAP or final SAP. The study team will remain blinded to all interim results.

7.1.1. Data Monitoring Committee

An Independent Oversight Committee (IOC) will be constituted and be responsible for review of unblinded safety and laboratory data on a regularly scheduled, periodic basis (eg, after 25%, 50% and 75% of participants have been randomized, and periodically thereafter), as well as on an ad hoc basis at the request of the study team. The IOC may also have access to unblinded efficacy data from any interim analyses (if performed) for ongoing assessment of overall benefit: risk. For this study, the IOC will be an Internal Review Committee (IRC) comprised of Sponsor personnel (at least one of whom is medically qualified and one of whom is a clinical statistician) not directly involved in study conduct or in interactions with site personnel or participants.

The IRC will be responsible for ongoing monitoring of safety of participants in the study according to the Charter. Any recommendations made by the IRC to alter the conduct of the study will be forwarded to the Executive Committee for final decisions. Pfizer will forward such decisions, which may include summaries of aggregate analyses of endpoint events and of safety data that are not endpoints, to regulatory authorities, as appropriate.

8. REFERENCES

1. ICH Harmonized Guideline E9 (R1), Estimands and sensitivity analysis in clinical trials. Step 4 version. November 20, 2019.

2. Mallinckrodt C., Molenberghs G., Lipkovich I., Ratitich B. Estimands, estimators and sensitivity analysis in clinical trials, 2020.

3. Mallinckrodt C, Lipkovich I. Analyzing Longitudinal Clinical Trial Data: A Practical Guide; 2017.

4. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis, 2nd edition. Hoboken, NJ: Wiley Interscience; 2011.

5. Blyth CR, Still HA. Binomial confidence intervals. J Am Stat Assoc 1983;78:108-16.

6. Casella G. Refining binomial confidence intervals. Can J Stat 1986;14:113-29.

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

9. APPENDICES

9.1. Appendix 1. Definition and Use of Visit Windows in Reporting

Visit windows will be used for efficacy variables, and for any safety data that display/summarize by study visit. For other endpoints (eg, vital signs, physical examinations), visit windows will be applied for summary statistics by study visits if required.

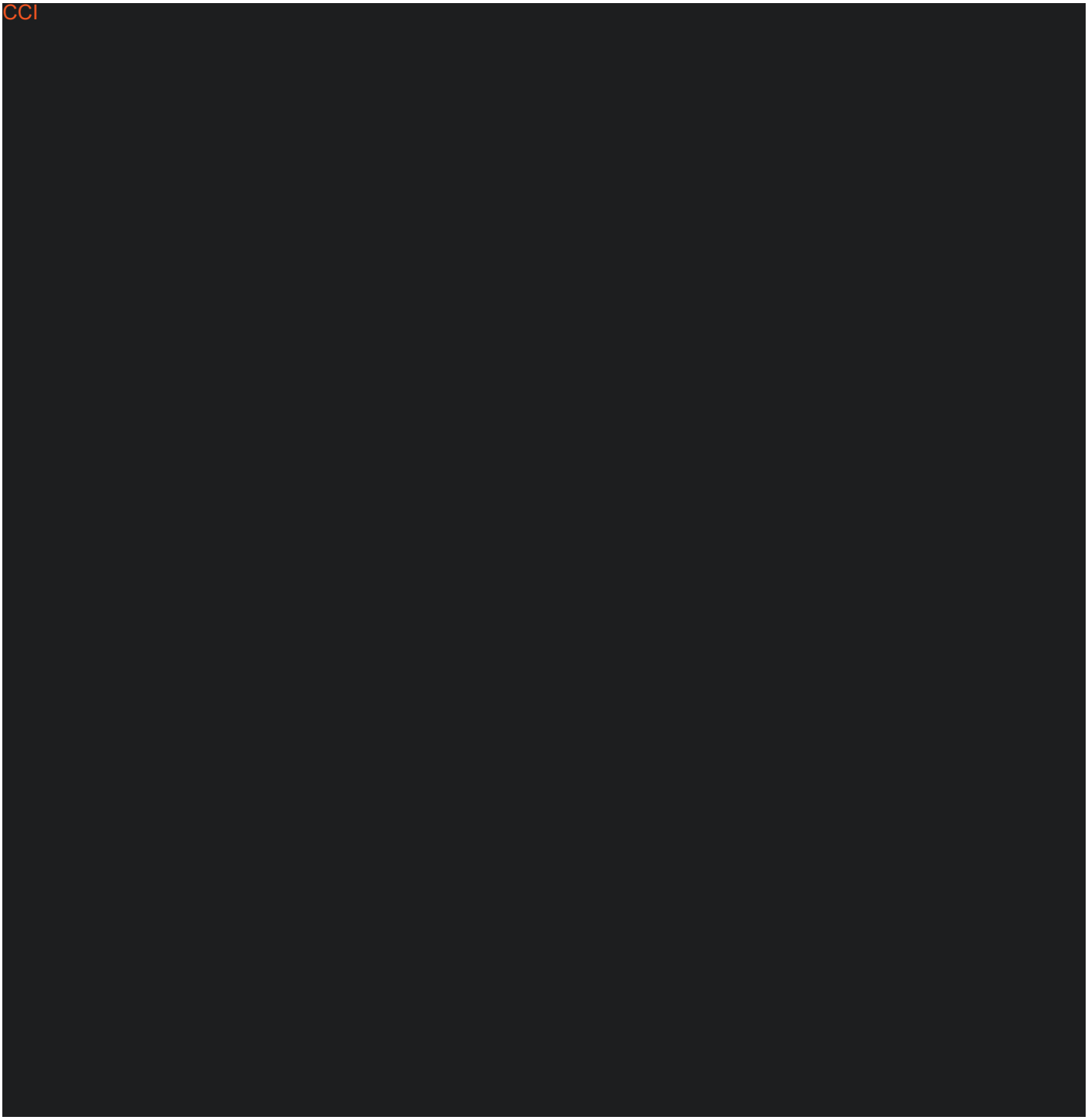
Table 12. Definition of Visit Windows

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day 0
Baseline	Day 1, Baseline	Day 1
Week 2	15	Days 2 to 22
Week 4	29	Days 23 to 43
Week 8	57	Days 44 to 71
Week 12	85	Days 72 to 99
Week 16	113	Days 100 to 127
Week 20	141	Days 128 to 155
Week 24	169	Days 156 to 183
Follow Up/End of Study		
Week 28		Days 184 to -

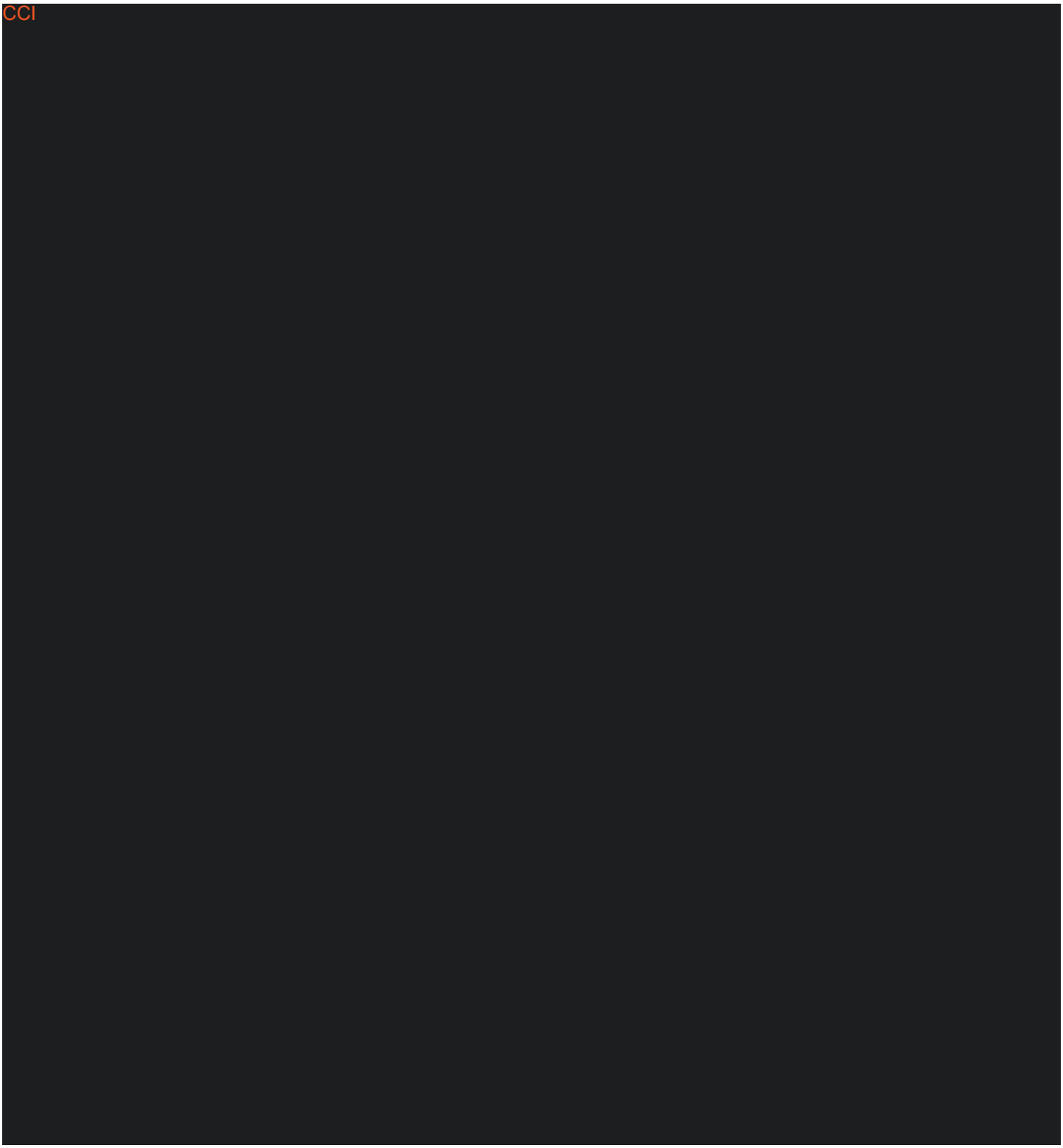
For the lab values, if the calculated study day for the labelled baseline visit is not study Day 1 but falls within 28 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

For the other values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

If two or more visits fall into the same window, keep the one closest to the Target Day. If two visits are equaled distant from the Target Day in absolute value, the later visit should be used.


CCI

CCI



CCI

[redacted]

[redacted]

CCI