Protocol B7841005

An Open-Label Study in Adolescent and Adult Severe (Coagulation Factor Activity <1%) Hemophilia A Participants With or Without Inhibitors or Moderately Severe to Severe Hemophilia B Participants (Coagulation Factor Activity ≤2%) With or Without Inhibitors Comparing Standard Treatment to PF-06741086 Prophylaxis

Statistical Analysis Plan (SAP)

Version: 6

Date: 31 January, 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/Date	Associated Protocol Amendment	Rationale	Specific Changes
Version 1 30 May 2019	Amendment 1 25 Apr 2019	Not Applicable (N/A)	N/A
Version 2 14 November 2019	Amendment 2 12 November 2019	Reflect Changes in the Protocol Amendment 2	• Interim analysis for futility was updated as the time when 50% of the planned participants either completes or discontinues from the study • In the negative binominal or Poisson regression model, the term "generalized estimating equation approach" was added for clarity • Safety endpoints were clarified as not having multiplicity control; 2 safety endpoints were added: Incidence and severity of thrombotic microangiopathy; and Disseminated intravascular coagulation/consumption coagulopathy • Several terminologies were updated following the Protocol Amendment 2: from the U.S. to the U.S., Canada, and all other regions outside EU; from subject(s) to participant(s); from AP to ATP; from preventive to preventative
		Per FDA recommendation	• Testing hierarchy for non-inhibitors with prior on-demand treatment was updated to be identical to that for inhibitors
		SAP clarification/ version control	• MedDRA terminology for thrombotic events and injection/infusion site reaction were removed; to be documented elsewhere
Version 3 15 April 2021	Amendment 5 20 November 2020	Reflect Changes in the Protocol Amendment 3, 4 and 5	• Relevant terminology changes in the protocol were followed • Section 4 Analysis Sets were updated per protocol
		Per correspondence with the FDA	• Section 2.1.3 Additional Estimand(s): an estimand for the primary endpoint was added -- the inclusion of data during the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours). Corresponding changes were made in Sections 6.1.1.2 & 6.1.3.2 Sensitivity/Supplementary Analyses • Section 5.1.1 Hypothesis and Decision Rules for Primary Endpoint: reformatted and clarified. sample size justification was added • Section 5.2.3 Analysis for Continuous Endpoints and 5.3 Methods to Manage Missing Data: how

Table 1. Summary of Changes

Version/Date	Associated Protocol Amendment	Rationale	Specific Changes
			to manage missing observation was added: multiple imputation or the key secondary endpoint of the physical health domain of Heam-A-QoL, a single value imputation using the last-observation-carried-forward for other continuous endpoints. Corresponding changes were made in Section 6.2.4 Hemophilia Joint Health Score and 6.3.1 Patient Reported Outcomes.
		SAP clarification/ change	• Typos and mistakes were corrected; some clarifications were added • Section 3.5.1 Adverse Events: definition of AE period for OP was updated to accommodate the On-Demand participants who are expected not to receive any treatment on OP Day 1 • Section 3.5.5 & 3.5.6: Laboratory Data/Vital Signs/ ECG: lag time was added as 28 days • Section 5.1.3 Hypothesis Testing Including Key Secondary Endpoints: testing hierarchy for non-inhibitors with prior on-demand treatment was augmented to include those participants on previously stable prophylaxis • Section 6.6 Safety Summaries and Analyses: analysis populations to be used for different endpoints were clarified reflecting the protocol change in the analysis population. Some analyses were modified or added. • Appendix 4.2.2 Superiority Criterion for Comparison versus On-demand Treatment: background information was added in the beginning. • Appendix 5 Statistical Methodology Details: SAS and R codes for Wilcoxon signed-rank test with multiple imputation were added.
Version 4 17 May 2022	Amendment 6 13 July 2021	Reflect changes in the Protocol Amendment 6	• Section 2.2.1 and 8 were added reflecting an allowance to continue to enroll participants in China after global enrollment is completed to enable China to meet CDE registration enrollment requirement • Descriptions of the study were updated where relevant: (a) to allow additional number of participants in order to provide sufficient enrollment into regions which have experienced delays created by the COVID-19 global pandemic, (b) to change the inclusion criteria for hemophilia B to participants with factor IX levels ≤2% to enable recruitment of hemophilia B participants

Table 1. Summary of Changes

Version/Date	Associated Protocol Amendment	Rationale	Specific Changes
			• Allowance of participants with inhibitors who are on routine *prophylaxis* to supplement the safety evaluation of inhibitor cohort on a case-by-case basis at the discretion of the sponsor. It is clarified that the efficacy data of these participants will be separately summarized using descriptive statistics without any hypothesis testing or inferential statistics • Other relevant changes in the protocol were followed
	Amendment 7 24 March 2022	Reflect changes in the Protocol Amendment 7	• Interim futility analysis was removed. • Percentage of participants with no bleeding episodes was removed from the efficacy endpoints; the number and percentage of such participants will be descriptively summarized. • Total coagulation factor or bypass product consumption was moved from a secondary efficacy endpoint to a tertiary endpoint; the analysis was updated as descriptive summary. • Analysis method for a tertiary efficacy endpoint of "number of target joints" were revised to be a descriptive summary
		SAP clarification/ change	• Sensitivity summaries were added to assess a potential impact on efficacy assessment due to a 1-way cross-over study design where some participants are excluded from MITT while included in All Safety Set (Sections 6.1.1.and 6.1.3) • COVID-19 related summary added (Section 6.7) • Enrolled set was added to Section 4 for clarification • A table was inserted to Section 2.2 following the latest SAP template dated 31-Jan-2022 • Typos and mistakes were corrected; some clarifications were added
Version 5 26 July 2022	None	Revise EU key secondary endpoints via adding Haem-A-QoL total score, EQ-VAS, and EQ-5D-5L index score	• For these three endpoints, the following changes were made: ✓ Added to the decision rule table (Table 4) ✓ Added to the testing hierarchy for the EU (Tables 5 & 7) ✓ Missing value handling was updated to multiple imputation; shift values for MNAR analysis added (Sections 5.2.3, 5.3 & 6.3.1); ✓ Non-inferiority margins were defined (Appendix 4.3.1)

Table 1. Summary of Changes

Version/Date	Associated Protocol Amendment	Rationale	Specific Changes
			• The look up country for EQ-5D-5L Index Score was changed from Spain to UK to apply the non-inferiority margin (Section 3.3.1.4) • Other editorial changes were made for clarification
Version 6 31 January 2023	None	Revise missing value handling in analyses of continuous endpoints (HJHS and PROs) per recommendations upon FDA SAP V5 review	• For endpoints that are key secondary in any of the regions, the main analysis of multiple imputation is based on the 'missing at random' assumption instead of 'missing not at random'. A sensitivity tipping point analysis is added for these endpoints. • For HJHS total score, Haemo-QoL total score, HAL/pedHAL total score, and PGIC, multiple imputation instead of LOCF is used. • For the 38 domain or component scores (9 in Haem-A-QoL, 12 in Haemo-QoL, 10 in HAL, and 7 in pedHAL), a single imputation procedure of the Month 6 value from the Observational Phase instead of LOCF is used. (sections 5.2.3 and 5.3)
		Revise the definition of "new untreated bleed" to avoid the under-counting of recurring bleeds as a single untreated bleed, and to match clinical judgement	• Updated definition: An untreated bleed is considered a "new untreated bleed" if occurring >48 hours after the previous "**new**" untreated bleed at the same site **and** occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated; an untreated bleed occurring at a different site is considered a "new untreated bleed" regardless of the time of the bleed. (Section 3.2.4); See Appendix 6 for the impact of this change on hypothetical bleeding scenarios • Other editorial changes were made for clarification

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7841005. 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. Modifications to the Analysis Plan Described in the Protocol

There are no changes to the analyses described in the protocol (Amendment 7). However, the revised definition of the "new untreated bleed" differs from the one stated in Appendix 10 of the protocol.

2.2. Study Objectives, Endpoints, and Estimands

Type	Objectives	Endpoints	Estimands
Primary Efficacy and Safety	To demonstrate the efficacy and safety of PF-06741086 (marstacimab) for routine prophylaxis in severe hemophilia A or moderately severe to severe hemophilia B (Factor VIII [FVIII] activity <1% or Factor IV [FIX] activity ≤2%, respectively) participants 12 to <75 years of age with or without inhibitors	Primary Efficacy Endpoint The primary efficacy endpoint is the annualized bleeding rate (ABR) of treated bleeding events.	Efficacy Estimands Population Summary • Inhibitor Cohort (participants with prior on-demand therapy) for all regions: the population summary will utilize the ratio of the mean ABR (PF-06741086 prophylaxis vs on-demand treatment) based on adult and adolescent participants meeting the entry criteria with prior on-demand therapy. • Non-Inhibitor Cohort for regions outside the European Union (EU): the population summary will utilize the ratio of the mean ABR (PF-06741086 prophylaxis vs on-demand treatment) based on adult and adolescent participants meeting the entry criteria with prior on-demand therapy. • Non-Inhibitor Cohort for the EU: the population summary will utilize the mean ABR difference (PF-06741086 prophylaxis – prior prophylaxis treatment) based on adult and adolescent participants meeting the entry criteria with prior prophylaxis therapy. Intercurrent Events • For all cohorts and regions, all data collected while receiving rescue medication in the form of factor replacement or bypass therapy are included. However, all data during the preventative treatment for medical/dental procedures, sport activity, or physical therapy (plus 72 hours), or collected after dose modification and/or discontinuation of treatment will not be included.

		Safety Endpoints Type I error control would not be applied to the following safety endpoints: • Adverse events (AEs) and serious adverse events (SAE)s • Incidence and severity of thrombotic events • Incidence and severity of thrombotic microangiopathy • Disseminated intravascular coagulation/consumption coagulopathy • Immunogenicity (incidence of anti-drug antibody [ADA] and clinically significant persistent neutralizing antibody [NAb] against PF-06741086) • Incidence and severity of injection site reaction • Changes in physical examination and vital signs • Incidence of clinically significant laboratory value abnormalities • Incidence of severe hypersensitivity and anaphylactic reactions	Safety Estimands • The safety estimand for AEs/SAEs will consider all safety data collected irrespective of rescue medication use, dose modification, and/or the preventative treatment for medical/dental procedures. Specifically, all data collected on or after obtaining informed consent/assent will be included in the clinical study report. • All other safety assessments will consider a "while on treatment" evaluation irrespective of rescue medication use, dose modification, and/or the preventative treatment for medical/dental procedures.
Secondary: Efficacy	To evaluate additional efficacy of PF-06741086	For the Non-Inhibitor Cohort, the primary endpoint with respect to comparisons with prior on-demand therapy for regions outside EU will be a secondary endpoint for the EU. Likewise, the primary endpoint with respect to comparisons with prior prophylaxis therapy for the EU will be a secondary endpoint for the regions outside EU. The following parameters will be assessed for comparison between PF-06741086 prophylaxis observed over the 12-month	• For incidences of joint bleeds, spontaneous bleeds, target joint bleeds, and total bleeds, the estimand will utilize the same population summary and intercurrent event handling as in the primary endpoint for each of the following: inhibitor cohort for all regions, non-inhibitor cohort for the regions outside of the EU, and non-inhibitor cohort for the EU. • For HJHS, the difference in mean changes from baseline at 6 months between the PF-06741086 prophylaxis in ATP versus the respective reference therapy in observational phase (OP) will be presented using the data irrespective of rescue

		Active Treatment Phase (ATP) versus a respective control group corresponding to each of the 3 statistical objectives (inhibitors [participants with prior on-demand therapy], non-inhibitors [with prior on-demand therapy], and non-inhibitors [with prior prophylaxis therapy]). • Incidence of joint bleeds • Incidence of spontaneous bleeds • Incidence of target joint bleeds • Incidence of total bleeds (treated and untreated). • Change in joints as measured by the Hemophilia Joint Health Score (HJHS)	medication use, dose modification, and/or the preventative treatment for medical/dental procedures. Observations after discontinuation of treatment, if collected, will not be included.
Secondary: Health-related quality of life (HRQoL)	To evaluate the effect of PF-06741086 on HRQoL	HRQoL: • Hemophilia Quality of Life Questionnaire for Adults (Haem-A-QoL) (≥17 years of age)/Hemophilia Quality of Life Questionnaire for Children (Haemo-QoL); (Adolescents 12 to <17 years of age) • Hemophilia Activities List (HAL) (Adult ≥17 years of age)/Pediatric Hemophilia Activities List (pedHAL) (Adolescents 12 to <17 years of age) • Patient Global Impression of Change – Hemophilia (PGIC-H) (OP and ATP) • Health Utilities Measure (EuroQol 5 Dimensions 5 Level [EQ-5D-5L])	The estimand for all HRQoL endpoints will utilize the same population summary and intercurrent event handling as in the HJHS.
Tertiary/ Exploratory: Pharmaco-kinetic (PK)/ Pharmaco-dynamic (PD)	Assess PK and PD parameters in this study population	• Analysis of PF-06741086 concentrations (trough as well as post-dose) over the duration of the study • Analysis of changes in biomarkers: tissue factor pathway inhibitor (TFPI)	Not Applicable

		(total and free), peak thrombin [PKT], prothrombin fragment 1+2 [PF1+2], D-dimer, and dilute prothrombin time (dPT) over duration of the study	
Tertiary/ Exploratory: HRQoL	To evaluate the effect of PF-06741086 on HRQoL	• Hemophilia Life Impacts Questionnaire (HLIQ)	Not Applicable
Tertiary/ Exploratory: efficacy	To compare joint health post PF-06741086 to baseline	• Number of target joints	Not Applicable
	To evaluate the effect of PF-06741086 on total coagulation factor or bypass product consumption	• Total coagulation factor or bypass product consumption	Not Applicable

2.2.1. Primary Estimand(s)

The estimand is defined according to the primary objective and is in alignment with the primary endpoint. It includes the following 4+1 attributes.

- Treatment: The Treatment condition of interest across all cohort is prophylaxis with PF-06741086. The alternative treatment condition to which comparison will be made is on-demand treatment with bypass agents for Inhibitor Cohort (participants with prior on-demand therapy), on-demand treatment with factor replacements for Non-Inhibitor Cohort for regions outside the EU, and prophylaxis with factor replacements for Non-Inhibitor Cohort for the EU

- Population: Adult and adolescent participants (>=12 years) with a diagnosis of severe hemophilia A or moderately severe to severe hemophilia B (FVIII activity <1% or FIX activity ≤2%, respectively) meeting the entry criteria.

- Variable: ABR

- Intercurrent event(s): All data collected while receiving rescue medication in the form of coagulation factor replacement or bypass therapy due to bleeding are included. However, all data during the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours) are excluded; see Protocol Section 6.3 and Protocol Appendix 10.10. Data collected after dose modification (dose increase as defined in Protocol Section 6.6) are not included in the primary analysis to avoid over-estimation of the effect of the intended dose. Data after discontinuation of treatment, if collected, will not be included.

- Population-level summary: For Inhibitor Cohort (participants with prior on-demand therapy) in all regions, the ratio of the mean ABRs for PF-06741086 prophylaxis over the on-demand treatment; For Non-Inhibitor Cohort based on participants who receive routine *on-demand* treatment during OP for regions outside the EU, the population-level summary is the same as for Inhibitor Cohort in all regions; For Non-Inhibitor Cohort based on participants who receive routine *prophylaxis* during for the EU, the population-summary is the difference in mean ABRs between PF-06741086 prophylaxis and standard prophylaxis.

2.2.2. Secondary Estimand(s)

The estimand associated with the secondary endpoints for the incidence of specific types of bleeding events (eg, joint bleeds) has the same attributes of treatment, population, intercurrent events, and population-level summary as in the primary estimand.

The estimand associated with HJHS and secondary HRQoL endpoints has the same attributes of treatment and population as the primary estimand (See Section 2.2.1). For intercurrent events, all data will be considered irrespective of rescue medication use, dose modification, and the preventative treatment for medical/dental procedures, sport activity, or physical therapy. However, observations after discontinuation of treatment, if collected, will not be included. The population-level summary will utilize the difference in the median between the treatments received during the OP versus the ATP.

2.2.3. Additional Estimand(s)

Additional estimands for the primary endpoint will be applied with alternative rules for handling the intercurrent event of PF-06741086 dose modification (via including the bleeding data after the dose modification) and the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours) (via including the bleeding data during this period).

2.3. Study Design

This is a 1-way, cross-over, open-label, multi-center study in approximately 145 adolescent and adult participants between ages 12 to <75 years with severe hemophilia A or moderately severe to severe hemophilia B (defined as FVIII <1%, or FIX activity ≤2%, respectively) with and without inhibitors, with approximately 20% of participants as adolescents (between ages 12 to <18 years old). Additional participants (approximately 20%) may be recruited for this study in order to account for attrition of enrolled participants and to provide for sufficient representation into regions which have experienced recruitment delays due to the corona virus disease 2019 (COVID-19) pandemic. This study is comparing treatment with their prescribed factor replacement therapy or bypass therapy during an OP with a 12-month ATP, during which participants will receive prophylaxis (defined as treatment by subcutaneous [SC] injection of PF-06741086). For simplicity the drug name will be referred to as PF-06741086 throughout the remainder of the document.

The dosing regimen for PF-06741086 is 300 mg SC for initial loading dose followed by 150 mg SC once weekly (QW). Individual participants who meet protocol specified dose

escalation criteria based upon breakthrough bleeding, (see Protocol Section 6.6 Dose Modification) may have the dose increased to 300 mg SC QW.

Participants will be enrolled into 1 of 2 cohorts:

- Inhibitor Cohort – individuals with inhibitors (approximately 45 participants, with approximately 35 hemophilia A and 10 hemophilia B participants)
- Non-Inhibitor Cohort – individuals without inhibitors (approximately 100 participants, with approximately 80 hemophilia A and 20 hemophilia B participants)

The Inhibitor Cohort (n=45) will include at least 5 adolescent participants and the Non-Inhibitor Cohort will include at least 20 adolescent participants.

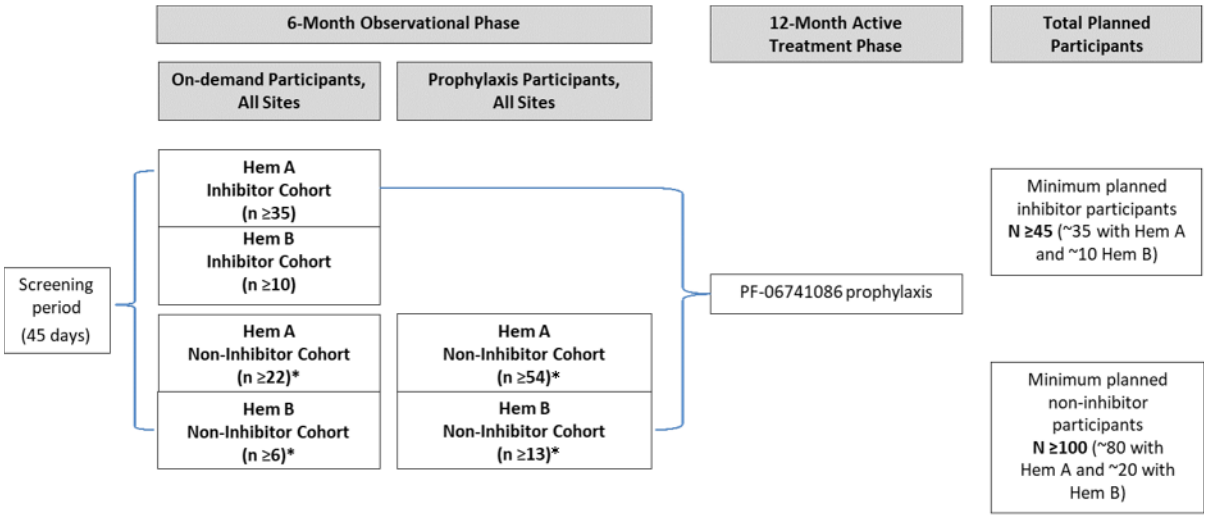
Prior On-Demand Treatment: Participants with either hemophilia A or B, with or without inhibitors, and who are prescribed an on-demand treatment regimen during the OP will transition to the ATPafter approximately 6 months. Individuals with inhibitors will receive on-demand treatment with recombinant factor VIIa (rFVIIa), Prothrombin Complex Concentrates (PCC), or activated Prothrombin Complex Concentrates (aPCC) agents eg, factor eight inhibitor bypassing activity (FEIBA) or in regions where available, Byclot (Inhibitor Cohort) and individuals without inhibitors will receive on-demand treatment with FVIII- or FIX-replacement therapy (Non-Inhibitor Cohort) during the OP before crossing over to the 12-month ATP.

Prior Prophylaxis Treatment: Participants without inhibitors who were on prior prophylactic treatment with FVIII- or FIX-replacement during the OP will transition to the ATPafter approximately 6 months.

Participants with inhibitors who are on routine prophylaxis (defined as treatment by IV injection of bypass agent to prevent bleeding) may be considered for eligibility on a case-by-case basis at the discretion of the sponsor, following discussion and agreement from the Pfizer medical monitor. These participants will be included in addition to the On-Demand Inhibitor participants and will supplement the safety evaluation of the inhibitor cohorts.

Figure 1 shows the overall study schematic.

Figure 1. Study Schematic



*Remaining 5 participants will be allocated to the non-inhibitor cohorts at the discretion of the sponsor.
Note: Participants with FVIII or FIX inhibitors who are treated with routine prophylaxis may be enrolled, on a case-by-case basis (to supplement the safety analysis of the inhibitor cohorts) after discussion and agreement from the Pfizer medical monitor.

2.3.1. China-specific Extension

Approximately [CCI]% of the overall planned enrollment into Study B7841005, which has an anticipated sample size of approximately 145 global participants, will come from China per local registration requirements.

Considering that global enrollment into Study B7841005 may be completed prior to enrollment of a sufficient number of Chinese participants, local participants will also continue to be enrolled into a China extension cohort after the study has achieved its global enrollment targets. Local Chinese B7841005 study participants enrolled during global enrollment and in the China extension cohort will be pooled together to form the China cohort and will be analyzed to support the registration in China only. Data analysis of the global trial will not include participants enrolled in the China extension cohort.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

The primary efficacy endpoint is the ABR of *treated* bleeding events, which is derived for each participant for each treatment period by using the following formula: ABR = number of bleeds *requiring treatments*/(days on treatment period/365.25). If a participant does not complete a treatment period, the days on treatment ends at the last dosing date + 6 days.

Occurrences of bleeding episodes will be obtained from participant diaries and medical records. Where relevant, investigators (or qualified designee) and monitors will ensure that there is consistency between the participant's medical record or participant diaries and the

Adverse Event case report forms, as applicable. The participant diary or medical record will serve as the source document for bleeding episodes during the study.

Data to be recorded on the electronic diary (eDiary), on a handheld provisioned device by the participant or caregiver, include weekly infusions of PF-06741086 and, when occurring, hemophilic bleeding episodes, including any exogenous factor replacement or bypass therapy that was required to treat the bleeds. The eDiary is programmed to provide reminders for the weekly infusions of PF-06741086 and also to contact the investigator if an injection site reaction occurs. Note: in the event the eDiary is not available, a paper diary may be used during the study.

The identification of a bleeding episode is clarified in the Protocol Appendix 10.10. Any sign or symptom of a bleed, regardless of whether or not medication/treatment is administered, should be reported as a "bleed" or "bleeding episode". Bleeds related to a procedure/surgery such as hematomas/bruising resulting from any surgeries or invasive procedures (eg, tooth extractions, venipuncture, or subcutaneous drug administrations) or invasive diagnostic procedures (eg, lumbar puncture, endoscopy with biopsy) would NOT be counted as bleeds. Accordingly, the duration on the preventative treatment for medical/dental procedures and also for sport activity or physical therapy (plus 72 hours in consideration for the half-life of the preventative treatment) will not be included in the derivation of the ABR.

A treated bleed from the same location occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated or a treated bleed occurring at a different site regardless of the time of the bleed will be considered a new treated bleed.

If there are multiple concurrent bleeds (spontaneous or traumatic) on the same date AND time but in different locations in a participant, they will be recorded as 1 bleed for that participant. However, the bleed will be counted towards all attributable locations in summaries of bleeds of specific site. If concurrent bleeds occurred on the same day but different times and different locations, they will be counted as independent bleeds. On the other hand, if concurrent bleeds occurred on the same day but different times at the SAME location, they will be counted as 1 bleed.

3.2. Secondary Endpoint(s)

3.2.1. Incidence of Joint Bleeds

Joint bleed is defined as a bleeding episode characterized by rapid loss of range of motion as compared with baseline that is associated with any combination of the following: pain or an unusual sensation in the joint, palpable swelling, and warmth of the skin over the joint. Only treated joint bleeds will be counted. The ABR of joint bleeds will be similarly derived as the derivation of the ABR of treated bleeds.

3.2.2. Incidence of Spontaneous Bleeds

Spontaneous bleed is bleeding for no apparent/known reason particularly into the joints, muscles, and soft tissues as opposed to traumatic bleed (ie, bleeding event occurring for an

apparent/known reason). Only treated spontaneous bleeds will be counted. The ABR of spontaneous bleeds will be similarly derived as the derivation of the ABR of treated bleeds.

3.2.3. Incidence of Target Joint Bleeds

Target Joint is defined as a major joint (eg, hip, elbow, wrist, shoulder, knee, and ankle) into which repeated bleeds occur (3 or more spontaneous bleeds into a single joint within a consecutive 6-month period). To derive the "incidence" of target joint bleeds in each of the OP and ATP, "baseline" target joints will be pre-determined before the start of the OP Only treated target joint bleeds will be counted. The ABR of target joint bleeds will be similarly derived as the derivation of the ABR of treated bleeds.

3.2.4. Incidence of Total Bleeds (Treated and Untreated)

Total bleeds include both treated and untreated bleeds. An untreated bleed is considered a "new untreated bleed" if occurring >48 hours after the previous **new** untreated bleed at the same site **and** occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated; an untreated bleed occurring at a different site is considered a "new untreated bleed" regardless of the time of the bleed. The ABR of total bleeds will be similarly derived as the derivation of the ABR of treated bleeds. Appendix 6 examines the impact of the updated definition on hypothetical bleeding scenarios in comparison with the definition provided in Appedix 10 of the protocol: an untreated bleed is considered a "new untreated bleed" if occurring >48 hours after the previous untreated bleed at the same site **or** occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated; an untreated bleed occurring at a different site is considered a "new untreated bleed" regardless of the time of the bleed.

3.2.5. HJHS

A qualified healthcare professional will assess joints using the HJHS version 2.1. The joint total score for each of 6 joints (left ankle, right ankle, left elbow, right elbow, left knee, right knee) and the global gait score will be derived. The joint total score per joint will range from 0 to 20 according to the following criteria with higher scores indicating worse joint health: swelling (0~3), duration of swelling (0~1), muscle atrophy (0~2), crepitus on motion (0~2), flexion loss (0~3), extension loss (0~3), joint pain (0~2), and strength (0~4). The global gait score will range from 0 to 4 based on walking, stairs, running, and hopping on 1 leg. A value of "Non-Evaluable (NE)" is allowed for each assessment in circumstances described in the instruction manual. The HJHS Total Score, ranging from 0 to 124, is obtained using the following formula:

$$\text{HJHS Total Score} = \frac{\text{total score of evaluated items (excluding NE's)}}{\text{possible maximum total score of evaluated items (excluding NE's)}} \times 124$$

HJHS Total Score will not be derived if >20% of items are missing (including NE's). HJHS will be assessed per Schedule of Activities (SoA) in the protocol, and visit windows will be applied to define visits. Change from Baseline in Total Score will be compared between 2 treatments. Baseline for Observational Phase is defined as the assessment measured at OP baseline; Baseline for Active Treatment Phase as the assessment measured at ATP day 0 predose. The ATP baseline assessment will also serve as the final assessment of OP.

3.3. Other Endpoint(s)

3.3.1. Patient Reported Outcomes (PROs)

The PROs described in this section should be completed at the scheduled office visit per SoA. An electronic tablet (PRO Tablet) will be provided at each site and pre-loaded with the instruments detailed below in this section. Sites will instruct the study participants on the proper use of the device to record their responses. In the event the electronic devices are not available, paper versions may be used during the OP for the initial visits. For PROs with adult and adolescent versions, participants aged 17 or older will be given the adult version. All participants will complete the same age group version of PROs throughout the study.

Baseline for OP is defined as the assessment measured at OP baseline; baseline for ATP, the assessment measured at ATP day 0 predose. Treatment between OP and ATP will be compared at 6 months. The estimated treatment difference will be compared to the minimum clinically important difference, if known, to aid in the interpretation of the result.

3.3.1.1. Haem-A-QoL and Haemo-QoL

The Haem-A-QoL and Haemo-QoL are disease specific measures of health-related quality of life in patients with hemophilia, depending on the age of the participant. The Haem-A-QoL will be completed for participants aged 17 years and older; and the Haemo-QoL for participants aged 12 to <17 years.

The adult version has 10 domains consisting of 46 items. The 10 domains and the number of items within each domain of the adult version are the following: Physical health (5); Feelings (4); View of yourself (5); Sports and leisure (5); Work and school (4); Dealing with haemophilia (3); Treatment (8); Future (5); Family planning (4); and Partnership and sexuality (3).

The adolescent version has 12 domains consisting of 77 items. The 12 domains and the number of items within each domain of the adolescent version are the following: Physical health (7); Feelings (8); Attitude/View (10); Family (8); Friends (4); Other People (6); Sport & School (9); Coping/Dealing (7); Treatment (8); Perceived Support (4); Future (4); and Relationship (2).

The scoring method is identical for both versions. Each item is scored in a 5-point Likert scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = all the time); positively worded items such as "*I was able to work/study like healthy colleagues*" are re-coded (as 1 = all the time, 2 = often, 3 = sometimes, 4 = rarely, and 5 = never) so that higher scores consistently represent greater impairment; a 'Not Applicable' response option is also available for the domains of 'Sports and leisure,' 'Work and school,' and 'Family planning' when the question may not apply to the participant in Haem-A-QoL. Scores are calculated by domain and a single total score by summing up the item scores. The transformed scaled scores (TSS) are then obtained as a scale of 0-100 using the following formula:

$$\text{TSS} = 100 \times \frac{\text{raw score} - \text{minimal possible raw score}}{\text{maximum possible raw score} - \text{minimum possible raw score}}$$

Each TSS reflects the average response of the valid items when there are at least a minimum required number of valid items. Otherwise, the TSS will be set as missing. The required minimum number of valid items for each TSS calculation are found in Appendix 1.

The change from baseline in TSS for each domain and the total will be analyzed by age group.

3.3.1.2. HAL2 and pedHAL

The Hemophilia Activities List, version 2 (HAL2), and pedHAL are disease-specific measures of HRQoL in participants with hemophilia, depending on the age of the participant. The HAL2 will be completed for a participant aged 17 years and older; and the pedHAL for a participant aged 12 to <17 years.

The HAL2 measures the impact of hemophilia on functional abilities in adults in the following 7 domains consisting of 42 items in total: Lying/sitting/kneeling/standing (8); lower (leg) functioning (9); upper (arm) functioning (4); Transportation (3); Self-care (5); Household tasks (6); and Sports/Leisure (7). Scoring will be obtained by domain; additionally, 3 components scores will be calculated (Activities involving the Upper Extremities, Basic activities involving the Lower Extremities and Complex activities involving the Lower Extremities) as well as an overall score. Items belong to each of the 3 components are defined in Appendix 2.

The pedHAL is a multiple domain measure of impact of hemophilia on functional capabilities in children. The 7 domains (and respective items) are Sitting/kneeling/standing (10); Functions of the legs (11); Functions of the arms (6); Use of transportation (3); Self-care (9); Household task (3); Leisure activities & sports (11). For pedHAL, only domain and total scores will be calculated.

The scoring method is identical for both versions. Each item is scored in a 6-point Likert scale (1 = impossible, 2 = always a problem, 3 = mostly a problem, 4 = sometimes a problem, 5 = rarely a problem, and 6 = never a problem); a '8 = n/a' response option is also available for all items in the domains of Transportation, Household tasks, and Sports/Leisure of HAL2 and for all items of pedHAL. The domain, component (for HAL2), and total scores are then obtained as normalized scores in a range of 0-100 using the following formula:

$$\text{Normalized Scores} = 100 \times \frac{\text{sum of valid item scores} - \#\text{ of valid items}}{5 * (\#\text{ of valid items})}$$

Higher normalized score represents better functional status and obtained when there are at least a minimum required number of valid items. Otherwise, the normalized score will be set as missing. The required minimum number of valid items for each normalized score calculation is found in Appendix 2.

The change from baseline in normalized score will be analyzed by age group.

3.3.1.3. PGIC-H

The PGIC-H is a single item assessment of the participant's overall impression of change in their life with hemophilia during the 2 distinct phases of Study B7841005 (ie, since being enrolled in the study [PGIC-H Observational Phase] and since being infused with PF-06741086 [PGIC-H Active Treatment Phase]). The recall period for the PGIC-H is reflective of the study phase (ie, "since the start of the study" or "since the infusion received as part of this study". The response scale is a 7-point ordinal scale centered around 'no change' with 3 grades of improvement and 3 grades of worsening.

The observed values will be analyzed.

3.3.1.4. EQ-5D-5L

The EQ-5D-5L is a brief self-administered standardized measure of health status in order to provide a simple, generic measure of health for clinical and economic appraisal comprising 2 parts. The first part consists of 5 dimensions of current health state (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). A patient is asked to rate each state on a 5-level scale (1=no problem, 2=slight problem, 3=moderate problem, 4=severe problem, 5=extreme problem) with higher levels indicating greater severity/impairment. The second part consists of the EuroQol 5 Dimensions (EQ-5D) general health state as measured by a visual analog scale (EQ-5D VAS, or EQ-VAS for short). The EQ-VAS measures the patient's self-rated health state on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state).

A single summary index is obtained from EQ-5D-5L health states by using the table lookup of health states where an EQ-5D-5L index value is assigned to each of the 3125 ($=5^5$) possible outcomes (5 dimensions and 5 possible levels within each dimension). A single country lookup value from UK will be applied to all subjects in this multi-regional study. Higher index scores indicate better health states with the maximum value of 1. This endpoint is considered missing if the value for any of its 5 dimensions is missing.

The change from baseline in EQ-5D-5L index and EQ-VAS will be analyzed.

3.3.1.5. HLIQ (Exploratory)

The HLIQ assesses, for all age groups, life impacts associated with living with and treating hemophilia. The HLIQ employs a 'past week' recall period in a 9-item assessment with higher scores indicating greater impact due to living with or treating hemophilia.

- Four items are assessed on a 5-point Likert scale scored from 0 to 4.
- Four items are gated such that responding 'yes' branches to a 5-point Likert scale and responding 'no' branches to a reason for not participating in the activity, ie, due to hemophilia or due to other reasons.
- One item is assessed on a 4-point Likert scale scored from 0 to 3.

The observed responses will be descriptively summarized.

3.3.2. PK Endpoints

- PF-06741086 concentrations will be determined over the duration of the study.

3.3.3. PD Endpoints

- Changes in following biomarkers during the ATP will be assessed over the duration of the study:
  - TFPI (total and free, "free" TFPI will only be analyzed if this assay is validated before the study completes)
  - PKT (other parameters to be reported *include lag time and endogenous thrombin generation potential*)
  - PF1+2
  - D-dimer
  - dPT

Baseline will be defined as the last measurement prior to the first dose in ATP.

3.3.4. Number of Target Joints

Target Joint is defined in Section 3.2.3 as a major joint (eg, hip, elbow, wrist, shoulder, knee, and ankle) into which repeated bleeds occur (3 or more spontaneous bleeds into a single joint within a consecutive 6-month period). Note that the endpoint of "incidence of target joint bleeding" counts *bleeding events* in "baseline" target joints identified at screening while "number of target joints" counts the number of *joints* where repeated bleeding occurs during a given treatment period. The number of target joints will be derived from eDiary for each of the OP and the ATP and descriptively summarized.

3.3.5. Total Coagulation Factor or Bypass Product Consumption

Total amount (international units [IU]) infused at each infusion recorded in eDiary and medical records, for prophylaxis and/or treatment of bleeding as well as preventative infusions for sport/physical activity (excluding preventative infusions for medical/dental), will be summed to calculate the Total coagulation factor or bypass product consumption for each participant for each treatment period. In addition, IU/ kilogram (kg) per month and IU/kg per year will also be derived for each participant for each treatment period in the following steps:

a. IU/kg: factor consumption at each infusion (dose in IU) is divided by the most recently recorded weight.

b. Total IU/kg is obtained by summing up all the IU/kg measurements.

c. IU/kg per year = B / (days in treatment duration) • 365.25.

d. IU/kg per month = C /12.

Total IU (annualized), IU/kg per year, and IU/kg per month will be summarized and analyzed.

3.4. Baseline Variables

Baseline for each of the treatment phase (OP and ATP) is defined for relevant endpoints when needed in the above sections.

Variables to be summarized at Screening/OP baseline include the following:

- Demographics (including age groups of adolescents [12-17] versus adults [>=18])
- Geographical region (See Appendix 3 for more information)
- Medical history
- Prior/current medication
- Hemophilia history including hemophilia type (A versus B)
- Number of target joints defined at screening

No covariates will be included in statistical modeling for treatment comparison.

3.5. Safety Endpoints

The following safety data will be summarized:

- AEs and SAEs
- Incidence and severity of thrombotic events
- Incidence and severity of thrombotic microangiopathy
- Disseminated intravascular coagulation/consumption coagulopathy
- Immunogenicity (incidence of ADA and clinically significantly persistent NAb against PF06741086)
- Incidence and severity of injection site reaction
- Changes in physical examination and vital signs
- Incidence of clinically significant laboratory value abnormalities
- Incidence of severe hypersensitivity and anaphylactic reactions

3.5.1. Adverse Events

For the purpose of summarizing AEs by treatment (OP versus ATP), all AEs observed on or after the first dose of OP until prior to the first dose of ATP will be attributed to OP for prior prophylaxis participants; for on-demand participants in OP, all AEs observed from the Visit 2 date until prior to the first dose of PF-06741086 at ATP will be attributed to OP; all AEs observed on or after the first dose of PF-06741086 at ATP until the last study contact day will be attributed to ATP.

An adverse event is considered treatment-emergent at each treatment phase if the event started during the effective duration of treatment regardless of whether a similar event of equal or greater severity existed in the baseline period.

Incidence and severity (Common Terminology Criteria for Adverse Events, (CTCAE) grade, see Protocol Section 10.3.3) will be summarized by treatment (OP versus ATP) for treatment-emergent AEs.

3.5.2. Thrombotic Event

Incidence and severity of thrombotic events including occurrence of thrombotic microangiopathy and disseminated intravascular coagulation/consumption coagulopathy will be obtained as part of adverse events collection. AE coding terms for endpoints related to thrombotic events will be provided prior to data base lock based on the most current version of Medical Dictionary for Regulatory Activities (MedDRA).

3.5.3. Immunogenicity

Blood samples will be collected for determination of ADA and NAb to PF-06741086 as specified in the SoA starting from ATP Day 0 Pre-dose. Only positive ADA samples will be tested in the NAb assay.

3.5.4. Injection Site Reaction

Injection site reactions will be obtained as part of adverse event collection. Study participants will be trained and reminded via the eDiary to report all injection site reactions to the investigator or designee.

Injection site reactions will be graded using CTCAE, Version 5.0, as shown in Table 2.

Table 2. Common Terminology Criteria for Adverse Events for Injection Site Reactions

Grade 1	Tenderness with or without associated symptoms (eg, warmth, erythema, itching)
Grade 2	Pain, lipodystrophy, edema, phlebitis
Grade 3	Ulceration or necrosis, severe tissue damage, operative intervention indicated
Grade 4	Life-threatening consequences, urgent intervention indicated
Grade 5	Death

Infusion site reaction during OP will be collected in a similar manner. AE coding terms for injection/infusion site reactions will be provided prior to data base lock based on the most current version of MedDRA.

3.5.5. Laboratory Data/Vital Signs

Safety laboratory data and vital signs will be collected per SoA. Baseline will be defined as the last measurement prior to respective first dose in each treatment. OP Baseline for participants receiving on-demand OP treatment is defined as the last measurement on or prior to OP Day 1. Lag time for PF-06741086 is set as 28 days.

Protocol, Appendix 2, Table 4 lists clinical laboratory tests to be performed by central and local laboratories.

3.5.6. Electrocardiogram (ECG)

ECG will be collected per SoA in the protocol. Abnormal ECG findings (See Protocol, Appendix 9) may be reported as adverse events. Lag time for PF-06741086 is set as 28 days.

3.5.7. Severe Hypersensitivity and Anaphylactic Reactions

Incidence of severe hypersensitivity and anaphylactic reactions will be obtained as part of adverse events collection and will be summarized separately as well as combined.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

For purposes of analyses, the following populations are defined for this open-label, single arm, 1-way crossover study:

Participant Analysis Set	Description	Applicable Analysis (for additional information refer to section 6)
Entered	All participants who sign the informed consent document.	Disposition
Enrolled	All entered participants assigned to a treatment in OP.	Important Protocol Deviation
All Safety	For participants with prior prophylaxis in OP, all participants who received at least 1 prophylaxis or on-demand treatment at OP; for participants with on-demand in OP, all participants who completed any of the procedures for visit 2 (OP baseline). Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from this population. Participants will be analysed according to the intervention they actually received.	Safety analyses

Participant Analysis Set	Description	Applicable Analysis (for additional information refer to section 6)
PF-06741086 Safety	All participants who received at least 1 dose of PF-06741086 in ATP	Safety analyses; PF-06741086 exposure
mITT (modified Intent-to-Treat)	All participants who completed OP and received at least 1 dose of PF-06741086 in ATP (excluding participants with inhibitors who are treated with routine prophylaxis in the OP). Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT; This population will be utilized for all efficacy analyses.	Efficacy analyses

5. GENERAL METHODOLOGY AND CONVENTIONS

Efficacy analyses will be performed separately for each of the 3 participant populations: Inhibitor Cohort; Non-Inhibitor Cohort with prior On-Demand therapy; Non-Inhibitor Cohort with prior Prophylaxis. These analyses will be conducted utilizing the Type I error rate allocated to that population, therefore not impacting the Type I error rate allocated to the remaining population(s) of interest.

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and Decision Rules for Primary Endpoint

The following primary hypotheses and decision rules for the primary endpoint will be separately applied to each Cohort/statistical objective; the Type I error rate will also be separately controlled at the 1-sided 0.025 level:

- Inhibitor Cohort (participants with prior on-demand therapy) in all regions: PF-06741086 prophylaxis will be considered *superior to on-demand treatment* with respect to ABR if the ABR ratio of PF-06741086 prophylaxis over on-demand treatment is lower than 0.5 at the 1-sided 0.025 level.

  Participants with inhibitors who are on routine *prophylaxis* (defined as treatment by IV injection of bypass factor to prevent bleeding) may be allowed to enroll to supplement the safety evaluation of inhibitor cohort. The efficacy of these participants, if included, will be separately summarized using descriptive statistics without any hypothesis testing or inferential statistics.

- Non-Inhibitor Cohort:

- For regions outside of the EU:
  PF-06741086 prophylaxis will be considered *superior to on-demand treatment* with respect to ABR if the ABR ratio of PF-06741086 prophylaxis over on-demand treatment is lower than 0.5 at the 1-sided 0.025 level. After establishing the superiority, non-inferiority of PF-06741086 versus prior prophylaxis will be tested as part of key secondary endpoints. Note: this endpoint will be evaluated as a secondary objective for the EU.

- For the EU:
  PF-06741086 prophylaxis will be considered *non-inferior to prior prophylaxis treatment* if the upper bound of the 2-sided 95% confidence interval for the difference (ATP minus OP) in ABR is less than the non-inferiority boundary of 2.5. Note: this endpoint will be evaluated as a secondary objective for regions outside of the EU.

For adequate powering, a sample size of 25 evaluable participants is planned when PF-06741086 prophylaxis is compared with routine on-demand treatment to demonstrate superiority; 60 evaluable participants when PF-06741086 prophylaxis is compared with routine prophylaxis for non-inferiority.

These sample sizes were derived via simulation under a range of scenarios. In the simulation, bleeding counts were generated from a negative binomial distribution with the mean bleeds over 6 months of 12.5 for routine on-demand treatment. The mean bleeds over 6 months for routine prophylaxis was initially assumed as 2.5 with an added scenario of 2.35 that was obtained as the weighted average of Hemophilia A and B ABRs based on the meta-analysis of Appendix 4.1.1 (5.1 • 0.8 + 3.1 • 0.2 = 4.7 ABR); the weights were based on the prevalence of each hemophilia type. The mean bleeds over 12 months for PF-06741086 prophylaxis was initially assumed as 4 with 2 additional scenarios of 2.67 and 3.46, the point estimate and the upper bound of the 1-sided 80% CI from Study B7841002, respectively. The variance was assumed as 6 times the mean for each scenario; the correlation between the 2 bleeding counts at OP and ATP from the same participant was assumed as 0.2.

The simulation further accounted for different follow-up times due to drop-out or lost-to-follow-up. To simulate 10% discontinuation during the ATP, a uniform (0, 1) random number U_i was generated for each subject; when $U_i < 0.1$, the duration of treatment was curtailed into $10 \cdot U_i$ years; the number of bleeding $\boldsymbol{X_i}$ in this reduced duration was newly generated from a negative binomial distribution with the expected value of $\boldsymbol{\mu_i = 10 \cdot U_i \cdot \mu}$ and variance of $\boldsymbol{6 \cdot \mu_i}$ where $\boldsymbol{\mu}$ represents the mean bleeding for PF-06741086 prophylaxis for complete follow up. Table 3 below demonstrates that the planned sample sizes maintain adequate power (>90%) for all scenarios except for the most pessimistic case.

Table 3. Power From Simulations With Different Follow-Up*

Testing/ Sample Size	Assumed PF-06741086 Mean Bleeding for 12 Months[1]	Assumed Reference Mean Bleeding for 6 Months[2]	Simulated Power[3] (%)
	4	12.5	98.9

Superiority vs on-demand with N=25	3.46		>99
	2.67		>99
Non-inferiority vs prior prophylaxis with N=60	4	2.5	90.1
	3.46		96.9
	2.67		>99
	4	2.35	86.9
	3.46		95.4
	2.67		>99

1. Initial assumption 4; then 3.46 and 2.67 as the upper bound of 80% CI and the point estimate from B7841002.
2. Initial assumption 12.5 and 2.5; then 2.35 as the weighted average (80% and 20%) of hemophilia A and B ABRs from an internal meta-analysis.
3. 5,000 simulations per each scenario where bleed counts were generated from a negative binomial distribution; each variance was assumed to be 6 times the mean with correlation coefficient of 0.2 between 2 bleed counts from the same participants. To simulate 10% discontinuation during the ATP, a uniform $(0, 1)$ random number U_i was generated for each subject; when $U_i < 0.1$, the year of observation was curtailed into $10 * U_i$ years with the number of bleeding X_i newly generated from a negative binomial distribution with the expected value of $\mu_i = 10 \cdot U_i \cdot \mu$ and variance $6 \cdot \mu_i$.

5.1.2. Key Secondary Endpoints & Decision Rules

The list of the key secondary endpoints and related decision rules regarding superiority and non-inferiority are defined below in Table 4. Appendix 4 includes derivation of these decision rules. In addition, the conventional superiority criterion against 0 difference will be applied when PF-06741086 is compared with prior prophylaxis after achieving non-inferiority.

Table 4. Key Secondary Endpoints and Decision Rules

Endpoint	Superiority vs On-Demand	Non-Inferiority vs Prior Prophylaxis for Non-inhibitors	Superiority vs Prior Prophylaxis for Non-inhibitors
Incidence of spontaneous bleeds	ABR Ratio <0.5	2.5 ABR Difference	ABR Difference < 0
Incidence of joint bleeds		2.5 ABR Difference	
Incidence of target joint bleeds		1.2 ABR Difference	
Incidence of total bleeds		2.9 ABR Difference	
Physical health domain in Haem-A-QoL at 6 months	Median Difference <0	Median Difference: 10 units	Median Difference < 0
Total score in Haem-A-QoL at 6 months*		Median Difference: 7 units	
EQ-5D-5L Index score at 6 months*		Median Difference: 0.1	
EQ-VAS score at 6 months*		Median Difference: 9.5 units	

*: These endpoints are key secondary only for the EU, but not for regions outside the EU

5.1.3. Hypothesis Testing Including Key Secondary Endpoints

Within each cohort and/or statistical objective, the familywise Type I error rate will be controlled using hierarchical testing method at the 1-sided 0.025 level. Table 5, Table 6, and Table 7 show the respective ordering of statistical testing.

Table 5. Ordering of Statistical Testing in Inhibitor Cohort (participants with prior on-demand therapy): in All Regions

Endpoint	Superiority vs On-Demand
Treated bleeds (primary)	1
Incidence of spontaneous bleeds	2
Incidence of joint bleeds	3
Incidence of target joint bleeds	5
Incidence of total bleeds	4
Physical function domain in Haem-A-QoL	6
Total score in Haem-A-QoL at 6 months*	7
EQ-5D-5L Index score at 6 months*	8
EQ-VAS score at 6 months*	9

*: These endpoints are key secondary only for the EU, but not for regions outside the EU

Table 6. Ordering of Statistical Testing in Non-Inhibitor Cohort for Regions Outside EU

Endpoints	Superiority vs On-Demand	Non-Inferiority vs Prior Prophylaxis	Superiority vs Prior Prophylaxis
Treated bleeds (primary)	1	2	12
Incidence of spontaneous bleeds	3	7	15
Incidence of joint bleeds	4	8	16
Incidence of target joint bleeds	6	10	18
Incidence of total bleeds	5	9	17
Physical function domain in Haem-A-QoL	11	13	14

Table 7. Ordering of Statistical Testing in Non-Inhibitor Cohort for EU

Endpoints	Non-Inferiority vs Prior Prophylaxis	Superiority vs Prior Prophylaxis
Treated bleeds (primary)	1	6
Incidence of spontaneous bleeds	2	12
Incidence of joint bleeds	3	13
Incidence of target joint bleeds	5	15
Incidence of total bleeds	4	14
Physical function domain in Haem-A-QoL	7	11
Total score in Haem-A-QoL at 6 months	8	16
EQ-5D-5L Index score at 6 months	9	17
EQ-VAS score at 6 months	10	18

5.2. General Methods

Participant disposition and descriptive summaries for endpoints pertaining to subject-level characteristics (eg, demographics, geographical region, medical history, prior/current medication, hemophilia history including hemophilia type, and number of target joints defined at screening) will be displayed using the following reporting groups in the same table:

- Inhibitors/On-demand at OP
- Inhibitors/Prophylaxis at OP (if included)
- Non-Inhibitors/On-demand at OP
- Non-Inhibitors/Prophylaxis at OP
- Overall

Efficacy endpoints obtained for treatment comparison of OP versus ATP will be presented in separate tables per cohort/treatment received during OP. For Inhibitors/On-demand at OP and Non-inhibitors/On-demand at OP, the reporting groups will be the following:

- On-demand during OP
- PF-06741086 Prophylaxis during ATP

For Non-Inhibitors/Prophylaxis at OP, the reporting groups will be the following:

- Routine Prophylaxis during OP
- PF-06741086 Prophylaxis during ATP

If any inhibitors with prior prophylaxis are included, the reporting groups will follow the above convention.

Reporting groups for safety endpoints will in general follow 1 of the above groupings according to the objective of the table presentation. Other groupings will be defined in Section 6.6 for each safety endpoint as needed.

Visit windows will be developed for specific data points in reference to SoA, so the data may be summarized by visit window. If there are multiple measures in 1 visit window, the average of the numeric measurements will be used for descriptive summary; the most abnormal value of the categorical measurements will be used for categorization.

5.2.1. Analyses for Count Endpoints

5.2.1.1. Endpoints Related to Bleeding Counts

This section describes analyses for various endpoints derived from bleeding counts during OP and ATP. Note that different analyses are applied when PF-06741086 prophylaxis during ATP is compared against on-demand treatment during OP versus routine prophylaxis during OP.

Superiority versus on-demand treatment

When PF-06741086 prophylaxis is compared with on-demand treatment for various bleeding count endpoints (treated, spontaneous, joint, target joint[s], and all bleeds), a repeated measure negative binomial regression model via generalized estimating equation (GEE) approach will be used with *log* link function. In the model, the number of bleeds will be a response variable, and treatment (PF-06741086 prophylaxis or on-demand) will be a factor. The time on treatment in log years will be an offset variable to account for different duration on treatment. The working correlation will be set as Unstructured. See Appendix 5 for example SAS codes. Superiority of PF-06741086 prophylaxis is demonstrated if the 2-sided 95% confidence interval (CI) for the mean ABR ratio (PF-06741086 prophylaxis/ routine on-demand) lies below the pre-set threshold of 0.5.

The estimated mean ABR ratio and its 2-sided 95% CI obtained from the analysis model will be presented along with the conventional p -value (for the null hypothesis that the ratio is 1). The following will also be presented by treatment for each endpoint: number of participants, the model-based mean ABR and its 2-sided 95% CI, the median and the first & third quartiles (Q1 and Q3) of the calculated ABR per treatment, and n (%) of participants with 0, 1, 2, >=3 treated bleeding.

Non-inferiority versus routine prophylaxis

When PF-06741086 prophylaxis treatment is compared with prior routine prophylaxis for various bleeding count endpoints (treated, spontaneous, joint, target joint[s], and all bleeds), a repeated measure negative binomial regression model via GEE approach will be used with *identity* link function. In the model, the number of bleeds will be a response variable, and duration (in years) and the interaction by treatment (PF-06741086 prophylaxis or routine prophylaxis) and duration will be factors with no intercept. The working correlation will be set as Unstructured. See Appendix 5 for example SAS codes. Treatment difference in the mean ABRs will be obtained using a contrast within the interaction term. Non-inferiority is demonstrated if the upper bound of the 2-sided 95% CI for the mean ABR difference

(PF-06741086 prophylaxis – routine prophylaxis) lies below the pre-set non-inferiority margin for each endpoint as defined in Sections 5.1.1 and 5.1.2. If the non-inferiority is established, subsequent testing for superiority may be conducted using the same statistical model in a way to control the Type I error rate as detailed in Section 5.1.3.

The estimated mean ABR difference and its 2-sided 95% CI obtained from the analysis model will be presented along with the conventional p-value (for the null hypothesis that the difference is 0). The following will also be presented by treatment for each endpoint: number of participants, the model-based mean ABR and its 2-sided 95% CI, the median and the Q1 & Q3 of the calculated ABR per treatment, and n (%) of participants with 0, 1, 2, >=3 treated bleeding.

5.2.1.2. Other Count Endpoint(s)

This study includes additional count efficacy endpoint of number of target joints as a tertiary endpoint. The following descriptive statistics will be presented by treatment: number of participants, mean, median, Q1, and Q3 of the observed counts over the duration of treatment per participant.

5.2.2. Analyses for Binary Endpoints

Descriptive statistics of binary data for baseline/demography and safety endpoints will include the number of non-missing observations, the frequency of the observed endpoint as well as the observed proportion.

5.2.3. Analyses for Continuous Endpoints

Descriptive statistics of continuous data for baseline/demography and safety endpoints will include the number of non-missing observations, mean and standard deviation. The median, minimum, and maximum will also be presented when appropriate.

For HJHS and all PRO outcomes except HLIQ, a non-parametric analysis will be applied. The change from baseline at 6 months will be compared between the treatments received during the OP and ATP using the Wilcoxon signed rank test (the exact version). The Hodges-Lehman location shift estimator and its 95% CI will be presented. In addition, change from baseline at the end of each treatment will be compared.

For the physical health domain and the total score of Heam-A-QoL as well as EQ-5D-5L index score and EQ-VAS score, the superiority and non-inferiority testing will be conducted as outlined in Sections 5.1.2 and 5.1.3 based on the Wilcoxon signed rank test. Missing observations for these endpoints will be handled via multiple imputation approach under the 'missing at random' assumption for the main analyses. Sensitivity tipping point analyses will also be conducted via successively applying pessimistic assumptions for missingness due to withdrawals due to AE or lack of efficacy.

For the following endpoints (HJHS total score, Haemo-QoL total score, HAL/pedHAL total score, and PGIC), missing observations will be handled via multiple imputation approach under the 'missing at random' assumption. For the domain or component scores (9 in Haem-A-QoL, 12 in Haemo-QoL, 10 in HAL, and 7 in pedHAL), a single imputation procedure

using the Month 6 OP value will be adopted. Further information regarding missing data is detailed in Section 5.3.

In addition, the following descriptive summary will be provided for all continuous efficacy endpoints for all visits for observed values as well as change from respective treatment baseline: the number of non-missing observations, mean (Standard Deviation [SD]), median, Q1, Q3, and the minimum/maximum.

For total coagulation factor or bypass product consumption, the following summary will be presented by treatment: number of participants, mean, standard deviation, median, Q1, and Q3.

5.2.4. Analyses for Categorical Endpoints

In general, counts and percentages for each level will be presented for categorical variables.

5.3. Methods to Manage Missing Data

Table 8 below describes possible reasons for *incomplete* treatment duration and how each of these are handled in analyses of endpoints related to bleeding counts.

Table 8. Handling Intercurrent Events in Bleed Count Analyses

Reason	Primary analysis of the primary endpoint; analyses of all key secondary endpoints	Sensitivity/supplementary analysis of the primary endpoint
Preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours)	Exclude	Include
Dose increase of PF-06741086	Exclude	1. Include the portion after dose increase. 2. Impute the prorated duration and the bleeding counts using the data in the primary analysis. 3. Tipping point analysis where the observed data are combined with multiples of successively larger (>1) portion to determine the tipping point.
Treatment discontinuation	Data collected after treatment discontinuation will be excluded; Analyze observed data during treatment only	1. Impute the prorated duration and the bleeding counts using the data in the primary analysis. 2. Tipping point analysis where the data in the primary analysis are combined with multiples of successively larger (>1) portion to determine the tipping point.

A tertiary endpoint of the number of target joints will be derived during the treatment duration excluding the period of preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours) and the increased dose of PF-06741086.

For the Wilcoxon signed rank test used to analyze physical health domain and the total score of Haem-A-QoL, as well as EQ-5D-5L index score and EQ-VAS scores, a multiple imputation approach with 10 imputations will be applied via monotone or fully conditional specification (FCS) regression approach utilizing the 'missing at random' assumption. Appendix 5 includes example SAS and R codes for multiple imputation, analysis of each imputed data set, and the final combined testing and inferences. The same approach will be used for the following endpoints: HJHS total score, Haemo-QoL total score, HAL/pedHAL total score, and PGIC. A single imputation procedure using the month 6 OP value will be used for the Wilcoxon signed rank test for the domain and component scores of PRO endpoints.

Handling of missing items deriving these scores at each measurement time is described in Sections 3.2.5 & 3.3.1 and Appendix 1 & Appendix 2.

Missing PK data will be handled as per a separate PK and PK/PD analysis plan ie, the Population Modeling and Analysis Plan (PMAP).

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. Inhibitor Cohort in All Regions (participants with prior on-demand therapy): ABR

The analyses in this section pertain to data from inhibitor participants with prior on-demand therapy. Efficacy data from inhibitors who are on routine *prophylaxis*, if included, will be separately summarized using descriptive statistics without any hypothesis testing or inferential statistics.

6.1.1.1. Main Analysis

- Analysis set: mITT (Section 4). All participants who completed OP and received at least 1 dose of PF-06741086 in ATP. Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT.

- Analysis methodology: see Section 5.2.1.1.

- Intercurrent events and missing data: Observations during the rescue medication use (in the form of coagulation factor replacement or bypass therapy) will be included. Observations during the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours), after PF-06741086 dose increase, or study drug discontinuation will be excluded; missing values will not be imputed.

- The estimated ABR ratio and its 2-sided 95% CI obtained from the analysis model will be presented along with the conventional p-value (for the null hypothesis that the ratio is 1). The following will also be presented by treatment for each endpoint: number of

subjects, the model-based mean ABR and its 95% CI; the mean (SD), median, Q1, Q3, and the minimum/maximum of the calculated ABR per participant; and n (%) of participants with 0, 1, 2, >=3 treated bleeding.

6.1.1.2. Sensitivity/Supplementary Analyses

Supplementary Analyses

Supplementary analyses will use the same methodology and summary as the main analysis but will use the following approaches after the intercurrent events of study drug dose increase or discontinuation.

1. To assess the impact of preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours): include data

2. To assess the impact of PF-06741086 dose increase, the main analysis will be repeated with the same model, using the following approaches:

    a. Including data after a dose increase

    b. Imputing the portion after a dose increase using the data before the dose increase to give full weight in the negative binomial regression to the observations from participants with such treatment

    c. Tipping point analysis using ABRs calculated combining bleeds during the intervention dose and assuming the bleed rate after the dose increases as multiples of successively larger (>1) factors of the former to determine the tipping point.

3. To assess the impact of treatment discontinuation, the main analysis will be repeated with the same model, using the following approaches:

    a. Imputing the portion after treatment discontinuation using the data during treatment to give full weight in the negative binomial regression to the observations from those participants.

    b. Tipping point analysis using ABRs calculated combining bleeds during the treatment and assuming the bleed rate after the treatment discontinuation as multiples of successively larger (>1) factors of the former to determine the tipping point.

Sensitivity Analysis

1. The following analyses will assess the impact of carryover effect from the OP.

    a. The main analysis will be repeated with the same model described in Section 5.2.1.1, excluding the first month data after initiation of PF-06741086.

    b. Bleed rates during the early part of the treatment period (Months 1-6 in ATP) and the latter part of the treatment period (Months 7-12 ATP) will be descriptively summarized and compared.

2. The following analysis will be performed to assess the seasonal effect on bleeding; it will use the same methodology and summary as the main analysis but will include only the observations that match the calendar time between OP and ATP.

3. The following summary will be provided to assess a potential impact on efficacy assessment due to a 1-way cross-over study design where some participants are excluded from mITT via not successfully completing the OP and/or not meeting the eligibility criteria to enter the ATP

   a. disposition of participants who are excluded from mITT

   b. demography by mITT status – All Safety Set

   c. subject characteristics by mITT status – All Safety Set

   d. descriptive summary of the ABR of treated bleeds during OP by mITT status – All Safety Set

6.1.2. Non-Inhibitor Cohort for Regions Outside the EU: ABR

All the ABR analyses described in Section 6.1.1 will be applied to the non-inhibitor participants who receive the on-demand treatment in OP.

6.1.3. Non-Inhibitor Cohort for the EU: ABR

6.1.3.1. Main Analysis

- Analysis set: mITT (Section 4). All participants who completed OP and received at least 1 dose of PF-06741086 in ATP. Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT.

- Analysis methodology: see Section 5.2.1.1.

- Intercurrent events and missing data: observations during the rescue medication use (in the form of coagulation factor replacement or bypass therapy) will be included. Observations during the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours), after PF-06741086 dose increase, or study drug discontinuation will be excluded; missing values will not be imputed.

- The estimated mean difference, its 2-sided 95% CI, and the p-values (conventional for the null hypothesis of no difference) will be presented for each endpoint. The following will also be presented by treatment for each endpoint: number of participants, the model-based mean ABR and its 95% CI; the mean (SD), median, Q1, Q3, and the minimum/maximum of the calculated ABR per participant; and n (%) of participants with 0, 1, 2, >=3 treated bleeding.

6.1.3.2. Sensitivity/Supplementary Analyses

Supplemental Analyses

Supplemental analyses will use the same methodology and summary as the main analysis but will use the following approaches after the intercurrent events of study drug dose increase or discontinuation.

1. To assess the impact of preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours): include data.

2. To assess the impact of PF-06741086 dose increase, the main analysis will be repeated with the same model, using the following approaches:

    a. Including the portion of dose increase

    b. Imputing the portion after a dose increase using the data before the dose increase to give full weight in the negative binomial regression to the observations from participants with such treatment.

    c. Tipping point analysis using ABRs calculated combining bleeds during the intervention dose and assuming the bleed rate after the dose increases as multiples of successively larger (>1) factors of the former to determine the tipping point.

3. To assess the impact of treatment discontinuation, the main analysis will be repeated with the same model, using the following approaches:

    a. Imputing the portion after treatment discontinuation using the data during treatment to give full weight in the negative binomial regression to the observations from those participants.

    b. Tipping point analysis using ABRs calculated combining bleeds during the treatment and assuming the bleed rate after the treatment discontinuation as multiples of successively larger (>1) factors of the former to determine the tipping point.

Sensitivity Analysis

1. The following analyses will assess the impact of carryover effect from the Observational Phase.

    a. The main analysis will be repeated with the same model described in Section 5.2.1.1, excluding the first month data after initiation of PF-06741086.

    b. Bleed rates during the early part of the treatment period (Months 1-6 in ATP) and the latter part of the treatment period (Months 7-12 ATP) will be descriptively summarized and compared.

2. The following analysis will be performed to assess the seasonal effect on bleeding; it will use the same methodology and summary as the main analysis but will include only the observations that match the calendar time between OP and ATP.

3. The following summary will be provided to assess a potential impact on efficacy assessment due to a 1-way cross-over study design where some participants are excluded from mITT via not successfully completing the OP and/or not meeting the eligibility criteria to enter the ATP

   a. disposition of participants who are excluded from mITT

   b. demography by mITT status – All Safety Set

   c. subject characteristics by mITT status – All Safety Set

   d. descriptive summary of the ABR of treated bleeds during OP by mITT status – All Safety Set

6.1.4. Control for Multiplicity

Type I error rate for the primary analysis of the primary endpoints will be separately controlled for each of the 3 statistical objectives as 2-sided 0.05, or equivalently, 1-sided 0.025: (1) inhibitor cohort (participants with prior on-demand therapy) in all regions, (2) non-inhibitor cohort for regions outside the EU based on participants who receive routine on-demand treatment during OP, and (3) non-inhibitor cohort for the EU based on participants who receive routine prophylaxis treatment during OP.

6.2. Secondary Endpoint(s)

6.2.1. Bleeding-Related Secondary Endpoints

Bleeding-related secondary endpoints include incidence of joint bleeds, incidence of spontaneous bleeds, incidence of target joint bleeds, and incidence of total bleeds (treated and untreated). The estimand has the same attributes of treatment, population, intercurrent events, and population-level summary as in the primary estimand. Analyses for inhibitor cohort (participants with prior on-demand therapy) in all regions, non-inhibitor cohort for regions outside the EU, and non-inhibitor cohort for the EU will be the same as described in Section 6.1.1, Section 6.1.2, and Section 6.1.3, respectively. The decision rule for each endpoint is specified in Table 4.

No sensitivity or supplemental analyses will be performed for the bleeding-related secondary endpoints.

Type I error rate for the secondary endpoints will be separately controlled for each of the 3 cohort/statistical objectives, as 2-sided 0.05, or equivalently, 1-sided 0.025. Within each cohort and/or statistical objective, the familywise Type I error rate will be controlled using hierarchical testing method at the 1-sided 0.025 level. Table 5, Table 6, and Table 7 show the respective ordering of statistical testing.

6.2.2. HJHS

- Analysis set: mITT (Section 4). All participants who completed OP and received at least 1 dose of PF-06741086 in ATP. Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT.

- Analysis methodology: The change from baseline at 6 months will be compared using the Wilcoxon signed rank test (Section 5.2.3). Change from baseline at the end of treatment will also be compared.

- Intercurrent events and missing data: All data will be considered irrespective of rescue medication use, dose modification, and the preventative treatment for medical/dental procedures, sport activity, or physical therapy. However, observations after discontinuation of treatment, if collected, or will not be included. Missing values will be imputed via last-observation-carried-forward approach for non-parametric analysis of Wilcoxon signed rank test.

- For Wilcoxon signed rank test (the exact version), the Hodges-Lehman location shift estimator, its 95% CI and p-value will be presented. In addition, the following descriptive summary will be provided for all visits for observed values as well as change from respective treatment baseline: the number of non-missing observations, mean (SD), median, Q1, Q3, and the minimum/maximum.

6.3. Other Endpoint(s)

6.3.1. PROs

In the non-parametric analysis of PRO endpoints of physical health domain and the total score of Haem-A-QoL as well as EQ-5D-5L index score and EQ-VAS scores, multiple imputation will be applied as detailed in Section 5.3.

The following PRO endpoints will be analyzed using the same analysis as used for HJHS in Section 6.2.2: the rest of the domain scores for Haem-A-QoL, the total and all domain scores for Haemo-QoL, HAL2/pedHal and PGIC-H.

Each item of HLIQ will be descriptively summarized.

6.3.2. PK Endpoints

- PF-06741086 concentration data will be summarized for all participants for each of the study days samples are collected based on PF-06741086 safety set when at least 1 measurable PK concentration is available.

- PF-06741086 concentration data will also be analyzed along with PK data from other studies using population PK methods to determine population PK parameters for PF-06741086. Please refer to the PMAP for further details. The population pharmacokinetic analysis will be reported separately from the main clinical study report.

- The impact of ADA status will be investigated if there are a sufficient number of participants testing positive for ADA to PF-06741086.

6.3.3. PD Endpoints

- Change from baseline will be summarized.

- PK/PD analysis for select biomarkers (eg, peak thrombin and/or dPT) will also be performed using an appropriate PK/PD model. Details of this modeling analysis will be described in the PMAP.

- Participants with detectable FVIII activity (hemophilia A) or FIX activity (hemophilia B) at the time of a PD specimen collection may have result(s) for the respective PD parameter censored from the PK/PD analysis.

Analysis of PD endpoints will be based on PF-06741086 safety set when at least baseline and/or any post-dose measurement for at least 1 PD endpoint is available. Exploratory analysis of the correlation of clinical efficacy and safety with exposure may be performed as data permits.

6.3.4. Number of Target Joints (Tertiary)

- Analysis set: mITT (Section 4). All participants who completed OP and received at least 1 dose of PF-06741086 in ATP. Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT.

- Analysis methodology: Descriptive summaries will be provided.

- Intercurrent events and missing data: Observations during the rescue medication use (in the form of coagulation factor replacement or bypass therapy) will be included. Observations during the preventative treatment for medical/dental procedures, sport activity or physical therapy (plus 72 hours), after PF-06741086 dose increase, or study drug discontinuation will be excluded; missing values will not be imputed.

- The following descriptive statistics will be presented by treatment: number of participants, mean, median, Q1, and Q3 of the observed counts over the duration of treatment per participant.

6.3.5. Total Coagulation Factor or Bypass Product Consumption (Tertiary)

- Analysis set: mITT (Section 4). All participants who completed OP and received at least 1 dose of PF-06741086 in ATP. Participants who changed from a non-inhibitor to an inhibitor on or before ATP Day -7 testing will be excluded from mITT.

- Analysis methodology: Descriptive summaries will be provided.

- Intercurrent events and missing data: Observations during the rescue medication use (in the form of coagulation factor replacement or bypass therapy) and the preventative infusions for sport/physical activity will be included. Observations during the

preventative treatment for medical/dental procedures (plus 72 hours), after PF-06741086 dose increase, or study drug discontinuation will be excluded; missing values will not be imputed.

- The following descriptive statistics will be presented by treatment: number of participants, mean, median, Q1, and Q3 of the observed values over the duration of treatment per participant for total IU (annualized), IU/kg per year, and IU/kg per month.

6.4. Subset Analyses

Subset analyses will be conducted for the primary endpoint for the following variables: Hemophilia A and B, age (12-17 and >=18), race, ethnicity, and geographic region (see Appendix 3 for description). Subset analyses aim to evaluate the consistency of treatment effect of PF-06741086 prophylaxis across various subsets. The primary analysis will be applied to each subset and the resulting treatment contrast (difference or ratio) will be tabulated along with the estimated ABR and 95% CI per treatment.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Baseline summaries will be based on All Safety population. Baseline variables to be summarized are listed in Section 3.4. These participant-level baseline variables will be summarized by cohort/treatment received during OP.

6.5.2. Study Conduct and Participant Disposition

The participant disposition will be summarized by cohort/treatment received during OP and overall for all screened participants.

6.5.3. Study Treatment Exposure

Treatment exposure for factor replacement or bypass treatment will be summarized separately per cohort/treatment received during OP for All Safety population. The following will be summarized by reasons for infusion (on demand, prophylactic, preventative for medical/dental procedure, preventative for sport/physical activity, and overall) separately for OP and ATP: total dose (IU, annualized) per participant, IU/kg per year, range of infusion for injection (IU), number of infusions per participant, exposure days per participant, and dose (IU/Kg) per infusion.

Treatment exposure to PF-06741086 during ATP will be summarized for PF-06741086 Safety population by cohort/treatment received during OP for the following: duration of treatment, exposure days, average weekly doses, number of missed dose days, and the compliance. In summarizing the treatment exposure to PF-06741086 during ATP, each partial injection will be considered as 75 mg, ie, 50% of the full intended dose.

6.5.4. Concomitant Medications and Nondrug Treatments

Concomitant and prior medications will be coded using the World Health Organization (WHO)-drug coding dictionary. Concomitant and prior non-drug treatments/procedures will

be coded using the MedDRA coding dictionary. The summary will be provided for the safety population.

6.6. Safety Summaries and Analyses

Safety will be summarized per cohort/treatment received during OP.

Safety analyses will be based on All Safety population when the safety of PF-06741086 during the ATP is compared to safety of the treatment given during the OP. Participants in the All Safety population who do not receive at least 1 dose of PF-06741086 will be included in columns summarizing safety from the OP but excluded from columns presenting safety during the ATP.

PF-06741086 Safety population will be utilized when safety of PF-06741086 during ATP across cohorts/treatment received during OP is summarized. These populations are defined in Section 4 Analysis Sets.

6.6.1. Adverse Events

Incidence of aEs and incidence/severity (CTCAE grade) of treatment-emergent aEs will be summarized by treatment (OP versus ATP) per cohort/treatment received during OP using All Safety population with data in the ATP column only including participants who received at least 1 dose of PF-06741086 in the denominator. In addition, aEs leading to treatment/study discontinuation, medication errors, deaths, SAEs, and aEs of special interest will be summarized.

6.6.2. Thrombotic Event

Incidence of thrombotic events including occurrence of thrombotic microangiopathy and disseminated intravascular coagulation/consumption coagulopathy will be summarized comparing treatments at OP versus ATP per cohort/treatment received during OP based on All Safety population with data in the ATP column only including participant who received at least 1 dose of PF-06741086 in the denominator; corresponding listing will also be provided.

6.6.3. Immunogenicity

Anti-marstacimab antibody and anti-marstacimab neutralizing antibody results will be summarized for each test at each timepoint and cumulatively during the ATP based on PF-06741086 Safety population. The anti-marstacimab neutralizing antibody status (positive or negative) is only available when a participant has positive marstacimab specificity, so any sample that has not shown marstacimab specificity (ie, ADA-) will be considered negative for the respective test. The time will also include “any time on or after ATP baseline” as well as “any time after the first dosing of PF-06741086.”

6.6.4. Injection Site Reaction

Incidence of participants experiencing any injection site reaction will be summarized by cohort/treatment received during OP and overall using PF-06741086 Safety population. The worst severity (Grade 1 to 5) per participant will also be summarized. Infusion site reaction

will be separately summarized using the same method by cohort/treatment received during OP and overall using All Safety population.

6.6.5. Vital Signs

Descriptive summary of observed values and change from baseline at each scheduled visit will be presented by cohort/treatment received during OP using All Safety population. Categorization of vital signs data will be presented by treatment (OP versus ATP) per cohort/treatment received during OP using All Safety population with data in the ATP column only including participant who received at least 1 dose of PF-06741086 in the denominator.

6.6.6. Laboratory Data

Shift tables will be presented by cohort/treatment received during OP using All Safety population. In addition, descriptive summary of observed values and change from baseline to the last on-treatment observation will be presented.

Incidence of laboratory abnormalities will be summarized by treatment (OP versus ATP) irrespective of the baseline abnormality. Laboratory test abnormalities by participant will also be reported.

6.6.7. ECG

Occurrence of abnormal ECG findings or abnormal cardiac troponin I (cTnI) results will be summarized by treatment (OP versus ATP) per cohort/treatment received during OP using All Safety population with data in the ATP column only including participant who received at least 1 dose of PF-06741086 in the denominator.

6.6.8. Physical Examination

Physical examination will be summarized by treatment (OP versus ATP) per cohort/treatment received during OP using All Safety population with data in the ATP column only including participant who received at least 1 dose of PF-06741086 in the denominator.

6.6.9. Severe Hypersensitivity and Anaphylactic Reactions

Incidence of severe hypersensitivity and anaphylactic reactions will be summarized using the same method applied to the summary of thrombotic event detailed in Section 6.6.2.

6.7. Impact of COVID-19 Summary

The following data listings and summary will be provided pertaining to COVID-19-related impact on the study:

- Listing of study discontinuation due to COVID-19 (All Safety Set)
- Listing of PF-06741086 dose adjustment due to COVID-19 (All Safety Set)
- Listing of missed Haem-A-QoL assessment due to COVID-19 (All Safety Set)
- Listing of Protocol Deviations related to COVID-19 (All Safety Set)

- Summary and listing of COVID-19 related aEs by SOC and PT (All Safety Set)
- Listing of study discontinuation due to COVID-19 related AE, COVID-19 related SAE, and deaths related to COVID-19 (All Safety Set)
- Listing of COVID-19 vaccination (prior and concomitant)

7. INTERIM ANALYSES

7.1. Introduction

Upon completion of the 12-month ATP for the Non-Inhibitor Cohorts (with prior On-Demand therapy as well as with prior Prophylaxis), analyses of the primary endpoint of interest along with secondary endpoints in the corresponding participant populations will be performed to assess respective efficacy and safety for a potential initial registration filing regardless of whether or not follow-up is ongoing in the Inhibitor cohort. These analyses will be considered the final analyses for these participant populations. The study will continue until the completion of participants from all cohorts.

Protocol Section 7.1 describes discontinuation of study intervention at individual level and study level due to safety concern throughout the study.

7.2. Interim Analyses and Summaries

Type I error rate of 2-sided 0.05 is allocated separately to each of the 3 populations (Inhibitor Cohort; Non-Inhibitor Cohort with prior On-Demand therapy as well as with prior Prophylaxis). Therefore, analyses upon completion of the Non-Inhibitor Cohorts (with prior On-Demand therapy as well as with prior Prophylaxis) do not impact the Type I error for the Inhibitor Cohort.

8. STATISTICAL ANALYSIS FOR CHINA COHORT

The purpose of enrolling additional participants in China is to support the registration of the indication in China. Approximately CCI % of the overall planned enrollment into Study B7841005, which has an anticipated sample size of approximately 145 global participants, should come from China per China's registration requirements.

Participants enrolled in China within the enrollment period of the Global cohort and in the China Extension cohort will be pooled together to form the China cohort to support the registration in China. Data from the China cohort will be analyzed separately per local regulatory requirement in China and reported in a separate document.

The analysis of the China cohort will use the same statistical methods as the Global cohort except that no hypothesis testing and decision rule will be performed. The analysis of the China cohort will be conducted when all participants enrolled in the China Extension cohort have completed 1 year of treatment with PF-06741086 or have otherwise withdrawn from treatment.

9. APPENDICES

Appendix 1. Haem-A-QoL/Heamo-QoL

Table A1.1 Haem-A-QoL: Minimum Number of Valid Items to Derive Normalized Transformed Scaled Scores

Scores to be derived	N. of Items (minimum N of valid items)
Domain	
Physical health	5 (3)
Feelings	4 (2)
View of yourself	5 (3)
Sports and leisure	5 (3)
Work and school	4 (2)
Dealing with haemophilia	3 (2)
Treatment	8 (4)
Future	5 (3)
Family planning	4 (2)
Partnership and sexuality	3 (2)
Total	46 (23)

Table A1.2 Haemo-QoL: Minimum Number of Valid Items to Derive Normalized Transformed Scaled Scores

Scores to be derived	N. of Items (minimum N of valid items)
Domain	
Physical health	7 (6)
Feelings	8 (6)
Attitude/View	10 (8)
Family	8 (6)
Friends	4 (3)
Other People	6 (5)
Sports & School	9 (7)
Coping/Dealing	7 (6)
Treatment	8 (6)
Perceived Support	4 (3)
Future	4 (3)
Relationship	2 (2)
Total	77 (63)

Appendix 2. HAL2/pedHAL

Appendix 2.1. Items for Each Component

Score		Items	Score range
Lying / sitting / kneeling / standing	LSKS	1-8 (8)	8 - 48
Functions of the legs	LEGS	9-17 (9)	9 - 54
Functions of the arms	ARMS	18-21 (4)	4 - 24
Use of transportation	TRANS	22-24 (3)	3 - 18
Self care	SELFC	25-29 (5)	5 - 30
Household tasks	HOUSEH	30-35 (6)	6 - 36
Leisure activities and sports	LEISPO	36-42 (7)	7 - 42
Upper Extremity Activities	UPPER	* (9)	9 - 54
Basic Lower Extremity Activities	LOWBAS	** (6)	6 -36
Complex Lower Extremity Activities	LOWCOM	*** (9)	9 - 54
Sum score	SUM	1-42 (42)	42 - 252

* Items for UPPER-component: 18, 19, 20, 21, 25, 26, 27, 28, 29. (9 items)
** Items for LOWBAS-component: 8, 9, 10, 11, 12, 13. (6 items)
*** Items for LOWCOM-component: 3, 4, 6, 7, 14, 15, 16, 17, 22. (9 items)

Appendix 2.2. Minimum Number of Valid Items to Derive Normalized Scores

Table A2.2.1 HAL2/pedHAL: Minimum Number of Valid Items to Derive Normalized Scores

	N. of Items (minimum N of available items)	
Scores to be derived	HAL2	pedHAL
Domain		
Lying/sitting/kneeling/standing (HAL2) Sitting/kneeling/standing (pedHAL)	8 (4)	10 (5)
lower (leg) functioning	9 (5)	11 (6)
upper (arm) functioning	4 (2)	6 (3)
Transportation	3 (2)	3 (2)
Self-care	5 (3)	9 (5)
Household tasks	6 (3)	3 (2)
Sports/Leisure	7 (4)	11 (6)
Component		
Activities involving the Upper Extremities	9 (5)	n/a
Basic activities involving the Lower Extremities	6 (3)	n/a
Complex activities involving the Lower Extremities	9 (5)	n/a
Total	42 (21)	53 (27)

Appendix 3. Country Mapping for Geographical Regions

Table A3.1 Country Mapping for Geographic Regions

Countries	Geographic Region[1]
Canada	A
China	C
Croatia	B
France	B
Hong Kong	C
India	B
Italy	B
Japan	C
Korea	C
Republic of Mexico	A
Oman	A
Russian Federation	B
Serbia	B
Spain	B
Taiwan	C
Turkey	B
United States	A

1. (A) North America and Middle East; (B) Europe and India, (C) Asia excluding India.

Appendix 4. Derivation of Decision Rules Regarding Superiority and Non-Inferiority

Appendix 4.1. Non-Inferiority Margin for Treated Bleeds

Appendix 4.1.1. Derivation of Non-Inferiority Margin for Treated Bleeds

The section describes the derivation of the non-inferiority margin for treated bleeds. The derivation followed the methodology outlined in FDA (Food and Drug Administration) Guidance for Non-Inferiority Clinical Trials to Establish Effectiveness.[1]

Historical data were selected among Pfizer hemophilia studies. This enabled selection of individual participant level data that would closely match the inclusion/exclusion criteria of Study B7841005. In selecting studies, the following aspects were considered: (1) having both prophylaxis and on-demand treatment, providing a within-study treatment difference estimate, (2) having an overlapping age group of 12-74 years old to match Study B7841005 age group, and (3) not being a regional study. The following 3 studies satisfied these criteria: B1821010 (BeneFIX), B1821002 (BeneFIX), and B1831004 (Xyntha/ReFacto).

B1821010 was an open-label 1-way crossover study intended to compare the efficacy and safety between the on-demand treatment and prophylactic therapy using BeneFIX in hemophilia B patients with ages 12-65 years old and the factor IX activities not greater

than 2%. This study comprised Period 1 (on-demand treatment at a dosage determined by principal investigator, 6 months) and Period 2 (100 IU/kg once-weekly dose, 12 months).

B1821002 was an open-label, randomized, 4-period crossover study on males with hemophilia B with ages 6-65 years old and the factor IX activities not greater than 2% to compare on-demand administration and 2 prophylaxis regimens. Study was conducted following 4 periods:

- Period 1: On-demand administration per dose directed by investigators, 16 weeks.
- Period 2: 50 IU/kg/ twice weekly or 100 IU/kg/once weekly (randomized), 16 weeks.
- Period 3: On-demand administration per dose directed by investigators, 8 weeks.
- Period 4: 50 IU/kg/twice weekly or 100 IU/kg/once weekly (different regimen than that applied during Period 2), 16 weeks.

B1831004 evaluated safety and efficacy of ReFacto AF in hemophilia A patients switching to ReFacto AF from ReFacto or other factor VIII products in usual care settings. The patients were of ages 12 and greater and had factor XIII activities <1%. All participants were treated under standard care with ReFacto AF at a dose and frequency prescribed by the treating physician. The duration of treatment depended on the frequency of infusions (to achieve 100 exposure days) with mean (SD) of 336 (223) days.

Table A4.1.1.1 presents analyses of the ABR of treated bleeds of subject level data from 3 Pfizer hemophilia studies with inclusion/exclusion criteria that closely matched those for Study B7841005. The analyses included subjects with matching age (12-74 years old) and factor level (<1%) criteria to Study B7841005. The mean ABR difference between prophylaxis and on-demand treatment was estimated using 3 analyses: t-test allowing for non-homogeneous variances, Hodges-Lehman location-shift parameter estimation, and additive negative binomial model. All analyses resulted in similar estimated mean differences of approximately 21 and the lower bound of 2-sided 95% confidence intervals ranging from 16.8 to 17.6.

Table A4.1.1.1 Analyses of Bleeding in 3 Pfizer Hemophilia Studies Used in Deriving the Non-Inferiority Margin

	BeneFIX B1821010	BeneFIX B1821002	Xyntha/ReFacto AF B1831004	Overall
On Demand				
N	22	9	52	83
Mean ABR (95% CI)[1]	31.5 (24.3, 40.7)	26.3 (16.9, 41.0)	28.5 (22.7, 35.7)	29.0 (24.6, 34.3)
Prophylaxis				
N	22	9	154	185
Mean ABR (95% CI)[1]	4.0 (2.2, 7.2)	2.6 (0.8, 8.4)	7.9 (6.2, 10.2)	7.2 (5.7, 9.0)
ABR Ratio (95% CI)[2]: Prophylaxis/On Demand				
	0.13 (0.08, 0.19)	0.10 (0.03, 0.29)	0.28 (0.18, 0.42)	0.25 (0.18, 0.34)
ABR Difference (95% CI): On Demand – Prophylaxis				
D1: t-test Satterthwaite	27.4 (20.6, 34.3)	23.7 (11.5, 36.0)	19.9 (14.0, 25.8)	21.3 (16.8, 25.9)
D2: Hodges Lehman	26.8 (20.3, 34.3)	24.4 (10.2, 39.6)	19.5 (16.0, 23.5)	20.8 (17.3, 24.8)
D3: additive NB	27.5 (21.1, 33.8)	23.7 (13.8, 33.6)	20.5 (10.3, 30.8)	21.9(17.6, 26.2)

1: Model-based from negative binomial regression
2: Negative binomial regression
Notes:
D1: t-test for unequal variances assuming independence
D2: HL location shift parameter estimate assuming independence except for BeneFIX studies
D3: additive negative binomial model using identity link

Based on the lower bound of the confidence interval of this estimate, the on-demand ABR was assumed to be higher by at least 16.8, which was considered M_1 in the non-inferiority test setting.

Given the large effect size of prophylaxis treatment, an appropriate value for M_2 was considered in order to preserve a sufficiently large proportion of this effect. A value of 2.5 for M_2 (~85% effect preserved) was selected as both clinically meaningful and yielding a reasonable sample size. M_1 and M_2 are derived from the notation used in the FDA Guidance document, Non-Inferiority Clinical Trials to Establish Effectiveness.[1]

The process of deriving the non-inferiority margin and the selected magnitude of the margin also satisfy the principles delineated in the EMA/CHMP Guideline on the Choice of the Non-inferiority Margin.[2] In particular, the margin of 2.5 establishes acceptable efficacy relative to the active comparator predicated in Section 4 of the aforementioned guideline:

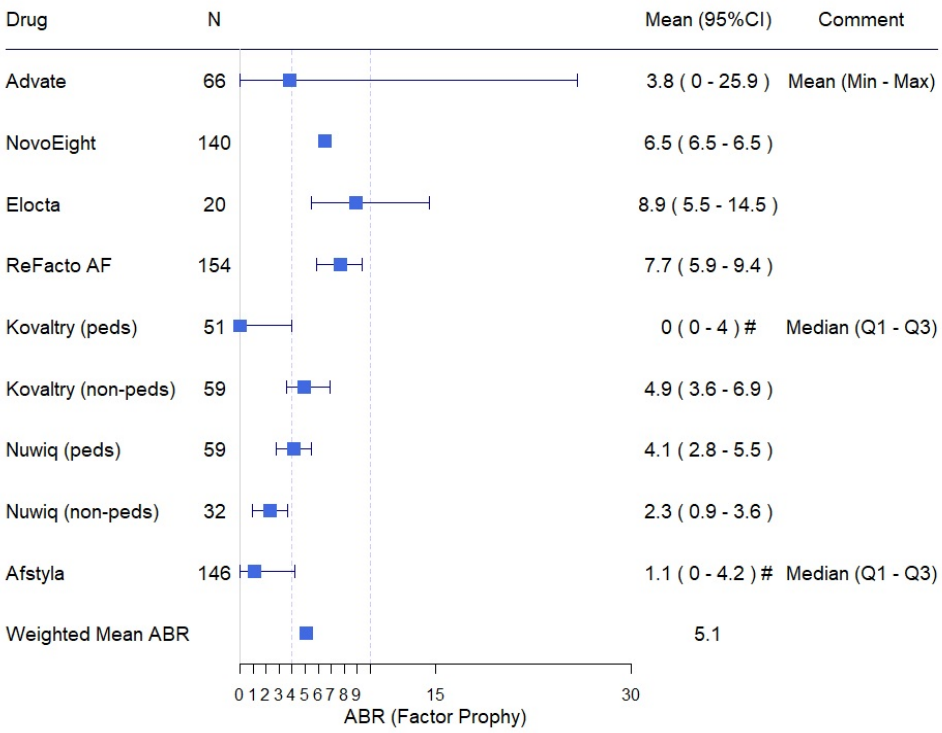
- The estimated effect of factor replacement therapy versus on-demand and its most conservative lower bound of the 95% CI is ~21 ABR and 16.8, respectively.
- The margin of 2.5 preserves 85% of the most conservative estimated effect of 16.8.

- In meta-analyses of factor replacement products to 2017, the weighted mean ABR was 5.1 for factor VIII prophylaxis (Figure A4.1.1.1) and 3.1 for factor IX prophylaxis (Figure A4.1.1.2); for factor IX prophylaxis, all products have participants with factor level activity ≤2%, the planned phase 3 <1%; therefore, the non-inferiority margin of 2.5 should be interpreted in the background comparator mean ABR of ~4.5.

- Expected half-width of 95% CI is 2.1 applying the parameter assumptions for the sample size calculation (ABR of 2.5 for 6 months for prior prophylaxis, 4 for 1 year for PF06741086 prophylaxis, variance = 6*mean, correlation = 0.2); therefore, meeting the non-inferiority margin of 2.5 requires that the point estimate for the difference is close to 0; further establishing acceptable efficacy of PF06741086 relative to factor replacement therapy.

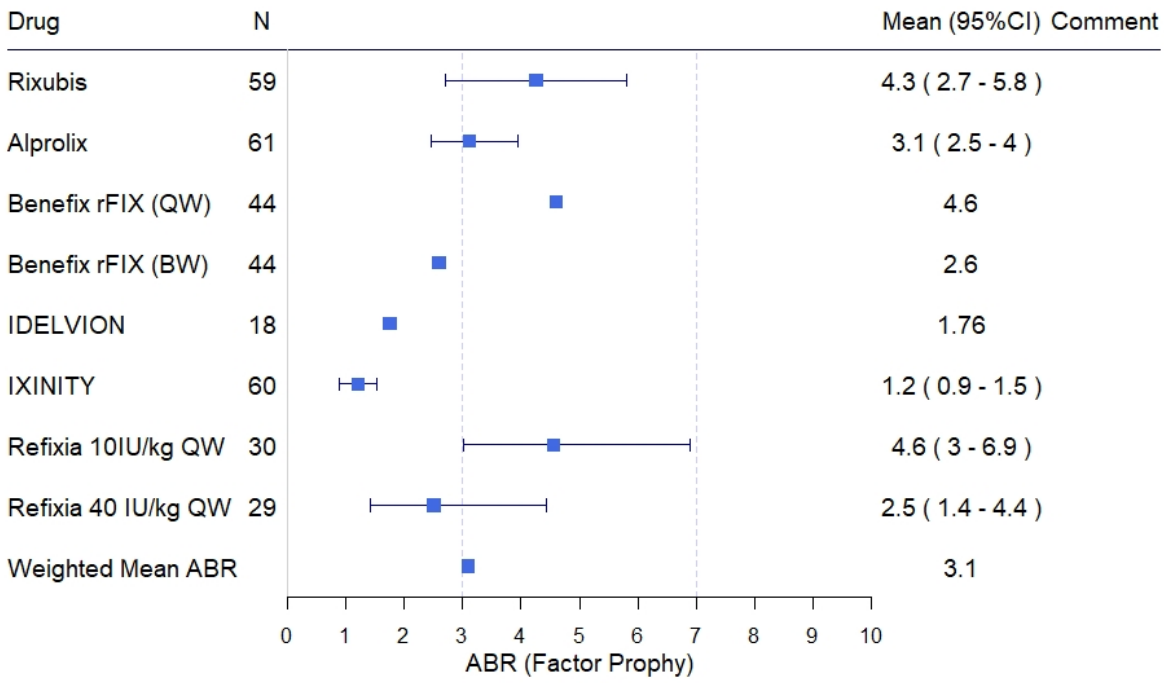
- In addition, PF-06741086 has an advantage via once weekly subcutaneous dosing versus up to 2 or 3 times weekly IV dosing of the active comparators. Per guidance, it may be possible to justify a wider non-inferiority margin for efficacy in this case.

Figure A4.1.1.1 Results for Prophylaxis With FVIII Products (to 2017)



The sources are the FDA summary basis of approvals for all factor replacement therapies that were approved post 2010 in the United States as well as Pfizer internal data (ReFactoAF).

Figure A4.1.1.2 Results for Prophylaxis With FIX Products (to 2017)



The sources are the FDA summary basis of approvals for all factor replacement therapies that were approved post 2010 in the United States.

Clinically, the planned non-inferiority hypothesis will be met if the difference between PF-06741086 and factor-based prophylaxis regimens does not exceed 2.5. The pooled ABR observed across the 4 cohorts studied in the 3 month phase 2 B7841002 study of PF-06741086-based prophylaxis was 2.6 (N=24).[3] Reproducing this result in B7841005 during the active treatment period, and assuming a factor-based prophylaxis ABR of ~4.5 during the observation period, would place the upper bound of the 95% CI difference at around and ABR of 5, which is comfortably within the range of ABR achieved with approved factor XIII and IX-based prophylaxis regimens depicted in Figures A4.1.1.1 and A4.1.1.2.

Appendix 4.1.2. Non-Inferiority Margin: ABR Difference Versus Ratio

The statistical strategy for the non-inferiority comparison based on ABR rate *difference* was provided to the FDA, and the FDA encouraged Pfizer to consider using the ABR rate *ratio* rather than the ABR *difference* for the non-inhibitor population. As per the FDA's request, Pfizer had provided more details regarding the derivation of non-inferiority margin for rate ratio and rate difference as follows.

Table A4.1.1.1 presented analyses of bleeding data using the metric of ABR rate difference as well as the rate ratio along with descriptive statistics for the 3 Pfizer hemophilia studies in which the original non-inferiority margin based on rate difference had been derived. While

the estimated ABR ratios of the standard prophylaxis therapy over the on demand treatment varied considerably from study to study (0.13, 0.10, and 0.28), the estimated differences were relatively stable (27.4, 23.7, and 19.9) across the 3 studies. Therefore, Pfizer believes that the rate difference provides a more consistent estimate of the average effect across studies and supports a more credible choice of M_1 as per the principle laid out in the FDA Guidance.[1]

The difference in ABR (rather than the rate ratio) also provides a more meaningful clinical context for interpretation of results for PF-06741086. Small differences in the frequency of joint bleeds have been shown to result in clinically meaningful differences in long term joint outcomes. Evidence to this point comes from clinical studies that have evaluated modest delays in implementation of prophylaxis. These delays are associated with the respective individuals incurring a limited number of joint bleeds and result in more orthopedic damage relative to individuals where these delays were not incurred. In a study comparing results for prophylaxis over a median treatment period of 11.25 years, individuals beginning treatment prior to age 2 with ≤1 prior joint bleed had better orthopedic and radiologic joint scores than individuals who began prophylaxis at the age of 3-6 years after incurring a median of 6 (range 1-8) joint bleeds.[4] The Scientific Subcommittee of the ISTH has reviewed studies reporting effect of age on prophylaxis outcomes.[5] By age 19 the Peterson radiographic score had increased by 8% for each year's delay in starting prophylaxis following the first joint bleed and joint outcomes were better for patients beginning prophylaxis before the second joint bleed than in patients who started prophylaxis later. These data show that even low numbers of joint bleeds are clinically meaningful as they result in decrements in joint health and underscore the importance of absolute differences in frequency of bleeding episodes when evaluating prophylaxis treatment regimens. Such small but clinically meaningful differences in the observed bleed rates can be readily apparent when the more clinically meaningful endpoint of absolute bleed rates is used for the comparison of prophylaxis efficacy.

In advocating the metric of rate difference, Pfizer also acknowledges the FDA's concern regarding the normality assumption of the t-test that had been utilized to derive the rate difference non-inferiority margin. To assess its impact on the magnitude of the derived non-inferiority margin, Pfizer conducted additional analyses: a non-parametric approach (Hodges-Lehmann estimate of location shift) and an additive negative binomial model accounting for within-subject correlation (parameterized similar to additive Poisson model). The non-parametric analysis and the additive negative binomial model based on all studies resulted in the estimated rate differences and 95% Cis that were very close to those from the t-test. In particular, the lower bounds of the 95% CI from these methods would result in an almost identical M_2, thus allaying the Agency's concern.

In conclusion, Pfizer advocates for the use of the rate difference in defining the non-inferiority margin compared with the rate ratio; the estimates in the former were more consistent across studies and the difference enables a clinically meaningful comparison of prophylaxis efficacy which is easier to interpret, especially when the mean ABR could be as low as 2 or 3 in the active control of standard prophylaxis. Furthermore, the derived non-inferiority margin of 2.5 in rate difference is stable under various distributional assumptions, preserving about 85% of the effect seen in the prophylaxis treatment. A sample size of

60 evaluable participants is projected to provide ≥90% of power to show non-inferiority using a repeated measure negative binomial regression with identity link function using type I error rate of 1-sided 0.025 when the mean ABR is assumed to be 4 for PF-06741086 and 5 for standard prophylaxis.

To further aid in the FDA's review, Table A4.1.2.1 presents several non-inferiority margins (M_2) in rate ratio and its required sample size to have 90% of power to demonstrate non-inferiority. The margins are derived based on the estimated rate ratio and the 95% CI of 0.25 (0.18, 0.34) from the 3 studies as presented in Table A4.1.1.1. It appears that equating the percent of effect preserved in defining M_2 for difference and ratio leads to very different required sample sizes.

Table A4.1.2.1 Various Non-Inferiority Margins and the Required Sample Size

Percent preserved for Ratio	M_2 for Ratio	N to have 90% power to show non-inferiority with 1-sided type I error rate of 0.025		
		ABR=4 for PF-06741086 ABR= 4 for Prophy	ABR=4 for PF-06741086 ABR= 4.5 for Prophy	ABR=4 for PF-06741086 ABR= 5 for Prophy
90%	1.20	>500	450	240
80%	1.41	400	200	120
70%	1.61	200	120	90
60%	1.81	130	90	60
50%	2.02	90	70	50

Note: sample size was derived based on simulation where data were generated using negative binomial distribution with the stated means (4, 4.5, or 5) and variance = 6*mean. The correlation between the 2 bleed counts from the same participant was assumed to be 0.2.

Appendix 4.2. Criteria for Other Bleed Count Endpoints

Appendix 4.2.1. Non-Inferiority Margin for Comparison Versus Prior Prophylaxis

Non-Inferiority Margin for Incidence of Joint Bleeds

The following describes derivation of a non-inferiority margin for incidence of joint bleeds. Table A4.2.1.1 presents analyses based on the 2 BeneFIX studies since the ReFacto study did not utilize this endpoint. All analyses resulted in similar estimated mean ABR differences of approximately 21 and the lower bound of 2-sided 95% confidence intervals ranging from 15.2 to 16.5. Therefore, the on-demand ABR was assumed to be higher than the prophylaxis ABR by at least 15.2, which was considered M_1 in the non-inferiority test setting. Given the large effect size of prophylaxis treatment, an appropriate value for M_2 was considered in order to preserve a sufficiently large proportion of this effect. A value of 2.5 for M_2 (~84% effect preserved) was selected as clinically meaningful for a planned sample size of 60.

Table A4.2.1.1 Analyses of Incidence of Joint Bleeds in 3 Pfizer Hemophilia Studies

	BeneFIX B1821010	BeneFIX B1821002	ReFacto 3082B2-4432*	Overall
On Demand				
N	22	9	NA	31
Mean ABR (95% CI)[1]	25.7 (19.2, 34.3)	19.6 (12.5, 30.8)	NA	24.0 (18.8, 30.7)
Prophylaxis				
N	22	9	NA	31
Mean ABR (95% CI)[1]	2.3 (1.1, 4.9)	2.0 (0.6, 6.5)	NA	2.2 (1.2, 4.2)
ABR Difference (95% CI): On Demand – Prophylaxis				
D1: t-test Satterthwaite	23.4 (16.4, 30.3)	17.7 (7.1, 28.3)	NA	21.7 (16.2, 27.3)
D2: Hodges Lehman	22.0 (15.7, 30.2)	17.9 (6.2, 32.0)	NA	20.6 (15.2, 26.5)
D3: additive NB	23.4 (17.0, 29.8)	17.6 (9.0, 26.2)	NA	21.7 (16.5, 27.0)

*: ReFacto study did not have the endpoint of incidence of joint bleeds.

1. Model-based from negative binomial regression.

Note:
D1: t-test for unequal variances assuming independence.
D2: HL location shift parameter estimate assuming independence except for BeneFIX studies.
D3: additive negative binomial model using identity link.

Non-Inferiority Margin for Incidence of Spontaneous Bleeds

The following describes derivation of a non-inferiority margin for incidence of spontaneous bleeds. Table A4.2.1.2 presents analyses based on the 2 BeneFIX studies since the ReFacto study did not utilize this endpoint. Results from the t-test and negative binomial regression were similar with the estimated ABR mean difference of 18.6 and similar 95% cIs. The result from Hodges-Lehman estimation was larger with 21.8 (16.0, 27.4) for the estimated ABR median difference and its 95% CI.

All analyses indicated that the on-demand ABR was higher than the prophylaxis ABR by at least 13.2, which was considered M_1 in the non-inferiority test setting. Given the large effect size of prophylaxis treatment, an appropriate value for M_2 was considered in order to preserve a sufficiently large proportion of this effect. A value of 2.5 for M_2 (~81% effect preserved) was selected as clinically meaningful for a planned sample size of 60.

Table A4.2.1.2 Analyses of Incidence of Spontaneous Bleeds in 3 Pfizer Hemophilia Studies

	BeneFIX B1821010	BeneFIX B1821002	ReFacto* 3082B2-4432	Overall
On Demand				
N	22	9	NA	31
Mean ABR (95% CI)[1]	23.5 (15.8, 35.0)	15.2 (7.0, 33.1)	NA	21.1 (14.7, 30.4)
Prophylaxis				
N	22	9	NA	31
Mean ABR (95% CI)[1]	2.9 (1.4, 6.0)	1.6 (0.4, 6.0)	NA	2.5 (1.3, 4.8)
ABR Difference (95% CI): On Demand – Prophylaxis				
D1: t-test Satterthwaite	20.6 (13.8, 27.4)	13.7 (3.9, 23.4)	NA	18.6 (13.2, 24.0)
D2: Hodges Lehman	23.7 (16.3, 31.2)	17.7 (6.4, 29.0)	NA	21.8 (16.0, 27.4)
D3: additive NB	20.6 (14.3, 26.9)	13.6 (5.8, 21.4)	NA	18.6 (13.4, 23.7)

*: ReFacto study did not have the endpoint of incidence of joint bleeds.

1. Model-based from negative binomial regression.

Note:
D1: t-test for unequal variances assuming independence.
D2: HL location shift parameter estimate assuming independence except for BeneFIX studies.
D3: additive negative binomial model using identity link.

Non-Inferiority Margin for Incidence of Target Joints Bleeds

The following describes derivation of a non-inferiority margin for incidence of target joint bleeds. The incidence of target joint bleeds was not an efficacy endpoint in any of the 3 Pfizer hemophilia studies in which the estimates for several bleed count endpoints were derived. In the absence of internal data, Pfizer relied on the results of emicizumab publications[6,7] of HAVEN1 and HAVEN3 studies. Pfizer assumed that the ratio of the effect expressed as the difference between the on-demand treatment versus emicizumab in target joint bleeds over that in joint bleeds would be applicable to characterize the PF-06741086 effect of the same endpoints. Table A4.2.1.3 shows that the effect in target joint bleeds is consistently 49% of that in joint bleeds. Therefore, the non-inferiority margin of 1.2, derived as 49% of the margin of 2.5 for the joint bleeds, was selected for target joint bleeds expecting that a similar magnitude (~84%) of the effect would be preserved.

Table A4.2.1.3 Estimated Effect of Target Joint Bleeds and Joint Bleeds From Emicizumab Studies

Mean Difference	HAVEN1 (inhibitors)	HAVEN3* (non-inhibitors)
Target joint bleeds	2.9	12.4
Joint bleeds	5.9	25.4
ratio	0.49	0.49

* difference between the on-demand treatment and the emicizumab once weekly group.

Non-Inferiority Margin for Incidence of Total Bleeds (Treated and Untreated)

The following describes derivation of a non-inferiority margin for incidence of total bleeds. The incidence of total bleeds was not an efficacy endpoint in any of the 3 Pfizer hemophilia

studies in which the estimates for several bleed count endpoints were derived. In the absence of internal data, Pfizer relied on the results of emicizumab publications[1,2] of HAVEN1 and HAVEN3 studies. Pfizer assumed that the ratio of the effect expressed as the difference between the on-demand treatment versus emicizumab in total bleeds over that in treated bleeds would be applicable to characterize the PF-06741086 effect of the same endpoints. Table A4.2.1.4 shows that the effect in total bleeds exceeds 12% to 23% that in treated bleeds, resulting in the geometric mean of 17% increase. Therefore, the non-inferiority margin of 2.9, derived as 117% of the margin of 2.5 for the treated bleeds, was selected for total bleeds expecting that a similar magnitude (~85%) of the effect would be preserved.

Table A4.2.1.4 Estimated Effect of Treated Bleeds and All Bleeds From Emicizumab Studies

Mean Difference	HAVEN1 (inhibitors)	HAVEN3* (non-inhibitors)
All bleeding	22.8	45.1
Treated bleeding	20.4	36.7
Ratio	1.12	1.23

*difference between the on-demand treatment and the emicizumab once weekly group

Appendix 4.2.2. Superiority Criterion for Comparison Versus On-Demand Treatment

Superiority criterion for comparison versus on-demand treatment for the following bleed count endpoints was set as 0.5: incidence of joint bleeds, incidence of spontaneous bleeds, incidence of target joint bleeds, and incidence of all bleeds. This criterion is consistent with the 1 set for treated bleeds.

Appendix 4.3. Criteria for Key Secondary HRQoL Endpoints

Appendix 4.3.1. Non-Inferiority Margin for Key Secondary HRQoL Endpoints

This section defines non-inferiority margins for HRQoL endpoints that are key secondary in any of the regions (physical health domain and the total score of Haem-A-QoL; EQ-5D-5L index and EQ-5D VAS scores). These endpoints were not measured in any of the 3 Pfizer hemophilia studies where non-inferiority margins for other endpoints were derived. In addition, emicizumab failed to consistently demonstrate treatment effect in these measures.[6,7] To define non-inferiority margins without reliable internal/external treatment effect estimates, Pfizer adopted a notion in McLeod[8] where an individual-level threshold based on a minimally important difference (MID) was applied as a group-level threshold to interpret the meaningfulness of group difference. This strategy effectively set the non-inferiority margins as MIDs.

Wyrwich[9] derived an MID of 10 and 7 points for physical health domain and the total score of Haem-A-QoL, respectively, following the procedures laid out in the FDA guidance for patient reported outcome measures.[10] Likewise, Pickard[11] derived an MID for EQ-5D based on UK index score as around 0.1 and the range of 7 to 12 for EQ-VAS. Therefore, the following non-inferiority margins were set reflecting the reported MIDs for each of these

endpoints: 10 points for the physical health domain of Haem-A-QoL, 7 points for the total score of Heam-A-QoL, -0.1 for the EQ-5D-5L index score using UK crosswalk values, and -9.5 (the averages of 7 and 12, added the negative sign as higher scores represent better health status) for EQ-VAS. These margins are interpreted as thresholds that would rule out important differences.

Appendix 4.3.2. Superiority Criterion for Comparison Versus On-Demand Treatment for Key Secondary HRQoL Endpoints

Superiority criterion for comparison versus on-demand treatment for each HRQoL endpoint was set as 0.

References for Appendix 4

1. FDA Guidance for Industry. Non-inferiority clinical trials to establish effectiveness. Nov 2016.

2. EMA/CHMP Guideline on the Choice of the Non-inferiority Margin. Issued 27 July 2005. Doc. Ref. EMEA/CPMP/EWP/2158/99.

3. B7841002 Clinical Study Report.

4. Kreuz W, Escuriola-Ettingshausen C, Funk M, et al. When should prophylactic treatment in patients with haemophilia A and B start? – The German experience. Haemophilia 1998;4:413-417.

5. Fischer K, Collins PW, Ozelo MC, et al. When and how to start prophylaxis in boys with severe hemophilia without inhibitors: communication from the SSC of the ISTH. Journal of Thrombosis and Haemostasis 2016;14:1105-1109.

6. Oldenburg J, Mahlangu JN, Kim B, et al. Emicizumab prophylaxis in Hemophilia A with inhibitors. N Engl J Med 2017;377:809-18.

7. Mahlangu J, Oldenburg J, Paz-Priel I, et al. Emicizumab prophylaxis in patients who have Hemophilia A without inhibitors. N Engl J Med. 2018;379(9):811-22.

8. McLeod L, Cappelleri J, Hays R. Best (but oft-forgotten) practices: expressing and interpreting associations and effect sized in clinical outcome assessments. Am J Clin Nutr. 2016;103(3):685-93.

9. Wyrwich KW, Krishnan S, Poon JL, et al. Interpreting important health-related quality of life change using the Haem-A-QoL. Haemophilia. 2015;21(5):578-84.

10. FDA Guidance for Industry: patient-reported outcome measures: use in medical product development to support labeling claims. Dec 2009.

11. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes 2007; 5:70

Appendix 5. Statistical Methodology Details

Example SAS Codes for Testing vs. On-demand for Bleed Count Endpoints

```
proc genmod data=alldata;
        class trt subjid;
        model bleeds = trt /offset=lyears dist=negbin link=log alpha=0.05;
        repeated subject=subjid/ Type=UN;
        estimate 'log diff' trt 1 -1;
run;
```

Example SAS Codes for Testing vs. Prior Prophylaxis for Bleed Count Endpoints

```
proc genmod data=alldata;
class trt subjid;
model bleeds = years years*trt/ noint dist=negbin link=identity;
repeated subject=subjid/ Type=UN;
estimate 'diff' years*trt 1 -1/alpha=0.05;
run;
```

Example SAS and R Codes for Analysis of the Physical Domain of Heam-A-QoL

Multiple Imputation

```
*create missing-at-random MI data;
*y0 y60 y120 y180 y240 y360: values observed at ATP baseline, day 60,…;
proc mi data=observed seed=1271 out=impute0 nimpute=10_impute;
     monotone reg(/ details); *if nomotone missing;
     *FCS nbiter=20 reg(/ details); *if not monotone missing;
     var y0 y60 y120 y180 y240 y360;
run;
*apply shift value of 10 (=MID) for missing due to AE or lack of efficacy;
data impute1; set impute0;

     array ys(5) y60 y120 y180 y240 y360;
     do i=1 to 5;
        if ys(i)<0 then ys(i)=0;
        if ys(i)>100 then ys(i)=100;
     end;
     drop I;
run;
```

Prepare Wilcoxon Signed Rank Test Statistics for Each Imputed Set

```
*get the treatment difference;
data get_dif; set XXXX;
     chg_op=x2-x1;
     chg_atp=y5-x3;
     trtdif=chg_atp-chg_op;
     abs_dif=abs(trtdif);
run;
data for_anal; set get_dif;
     if abs_dif=0 then delete;
```

```
        *zero is ignored in Wilcoxon Signed Rank Test;
run;
*get the exact statistic using PROC RANK;
proc rank data=for_anal out=ranks ties=mean; by studyID _imputation_;
     var abs_dif;
     ranks r_dif;
run;
data ranks2; set ranks;
     if trtdif>0 then sg_rank=r_dif; else sg_rank=0;
     r2=(r_dif)**2;
run;
proc means noprint data=ranks2; by _imputation_;
     var r_dif sg_rank r2;
     output out=outr sum=sumr sumpos sumr2 n=n;
run;
*centered rank sum and variance, exact values;
data mytest; set outr;
     mys0=sumpos-sumr/2;
     num_var=n*sumr2/4-mys0**2;
     se_e=sqrt(sumr2/4);
run;
```

Combine the Results

```
*Test: combining results across MI sets via Rubin's rule;
proc mianalyze data=mytest;
     modeleffects mys0;
     stderr se_e;
run;

*prepare data to be sent to R in a horizontal format;
proc transpose data=get_dif out=60or(drop=_name_) prefix=g;
     by _imputation_;
     var trtdif;
     id subjid;
run;
data testdat_to_R;
     merge 60or mytest(keep=studyID _imputation_  mys0 se_e);
     by _imputation_;
run;
```

Estimation Implemented in R

Get the Exact Estimation in R using wilcox.exact() for Each Imputed Set

```
#perform exact Wilcoxon signed rank test per imputed set
#difs: dim (n_imputation x n_subject) matrix containing only the treatment differences
#outwilc: +rank sum(not centered), p-value, HL estimate, low bd, hi bd
outwilc=matrix(NA, nrow=n.test, ncol=5) #to store the wilcox.exact result
for (I in 1:n.test){
  w=wilcox.exact(difs[I,],exact=T,conf.int=T)
  outwilc[I,]=c(w$statistic, w$p.value,w$estimate,w$conf.int)
}
```

Derive Needed Statistics for Overall Estimation

```
ci_len=outwilc[,5]-outwilc[,4] #length of CI
```

```
se2=ci_len/2/qnorm(.975)
within.var.med=se2^2
s0=XXXX #centered rank sum, EH calculated in SAS
allwilc=data.frame(id,outwilc,ci_len,se2,within.var.med,s0)
library(dplyr)
grp <- group_by(allwilc, studyID)
out=summarise(grp, mean.rank=mean(s0), bet.var.rank=var(s0),
      mean.med=mean(X3), bet.var.med=var(X3),
      mean.within.var.med=mean(within.var.med))
```

Overall Estimation on Median and 95% CI

```
oneuse=(1+1/n_imputation)*out$bet.var.med
tot=out$mean.within.var.med+oneuse #total variance for inference on median
se.mi.med=sqrt(tot)
df=(n_imputation-1)*(1+out$mean.within.var.med/oneuse)^2
half=qt(.975,df)*sqrt(tot)
#out$mean.med. lo.med & hi.med are the overall median estimate & 95% CI
lo.med=out$mean.med-half
hi.med=out$mean.med+half
```

Appendix 6. Justification for Updated Definition of "NEW" Untreated Bleed

Table A6.1 compares several definitions for new untreated bleed using hypothetical scenarios in reference to clinical judgement. The last line in each example provides justification for the clinical judgement. In cases where the algorithm looks back at all prior bleeds, as in columns (8) and (9), checking the >48 hour previous untreated bleed criteria against the previous "NEW" untreated bleeds, as well as the replacement of the "or" condition with "and", results in the conclusion that matches the clinical judgement in column (4). Notice that when only an immediate prior bleed is considered, the introduction of the "NEW" untreated bleed condition is sufficient without changing the "or" condition into "and". That is, column (10) results in the same conclusion as those decided by clinical judgement.

Table A6.1 Comparison of New Untreated Bleed Definitions via Hypothetical Examples

Bleeding Time (1)	Bleeding Site (2)	Last Infusion, if NEW Treated (3)	"New" Untreated Bleed (Y/N): Clinical Judgement (4)	>72 hr Infusion (Y/N) (5)	>48 hr Untreated (Y/N) (6)	>48 hr NEW Untreated (Y/N) (7)	New Untreated Bleed (Y/N) Look back ALL prior bleeds		New Untreated Bleed (Y/N) IMMEDIATE prior bleed only	
							Without NEW* (8)	With NEW# (9)	NEW Bleed (10)	Any Bleed (11)
Example 1										
2020-11-25T18:25	Shoulder_L				1st untreated		Y	Y		
2020-11-27T18:34	Shoulder_L	2020-11-27T18:34								
2020-11-28T18:28	Shoulder_L		**No**	N	Y	Y	**Y**	**N**	N	N
		No, while >48 hrs from the previous untreated bleed, it is NOT >72 hours from the infusion of the same site.								
Example 2										
2020-11-25T18:25	Shoulder_L	2020-11-25T18:25								
2020-11-26T16:30	Shoulder_L	NA		N	1st untreated		N	N		
2020-11-28T18:24	Shoulder_L	NA	**No**	N	Y	N	**Y**	**N**	N	Y
		No, while >48 hours from the previous untreated bleeds, it does NOT meet the >72 hour infusion criteria; a good amount of drug is still in the body.								
Example 3										
2020-11-25T18:25	Shoulder_L	2020-11-25T18:25								
2020-11-29T18:34	Shoulder_L	NA		Y	1st untreated		Y	Y		
2020-11-30T18:28	Shoulder_L	NA	**No**	Y	N	N	**Y**	**N**	N	N
		No, while >72 hours from the previous infusion, it is NOT >48 hours from (new) previous untreated bleeds.								
Example 4										
2020-11-25T18:25	Shoulder_L	2020-11-25T18:25								

2020-11-27T18:34	Shoulder_L	NA		N	1st untreated		N	N		
2020-11-28T18:28	Shoulder_L	NA	**Yes**	Y	N	Y	**Y**	**Y**	Y	N
		Yes, (1) >72 hours from the previous infusion (2) while NOT >48 hours from previous untreated bleed but that untreated bleed was NOT counted as NEW.								
Example 5										
2020-11-**25**T18:25	Shoulder_L	2020-11-**27**T18:24								
2020-11-27T18:28	Shoulder_L	NA		N	1st untreated					
2020-11-28T18:28	Shoulder_L	NA		N	N	Y				
2020-11-29T18:28	Shoulder_L	NA		N	N	Y				
2020-11-30T18:28	Shoulder_L	NA	**Yes**	Y	N	Y	**N**	**Y**	Y	N
		Yes, (1) >72 hours from the previous infusion (2) while NOT >48 hours from previous untreated bleed but that untreated bleed was NOT counted as NEW.								

*: Previous definition: An untreated bleed is considered a "new untreated bleed" if occurring >48 hours after the previous untreated bleed at the same site **or** occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated; an untreated bleed occurring at a different site is considered a "new untreated bleed" regardless of the time of the bleed.

#: Revised definition: An untreated bleed is considered a "new untreated bleed" if occurring >48 hours after the previous "**new**" untreated bleed at the same site **and** occurring >72 hours after stopping treatment from the original bleed for which treatment was initiated; an untreated bleed occurring at a different site is considered a "new untreated bleed" regardless of the time of the bleed.

Appendix 7. List of Abbreviations

Abbreviation	Term
ABR	annualized bleeding rate
ADA	antidrug antibody
AE	adverse event
ATP	active treatment phase
aPCC	activated Prothrombin Complex Concentrates
CI	confidence interval
COVID-19	coronavirus disease 2019
CTCAE	Common Terminology Criteria for Adverse Events
cTnI	cardiac troponin I
dPT	dilute prothrombin time
ECG	electrocardiogram
eDiary	electronic Diary
EQ-5D	EuroQol 5 Dimensions
EQ-5D-5L	EuroQol 5 Dimensions 5 Level
EQ-VAS	EQ-5D general health state as measured by a visual analog scale
EU	European Union
FCS	fully conditional specification
FDA	Food and Drug Administration (United States)
FEIBA	Factor Eight Inhibitor Bypassing Activity
FIX	Factor IX
FVIII	Factor VIII
GEE	generalized estimating equation
Haem-A-QoL	Hemophilia Quality of Life Questionnaire for Adults
Haemo-QoL	Hemophilia Quality of Life Questionnaire for Children
HAL	Hemophilia Activities List
HAL2	Hemophilia Activities List, version 2
HJHS	Hemophilia Joint Health Score
HLIQ	Hemophilia Life Impacts Questionnaire
HRQoL	health-related quality of life
IU	International units
kg	kilogram
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent-to-treat
N/A	not applicable
NAb	neutralizing antibodies
NE	non-evaluable
OP	observational phase
PCC	prothrombin complex concentrates
PD	pharmacodynamic(s)
pedHAL	Pediatric Hemophilia Activities List
PF1+2	prothrombin fragment 1+2

Abbreviation	Term
PGIC-H	Patient Global Impression of Change–- Hemophilia
PK	pharmacokinetic(s)
PKT	peak thrombin
PMAP	Population Modeling and Analysis Plan
PRO	patient reported outcome
rFVIIa	recombinant FVIIa
Q1	first quartile
Q3	third quartile
QW	once weekly
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
SD	standard deviation
SoA	schedule of activities
TFPI	tissue factor pathway inhibitor
TSS	transformed scaled score
WHO	World Health Organization