STATISTICAL ANALYSIS PLAN FOR

STUDY CSL222_3001

Protocol Number:	CSL222_3001 (formerly CT-AMT-061-02)
Investigational Drug and Drug Number:	CSL222 (formerly AMT-061)
Indication:	Hemophilia B
Dosage Form/Dose:	2×10^{13} gc/kg CSL222 (formerly AMT-061)
Sponsor:	CSL Behring

Protocol Title: Phase III, open-label, single-dose, multi-center multinational trial investigating a serotype 5 adeno-associated viral vector containing the Padua variant of a codon-optimized human factor IX gene (AAV5-hFIXco-Padua, AMT-061) administered to adult subjects with severe or moderately severe hemophilia B

Date of Issue: 04 Oct 2024

Version: Version 6.0

APPROVAL SIGNATURES:

Biostatistics TA Lead		
Name/Title	**Signature**	**Date (dd/mmm/yyyy)**
PPD PPD	PPD	10-Oct-2024

Clinical Lead		
Name/Title	**Signature**	**Date (dd/mmm/yyyy)**
PPD	PPD	09-Oct-2024

Change Log			
Version No.	**Effective Date*** **(dd-mmm-yyyy)**	**Reason for the Change / Revision**	**Supersedes**
6.0	29 Nov 2023	Including the "early discontinuation visit" in the visit mapping based on visit window.	5.0
6.0	15 Jan 2024	Including the analyses from the durability recommended by pharmacokinetic group in Section 7.6.7	5.0
6.0	12 Jan 2024	Generalizing the analysis time points to Month 60 (5 Year)	5.0
6.0	01 Mar 2024	Adding "premature discontinuation from PROBE sub-study to the disposition table in Section 7.1.	5.0
6.0	04 Mar 2024	Adding Glossary of Terms table for common working definitions	5.0
6.0	04 Mar 2024	Adding extra column in the table of demographic and baseline characteristics for PROBE to Section 7.2.	5.0
6.0	05 Aug 2024	Adding Intercurrent Events to Section 7.5.1	5.0
6.0	15 Aug 2024	Updating Section 7.2.2.4 title by replacing Antibody with Immunologic since AAV5 capsid-specific T-cells is not an antibody	5.0
6.0	15 Aug 2024	Removing 10-day contamination rule and analysis	5.0

* This is the date when the given change was entered into the SAP text document.
Note: for Change Log from earlier versions, please refer to Statistical Analysis Plan V5.0, 15 Apr 2022.

Table of Contents

1. INTRODUCTION 14
1.1 GENERAL INTRODUCTION 14
1.2 SAFETY RESULTS FOR CSL222_2001 15
1.3 EFFICACY RESULTS FOR CSL222_2001 16
1.4 THIS STATISTICAL ANALYSIS PLAN (SAP) 17
2. STUDY OBJECTIVES AND ENDPOINTS 17
2.1 STUDY OBJECTIVES 17
2.1.1 Primary Objectives 17
2.1.2 Secondary Objectives 18
2.1.2.1 Secondary Efficacy Objectives 18
2.1.2.2 Exploratory Efficacy Objectives 19
2.1.3 Safety Objectives 19
2.2 STUDY ENDPOINTS 20
2.2.1 Efficacy Endpoints 20
2.2.1.1 Primary Efficacy Endpoint 20
2.2.1.2 Secondary Efficacy Endpoints 20
2.2.1.3 Exploratory Efficacy Endpoints 21
2.2.1.4 Patient Reported Outcomes, Burdens, and Experiences (PROBE) Sub-Study Endpoints 24
2.2.1.5 Musculoskeletal Ultrasound (MSKUS) Sub-study Endpoints 24
2.2.2 Safety Endpoints 24
3. STUDY DESIGN AND ANALYTICAL CONSIDERATIONS 26
3.1 STUDY DESIGN 26
3.1.1 Overall Study Design and Plan 26
3.2 INTERIM ANALYSIS 28
3.3 SAMPLE SIZE 28
3.3.1 Non-inferiority Margin for Rate Ratio Analysis 28
3.4 RANDOMIZATION AND BLINDING 30
4. DATA DEFINITIONS AND PRE-PROCESSING 31
4.1 BASELINE DEFINITIONS 31
4.2 DATA HANDLING RULES AND DEFINITIONS, INCLUDING HANDLING OF MISSING DATA 31
4.3 BLEED COUNTING RULES 35
5. DATA AND ANALYTICAL QUALITY ASSURANCE 36
6. ANALYSIS POPULATIONS 36
6.1 POPULATION DEFINITIONS 36
6.1.1 Screen Failures 36
6.1.2 Lead-in Discontinuers 37
6.1.3 Safety Population 37
6.1.4 Full Analysis Set (FAS) 37
6.1.5 Per-Protocol Population 37
7. STATISTICAL ANALYSIS 37
7.1 SUBJECT DISPOSITION 37
7.2 DEMOGRAPHIC AND BASELINE CHARACTERISTICS 38
7.2.1 Demography 38
7.2.2 Baseline Disease Characteristics 38
7.2.2.1 Hemophilia B History 38
7.2.2.2 Medical and Surgical History 38
7.2.2.3 Target Joints at Screening 38
7.2.2.4 Baseline Immunologic Parameters 39
7.2.2.5 FIX Gene Mutation 39
7.2.2.6 Prior and Concomitant Medications 39
7.2.2.7 Prior FIX Therapy Use 39
7.3 INVESTIGATIONAL PRODUCT EXPOSURE 39
7.4 BLOOD SAMPLE FOR FUTURE RESEARCH 39

7.5 EFFICACY ANALYSES....40
7.5.1 Intercurrent Events....40
7.5.2 Primary Endpoint....40
7.5.2.1 ABR Comparison Between CSL222 Post-treatment Period Following Stable FIX Expression (Beginning at Month 7) and Prophylaxis During the Lead-in Period for Non-Inferiority....41
7.5.2.2 Sensitivity Analysis 1: PP Population....41
7.5.2.3 Sensitivity Analysis 2: Including (Not Excluding) Periods Subsequent to Exogenous FIX Use....42
7.5.2.4 Sensitivity Analysis 3: Bleeds Treated with Exogenous FIX....42
7.5.2.5 Sensitivity Analysis 4: Cumulative Responder Analysis using Subject-Specific Bleeding Rates....42
7.5.2.6 Sensitivity Analysis 5: New and True Bleeds....42
7.5.2.7 Sensitivity Analysis 6: New and True Bleeds Treated with Exogenous FIX....42
7.5.2.8 Sensitivity Analysis 7: Excluding Periods Contaminated by Systemic Corticosteroid Exposure....42
7.5.3 Secondary Efficacy and Quality of Life Endpoints....43
7.5.3.1 Sensitivity Analysis: PP Population....44
7.5.4 Endogenous FIX Activity After CSL222 Dosing....45
7.5.4.1 Sensitivity Analysis 1: PP Population....46
7.5.4.2 Sensitivity Analysis 2: To Account for Missing Data....46
7.5.4.3 Sensitivity Analysis 3: Cumulative Responder Analysis....46
7.5.5 Annualized Consumption of FIX Replacement Therapy, Excluding FIX Replacement for Invasive Procedures, Following Stable FIX Expression (6 Months) during the Post-treatment Follow-up Compared to the Lead-in Period....46
7.5.5.1 Sensitivity Analysis 1: Including FIX Consumption for Invasive Procedures....47
7.5.5.2 Sensitivity Analysis 2: Excluding Subjects who Returned to Continuous Prophylaxis....47
7.5.6 Annualized Infusion Rate of FIX Replacement Therapy, Excluding FIX Replacement for Invasive Procedures, Following Stable FIX Expression (6 Months) during the Post-treatment Follow-up Compared to the Lead-in Period....47
7.5.7 Proportion of Subjects Remaining Free of Previous Continuous Routine Prophylaxis during the Post-treatment Follow-up Period....48
7.5.8 Comparison of the Percentage of Subjects with FIX Activity < 12% of Normal between the Lead-in Period and after Treatment with CSL222 Following Stable FIX Expression (6 Months)....48
7.5.8.1 Sensitivity Analysis 1: PP Population....49
7.5.8.2 Sensitivity Analysis 2: Cumulative Responder Analysis Comparing the Lead-in Period with the Post-treatment Period Following Stable FIX Expression (6 Months)....49
7.5.9 ABR Comparison between CSL222 and Prophylaxis for Superiority between the Lead-in Period and the Post-treatment Period Following Stable FIX Expression (6 Months) Post-treatment (CSL222) Follow-up....50
7.5.10 Rate of Spontaneous Bleeding Events during the Post-treatment Period Following Stable FIX Expression (6 Months) Compared to the Lead-in Period....50
7.5.11 Rate of Joint Bleeding Events Following Stable FIX Expression Post-treatment Follow-up Compared to the Lead-in Period....51
7.5.12 Estimated ABR – during the Post-treatment Period Following Stable FIX Expression (6 Months) as a Function of Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay (as a “Correlation” Analysis)....51
7.5.13 Correlation of FIX Activity Levels during the Post-treatment Period Following Stable FIX Expression (6 Months) with Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay....51
7.5.14 Occurrence of (and Resolution of) New Target Joints during the Post-treatment Follow-up Period Following Stable FIX Expression (6 Months)....52
7.5.15 Time to Resolution of Pre-existing Target Joints during the Post-treatment Follow-up Period....52
7.5.16 Proportion of Subjects with Zero Bleeds in the Post-treatment Period Following Stable FIX Expression (6 Month)....52
7.5.17 Patient Reported Outcome (PRO) Questionnaire Scores from the iPAQ (Total Physical

Activity Score) during the 12 Months Following CSL222 Dosing Compared with the Lead-in Period....53
7.5.17.1 Sensitivity Analysis 1: PP Population....54
7.5.18 PRO Questionnaire Scores from the EQ-5D-5L (VAS Score) during the 12 Months Following CSL222 Dosing Compared with the Lead-in Period....54
7.5.18.1 Sensitivity Analysis 1: PP Population....55
7.5.19 Additional Sensitivity Analysis....55
7.5.20 Type I Error Control and Simultaneous Confidence Intervals....56
7.6 EXPLORATORY EFFICACY ENDPOINTS....58
7.6.1 Analysis of Exploratory Endpoints....58
7.6.2 FIX Protein Levels during Following Stable FIX Expression (7-to 60 Months) Following CSL222 Dosing....58
7.6.3 Analysis of PRO Questionnaire Scores....58
7.6.3.1 iPAQ Subscale Scores....59
7.6.3.2 HJHS Total Score....59
7.6.3.3 EQ-5D-5L Index Score and Categorical Responses....59
7.6.3.4 Additional PRO Questionnaire Scores: WPAI....59
7.6.3.5 Additional PRO Questionnaire Scores: BPI....60
7.6.3.6 Additional PRO Questionnaire Scores: HAL....61
7.6.3.7 Additional PRO Questionnaire Scores: Hem-A-QoL....61
7.6.4 Impacted Responders Analysis: Correlation of FIX Activity Levels at Month 18 with Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay....61
7.6.5 Estimated ABR over the Post-treatment Follow-up as a Function of Mean FIX Activity ("Correlation" Analysis)....62
7.6.6 Rate of Traumatic Bleeding Events Following Stable FIX Expression (6 Months) Post-treatment Follow-up Compared to the Lead-in Period....63
7.6.7 Subgroup Analyses....63
7.6.8 Sub-Study Efficacy and Quality of Life Endpoint Analyses....65
7.6.8.1 PROBE Questionnaire Summary Scores....65
7.6.8.2 Musculoskeletal Ultrasound Results....65
7.7 SAFETY ANALYSES....65
7.7.1 Adverse Events....66
7.7.1.1 Adverse Events of Special Notification....68
7.7.1.2 Severity of Adverse Event....68
7.7.1.3 Relationship Between IMP and Adverse Event....68
7.7.2 Changes in Abdominal Ultrasound....69
7.7.3 Anti-AAV5 Antibodies, Anti-FIX Antibodies, and FIX Inhibitors....69
7.7.4 AAV5 Capsid-specific T Cells....69
7.7.5 Clinical Laboratory Measurements....69
7.7.6 ALT Levels, AST Levels and Corticosteroid Use for ALT and AST Increases....73
7.7.7 Vector Genome Detection....73
7.7.8 Inflammatory Markers....74
7.7.9 Alpha-fetoprotein....74
7.7.10 Physical Examination (Including Height and Weight)....74
7.7.11 Vital Signs....75

8. INTERIM 6 MONTH ANALYSIS....75
9. INTERIM 12 MONTH ANALYSIS....76
10. CSR....77
11. CSR ADDENDUM....77
12. CHANGES FROM METHODS PLANNED IN THE PROTOCOL....78
13. STATISTICAL SOFTWARE....78
14. REFERENCES....78
APPENDIX 1: DATA HANDLING RULES....81
APPENDIX 2: ANALYSIS DATASET SPECIFICATIONS....87

APPENDIX 3: CENTRAL LABORATORY REFERENCE RANGES FOR USE IN FLAGGING ABNORMAL VALUES 87
APPENDIX 4: EQ-5D-5L US VALUE SET 88
APPENDIX 5: SAS CODE FOR STATISTICAL ANALYSES 88

List of Tables

Table 1: Recent Results of Prophylaxis Compared to On-demand 29
Table 2: Non-inferiority Margin (for the ABR rate ratio) for Retention of p%.....30
Table 3: Relation of Pre-treatment ABR and NI Margin using a 1.8 Rate Ratio 30
Table 4: Protocol-Recommended Study Time Intervals for Collection of Efficacy and Safety Evaluation and Laboratory Parameters.....33
Table 5: Adverse Events of Special Notification.....68
Table 6: Safety Lab Parameters.....70
Table 7: Potentially Clinically Significant (PCS) Laboratory Parameter Criteria 72

List of Figures

Figure 1: Study Overview.....26
Figure 2: Study Design.....27

GLOSSARY OF ABBREVIATIONS

Abbreviation	Term
AAV	adeno-associated viral
AAV5	adeno-associated viral vector serotype 5
AAV5-hFIXco	recombinant adeno-associated viral vector serotype 5 containing the wild type human FIX gene, codon-optimized for optimal expression in humans, under control of a liver-specific promoter (CSL220 [formerly AMT-060])
AAV5-hFIXco-Padua	recombinant adeno-associated viral vector serotype 5 containing a codon-optimized Padua derivative of human coagulation FIX cDNA (CSL222 [formerly AMT-061])
ABR	annualized bleeding rate
AE	adverse event
AFP	alpha-fetoprotein
ALP	alkaline phosphatase
ALT	alanine aminotransferase
aPTT	activated partial thromboplastin time
AR	autoregressive
AST	aspartate aminotransferase
BPI	Brief Pain Inventory
cDNA	complementary deoxyribonucleic acid
CI	confidence interval
COVID-19	Coronavirus Disease (discovered in) 2019
CRP	c-reactive protein
CSR	Clinical Study Report
DNA	deoxyribonucleic acid

eCRF	electronic case report form
ELISA	enzyme-linked immunosorbent assay
EQ-5D-5L	EuroQol-5 dimensions-5 levels: Refers to both the EQ-5D-5L descriptive system and the EQ Visual Analogue Scale (VAS)
FAS	Full Analysis Set
FDA	Food and Drug Administration
FIX	coagulation factor IX
GAM	Generalized Additive Model
GEE	Generalized Estimating Equations
GTWP	Gene Therapy Working Party
HAL	Hemophilia Activities List
HBeAg	hepatitis B extracellular antigen
HBsAg	hepatitis B surface antigen
HBV DNA	hepatitis B virus deoxyribonucleic acid
HCV RNA	hepatitis C virus ribonucleic acid
Hem-A-QoL	hemophilia specific quality of life index questionnaire for adults
hFIX	human coagulation FIX
HJHS	Hemophilia Joint Health Score
IFNγ	interferon gamma
IgG	immunoglobulin G
IgM	immunoglobulin M
IL-1β	interleukin-1beta
IL-2	interleukin-2
IL-6	interleukin-6

IMP	investigational medicinal product
INR	international normalized ratio
iPAQ	international Physical Activity Questionnaire
IU	international unit
J.A.D.E.	Joint Tissue Activity and Damage Exam
LOD	limit of detection
MET	metabolic equivalent of task
MCP-1	monocyte chemotactic protein-1
MedDRA	Medical Dictionary for Regulatory Activities
MSKUS	Musculoskeletal Ultrasound
NAB	neutralizing antibody
PCS	potentially clinically significant
PP	per-protocol
PRO	patient reported outcome
PROBE	Patient Reported Outcomes, Burdens, and Experiences
Q1	first quartile
Q3	third quartile
QoL	quality of life
rAAV5	recombinant adeno-associated viral vector serotype 5
SAE	serious adverse event
SAP	Statistical Analysis Plan
SDTM	Study Data Tabulation Model
SOC	System Organ Class

TEAE	treatment-emergent adverse event
US	United States
VAS	visual analogue scale
WFH	World Federation of Haemophilia
WPAI	Work Productivity and Activity Impairment Questionnaire

GLOSSARY OF TERMS

Term	**Definition**
Contamination rule	Data collected Post-treatment that are within 5 half-lives of exogenous FIX use.
New bleed	Any bleed occurring more than 72 hours after stopping treatment for the original bleed for which treatment was initiated.
Person-time at risk	Person-time within 5 half-lives subsequent to exogenous FIX use is not considered to be time at risk, i.e., contaminated.
Responders	Subjects that have an endpoint (e.g., ABR or FIX level) equal to or greater than the level specified.
Return to continuous FIX prophylaxis	At least 80% of time being contaminated by exogenous FIX by the 5 half- lives rule during a continuous 3- month period during this study.
Subject-specific ABR	Number of bleedings divided by person-time at risk (in year) at subject level
Suffer from FIX inhibitors	A subject is said to suffer from FIX inhibitors if the subject tests positive (≥ 0.6 BU/mL) for FIX inhibitors at 2 consecutive tests from the central laboratory, performed preferably within 2 weeks.
Unique bleed	Any single or multiple occurrence(s) of bleeds of the same type that happen on the same study day.

Trademark Information

SAS

nQuery Advisor

1. INTRODUCTION

1.1 General Introduction

Congenital hemophilia B is an inherited bleeding disorder characterized by an increased bleeding tendency due to either a partial or complete deficiency of the essential blood coagulation factor IX (FIX). The deficiency is the result of mutations in the FIX gene. Hemophilia B is an X-linked, recessive condition since it occurs almost exclusively in males. Females typically are asymptomatic carriers. The number of people with hemophilia B worldwide is approximately 30,000 and in the United States (US) alone is approximately 4,000 (World Federation of Haemophilia [WFH], 2017). Approximately 1 in 20,000 – 50,000 live male newborns has hemophilia B.

The severity of symptoms can vary and the severe forms become apparent early in life. About one-third of individuals with hemophilia B have a severe disorder characterized by FIX levels that are less than 1% of normal (Kessler & Mariani, 2006). Moderate and mild hemophilia B, with 1 – 5% or 5 – < 40% of normal FIX level, respectively, are each observed in about one-third of subjects (Kessler & Mariani, 2006).

There is no cure for hemophilia B. The primary goals of hemophilia B therapy are the prevention of bleeding episodes, rapid and definitive treatment of bleeding episodes (breakthrough bleeds) that occur even while on a regular prophylactic regimen and provision of adequate hemostasis during surgery and emergencies. Currently, these goals are essentially met for hemophilia B subjects by intravenous (IV) injections of commercially available recombinant- or plasma-derived FIX products, either at the time of a bleed (on-demand) or by regular infusions up to several times a week (prophylactically).

Somatic gene therapy offers the potential for a shift of the disease severity phenotype from severe to a moderate or mild hemophilia phenotype or complete amelioration through continuous production of stable FIX levels after a single administration of vector, especially since a small rise in circulating FIX to at least 1% of normal levels can substantially ameliorate the bleeding phenotype and potentially improve the quality of life (QoL) for subjects.

CSL222 (formerly AMT-061) has been developed for the treatment of hemophilia B. CSL222 is a recombinant adeno-associated viral vector serotype 5 (rAAV5) containing the coding sequence for Padua variant of the human coagulation FIX (hFIX Padua), codon-optimized, under control of a liver-specific promoter (also known as AAV5-hFIXco Padua).

CSL222 is a derivative of CSL220 (formerly AMT-060), which has been studied in a Phase I/II clinical trial in humans with severe and moderately severe hemophilia B. Both CSL220 and CSL222 have the same rAAV5 containing the codon-optimized wild-type human FIX gene, but the latter incorporates two-nucleotide change in order to encode the naturally occurring Padua variant of human coagulation FIX. The FIX-Padua protein differs from the ‘wild-type’ human FIX protein by a single amino acid and it is responsible for the observed increased FIX activity per unit of dose achieved with CSL222 as compared to its predecessor CSL220 (Simioni et al, 2009).

This Phase III trial is to demonstrate the efficacy of CSL222 in terms of endogenous FIX activity and annualized bleed rate (ABR), and to further describe its safety profile. The strong efficacy (protein expression) and safety results obtained during the Phase I/II trial with AMT-060 demonstrate 2×10^{13} gc/kg to be the selected dose for use in future trials. In addition, non-clinical data support the implementation of the prospectively defined product enhancement of CSL220 (i.e., incorporation of the Padua mutation to form the new construct CSL222) at a dose of 2×10^{13} gc/kg for this pivotal Phase III trial. At the time of preparing the original SAP, interim data analysis from the dose confirmation trial (CSL222_2001 [formerly CT-AMT-061-01]) were available and the dose for CSL222 of 2×10^{13} gc/kg has been confirmed for the treatment period of this Phase III trial.

1.2 Safety Results for CSL222_2001

Three subjects dosed with 2×10^{13} gc/kg of CSL222 completed dose confirmation Phase IIb trial CSL222_2001. Final Clinical Study Report (CSR) is available now.

There were no deaths or treatment-emergent adverse events (TEAEs) resulting in early discontinuation from CSL222_2001. One subject with a history of avascular necrosis (bilateral hip) had 2 serious adverse events (SAEs) of osteonecrosis (verbatim term: worsening avascular necrosis – left hip and worsening of avascular necrosis of right hip) resulting in 3 major hip surgeries; the SAEs were assessed by the Investigator as not treatment-related. All subjects had at least 1 TEAE in the 5 years (60 months) post-administration of CSL222, with a total of 84 TEAEs reported. One subject had mild AEs of headache and paresthesia on the day of CSL222administration; both events resolved the same day without treatment. Most TEAEs were mild (55 events)or moderate (26 events) in severity; 3 TEAEs were severe and all 3 were reported between the 3-year(36-month) and 5-year (60-month) analyses. A total of 2 TEAEs were assessed as treatment-related. One subject experienced 2 mild treatment-related TEAEs of C-reactive protein increased and headache, which resolved without intervention. The most common TEAE was arthralgia, which was reported in all3 subjects. All subjects experienced at least 1 TEAE between the 3-year (36-month) and 5-year (60-month)analyses. There were 28 TEAEs reported between the 3- and 5-year analyses; 2 of these events were SAEs. The most common TEAEs were joint swelling and dental caries (2 subjects each). Transient mild increases in ALT levels post-CSL222 administration were reported in 2 subjects who did not receive treatment with steroids; the increases were not associated with loss of FIX activity. No subjects experienced clinically significant increases in ALT levels post-CSL222 administration. All 3 subjects were positive for AAV5 NAbs at screening with a rapid increase at Week 2 post-CSL222administration to titers > upper limit of quantification and titers remained at that level for the duration of the study period. No subjects developed FIX inhibitors. One subject had a single, transient AAV5capsid-specific T-cell response at Week 48. There were no clinically relevant findings for inflammatory markers (IFNγ, IL-1β, IL-2, IL-6, andMCP-1). Clearance of vector DNA was achieved from semen and blood in 2 and 3 subjects, respectively,post-CSL222 administration. The earliest clearance of vector DNA was achieved at 26.1 weeks from semen and 31.1 weeks from blood. Minor findings were observed in the abdominal ultrasound performed at Month 36 through Month 60visits but were assessed by the Investigator as not clinically significant; there was no pattern or common root cause noted for these findings.

1.3 Efficacy Results for CSL222_2001

All 3 subjects had been diagnosed with severe hemophilia B with corresponding circulating FIX activity levels <1% of normal. All subjects experienced clinically relevant increases in FIX activity after administration of CSL222 that were maintained through to Month 60. At 6 weeks, 52 weeks, and 60 months post-CSL222 administration, the 3 subjects expressed a mean ± SD uncontaminated endogenous FIX activity level of 30.6 ± 6.97%, 40.8 ± 9.45%, and 45.7 ± 6.18% of normal, respectively, as measured by the one-stage (aPTT-based) assay. FIX activity levels at these and other time points are shown below.

Visit	n	Mean (SD)
Week 6	3	30.6 (6.97)
Week 52	3	40.8 (9.45)
Month 24	3	44.2 (7.66)
Month 30	3	50.0 (11.40)
Month 36	2	36.9 (6.51)
Month 48	3	45.0 (2.76)
Month 60	3	45.7 (6.18)

Two of 3 subjects did not experience bleeding episodes post-CSL222 administration. The third subject had 2 lower leg muscle bleeding episodes that were treated with FIX; one was spontaneous, and the other was traumatic. The ABR over 5 years (60 months) of follow-up was 0.14 and the ABRs for spontaneous and traumatic bleeding episodes over 5 years (60 months) of follow-up were both 0.07. Both bleeding episodes occurred in the first 18 months post-CSL222 administration and there were no other bleeding episodes for the duration of the study. All 3 subjects discontinued use of routine continuous FIX prophylaxis during the study, with the end of contamination from routine continuous prophylaxis on Days 10; 10; and 22 post-CSL222 administration. The annualized mean FIX use (excluding use for invasive procedures) was 342.1 IU/year over 5 years(60 months) of follow-up for the post-continuous prophylaxis period, with all of this use by a single subject. This subject required FIX replacement therapy post-CSL222 administration per protocol for the following: due to elective surgeries (3 major hip surgeries associated with ongoing SAEs of worsening avascular necrosis – left hip [preferred term: osteonecrosis] and worsening of avascular necrosis of right hip [preferred term: osteonecrosis]); for 2 reported bleeding episodes; and for 6 separate times as prophylaxis prior to dental surgery or acupuncture or due to unreported reasons. Factor IX activity measured using the chromogenic assay was lower than using the one-stage(aPTT-based) assay. Using the chromogenic assay, mean ± SD uncontaminated FIX activity levels were17.5 ± 3.64%, 22.2 ± 5.98%, and 24.5 ± 4.17% at Week 6, Week 52, and Month 60, respectively. FIX activity levels at these and other timepoints are shown below.

Visit	n	Mean (SD)
Week 6	3	17.5 (3.64)
Week 52	3	22.2 (5.98)
Month 24	3	19.8 (3.23)
Month 30	3	22.3 (5.90)

Month 36	2	19.9 (4.10)
Month 48	3	20.7 (6.58)
Month 60	3	24.5 (4.17)

HJHS scores decreased for 2 subjects from 35 and 36 at Baseline to 27 and 34 at Month 48; scores were41 and 28 at Month 60, respectively. The third subject had scores of 1, 1, and 9 at Baseline, Month 48,and Month 60, respectively. Generally, QoL improved for 2 subjects based on responses to the patient-reported outcome questionnaires. The third subject showed a decrease in QoL at Week 26 and Week 52, including worsening pain and problems with pain and mobility as indicated via the BPI and EQ 5D 5L, which was likely due to pre-existing avascular necrosis of the bilateral hip (preferred term: osteonecrosis). The subject had2 associated elective major hip surgeries on Day 197 and Day 720; improvement of BPI and EQ 5D 5Loverall scores was noted at Month 24 (Day 764). This subject also showed a decrease in QoL at Month 36, including worsening pain and problems with pain/discomfort as indicated via the BPI andEQ-5D-5L; he received a cortisone injection to treat an ongoing TEAE of sciatica the day prior to the Month 36 visit. The QoL for this subject decreased after Month 48, which was likely due to the worsening avascular necrosis of both hips; the subject had a third major hip surgery on Day 1672. Pain was reported to have interfered with all aspects of QoL at Month 60 via the BPI; however, no problems were noted via the EQ-5D-5L.

1.4 This Statistical Analysis Plan (SAP)

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary, and analysis of data to be performed for the interim analysis at 26 weeks Post-treatment, the primary analysis at 52 weeks Post-treatment following the last subject reaching stable FIX, and the CSR addendum covering the long-term follow-up period from Month 18 to Month 60. The SAP should be read in conjunction with the study protocol. This version of the SAP has been developed using the CSL222_3001 Protocol (Version 9.0, 12 Feb 2024), Ultrasonography and Hemophilia Joint Tissue Activity and Damage Exam (J.A.D.E.) Protocol (Edition 1, 2017), and the CSL222_3001 CRF (Version 12.1, 19 Aug 2022).

2. STUDY OBJECTIVES AND ENDPOINTS

2.1 Study Objectives

2.1.1 Primary Objectives

The primary objective is:

- to demonstrate the non-inferiority of CSL222 (2×10^{13} gc/kg) during the 52 weeks following establishment of stable FIX expression (months 7 to 18) Post-treatment (CSL222) follow-up compared to standard of care continuous routine FIX prophylaxis during the Lead-in period, as measured by the ABR.

2.1.2 Secondary Objectives

The secondary objective is to demonstrate additional efficacy and safety aspects of systemic administration of CSL222.

2.1.2.1 Secondary Efficacy Objectives

To investigate the effect of 2×10^{13} gc/kg CSL222 on the following:

- Endogenous FIX activity 6 months after a single CSL222 treatment
- Endogenous FIX activity 12 months after a single CSL222 treatment
- Endogenous FIX activity 18 months after a single CSL222 treatment
- Annualized consumption of FIX replacement therapy
- Annualized infusion rate of FIX replacement therapy
- Discontinuation of previous continuous routine prophylaxis
- FIX activity
- Prevention of bleedings (comparison for superiority)
- Prevention of spontaneous bleeding
- Prevention of joint bleeding
- Estimated ABR – during the 52 weeks following stable FIX expression (7-18 months) – as a function of pre-investigational-medical-product (IMP) anti-AAV5 antibody titers using the luciferase based NAB assay (as a “correlation” analysis)
- Correlation of pre-investigational medicinal product (IMP) anti-AAV5 antibody titers using the luciferase based neutralizing antibody (NAB) assay on FIX activity levels after CSL222 dosing
- Occurrence and resolution of target joints
- Proportion of subjects with zero bleeds during the 52 weeks following stable FIX expression (7-18 months) after CSL222 dosing
- International Physical Activity Questionnaire (iPAQ)
- EuroQol-5 dimensions-5 levels (EQ-5D-5L) Visual Analog Scale (VAS)

2.1.2.2 Exploratory Efficacy Objectives

To investigate the effect of CSL222 on the following:

- FIX protein levels during the 18 months following CSL222 dosing
- Hemophilia Joint Health Score (HJHS) scores
- Other Patient Reported Outcome (PRO) questionnaires: Work Productivity and Activity Impairment Questionnaire (WPAI), Brief Pain Inventory (BPI), hemophilia activities list (HAL), and Hemophilia Quality of Life Questionnaire for Adults (Hem-A-QoL) during the Lead-in period (prophylaxis) and during the 12 months following CSL222 dosing
- Estimated ABR over time as a function of mean FIX activity (as a "correlation" analysis) over the 18 month post-treatment follow-up
- Rate of traumatic bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period
- Subgroup analyses will be carried out for the following endpoints (the subgroups will be mentioned farther below in this SAP):
    - Endogenous FIX activity at 18 months
    - Annualized consumption of FIX replacement therapy, excluding replacement for invasive procedures
    - Annualized infusion rate of FIX replacement therapy
    - ABR comparison between CSL222 and FIX prophylaxis
    - Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 over the 52 weeks following stable FIX expression (7-18 months)
    - Proportion of subjects remaining free of previous prescribed continuous routine prophylaxis.
- All efficacy endpoints (as exploratory endpoints) at 2, 3, 4, and 5 years after CSL222 dosing, as part of the CSR addendum (see Section 11 for additional details)

2.1.3 Safety Objectives

The safety objectives include evaluating the following:

- Adverse events (AE)

- Changes in abdominal ultrasound
- Formation of anti-AAV5 antibodies (total immunoglobulin M and immunoglobulin G [IgM and IgG], neutralizing antibodies)
- AAV5 capsid-specific T cell response
- Formation of anti-FIX antibodies
- Formation of FIX inhibitors and recovery
- Hematology and serum chemistry parameters
- AST and alanine aminotransferase (ALT) level increases and use of corticosteroids
- Shedding of vector deoxyribonucleic acid (DNA) in blood and semen
- Inflammatory markers: interleukin-1beta (IL-1β), interleukin-2 (IL-2), interleukin-6 (IL-6), interferon gamma (IFNγ), and monocyte chemotactic protein-1 (MCP-1)
- Alpha-fetoprotein (AFP)

2.2 Study Endpoints

2.2.1 Efficacy Endpoints

2.2.1.1 Primary Efficacy Endpoint

- ABR comparison between CSL222 and prophylaxis for non-inferiority between the Lead-in period and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up

2.2.1.2 Secondary Efficacy Endpoints

- Endogenous FIX activity at Month 6 after CSL222 dosing
- Endogenous FIX activity at Month 12 after CSL222 dosing
- Endogenous FIX activity at Month 18 after CSL222 dosing
- Annualized consumption of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period
- Annualized infusion rate of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period

- Proportion of subjects remaining free of previous continuous routine prophylaxis during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up
- Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 over the 52 weeks following stable FIX expression (7-18 months)
- ABR comparison between CSL222 and prophylaxis for superiority between the Lead-in and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up
- Rate of spontaneous bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to Lead-in period
- Rate of joint bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period
- Estimated ABR – during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up – as a function of pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (as a "correlation" analysis)
- Correlation of FIX activity levels during the 7-18 months Post-treatment follow-up with pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay
- Occurrence of (and resolution of) new target joints during the 52 weeks following stable FIX expression (7-18 months) following CSL222 dosing and resolution of pre-existing target joints following CSL222 dosing
- Proportion of subjects with zero bleeds during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up
- Patient reported outcome (PRO) questionnaire scores from the iPAQ (total physical activity score) during the 12 months following CSL222 dosing compared with the Lead-in period
- Patient reported outcome (PRO) questionnaire scores from EQ-5D-5L VAS score during the 12 months following CSL222 dosing compared with the Lead-in period

2.2.1.3 Exploratory Efficacy Endpoints

- FIX protein levels during the 18 months following CSL222 dosing
- HJHS scores during the Lead-in period (prophylaxis) and during the 12 months following CSL222 dosing

- Other PRO questionnaires: WPAI, BPI, HAL, and Hem-A-QoL questionnaire scores during the Lead-in period (prophylaxis) and during the 12 months following CSL222 dosing
- EQ-5D-5L Index scores during the Lead-in period (prophylaxis) and during the 12 months following CSL222 dosing
- Estimated ABR as a function of mean FIX activity (as a "correlation" analysis) over the 18 months post-treatment follow-up
- Rate of traumatic bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period
- Subgroup analyses will be carried out for the following endpoints (the subgroups will be mentioned farther below in this SAP):
    - Endogenous FIX activity at Month 18
    - Annualized consumption of FIX replacement therapy during the 52-weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding replacement for invasive procedures, compared to the Lead-in period
    - Annualized infusion rate of FIX replacement therapy during the 52-weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding replacement for invasive procedures, compared to the Lead-in period
    - ABR comparison between CSL222 during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up and FIX prophylaxis (during the Lead-in period)
    - Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 during the 52 weeks following stable FIX expression (7-18 months)
    - Proportion of subjects remaining free of previous prescribed continuous routine prophylaxis during the 52 weeks following stable FIX expression (7-18 months) post-treatment follow-up

All efficacy endpoints will be analyzed as exploratory endpoints as part of the CSR addendum (see Section 11 for details):

- Endogenous FIX activity at 2, 3, 4, and 5 years after CSL222 dosing
- ABR comparison between CSL222 and prophylaxis for non-inferiority between the 2-year, 3-year, 4-year, and 5-year post treatment (CSL222) follow-up and the lead-in period

- Annualized consumption of FIX replacement therapy during the 2-year, 3-year, 4-year, and 5-year Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period
- Annualized infusion rate of FIX replacement therapy during the 2-year, 3-year, 4-year, and 5-year Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period
- Proportion of subjects remaining free of previous continuous routine prophylaxis during the 2-year, 3-year, 4-year, and 5-year Post-treatment follow-up
- Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 at 2, 3, 4, and 5 years after CSL222 dosing
- ABR comparison between CSL222 and prophylaxis for superiority between the Lead-in and the 2-year, 3-year, 4-year, and 5-year Post-treatment (CSL222) follow-up
- Rate of spontaneous bleeding events during the 2-year, 3-year, 4-year, and 5-year post-treatment (CSL222) follow-up compared to Lead-in period
- Rate of joint bleeding events during the 2-year, 3-year, 4-year, and 5-year Post-treatment (CSL222) Post-treatment follow-up compared to the Lead-in period
- Estimated ABR – at 2 years, 3 years, 4 years, and 5 years post-treatment – as a function of pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (as a "correlation" analysis)
- Correlation of FIX activity levels at 2 years, 3 years, 4 years, and 5 years post-CSL222 treatment with pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay
- Occurrence and resolution of target joints during the 2 years, 3 years, 4 years, and 5 years following CSL222 dosing
- Proportion of subjects with zero bleeds in the 2 years, 3 years, 4 years, and 5 years post-treatment (CSL222) follow-up
- Patient reported outcome (PRO) questionnaire scores from the iPAQ at 2 years, 3 years, 4 years, and 5 years post-treatment compared to the Lead-in period
- PRO questionnaire scores from the EQ-5D-5L at 2 years, 3 years, 4 years, and 5 years post-treatment compared to the Lead-in period
- FIX protein levels during the 2 years, 3 years, 4 years, and 5 years following CSL222 dosing

- HJHS scores during the Lead-in period (prophylaxis) and during the 2 years, 3 years, 4 years, and 5 years following CSL222 dosing

- Other PRO questionnaires: WPAI, BPI, HAL, and Hem-A-QoL questionnaires during the Lead-in period (prophylaxis) and during the 2 years, 3 years, 4 years, and 5 years

- Estimated ABR over time as a function of mean FIX activity (as a "correlation" analysis) over the 2 years, 3 years, 4 years, and 5 years post-treatment follow-up

- Rate of traumatic bleeding events during the 2 years, 3 years, 4 years, and 5 years post-treatment follow-up compared to the Lead-in period

2.2.1.4 Patient Reported Outcomes, Burdens, and Experiences (PROBE) Sub-Study Endpoints

PROBE is an optional sub-study of Study CSL222_3001. Subjects who consent to this sub-study will complete the PROBE questionnaires at the same visits as the other PRO questionnaires. The endpoints of this sub-study are PROBE summary scores and individual item responses.

2.2.1.5 Musculoskeletal Ultrasound (MSKUS) Sub-study Endpoints

MSKUS sub-study endpoints will be addressed in a separate SAP.

2.2.2 Safety Endpoints

All adverse event (AE) data will be collected from signing of the informed consent form until the end of the five-year follow-up. An AE, adverse drug reaction (ADR), and SAE are defined according to the ICH Guidelines E2A.

Safety analyses will be based on the safety population and described as descriptive analyses. Safety set is defined as any study subjects receiving at least some amount of study treatment even when the full dose was not administered (including partial dose).

Safety data will be analysed per study period i.e., lead-in period, treatment-emergent adverse events (TEAEs) during 26 and 52 weeks after study treatment start and during the follow up period. The overall safety profile of CSL222 will be assessed given the below safety and tolerability criteria.

Secondary safety endpoints include the following:

- AEs

- Changes in abdominal ultrasound

- Anti-AAV5 antibodies (total [IgM and IgG], neutralizing antibodies)

- AAV5 capsid-specific T cells
- Anti-FIX antibodies
- FIX inhibitors and recovery
- Hematology and serum chemistry parameters
- ALT and AST levels, and corticosteroid use for ALT and AST increases
- Vector DNA in blood and semen
- Inflammatory markers:
  - IL-1β
  - IL-2
  - IL-6
  - IFNγ
  - MCP-1
- AFP

Other laboratory evaluations include coagulation and serology parameters. In addition, the following (S)AEs qualify for special notification as they are seen as safety issues of particular concern for Advanced Therapy (ENTR/F/2/SF/dn D(2009) 35810. Brussels, 03/12/2009) and gene therapy medicinal products (EMA / CHMP / Gene Therapy Working Party (GTWP) / 60436/2007):

- AEs related to the IMP administration procedure
- Suspected or confirmed cases of opportunistic or serious infections that in the investigator's opinion might be related to the IMP
- Unexpected reactions (e.g., hypersensitivity, immunological, toxic or other as consequence of a change in the construction or function of the viral vector [e.g., generation of replication competent virus])
- AEs related to product failure (including lack of efficacy), mandatory concomitant medication (e.g., immunosuppression), and medical devices which form part of the product or are used for application of the product
- Development of any new/recurrent cancer

All TEAEs are tabulated displaying the number of subjects (and percentage) experiencing an event and the number of events by SOC and preferred terms within each SOC according to the Medical Dictionary for Regulatory Activities (MedDRA) terminology. TEAEs will also be tabulated by severity and by relationship to trial medication, using frequency counts (number of subjects with event and number of events) and percentages. Similar tables will be created for TEAEs leading to premature discontinuation or interruption, AESIs, deaths, seriousness, infusion related and hypersensitivity reactions if applicable. These summary tables will be presented by decreasing frequency of occurrence based on SOC and preferred term.

The summary tables will be accompanied by individual subject listings of all AEs, including information on AE number, actual AE description, date/time of start and end of AE, preferred term (MedDRA), SOC (MedDRA), severity, relationship/causality, type of AE, seriousness, and outcome. Pre-existing AEs will be flagged. Pre-existing AEs are not considered to be treatment emergent, except in case of worsening during/after trial treatment (to be collected as separate AEs). Separate listings will be created for AEs for special notification, deaths, and SAEs, if applicable. Other safety data will be presented using graphical displays, as applicable, descriptive statistics (including change from baseline, if applicable), and/or individual data listing. The number of days until vector DNA can no longer be detected in semen and blood will be tabulated. The number of days is calculated using the date of collection of the first of three consecutive negative samples for each matrix.

3. STUDY DESIGN AND ANALYTICAL CONSIDERATIONS

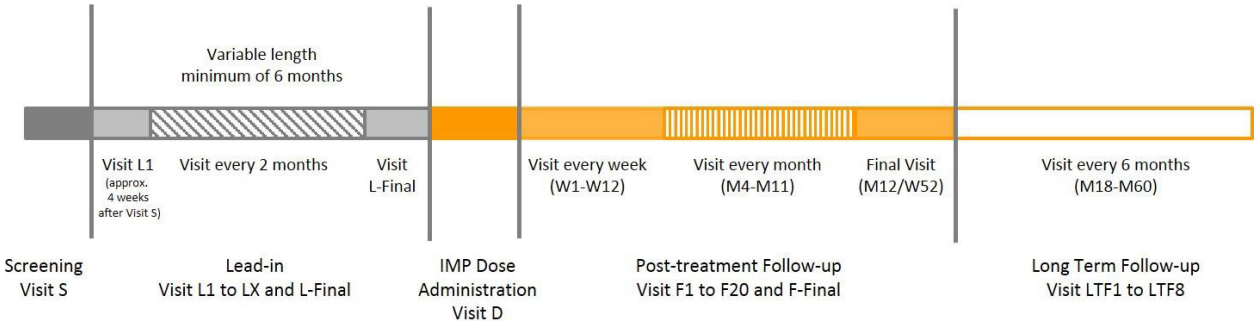
3.1 Study Design

3.1.1 Overall Study Design and Plan

CSL222_3001 is an open-label, single-dose, multi-center, multi-national trial, with a Screening period, a Lead-in period, a treatment + Post-treatment follow-up period, and a long-term follow-up period. Overviews of the trial and its design are presented in Figure 1 and Figure 2.

Figure 1: Study Overview

Figure 2: Study Design



Refer to the protocol for the schedule of events for efficacy and safety evaluations and laboratory parameters during Screening, Lead-in, treatment, Post-treatment, and long-term follow-up visits.

After Screening, eligible subjects will enter a Lead-in period prior to the start of CSL222 treatment. Visits will occur every 2 months during the Lead-in period with the final visit occurring a month prior to dosing. During the Post-treatment follow-up, visits will occur weekly up to Week 12 and then monthly up to Month 12, after which subjects will enter a long-term follow-up with visits every 6 months.

Six months after IMP administration, the first secondary endpoint, endogenous FIX activity at 6 months after IMP administration, will be analyzed and reported via an interim analysis once the last subject has achieved 6 months after CSL222 treatment. This assessment will be based on clean data and a partially locked database.

Twelve months after IMP administration, the second secondary endpoint, endogenous FIX activity at 12 months after IMP administration, will be analyzed and reported via another interim analysis once the last subject has achieved 12 months after CSL222 treatment. This assessment will be based on clean data and a partially locked database.

After 52 weeks following stable FIX expression (Month 18 post-dose), all available efficacy and safety data collected between Screening and Month 18 Post-treatment will be analyzed and reported in a full CSR, including (but not limited to) the primary ABR endpoint and the third secondary endpoint. The first and second secondary efficacy endpoints (which will have been analyzed in their respective data cuts) will also be added to the full (18-month) CSR. Data up to each analysis time point will be considered locked and will not be changed (with the exception of ending dates and outcomes for continuing events and treatments) without explicit authorization. The subjects will be followed for an additional 3.5 years for evaluation of efficacy parameters and safety. At the end of that 3.5-year period, all safety and efficacy data will be reported in a CSR addendum covering the entire study duration, including the later 3.5-year period.

3.2 Interim Analysis

There was an interim efficacy analysis of the first secondary endpoint, 6-month endogenous FIX activity levels, after all subjects completed the 6-month assessment and the database was (partially) locked.

There was an interim efficacy analysis of the second secondary endpoint, 12-month endogenous FIX activity levels, after all subjects completed the 12-month assessment and the database was (partially) locked.

3.3 Sample Size

The study sample size is constrained by the non-inferiority analysis of the primary endpoint, ABR.

Based on a literature search of trials in a similar clinical setting and the same underlying disease, as well as the previous CSL220 Phase I/II trial, a non-inferiority margin of 1.8 is assessed for the rate ratio of ABR between CSL222 (Post-treatment) and FIX prophylaxis (Lead-in). For establishing the non-inferiority margin, an ABR of 2.4 between FIX prophylaxis and placebo treatment has been assumed. Via simulation of ABR under a negative binomial distribution with a yearly rate of 2.4 events for Lead-in and 1.9 for Post-treatment, with a Pearson correlation of 0.05 for the number of events between the two periods, and with a common negative binomial dispersion parameter of 1.5, a sample size of N=50 will demonstrate non-inferiority with a non-inferiority margin of 1.8 and a power of 82.0%. Therefore, the study should consist of at least 50 analyzable subjects.

Given the sample size needed for ABR, this will produce a power >95% for the secondary statistical analysis of endogenous FIX activity. For the secondary statistical analyses of FIX activity at Months 6, 12, and 18, assuming a mean of 30.6 percent of normal (as observed at 6 weeks in Study CSL222_2001) and assuming a standard deviation of 6.97 (as observed at 6 weeks in Study CSL222_2001), assuming conservatively that the baseline FIX activity is 2%, and assuming that the sample size is 50 subjects, for a one-sample t-test at the 0.025 one-sided level of significance to test whether the change from baseline is > 0, the statistical power is > 99%. Alternatively, assuming that the standard deviation is 6.95, which is half of the range of FIX activity values (23.9 to 37.8) observed at 6 weeks in Study CSL222_2001, the statistical power is still > 99%. The nQuery Advisor software was employed for this power calculation.

3.3.1 Non-inferiority Margin for Rate Ratio Analysis

Recent studies in similar Hemophilia B populations have shown the results – presented in the table below – relative to on-demand treatment. In the Idelvion publication, Group 2 (n=19) was followed for 6 months using on-demand treatment and then for 6 months using 7-day prophylaxis treatment. The ABR rates were estimated as 365.25 x (number of bleeding episodes)/(number of days in the observed treatment period of interest). The rate reduction, 0.11, was estimated from a Poisson distribution, with a corresponding two-sided, 95% confidence interval (CI) for the rate reduction of (0.051, 0.238).

Table 1: Recent Results of Prophylaxis Compared to On-demand

Publication	ABR On-demand	ABR Prophylaxis	Rate Reduction
Alprolix (Powell, 2013)	18.67 (N=27)	3.12 (N=61)	0.17
Idelvion (Santagostino, 2016)	20.09 (N=19)	2.22 (N=19)	0.11
Nonacog (Collins, 2014)	15.58 (N=15)	40 IU: 2.51 (N=29)	0.16

Using a rate ratio analysis, as opposed to a difference in rates, results in an evaluation that is relatively independent of the magnitude of the baseline ABR. Thus, a rate reduction of 0.50 for a subject with 20 events during Lead-in has the same meaning as for a subject with 4 events during Lead-in. However, the difference in rates is quite different between such subjects (10 events and 2 events, respectively).

As currently proposed, the null and alternative hypotheses can be written as:

$$H_0: \frac{\lambda_{\text{CSL222}}}{\lambda_{Prophy}} \geq M \; vs \; H_1: \frac{\lambda_{\text{CSL222}}}{\lambda_{Prophy}} < M$$

where M represents the non-inferiority margin and λ_{XXX} represents the rate of events in the XXX group (CSL222 representing the Post-treatment period, Prophy representing the Lead-in period). This can be equivalently rewritten as the difference between the rates on the natural logarithmic (base e) scale:

$$H_0: \log \lambda_{\text{CSL222}} - \log \lambda_{Prophy} \geq \log M \; \; vs \; H_1: \log \lambda_{\text{CSL222}} - \log \lambda_{Prophy} < \log M$$

If M1 represents the entire effect of prophylactic treatment compared to on-demand treatment on the natural logarithmic scale, then M1 = log(0.238) = -1.4354846, the upper limit of the CI from the Idelvion publication. If M2 represents the percentage of treatment effect relative to on-demand to be preserved, then M2 can be stated as p*M1 = p*log(0.238). Then, the hypotheses of interest can be restated as:

$$H_0: \frac{\lambda_{\text{CSL222}}/\lambda_{on-d}}{\lambda_{Prophy}/\lambda_{on-d}} \geq M \; vs \; \; H_1: \frac{\lambda_{\text{CSL222}}/\lambda_{on-d}}{\lambda_{Prophy}/\lambda_{on-d}} < M$$

or equivalently as differences on the natural logarithmic scale, where “on-d” represents on-demand treatment. Substitution of M1 in the denominator and M2 in the numerator gives the following equation:

$$(p-1)M1 \geq \; \log M$$

$$(1-p)x1.4354846 \; \geq \log M$$

This can then be solved for M or for p. This approach will be called Approach (1).

Alternatively, if M1 represents the entire effect of prophylactic treatment on the efficacy scale, then M1 = 1 – 0.238 = 0.762, using the relationship that efficacy = 1 – rate ratio, and M2 = pM1. Substituting 1-M2 into the numerator and 1-M1 into the denominator (RR = 1-efficacy), yields the following equation:

$$\frac{1 - p * 0.762}{1 - 0.762} < M$$

This approach will be called Approach (2).

The two approaches provide slightly different interpretations.

Table 2: Non-inferiority Margin (for the ABR rate ratio) for Retention of p%

	Percentage Retention				
	60%	**70%**	**75%**	**80%**	**90%**
Approach (1)	1.776[a]	1.538	1.43	1.33	1.15
Approach (2)	2.28	1.96	1.80	1.64	1.32

a. A margin of 1.8 using difference in log values corresponds to 59% retention of treatment effect.

NI margin of 1.8 was selected as it retained a sufficiently large and clinically meaningful proportion of treatment effect [60% for Approach (1) and 75% for Approach (2)] of Prophylaxis over On-demand treatment.

Hypothetical scenarios illustrating the upper limit of the non-inferiority margin under observations of varying pre-treatment annualized bleed rates are shown in the table below.

Table 3: Relation of Pre-treatment ABR and NI Margin using a 1.8 Rate Ratio

ABR observed during Lead-in period	2.0	2.5	5.0	10
Maximum ABR Post-treatment to maintain NI (bleeds)	<3.6	<4.5	<9.0	<18
Permissible increase in number of bleeds	<1.6	<2.0	<4.0	<8

NI: non-interiority

3.4 Randomization and Blinding

Not applicable, as this is an open-label trial with one treatment arm.

4. DATA DEFINITIONS AND PRE-PROCESSING

4.1 Baseline Definitions

The baseline FIX activity will be imputed based on the subject's historical hemophilia B severity that is documented on the CRF. If the subject has documented severe FIX deficiency (FIX plasma level <1%) their baseline FIX activity level will be imputed as 1%. If the subject has documented moderately severe FIX deficiency (FIX plasma level >= 1% and <= 2%) their baseline FIX activity level will be imputed as 2%.

Baseline age is the age in years at the time of the Screening Visit.

For patient reported (quality of life) outcomes and HJHS that are assessed at visits, the Baseline value is period-specific. For the Lead-in period, the Baseline value is the latest value prior to the Lead-in period that is not within 2 weeks of a bleed. For the Post-treatment period, the Baseline value is the latest value prior to IMP that is not within 2 weeks of a bleed. Note that the Baseline value is not used as a covariate in statistical-modeling-based treatment-period-comparative analyses for this study, because each subject serves as one's own control.

For vital signs and safety laboratory values, Baseline is period-specific. The Baseline value for the Lead-in period is the last non-missing central laboratory value or vital signs value on or prior to Visit L1. The Baseline value for the Post-treatment period is the last central laboratory value or vital signs value prior to the first dose of CSL222.

For height, weight, and BMI, the baseline value is the last value prior to the start of the Lead-in period.

4.2 Data Handling Rules and Definitions, Including Handling of Missing Data

Missing data will be maintained as missing in the safety and efficacy datasets, unless specified otherwise.

Data for Adverse Events Summaries by Severity and Relationship to Study Drug

For the AE summaries by severity (mild, moderate, or severe), an AE with missing severity will be deemed as severe. For the AE summaries by relationship to study drug, an AE with a missing relationship to study drug will be deemed as related. Seriousness cannot be imputed as 'Yes' by default, since this would affect the reconciliation between trial database and registry of SAEs.

Data for Laboratory Summaries (Continuous Parameters)

Data from unscheduled visits or early discontinuation visits will not be used for by-visit summaries (unless they have been assigned to a scheduled visit according to the Time Windows for Statistical Analysis). Data from both scheduled and unscheduled visits (or early discontinuation visits) will be used for determining incidence of clinically significant values.

Data for All Laboratory Summaries

Definitions are provided in Appendix 1 Data Handling Rules.

Study Dates and Day of Assessment or Event

Study Day and Day of Assessment or Event definitions are provided in Appendix 1 Data Handling Rules.

Duration of Event

Definitions are provided in Appendix 1 Data Handling Rules.

Distance between Events

Definitions are provided in Appendix 1 Data Handling Rules.

The Use of Lead-in Period Month 6 Visit versus Lead-in Final Visit for Analyses of FIX Activity

For FIX activity, for the Lead-in Month 6 assessment and the Lead-in Final assessment, some specific instructions are as follows. For FIX activity (aPTT and chromogenic from the central laboratory) for analysis purposes if a Lead-in Month 6 value is available, then it will be used (for analysis); otherwise, if a Lead-in Final value is available, then it will be used for the purpose of (i.e., as if it were) the Month 6 value. The rationale is that (1) Lead-in Month 6 is a planned assessment (for FIX activity) and that in the presence of a Month 6 assessment, a Lead-in Final value is not essential for analysis and (2) the planned duration of the Lead-in period is approximately 6 months.

Missing Data

If causality is missing for a TEAE, the TEAE will be regarded as ‘Related’. If causality is missing for an AE with onset before administration of trial drug, the AE will be regarded as ‘Not related’. If the intensity is missing, the intensity of the AE will be regarded as ‘Severe’. In the case where seriousness is missing, this should be queried. Seriousness cannot be imputed as ‘Yes’ by default, since this would affect the reconciliation between trial database and registry of SAEs.

Time Windows for Data Collection:

The below assessment time windows (in Table 4) are given in the protocol as ‘target’ time windows for assessments to be carried out. However, these are not the time windows to be applied for statistical analysis, which are described farther below.

Table 4: Protocol-Recommended Study Time Intervals for Collection of Efficacy and Safety Evaluation and Laboratory Parameters

Nominal Time for Visits or Assessment	Protocol-Recommended Time Interval for the Visit or Assessment
Screening	Approximately -28 days prior to Visit L1
Lead-in	
Visit L1 (L-W0)	0 days
Visit L2 to LX (every 2 months, starting at L-W8)	±14 days
Visit L-Final	-28 days (± 7 days) from Visit D
IMP Dose	
Post-IMP 3 hours	±15 minutes
Post-treatment Follow-Up	
Week 1 to Week 12	±2 days
Month 4 to Month 11	±5 days
Month 12/Week 52	±5 days
Long-Term Follow-up	
Month 18	±2 weeks
Month 24	±2 weeks
Month 30	±2 weeks
Month 36	±2 weeks
Month 42	±2 weeks
Month 48	±2 weeks
Month 54	±2 weeks
Month 60	±2 weeks

Time Windows for Statistical Analysis:

Scheduling difficulties due to the coronavirus disease (COVID-19) pandemic may result in an increased number of missed, delayed, or unscheduled visits. Scheduling difficulties may also result in the performance of assessments at a scheduled visit where performance of the assessment was not originally planned. As an action to mitigate risk, analysis windows will utilize such unplanned assessments as follows. A schema for the assignment of such unplanned assessments to scheduled time points for visit-based endpoint analysis and summary (for visit-based efficacy and safety endpoints) will be defined as follows:

- An unplanned assessment will be assigned to a scheduled visit only if that visit has a missing value for the relevant endpoint and the visit was a scheduled time point for the performance of the assessment per the study protocol.

- Analysis windows for the assignment of unplanned assessments will range from the previous visit at which the endpoint is planned to be collected to the next visit at which the endpoint is planned to be collected.

- The unplanned assessment closest in time within the analysis window (either before or after) will be used to replace a missing assessment for a scheduled visit. If two unplanned assessments are both the closest in time, with one being before and the other being after, the earlier assessment will be used.

- For efficacy endpoints, values obtained on a Post-treatment actual study day < Day 21 will not be candidates to be used for imputing missing values for Post-treatment visits at a nominal visit time subsequent to Post-treatment Study Day 21.

- Only values from unplanned assessments (not planned assessments) will be used to replace a missing scheduled assessment.

- An unplanned assessment may be used more than once provided it lies within the analysis window for two consecutive scheduled time points at which the assessment was not performed as planned.

- Analysis windows will be defined separately for the Lead-in and Post-treatment periods. This means that values will not be carried from one treatment period to the other.

- For laboratory-based efficacy assessments, the above instructions (in the first 8 bullet points above) are applicable pertaining to all planned assessments; except, however, for the Lead-in Month 6 assessment and the Lead-in Final assessment, some specific instructions are as follows. For relevant laboratory-based efficacy endpoints – i.e., FIX activity (aPTT and chromogenic) – and for the laboratory endpoints of Total (IgM and IgG) and Neutralizing Antibodies to AAV5 –, eligible unplanned assessment values will be used (as applicable) to replace a missing assessment for the Lead-in Month 6 Visit (but not used to provide a value for the Lead-in Final Visit). The rationale is that Lead-in Month 6 is a planned assessment and that – for such endpoints – in the presence of a Month 6 assessment, a Lead-in Final value is not essential for analysis. For these endpoints, the analysis window for the assignment of unplanned assessments to the Lead-In Month 6 Visit will range from the previous visit at which the endpoint is planned to be collected (i.e., the planned date of the Lead-in Month 4 Visit) to the actual Day before CSL222 dosing. If both the Lead-in Month 6 Visit and the Lead-in Final Visit have no value (for the endpoint), then an unplanned-assessment value will be sought to be assigned to the Lead-in Month 6 Visit; if either the Lead-in Month 6 Visit or the Lead-In Final Visit has a value (for the endpoint), then no such assignment (of an unplanned

assessment to Lead-in Month 6) is needed. Unscheduled-visit values should never be assigned to the Lead-in Final Visit.

- For quality-of-life efficacy endpoints, the above instructions (in the first 8 bullet points above) are applicable pertaining to all planned assessments; however, for certain quality-of-life endpoints some specific instructions for the imputation of the Lead-in final value are provided as follows. For quality-of-life efficacy endpoints that are scheduled to be collected at the Lead-in Month 4 and Lead-in Final Visits – i.e., EQ-5D-5L, IPAQ, WPAI, BPI, HAL, Hem-A-QoL, HJHS, and PROBE – eligible unplanned assessment values will be used (as applicable) to replace a missing assessment for the Lead-in Final Visit. For these endpoints, the analysis window for the assignment of unplanned assessments to the L-Final Visit will range from the previous visit at which the endpoint is planned to be collected (i.e., the planned date of the Lead-in Month 4 Visit) to the actual Day before CSL222 dosing. The latest available eligible unplanned assessment will be assigned to the Lead-in Final Visit. By the way, the Lead-in Month 6 Visit is not a planned assessment time and therefore is not to be receiving values from unplanned assessments.

- For quality-of-life efficacy endpoints, the above instructions (in the first 8 bullet points above) are applicable pertaining to all planned assessments; however, for certain (other) quality-of-life-endpoints some specific instructions for the imputation of the Lead-in final value are provided as follows. For visit-based endpoints that are scheduled to be collected for the first time (post-Screening) at the Lead-in Final Visit – i.e., Abdominal Ultrasound (safety) – eligible unplanned assessment values will be used (as applicable) to replace a missing assessment for the Lead-in Final Visit. For these endpoints, the analysis window for the assignment of unplanned assessments to the L-Final Visit will range across the entire duration of the Lead-in Period. The latest available eligible unplanned assessment will be assigned to the Lead-in Final Visit. By the way, the Lead-in Month 6 Visit is not a planned assessment time and therefore is not to be receiving values from unplanned assessments.

Unscheduled or unplanned assessment or early discontinuation values will not be assigned to scheduled (analysis) visits for vector DNA (genome) assessments.

Any other rules for missing data handling will be given in the endpoint-specific sections.

4.3 Bleed Counting Rules

Bleeds will be counted irrespective of assessments by the investigator as to the trueness or newness of the bleed (except for a small number of designated sensitivity analyses).

For designated supportive analyses, only exogenous-factor-IX-treated bleeds will be counted. If the field for whether the bleed was treated is missing, then (for conservativeness) it will be assumed that the bleed was treated with FIX.

For a small number of designated sensitivity analyses, only bleeds that are assessed to be new and true bleeds will be counted. If the assessment field for newness of the bleed is missing, then (for conservativeness) it will be assumed that the bleed is new. If the assessment field for trueness of the bleed is missing, then (for conservativeness) it will be assumed that the bleed is true. The rationale for these sensitivity analyses is given in the following bullet points:

- The bleeding events are evaluated by the Principal Investigator or designee through the review of the signs and symptoms self-reported in the diary and/or during discussions with the subject and it is determined whether the reported event was a true bleed and whether the reported event was a new bleed. For example, based on such sign and symptom evaluations, the investigator in some cases may need to distinguish whether there is a new bleed or whether the subject is experiencing pain (due e.g., to previous chronic joint bleeds and damage) that is not really a new bleed.

- When subjects are next at the study site, the physician may elect to use a diagnostic scan (X-ray, ultrasound, MRI, CT scan, etc.) to confirm the presence of blood or signs of acute inflammation. Blood or signs of acute inflammation observed using one or more of these confirmatory methods coupled with the physician's assessment will serve as sufficient confirmation to identify an event as a true bleed.

- The commercial names and half-lives of exogenous FIX medications that may contaminate endogenous FIX data are list in Table 8 of the protocol. In the analysis, any person-time during the Post-treatment period within 5 half-lives subsequent to exogenous FIX use will not be counted in the time at risk of (having) a bleeding event. Any bleeds occurring on or after stable FIX expression (Month 6 Post-treatment; on or after Day 21 for pre-Month 18 data cuts) should still be counted as events, even if they occurred during a time interval of "contamination".

5. DATA AND ANALYTICAL QUALITY ASSURANCE

The overall quality assurance procedures for the study data, statistical programming and analyses are described in Standard Operating Procedures of Everest Clinical Research and CSL Behring. Detailed data management procedures are documented in the study Data Management Plan, Data Validation Check Specifications, and Integrated Safety Data Review Plan.

6. ANALYSIS POPULATIONS

6.1 Population Definitions

6.1.1 Screen Failures

The screen failure population will include all subjects who were screened but never entered the Lead-in period.

6.1.2 Lead-in Discontinuers

The Lead-in discontinuers population will include all subjects who entered the Lead-in period but discontinued from the study prior to CSL222 dosing.

6.1.3 Safety Population

The Lead-in safety population will consist of all subjects who are enrolled into the Lead-in period. The Post-treatment safety population will consist of all subjects who receive CSL222, irrespective of any protocol deviations. Period-specific safety tabulations will use the period-specific safety population for the "N" and denominator (for percentages). The safety population will consist of all subjects who are in either the Lead-in safety population or the post-treatment safety population.

6.1.4 Full Analysis Set (FAS)

The FAS will include all subjects who are enrolled, entered the Lead-in period, were dosed with CSL222, and provide at least one assessment for any efficacy endpoint after CSL222 dosing. The FAS population will be the primary population for all efficacy statistical analyses.

6.1.5 Per-Protocol Population

The PP population will include all subjects from the FAS population who adhere to a stable and adequate prophylaxis use during the Lead-in period, and who have no major protocol deviations that impact the interpretation of efficacy. If subjects have major protocol deviations Post-treatment that impact the interpretation of efficacy, the subjects will be included in PP population, but will be excluded from the analysis starting from the time of the protocol deviations and thereafter. The PP population will be used for sensitivity analyses.

7. STATISTICAL ANALYSIS

7.1 Subject Disposition

A disposition table for CSL222_3001 for all screened subjects will be provided. This table will include the number of subjects who were screen failures, Lead-in discontinuers (i.e., those enrolled but not treated with CSL222), those who prematurely discontinued from treatment, who completed treatment (i.e., received the full dose of study treatment), who withdrew early from the study post dose of CSL222, and who completed the study. The number and percentage of subjects included in the FAS, PP, Lead-in safety population, and Post-treatment safety population will also be tabulated. The number and percentage of subjects in the PROBE sub-study, as well as the number and percentage of subjects who prematurely discontinued from the PROBE sub-study, will be tabulated.

The data on subject disposition, missed visits, and protocol deviations (including those related to COVID-19) will be listed.

7.2 Demographic and Baseline Characteristics

Descriptive statistics of demographics and baseline characteristics will be presented for the FAS, PP, Lead-in safety population, and Post-treatment safety populations. Similar demographics and baseline characteristics will also be presented for subjects who enroll into the PROBE sub-study. For quantitative variables, all summaries will include the number of non-missing observations, mean, standard deviation (SD), first quartile (Q1), median, third quartile (Q3), minimum, and maximum. For the qualitative variables, the summaries will include the number and percentage of subjects in each category or level. All data will be included in listings.

7.2.1 Demography

Demographics collected at Screening include year of birth, race, ethnic group, and gender according to local regulations. According to inclusion criterion 1, all subjects are males.

7.2.2 Baseline Disease Characteristics

Baseline disease characteristics include duration of disease, endogenous FIX activity level at time of diagnosis, severity of disease, indicator of family history of hemophilia B disease, number of bleeds in the year prior to Screening (total, spontaneous, traumatic, joint, and unknown), and the type of FIX therapy used. The baseline disease characteristics are tabulated according to the information collected on the electronic Case Report Form (eCRF).

Severity of hemophilia B will be categorized as severe (FIX plasma level < 1%) or moderately severe (FIX plasma level >= 1% and <= 2 %).

7.2.2.1 Hemophilia B History

All hemophilia B history data will be listed, and the listing will include the following: date first presented symptoms, date of initial diagnosis, duration of disease, endogenous FIX activity level at diagnosis (if available), severity of hemophilia B at time of diagnosis, number of FIX exposure days (an exposure day is defined as a day when the subject received at least one injection of FIX treatment), and family members with a history of FIX inhibitors.

7.2.2.2 Medical and Surgical History

All medical and surgical history will be listed, including the following information: surgical or medical history event, start date and end date or current status. Medical history will be coded using the most recent version of the MedDRA at the time of the database lock.

7.2.2.3 Target Joints at Screening

A target joint is defined as any joint with 3 or more spontaneous bleeds within a consecutive 6-month period. Once there have been ≤ 2 bleeds into the joint within a consecutive 12-month period, the joint is no longer considered a target joint, and the target joint is then considered to have resolved. Target joints at Screening will be listed.

7.2.2.4 Baseline Immunologic Parameters

The baseline immunologic parameters include anti-FIX antibody titer levels, the presence of FIX inhibitors, total (IgG and IgM) antibodies to AAV5, neutralizing antibody levels to AAV5, and AAV5 capsid-specific T-cells. These data will be listed.

Box and whisker plots over Post-treatment study time will also be produced for these parameters, if applicable.

Please refer to Appendix 1 Data Handling Rules on how special laboratory values will be handled in quantitative analyses.

7.2.2.5 FIX Gene Mutation

FIX gene sequence analyses will be performed for all subjects who provide consent during the Screening Visit, even if they already have FIX gene mutation information available. The data will be presented in a listing.

7.2.2.6 Prior and Concomitant Medications

Prior and concomitant medications will be collected and coded using the most recent World Health Organization drug dictionary at the time of database lock. Prior and concomitant medications will be listed.

Prior medications are defined as those treatments with a start date before Visit L1 for the Lead-in period. A medication/therapy will be identified as a "Lead-in" concomitant medication/therapy if it is being continued by the subject at the date of the L1 Visit or is any new medication/therapy received during the Lead-in period prior to the date of CSL222 dosing. A medication/therapy will be identified as a "Post-treatment concomitant" medication/therapy if it is being continued by the subject at the date of CSL222 dosing or is any new medication/therapy received during the Post-treatment period. A medication with end date that is the same as the CSL222 dosing date will not be considered to be "Post-treatment concomitant".

7.2.2.7 Prior FIX Therapy Use

FIX therapy use during the year prior to Screening will be summarized and listed. FIX therapy use during the Screening period will be listed.

7.3 Investigational Product Exposure

A listing for exposure to investigational product will be provided showing the date of exposure and dose received. The listing will also state whether the full dose was received.

7.4 Blood Sample for Future Research

Data regarding the additional blood samples drawn at Screening (Visit S), Baseline (Visit D pre-IMP), Visit F12 (Week 12), and Visit F-Final (Month 12/Week 52), for the purpose of

potential future research in the hemophilia B disease area will be listed with the corresponding informed consent date.

7.5 Efficacy Analyses

All efficacy analyses detailed in this section will be performed using the FAS, unless otherwise stated. The primary efficacy endpoint was evaluated for the 6- and 12-month interim analyses, and again after the last subject completed the 18-month analysis visit (18-month data cut: 18 Oct 2021; refer to the 18-month SAP for Study CSL222_3001, Version 4.0, 10 Jun 2021).

7.5.1 Intercurrent Events

Intercurrent events of the efficacy analyses can be found in the following table.

Use of exogenous FIX therapy including continuous FIX prophylaxis during post-dose period	Contamination period will not be counted as time at risk for a bleeding event; however, bleeding events during such exogenous FIX exposure will be included in the ABR calculation. FIX activity measured from initiation of exogenous FIX therapy to 5 half-lives after the initation of FIX therapy will be considered contaminated and will be excluded from the analysis.
Liver Transplant	Efficacy data collected after liver transplantation will be excluded from the applicable analyses (eg. bleeding events, at-risk time, questionnaire, etc.)

7.5.2 Primary Endpoint

The primary efficacy endpoint is as follows:

- ABR comparison between CSL222 and prophylaxis for non-inferiority between the Lead-in period and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up.

For the 18-month data cut, bleeding events over the 52 weeks following stable FIX expression (months 7-18 Post-treatment) will be used in the analysis. For 6-month and 12-month interim analyses described in Section 3.2, bleeding events beginning at Day 21 of the Post-treatment period will be counted.

The ABR will continue to be analyzed using the same method for the remaining Post-treatment epoch through Month 60.

7.5.2.1 ABR Comparison Between CSL222 Post-treatment Period Following Stable FIX Expression (Beginning at Month 7) and Prophylaxis During the Lead-in Period for Non-Inferiority

ABR will be estimated for the lead-in and post-treatment periods (beginning from the first day of Month 7; e.g., months 7 to 60) using generalized estimating equations (GEE), assuming bleeding events follow a negative binomial distribution, including an offset parameter equal to the natural log of the collection periods (years). An unstructured covariance matrix will be specified and treatment (i.e., period) will be included as a categorical variable. If the convergence fails, then a compound symmetry covariance structure will be used. If convergence is not attained, then initial parameter estimates will be provided. The estimated rate ratio and one-sided 97.5% Wald CI and the corresponding p-value will be determined. The upper limit of the resultant CI of the rate ratio will be compared to the non-inferiority margin of 1.8 (see Section 3.3).

The Post-treatment time at risk of (having) a bleeding event is the subject's time on the study between stable FIX expression (Month 6) and the time of the last observed study day (e.g., study completion, early withdraw, death etc.) of the Post-treatment analysis period (e.g., months 7 to 60). Any bleeds prior to stable FIX expression (Month 6) of the Post-treatment period are not considered in the analysis. All bleeding events from the lead-in period will be counted, and the entire Lead-in period is assumed to be time-at-risk.

In the analysis, any person-time during the Post-treatment period within 5 half-lives subsequent to exogenous FIX use will not be counted in the time at risk of (having) a bleeding event. Nevertheless, any unique bleeding events (see Glossary of Terms) on or after stable FIX expression should still be counted as events, even if they occurred during a time interval of "contamination". See further details in the Data Handling Rules Appendix (Appendix 1 Data Handling Rules) in this document [under the category of "Contamination due to exogenous FIX (infusion) use"].

Descriptive statistics will be provided for the unadjusted ABR during each year since the first dose of CSL222 and during the lead-in period. The unadjusted ABR is the number of bleeds divided by the sum of person-time at risk during a given time period across all subjects. Bleeds prior to stable FIX expression (Month 6) post-CSL222 will be excluded from the calculation.

Bleeding events will be listed by subject. The number of bleeding events and the time-at-risk of bleeding events will be listed by period for each subject.

7.5.2.2 Sensitivity Analysis 1: PP Population

A sensitivity analysis using the PP population will be performed. The analysis will be similar to the main analysis described in Section 7.5.2.1.

7.5.2.3 Sensitivity Analysis 2: Including (Not Excluding) Periods Subsequent to Exogenous FIX Use

A second sensitivity analysis, using the FAS population, will be conducted for ABR to evaluate the robustness of the analysis findings to inclusion (i.e., non-exclusion) of time intervals with exogenous FIX use during the Post-treatment period. In this analysis, person-time during the Post-treatment period (that is) within 5 half-lives subsequent to exogenous FIX use will not be excluded from (i.e., will be included in) the time at risk for a bleeding event. However, as with the primary ABR analysis, bleeds and person-time on or after Day 1 and prior to stable FIX expression (Month 6) (to Day 21 for pre-Month 18 data cuts) Post-treatment will not be included in the analysis.

7.5.2.4 Sensitivity Analysis 3: Bleeds Treated with Exogenous FIX

A third sensitivity analysis will repeat the main analysis for ABR using the FAS population while considering only bleeds treated with exogenous FIX. Bleeds and person-time on or after Day 1 and prior to stable FIX expression (Month 6) (to Day 21 for pre-Month 18 data cuts) post-treatment will not be included in the analysis.

7.5.2.5 Sensitivity Analysis 4: Cumulative Responder Analysis using Subject-Specific Bleeding Rates

A cumulative responder analysis (as described in Farrar 2006) will be performed characterizing ABR in the Lead-in period and the Post-treatment period using the FAS population. The cumulative responder is defined as the percentage of subjects with an ABR being less than or equal to the listed criteria. The subject-specific ABR for the Lead-in and Post-treatment (beginning Month 7) periods will be plotted on the *x*-axis, and the proportion of "responders" will be plotted on the *y*-axis. Thus, a cumulative distribution plot will be produced where the proportion of responders can be compared by treatment period across a continuous range of ABR values.

7.5.2.6 Sensitivity Analysis 5: New and True Bleeds

A fifth sensitivity analysis will repeat the main analysis for ABR using the FAS population while considering only bleeds that are assessed to be new and true by the investigator.

7.5.2.7 Sensitivity Analysis 6: New and True Bleeds Treated with Exogenous FIX

A sixth sensitivity analysis will repeat the main analysis for ABR using the FAS population while considering only bleeds treated with exogenous FIX that are assessed to be new and true by the investigator.

7.5.2.8 Sensitivity Analysis 7: Excluding Periods Contaminated by Systemic Corticosteroid Exposure

A seventh sensitivity analysis will repeat the main analysis for ABR using the FAS population and will be conducted to evaluate the robustness of the analysis findings to exclusion of time

intervals with systemic corticosteroid use during the Post-treatment period. In this analysis, person-time during the Post-treatment period (that is) during systemic corticosteroid use or within 5 half-lives subsequent to the end of systemic corticosteroid use will be excluded from the time at risk for a bleeding event.

7.5.3 Secondary Efficacy and Quality of Life Endpoints

Secondary endpoints of the trial will focus on investigating the effect of 2×10^{13} gc/kg CSL222 on annualized consumption (and infusion rate) of FIX replacement therapy, remaining free of previous continuous routine prophylaxis, assessment of FIX activity, bleeding events, estimated ABR as a function of pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (as a "correlation" analysis), correlation of FIX activity levels and observed anti-AAV5 antibody titers using the luciferase based NAB assay after CSL222 dosing, occurrence and resolution of target joints, iPAQ, and EQ-5D-5L.

The secondary efficacy endpoints (listed below) were assessed at 12- and 18-Month Post-treatment. They will continue to be analyzed as described in Sections 7.5.4 through 7.5.18 for the remaining Post-treatment epoch through Month 60 (refer to the 18-month SAP for Study CSL222_3001, Version 4.0, 10 Jun 2021):

1. Endogenous FIX activity at 6 Month after CSL222 dosing
2. Endogenous FIX activity at 12 Month after CSL222 dosing
3. Endogenous FIX activity at 18 Month after CSL222 dosing
4. Annualized consumption of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up, excluding FIX replacement for invasive procedures compared to the Lead-in period
5. Annualized infusion rate of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up, excluding FIX replacement for invasive procedures compared to the Lead-in period
6. Proportion of subjects remaining free of previous continuous routine prophylaxis during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up
7. Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 over the 52 weeks following stable FIX expression (7-18 months).
8. ABR comparison between CSL222 and prophylaxis for superiority between the Lead-in and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up

9. Rate of spontaneous bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up compared to the Lead-in period

10. Rate of joint bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up compared to the Lead-in period

11. Estimated ABR – during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up – as a function of pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (as a "correlation" analysis) (this endpoint will not have hypothesis testing and therefore is not included in the Type I error control)

12. Correlation of FIX activity levels during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up with pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (this endpoint will not have hypothesis testing and therefore is not included in the Type I error control)

13. Occurrence of (and resolution of) new target joints during the 52 weeks following stable FIX expression (7-18 months) following CSL222 dosing and resolution of pre-existing target joints following CSL222 dosing (these endpoints will not have hypothesis testing and therefore are not included in the Type I error control)

14. Proportion of subjects with zero bleeds in the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up (this endpoint will not have hypothesis testing and therefore is not included in the Type I error control)

15. Patient reported outcome (PRO) questionnaire scores from the iPAQ (total physical activity score) during the 12 months following CSL222 dosing compared with the lead-in period

16. PRO questionnaire scores from the EQ-5D-5L (visual analogue scale (VAS) score) during the 12 monthsfollowing CSL222 dosing compared with the Lead-in period.

The secondary endpoints (#1 to 16 listed above) were addressed specifically with the 18-Month data cut submission (dated 31 Aug 2021). Each secondary endpoint analysis is described in the sub-sections below and will be continued for the remaining Post-treatment epoch throughout Month 60.

For the secondary efficacy endpoints, the main analyses will be performed using the FAS.

All data will be listed.

7.5.3.1 Sensitivity Analysis: PP Population

The sensitivity analyses using the PP population will be considered to be supportive analyses.

7.5.4 Endogenous FIX Activity After CSL222 Dosing

Uncontaminated endogenous central laboratory one-stage aPTT FIX activity after CSL222 dosing is the first secondary efficacy endpoint and will be analyzed using the FAS.

Change from baseline in endogenous FIX activity levels will be assessed over the post-treatment period. For the derivation of change, baseline FIX activity will be imputed as described in Section 4.1.

The change from baseline in FIX activity (FIX_{DIFF}) will be tested:

$$H_o: FIX_{DIFF} = 0 \text{ (no effect of treatment)}$$

$$H_1: FIX_{DIFF} > 0.$$

The hypothesis that $FIX_{DIFF} = 0$ (i.e., that the change from baseline is zero) will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant.

The change from baseline in FIX activity (%) will be analyzed using a repeated measures linear mixed model. The model will include visits as a categorical covariate, beginning at Week 3 (visits prior to Week 3 will be excluded from this analysis). A Toeplitz covariance matrix will be used to model correlation within a subject. If the Toeplitz model fails to converge, then a first-order autoregressive [AR(1)] structure will be used instead. In the AR(1) model, subject will be included as a random effect. If the AR(1) model also fails to converge, then subject will be modeled as a random effect in the absence of an AR(1) model. If convergence is still not attained, then initial parameter estimates will be provided. The change from baseline at each Post-treatment visit, the two-sided, 95% CI for the LS mean change, and the corresponding p-value for the comparison to zero will be obtained from the model and provided in a table. The LS mean change and CIs for each visit will be displayed graphically.

If a subject has no uncontaminated central laboratory Post-treatment FIX activity values, FIX activity will be imputed based on the historical hemophilia B severity as documented on the CRF in a manner identical to that used for baseline FIX activity, thus resulting in a change of zero.

Visits Post-treatment that are within 5 half-lives of exogenous FIX use are considered contaminated and will be excluded from this analysis. Details on the derivation of contaminated FIX samples can be found in the Data Handling Rules Appendix (Appendix 1 Data Handling Rules) in this document [under the category of "Contamination due to exogenous FIX (infusion) use"].

Descriptive statistics of actual value and change from baseline (for uncontaminated FIX activity) will be provided for all Post-treatment visits from Week 3 through Month 60.

Details of infused exogenous FIX medications used during the Lead-in and post-treatment periods will be provided in the listings.

7.5.4.1 Sensitivity Analysis 1: PP Population

The first sensitivity analysis will be to repeat the secondary efficacy analysis of FIX activity using the PP population.

7.5.4.2 Sensitivity Analysis 2: To Account for Missing Data

A second sensitivity analysis will be conducted for FIX activity after CSL222 treatment to evaluate the robustness of the main analysis findings to missing data. Any missing values (that are still missing even after the use of windowing to allow assignment of unplanned assessments to planned-assessment visits) will be imputed using the most recent previous Post-treatment uncontaminated FIX activity value that was observed for the subject.

7.5.4.3 Sensitivity Analysis 3: Cumulative Responder Analysis

A third sensitivity analysis, a (single-treatment) cumulative responder analysis, will be conducted for the change from baseline in uncontaminated central-laboratory FIX activity after CSL222 treatment. A cumulative distribution plot will also be produced. The observed change from baseline in FIX activity at each visit will be plotted on the *x*-axis and the proportion of responders will be plotted on the *y*-axis. If a subject lacks a visit value (being still missing even after the use of windowing to allow assignment of unplanned assessments to planned-assessment visits), then the (scheduled or unscheduled) uncontaminated central-laboratory Post-treatment value that is closest in time to (either before or after) that visit Post-treatment will be employed in the analysis. If two such assessments are both the closest in time, with one being before and the other being after, the earlier assessment will be used.

7.5.5 Annualized Consumption of FIX Replacement Therapy, Excluding FIX Replacement for Invasive Procedures, Following Stable FIX Expression (6 Months) during the Post-treatment Follow-up Compared to the Lead-in Period

Yearly (annualized) consumption of FIX replacement therapy, excluding replacement for invasive procedures, will be computed for each period by dividing the total consumption by the time of observation (in years). For the Lead-in period the time of observation will be the total number of days during which the subject is in the Lead-in period divided by 365.25. For the post-treatment period, the time of observation will be the number of days from stable FIX expression (beginning at Day 183) after CSL222 dosing, to the time of last observed study day (e.g., study completion at Month 60, early withdraw, death, etc.). Yearly (annualized) consumption of FIX replacement therapy will be compared between the treatment period and the Lead-in period using one-sided paired t-tests to test whether there is a significant decrease in annualized consumption (IU/Yr):

H_o: *(Post-treatment) – (Lead-in) = 0 (no effect of treatment)*
H_1: *(Post-treatment) – (Lead-in) < 0.*

One-sided p-values and treatment mean differences with two-sided 95% CIs will be produced for the treatment comparison and presented in a table. The treatments will be compared for superiority.

Descriptive statistics for annualized FIX consumption will be provided. All FIX therapy use will also be listed. A listing of annualized FIX therapy use (IU/kg/yr) will be provided.

7.5.5.1 Sensitivity Analysis 1: Including FIX Consumption for Invasive Procedures

A sensitivity analysis for annualized FIX consumption using the FAS will include FIX consumption for invasive procedures.

7.5.5.2 Sensitivity Analysis 2: Excluding Subjects who Returned to Continuous Prophylaxis

As another sensitivity analysis, the analysis (still excluding consumption for invasive procedures) will be repeated for the FAS excluding subjects who returned to continuous prophylaxis (refer to Glossary of Terms).

7.5.6 Annualized Infusion Rate of FIX Replacement Therapy, Excluding FIX Replacement for Invasive Procedures, Following Stable FIX Expression (6 Months) during the Post-treatment Follow-up Compared to the Lead-in Period

A similar negative binomial GEE with an offset term and unstructured covariance matrix, as described for the primary efficacy analysis (Section 7.5.2.1), will be used, analyzing the number of infusions of FIX replacement therapy as the response. The annualized infusion rate of FIX replacement (excluding use for invasive procedures) will be estimated for the Lead-in and post-treatment periods along with the rate ratio comparing them. The estimated rate ratio (between Post-treatment and lead-in) will be tested using the following hypotheses:

H_o: rate ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)
H_1: rate ratio (Post-treatment)/(lead-in) < 1.

The hypothesis will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant. The one-sided p-value and two-sided 95% CIs for the rate ratio will be presented in a table. The treatments will be compared for superiority.

If the model fails to converge, then a compound symmetry covariance structure will be used. If convergence is still not attained, then initial parameter estimates will be provided. The model will include treatment (i.e., period) as a categorical variable.

The Post-treatment time at risk of (having) a day-with-infusion event is the subject's time on the study on or subsequent to stable FIX expression (Month 6) (until the time of last observed study day (e.g., study completion at Month 60, early withdraw, death, etc.).

Descriptive statistics for annualized FIX infusion rate will be provided by period. All FIX therapy use will be listed. Details on exogenous FIX infusions for each subject including the number of exogenous FIX infusions and the time-at-risk, will be listed for the Lead-in period and for the Post-treatment period.

7.5.7 Proportion of Subjects Remaining Free of Previous Continuous Routine Prophylaxis during the Post-treatment Follow-up Period

The number and percentage of subjects remaining free of previous continuous routine FIX prophylaxis beginning from stable FIX expression (Month 6) until the time of last observed study day (e.g., study completion at Month 60, early withdraw, death, etc.) will be summarized. This endpoint will not have hypothesis testing and, therefore, will not be included in the Type I error control.

The date and study day (relative to CSL222 treatment) of return to continuous routine FIX prophylaxis will be listed for subjects who returned to such prophylaxis.

7.5.8 Comparison of the Percentage of Subjects with FIX Activity < 12% of Normal between the Lead-in Period and after Treatment with CSL222 Following Stable FIX Expression (6 Months)

Frequency counts and percentages of subjects with FIX activity < 12% of normal will be provided for all Lead-in time points, and Post-treatment time points beginning at Week 12 using the FAS population. Only uncontaminated values from the Post-treatment period will be used, whereas all values during the lead-in will be considered. The percentage of subjects attaining central laboratory one-stage aPTT FIX activity < 12% of normal will be compared between the Lead-in period and Post-treatment period (beginning at Month 6) using a generalized linear mixed model logistic regression with an assumed unstructured covariance matrix, adjusted by visit as a categorical covariate. The model will include visits Lead-in Months 2, 4, 6 and all Post-treatment beginning at Month 6. If there is no assessment at the nominal Lead-in Month 6 Visit but there is a nominal Lead-in Final Visit assessment, then the Lead-in Final Visit value will be used in the place of the Lead-in Month 6 Visit value.

The following hypotheses will be tested:

H_o: *Odds ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)*
H_1: *Odds ratio (Post-treatment)/(lead-in) < 1.*

A contrast will be constructed to compare the Lead-in and Post-treatment periods. The hypothesis that *odds ratio (Post-treatment)/(lead-in) = 1* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant. One-sided p-values and odds ratios with two-sided 95% CIs will be produced for the treatment comparison and presented in a table. The treatments will be compared for superiority.

If convergence is not attained, then a compound symmetry covariance matrix will be specified.

If a subject lacks a visit value (being still missing even after the use of windowing to allow assignment of unplanned assessments to planned-assessment visits), then the (scheduled or unscheduled) uncontaminated central-laboratory Post-treatment value that is closest in time to (either before or after) that visit Post-treatment will be employed in the analysis. If two such assessments are both the closest in time, with one being before and the other being after, the earlier assessment will be used.

If the above analysis does not converge, then the following steps will be employed in succession until convergence is met:

1. The 6 Month timepoint from the post-treatment period is removed
    i. The analysis is run assuming an unstructured covariance matrix
    ii. The analysis is run assuming a compound symmetry covariance matrix
2. Only the 'Lead-in Month 6' value will be included from the Lead-in Period
    i. The analysis is run assuming an unstructured covariance matrix
    ii. The analysis is run assuming a compound symmetry covariance matrix
3. Repeat Step 1 by removing Post-treatment time points iteratively from the analysis

7.5.8.1 Sensitivity Analysis 1: PP Population

The first sensitivity analysis will be to repeat the above analysis using the PP population.

7.5.8.2 Sensitivity Analysis 2: Cumulative Responder Analysis Comparing the Lead-in Period with the Post-treatment Period Following Stable FIX Expression (6 Months)

As a sensitivity analysis, a cumulative responder analysis (as described in Farrar 2006) using the FAS population will be conducted comparing the mean FIX activity from the central-laboratory between the Lead-in period (which is calculated using the mean of Lead-in Months 2, 4, and 6; if Month 6 value is missing then it will be substituted with Lead-in Final value) and the Post-treatment period beginning at Month 6. Only uncontaminated values from the Post-treatment period will be used, whereas all values during the lead-in will be considered. The mean FIX activity across visits for the Lead-in and Post-treatment periods will be plotted on the *x*-axis, and the proportion of "responders" will be plotted on the *y*-axis. The cumulative responder is defined as the percentage of subjects with a mean FIX activity being less than or equal to the listed criteria. Thus, a cumulative distribution plot by treatment period will be produced. The cumulative responder curves for each treatment period will then be compared using a two-sample Kolmogorov-Smirnov test. Frequency counts, percentages and p-value will be presented in the table.

If a subject lacks having at least one uncontaminated FIX activity value at least one of the time points during the Post-treatment period beginning at Month 6 (that is still missing even after the use of windowing to allow assignment of unplanned assessments to planned-assessment visits), then the single (scheduled or unscheduled) uncontaminated central-laboratory Post-treatment value that is closest in time to (either before or after) will be employed in the analysis. If two such assessments are both the closest in time, with one being before and the other being after, the earlier assessment will be used.

7.5.9 ABR Comparison between CSL222 and Prophylaxis for Superiority between the Lead-in Period and the Post-treatment Period Following Stable FIX Expression (6 Months) Post-treatment (CSL222) Follow-up

ABR will be determined for the Lead-in period and Post-treatment period. The same GEE detailed in Section 7.5.2.1 for the primary analysis (total bleeds ABR) will be used. Estimated ABR for the lead-in and post-treatment (beginning at Month 7) periods along with the associated rate ratio will be provided.

The estimated rate ratio (between Post-treatment and lead-in) will be tested using the following hypotheses:

H_o: *rate ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)*
H_1: *rate ratio (Post-treatment)/(lead-in) < 1.*

The hypothesis that *(Post-treatment)/(lead-in) = 1* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant. The one-sided p-value and two-sided 95% CIs for the rate ratio will be presented in a table. The treatments will be compared for superiority. The main population will be the FAS.

7.5.10 Rate of Spontaneous Bleeding Events during the Post-treatment Period Following Stable FIX Expression (6 Months) Compared to the Lead-in Period

The same negative binomial GEE detailed in Section 7.5.2.1 for the primary analysis (total bleeds ABR) will be repeated with spontaneous bleeding events as the response. Estimated ABR for the lead-in and post-treatment (beginning at Month 7) periods along with the associated rate ratio will be provided.

The estimated rate ratio (between Post-treatment and Lead-in) will be tested using the following hypotheses:

H_o: *rate ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)*
H_1: *rate ratio (Post-treatment)/(lead-in) < 1.*

The hypothesis that *(Post-treatment)/(lead-in) = 1* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant. The one-sided p-value and two-sided 95% CIs for the rate ratio will be obtained and

presented in a table. The treatments will be compared for superiority. The main population will be the FAS.

The number of spontaneous bleeding events and the time-at-risk of spontaneous bleeding events will be listed by period for each subject.

7.5.11 Rate of Joint Bleeding Events Following Stable FIX Expression Post-treatment Follow-up Compared to the Lead-in Period

The same negative binomial GEE detailed in Section 7.5.2.1 for the primary analysis (total bleeds ABR) will be repeated with joint bleeding events as the response. Estimated ABR for the lead-in and post-treatment (beginning at Month 7) periods along with the associated rate ratio will be provided.

The estimated rate ratio (between Post-treatment and Lead-in) will be tested using the following hypotheses:

H_o: rate ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)
H_1: rate ratio (Post-treatment)/(lead-in) < 1.

The hypothesis that *(Post-treatment)/(lead-in) = 1* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <=0.025 will be regarded as statistically significant. The one-sided p-value and two-sided 95% CIs for the rate ratio will be obtained and presented in a table. The treatments will be compared for superiority. The main population will be the FAS.

7.5.12 Estimated ABR – during the Post-treatment Period Following Stable FIX Expression (6 Months) as a Function of Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay (as a "Correlation" Analysis)

To examine the relationship (i.e., "correlation") between ABR and baseline anti-AAV5 neutralizing antibodies (NABs), the following analysis will be carried out. A nonparametric, generalized additive model (GAM) will be implemented to graph the relationship of ABR to the natural logarithm of baseline anti-AAV5 NABs with a negative binomial model. Values of "< LOD" will be set to LOD/2 for the purpose of this analysis. Information for each subject across the Post-treatment period following stable FIX expression (6 months) Post-treatment will be employed. This endpoint will not have hypothesis testing and therefore is not included in the Type I error control.

7.5.13 Correlation of FIX Activity Levels during the Post-treatment Period Following Stable FIX Expression (6 Months) with Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay

The arithmetic mean of uncontaminated central laboratory one-stage aPTT FIX activity level across all visits during the Post-treatment period beginning from stable FIX expression (e.g., months 6-60). The Pearson and Spearman correlation between this mean FIX activity and the

pre-IMP anti-AAV5 antibody titers (using the luciferase based NAB assay) will be tabulated as well as their 95% confidence intervals. The Pearson product-moment correlation coefficient (rp) will provide a measure of the strength of a linear association between FIX activity levels and pre-IMP Anti-AAV5 antibody titers and the Spearman correlation coefficient (rs) will provide a measure of the strength of a monotone association between FIX activity levels and pre-IMP Anti-AAV5 antibody titers. A scatter plot will be produced, with an overlaid linear regression line. This endpoint will not have hypothesis testing and therefore is not included in the Type I error control.

7.5.14 Occurrence of (and Resolution of) New Target Joints during the Post-treatment Follow-up Period Following Stable FIX Expression (6 Months)

The rate of occurrences of new target joints per person-time of follow-up starting from stable FIX expression (Day 21 for the 12-month data cut) Post-treatment (e.g., months 6-60) will be summarized descriptively across subjects. Target joints that are being counted are ones that did not exist prior to stable FIX expression (Post-treatment Month 6) (prior to Day 21 for the 12-month data cut). The percentage resolution of such new target joints will also be tabulated.

This endpoint will not have hypothesis testing and therefore is not included in the Type I error control. The main population will be the FAS.

7.5.15 Time to Resolution of Pre-existing Target Joints during the Post-treatment Follow-up Period

The time to resolution of pre-existing target joints (existing immediately prior to CSL222 dosing) will be summarized using a Kaplan-Meier curve. Time to resolution will be presented using the date of CSL222 dosing as the reference date. Each target joint will be handled as the experimental unit for this analysis, irrespective of subject. This analysis is censored at the subject's data cut-off date for the data cut.

This endpoint will not have hypothesis testing and therefore is not included in the Type I error control.

7.5.16 Proportion of Subjects with Zero Bleeds in the Post-treatment Period Following Stable FIX Expression (6 Month)

The number and percentage of subjects with zero bleeds during the Post-treatment period from stable FIX expression (Month 6) (from Day 21 for the 12-month data cut) Post-treatment will be tabulated with descriptive statistics and presented in a table. This endpoint will not have hypothesis testing and therefore is not included in the Type I error control.

7.5.17 Patient Reported Outcome (PRO) Questionnaire Scores from the iPAQ (Total Physical Activity Score) during the 12 Months Following CSL222 Dosing Compared with the Lead-in Period

iPAQ total physical activity score (MET-minutes/wk) will be analyzed using a repeated measures linear mixed model based on the FAS population. MET denotes metabolic equivalent of task. The model will include visit as a categorical covariate. An unstructured covariance matrix will be employed to account for the correlation across visits for the same subject (without modeling subject explicitly as a random effect); if that model fails to converge then the Toeplitz covariance structure will be employed (without modeling subject explicitly as a random effect); if both the unstructured and Toeplitz covariance structures fail to converge, then the AR(1) (autoregressive order 1) covariance structure will be employed and subject will be modeled explicitly as a random effect; if the unstructured and Toeplitz and AR(1) covariance structures fail to converge, then the compound symmetry covariance structure will be employed (without modeling subject explicitly as a random effect).

The following visits during the Lead-in period are when iPAQ is scheduled to be collected and will be employed in this analysis: Month 4 Visit (L3) and L-Final Visit. The following visits during the Post-treatment period are when iPAQ is scheduled to be collected and will be employed in this analysis: Month 6 Visit (F15), Month 12 Visit (F-Final), and Month 24 Visit (LTF2). Within the context of the repeated-measures analysis, the mean across the visits from the Post-treatment period will be compared to the mean across the visits from the Lead-in period, using a contrast. Visits will be weighted equally. For the comparison of the Post-treatment period to the lead-in period, the contrast is between the mean of Post-treatment months and the mean of Lead-in Month 4 and Lead-in Final. Further details about the contrasts are given in the SAP SAS Code Appendix (Appendix 5). Higher values of the total physical activity score are considered to be favorable.

If convergence was still not attained using the models above, then the arithmetic mean will be computed across visits within a period (for each subject) and a mixed model analysis will be carried out using period as a categorical covariate and with an unstructured covariance matrix to account for the correlation across periods for the same subject; if that model fails to converge then the compound symmetry covariance structure will be employed across periods for the same subject. If convergence is still not attained, initial parameter estimates will be provided. With these approaches, it will not be necessary to add subject as an additional random effect.

For the iPAQ questionnaire and other quality-of-life (patient reported outcomes) endpoints, questionnaires that are completed within 2 weeks of a bleed will not be included in the analysis.

The estimated treatment period difference (between Post-treatment and the Lead-in) will be tested using the following hypotheses:

H_o: (Post-treatment) – (Lead-in) = 0 (no effect of treatment)
H_1: (Post-treatment) – (Lead-in) > 0.

The hypothesis that *(Post-treatment) – (Lead-in) = 0* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <= 0.025 will be regarded statistically significant.

The one-sided p-value and two-sided 95% CI will be obtained and presented in a table. The treatments will be compared for superiority.

Also, all available visits will be summarized descriptively and listed. Both raw scores and summary scores will be listed. For descriptive and inferential analyses, only data from the short form IPAQ will be considered.

7.5.17.1 Sensitivity Analysis 1: PP Population

The sensitivity analysis will be to repeat the above analysis using the PP population.

7.5.18 PRO Questionnaire Scores from the EQ-5D-5L (VAS Score) during the 12 Months Following CSL222 Dosing Compared with the Lead-in Period

VAS Scores from the EQ-5D-5L will be analyzed using a repeated measures linear mixed model based on the FAS population. The VAS will be scored from 0 (worst imaginable health state) through 100 (best imaginable health state) to represent the subject's self-report concerning how bad or how good their health was during that day. The model will include visit as a categorical covariate. An unstructured covariance matrix will be employed to account for the correlation across visits for the same subject (without modeling subject explicitly as a random effect); if that model fails to converge then the Toeplitz covariance structure will be employed (without modeling subject explicitly as a random effect); if both the unstructured and Toeplitz covariance structures fail to converge, then the AR(1) (autoregressive order 1) covariance structure will be employed and subject will be modeled explicitly as a random effect; if the unstructured and Toeplitz and AR(1) covariance structures fail to converge, then the compound symmetry covariance structure will be employed (without modeling subject explicitly as a random effect).

The following visits during the Lead-in period are when EQ-5D-5L is scheduled to be collected and will be employed in this analysis: Month 4 Visit (L3) and L-Final Visit. The following visits during the Post-treatment period are when EQ-5D-5L is scheduled to be collected and will be employed in this analysis: Month 6 Visit (F15), Month 12 Visit (F-Final), and Month 24 Visit (LTF2). Within the context of the repeated-measures analysis, the mean across the visits from the Post-treatment period will be compared to the mean across the visits from the Lead-in period, using a contrast. Visits will be weighted equally. For the comparison of month 6-24 to Lead-in, the contrast is between the mean of Post-treatment Months 6, 12, and 24 and the mean of Lead-in Month 4 and Lead-in Final. Further details about the contrasts are given in the SAP SAS Code Appendix (Appendix 5).

If convergence was still not attained using the models above, then the arithmetic mean will be computed across visits within a period (for each subject) and a mixed model analysis will be carried out using period as a categorical covariate and with an unstructured covariance matrix to account for the correlation across periods for the same subject; if that model fails to converge then the compound symmetry covariance structure will be employed across periods for the same

subject. If convergence is still not attained, initial parameter estimates will be provided. With these approaches, it will not be necessary to add subject as an additional random effect.

For the EQ-5D-5L questionnaire and other quality-of-life (patient reported outcomes) endpoints, questionnaires that are completed within 2 weeks of a bleed will not be included in the analysis.

The estimated treatment difference (between Post-treatment and the Lead-in) will be tested using the following hypotheses:

H_o: *(Post-treatment) – (Lead-in) = 0 (no effect of treatment)*
H_1: *(Post-treatment) – (Lead-in) > 0.*

The hypothesis that *(Post-treatment) – (Lead-in) = 0* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <= 0.025 will be regarded as statistically significant.

The one-sided p-value and two-sided 95% CIs will be obtained and presented in a table. The treatments will be compared for superiority.

Results of the VAS as a measure of overall self-rated health status – baseline scores, scores at each visit, and changes from baseline at each visit will be summarized descriptively.

EQ-5D-5L VAS data will also be listed.

Both raw scores and summary scores will be listed. The compliance of completing the EQ-5D-5L questionnaires is a critical issue in the QoL and health-state evaluation, and will be described by visit, by displaying the number and percentage of subjects who were assessed (per subject, at least 1 question answered) at each visit for each period.

7.5.18.1 Sensitivity Analysis 1: PP Population

The sensitivity analysis will be to repeat the above analysis using the PP population.

7.5.19 Additional Sensitivity Analysis

The following secondary efficacy analyses will be repeated using the PP population as a sensitivity analysis:

- Annualized consumption of FIX replacement therapy, excluding FIX replacement for invasive procedures, following stable FIX expression (6 months) during the post-treatment follow-up compared to the Lead-in period (Section 7.5.5)
- Annualized infusion rate of FIX replacement therapy, excluding FIX replacement for invasive procedures, following stable FIX expression (6 months) during the post-treatment follow-up compared to the Lead-in period (Section 7.5.6)

- ABR comparison between CSL222 and prophylaxis for superiority between the Lead-in period and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up (Section 7.5.9)

- Rate of spontaneous bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period (Section 7.5.10)

- Rate of joint bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period (Section 7.5.11)

- Correlation of FIX activity levels during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up with pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay (Section 7.5.13)

- Occurrence of (and resolution of) new target joints during the 7-18 month Post-treatment follow-up (Section 7.5.14)

- Time to resolution of pre-existing target joints during the Post-treatment follow-up (Section 7.5.15)

7.5.20 Type I Error Control and Simultaneous Confidence Intervals

Formal statistical testing of the efficacy endpoints were performed using the closed testing principle (for Type I error control for multiple testing). Due to the closed testing principle, no correction for multiplicity was necessary. Among the endpoints formally tested for statistical significance, all were tested for superiority at a one-sided alpha level of 0.025 (except as otherwise noted).

Fixed sequential testing was performed using a hierarchical approach and would be continued until a non-significant result was obtained (except as otherwise noted). The order of fixed sequential tests is specified below:

1. ABR comparison between CSL222 and prophylaxis for non-inferiority between the lead-in and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up (primary efficacy endpoint)

2. Endogenous FIX activity at Month 6 after CSL222 dosing (first secondary efficacy endpoint)

3. Endogenous FIX activity at Month 12 after CSL222 dosing (second secondary efficacy endpoint)

4. Endogenous FIX activity at Month 18 after CSL222 dosing (third secondary efficacy endpoint)

5. Annualized consumption of FIX replacement therapy during the week52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period (secondary efficacy endpoint)

6. Annualized infusion rate of FIX replacement therapy during the week52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding FIX replacement for invasive procedures, compared to the Lead-in period (secondary efficacy endpoint)

7. Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 52 weeks following stable FIX expression (7-18 months) (secondary efficacy endpoint)

8. ABR comparison between CSL222 and prophylaxis for superiority between the Lead-in and the 52 weeks following stable FIX expression (7-18 months) Post-treatment (CSL222) follow-up (secondary efficacy endpoint)

9. Rate of spontaneous bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to Lead-in period (secondary efficacy endpoint)

10. Rate of joint bleeding events during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up compared to the Lead-in period (secondary efficacy endpoint)

11. Patient reported outcome (PRO) questionnaire scores from the IPAQ (total physical activity score) during the 12 months following CSL222 dosing compared with the lead-in period (secondary efficacy endpoint)

12. PRO questionnaire scores from the EQ-5D-5L (VAS score) during the 12 months following CSL222 dosing compared with the Lead-in period (secondary efficacy endpoint)

Simultaneous one-sided 97.5% CIs based on a graphical approach to multiple testing (Bretz et al, 2015; Guilbaud 2008; Strassburger and Bretz 2008) will be provided for the Type I error controlled efficacy endpoints as a supportive analysis. For endpoints for which an increase is favorable, the lower one-sided 97.5% confidence bound will be provided; for endpoints for which an increase is unfavorable, the upper one-sided 97.5% confidence bound will be provided.

For any data cuts (i.e., analysis times) that are not the main data cut for a given endpoint, the p-values and CIs will be considered to be descriptive rather than inferential.

Formal Type I error control ended with the 18-months data cut. No tabulations pertaining to Type I error control are required for the Post-treatment 24-months data cut.

7.6 Exploratory Efficacy Endpoints

Exploratory endpoints consist of:

- FIX protein levels during the 18 months following CSL222 dosing
- HJHS Total Score during the 12 months following CSL222 dosing
- EQ-5D-5L Index Score
- Additional PRO questionnaire scores (during the 12 months following CSL222 dosing)
    - WPAI impairment percentages: absenteeism, presenteeism, work productivity loss, and activity impairment
    - BPI pain intensity (severity) and the impact of pain on functioning (interference) scale scores
    - HAL overall summary score and component scores relating to upper extremity activities, basic lower extremity activities, and complex lower extremity activities
    - Hem-A-QoL Total Score and each of the 10 domains separately
- Estimated ABR comparisons as a function of mean FIX activity (a “correlation” analysis) over the 18-month Post-treatment follow-up.

7.6.1 Analysis of Exploratory Endpoints

All exploratory endpoints will be presented using descriptive statistics, where applicable. Continuous endpoints will be summarized descriptively by treatment and visit (and by subset for the subset analyses). Categorical endpoints will be summarized descriptively by treatment (and by subset for the subset analyses). Analysis will be performed using both the FAS and PP populations.

7.6.2 FIX Protein Levels during Following Stable FIX Expression (7-to 60 Months) Following CSL222 Dosing

FIX protein levels will be summarized descriptively by visit.

7.6.3 Analysis of PRO Questionnaire Scores

For all of the quality-of-life (patient reported outcomes) endpoints, questionnaires that are completed within 2 weeks of a bleed will not be included in the analysis. All available visits will be summarized descriptively and listed. Both raw scores and summary scores will be listed.

7.6.3.1 iPAQ Subscale Scores

iPAQ subscale scores (for vigorous physical activity, moderate physical activity, and walking) (MET-minutes/wk) (as exploratory endpoints) will be summarized by visit and period, displaying n, mean (SE), SD, Q1, median, Q3, minimum, and maximum. The subscale scores will be listed.

7.6.3.2 HJHS Total Score

HJHS Total Score will be analyzed in a manner similar to the EQ-5D-5L VAS score (section 7.5.18), except that the set of visits for scheduled collection of HJHS is different. HJHS is collected at Lead-In Final, Month 12, Month 24, Month 36, Month 48, and Month 60. HJHS questionnaires that are completed within 2 weeks of a bleed will be excluded in the analysis. Data will be summarized descriptively at each visit.

Further details about the contrasts are given in the SAP SAS Code Appendix (Appendix 5).

All available data will be listed.

7.6.3.3 EQ-5D-5L Index Score and Categorical Responses

The EQ-5D-5L index score will be analyzed in the same manner as the EQ-5D-5L VAS score. Analysis will be based on the FAS and the PP population.

The data will be weighted to calculate an index score based upon subjects' responses to the 5 dimensions. A higher score indicates a higher health utility.

Results from the EQ-5D-5L index score (using US value set) at baseline, at each visit, changes from baseline to each visit, and the mean index score over each treatment period will be presented. Descriptive statistics for the index score will be presented by treatment, and the index score will also be listed.

For calculations of index score, the following method – referenced in the EuroQol group's web site – will be employed (Pickard et al, 2019). Appendix 4 specifies the US Value Set for calculation of the index score. The subjects' categorical responses to each of the 5-dimensions will also be summarized. The following statistics will be displayed: n, frequency (number of subjects with the category), and percentage.

EQ-5D-5L data will also be listed.

No imputation will be made for missing data in the EQ-5D-5L responses.

7.6.3.4 Additional PRO Questionnaire Scores: WPAI

WPAI outcome scores will be presented as impairment percentages for: absenteeism, presenteeism, work productivity loss, and activity impairment. Higher percentages indicate greater impairment and less productivity, i.e., worse outcomes. The percentages will be

summarized descriptively by treatment and visit and will be listed. The following statistics will be displayed: n, mean (SE), SD, Q1, median, Q3, minimum, and maximum. Analysis will be based on the FAS and the PP population.

The WPAI will be analyzed in the same manner as the EQ-5D-5L VAS score.

The outcomes are calculated as follows.

Questions:

1 = currently employed

2 = hours missed due to health problems

3 = hours missed other reasons

4 = hours actually worked

5 = degree health affected productivity while working

6 = degree health affected regular activities

Scores:

Scores will be multiplied by 100 to be expressed as percentages.

- Percent work time missed due to health = Q2/(Q2+Q4)
- Percent impairment while working due to health = Q5/10
- Percent overall work impairment due to health =

  Q2/(Q2+Q4) + [(1-(Q2/(Q2+Q4))) × (Q5/10)]
- Percent activity impairment due to health = Q6/10.

7.6.3.5 Additional PRO Questionnaire Scores: BPI

The BPI mean pain intensity (severity) score is the mean of the four pain items and will be analyzed in the same manner as EQ-5D-5L VAS. A higher score indicates worse pain.

The four pain items of “worst”, “least”, “average”, and “now” and the mean pain intensity (severity) score will be summarized descriptively by treatment and visit. The following statistics will be displayed: n, mean, SD, Q1, median, Q3, minimum, and maximum. Analysis will be based on the FAS and the PP population. The raw scores and summary scores will be listed.

The BPI pain interference score is the mean of the seven interference items and will be analyzed in the same manner as the EQ-5D-5L VAS. A higher score indicates worse pain. The seven interference items will be summarized descriptively by treatment and visit. The following statistics will be displayed: n, mean, SD, Q1, median, Q3, minimum, and maximum. Analysis will be based on the FAS and the PP population. The raw scores and summary scores will be listed.

7.6.3.6 Additional PRO Questionnaire Scores: HAL

The normalized HAL overall summary score is scored using the HAL Scoring Sheet and will be analyzed in the same manner as EQ-5D-5L VAS. A higher score indicates a worse condition.

The component scores relating to upper extremity activities, basic lower extremity activities, and complex lower extremity activities will be summarized descriptively by treatment and visit. The following statistics will be displayed: n, mean (SE), SD, Q1, median, Q3, minimum, and maximum. Analysis will be based on the FAS and the PP population. The HAL scores and overall summary score will also be listed.

7.6.3.7 Additional PRO Questionnaire Scores: Hem-A-QoL

The normalized Hem-A-QoL total score will be derived using the Hem-A-QoL Scoring Manual and will be analyzed in the same manner as EQ-5D-5L VAS. A higher score represents a lower quality of life.

Each of the 10 domains will be summarized descriptively by treatment and visit. The following statistics will be displayed: n, mean (SE), SD, Q1, median, Q3, minimum, and maximum. Analysis will be based on the FAS and the PP population. The Hem-A-QoL scores and total score will also be listed.

Individual Hem-A-QoL domains will also be analyzed in the same manner as EQ-5D-5L VAS.

7.6.4 Impacted Responders Analysis: Correlation of FIX Activity Levels at Month 18 with Pre-IMP Anti-AAV5 Antibody Titers Using the Luciferase Based NAB Assay

An impacted-response curve will be developed – as an exploratory analysis – to examine the association between these two variables (FIX activity levels and pre-IMP anti-AAV5 antibody titers using the luciferase based NAB assay). An impacted response (for the purpose of NAB effect determination) is defined as a subject's having an uncontaminated one-stage aPTT assay for FIX activity (%) to be < 5% of normal at any post-treatment period visits. The percentage of subjects having impacted response for the group of subjects with NAB titer >= x will be plotted as a function of x. The number of subjects with NAB titer >=x will also be indicated as a function of x on the graph. If, based on this graph, there exists a value "x" of NAB titer above which > 25% of subjects have impacted response for a group (and if 12 or more subjects have NAB titer above that value "x"), then that NAB titer value "x" is considered to be a potential candidate for being a meaningful NAB cutoff; otherwise, no such candidate NAB cutoff titer will

have been identified. If a subject lacks a visit value (being still missing even after the use of windowing to allow assignment of unplanned assessments to planned-assessment visits), then the (scheduled or unscheduled) uncontaminated central-laboratory Post-treatment value that is closest in time to (either before or after) that visit Post-treatment will be employed in the analysis. If two such assessments are both the closest in time, with one being before and the other being after, the earlier assessment will be used.

Specific levels of pre-treatment (i.e., baseline) NAB titer may be identified and used as the basis for subgroups in evaluating their potential relationship with FIX activity levels. The pre-treatment NAB titer taken pre-dose on the day of dosing will be used. If this result is not available, the value closest in time pre-day of dosing (while being prior to the dose) will be used.

7.6.5 Estimated ABR over the Post-treatment Follow-up as a Function of Mean FIX Activity ("Correlation" Analysis)

To examine the relationship (i.e., "correlation") between ABR and FIX activity, the following analysis will be carried out. A nonparametric, GAM will be implemented to graph the relationship of ABR to mean FIX activity levels with a negative binomial model. Information for each subject across the Post-treatment period will be employed. Both scheduled and unscheduled central laboratory values will be employed, using their actual date of assessment. Let I denote the number (i.e., count) of uncontaminated during-the-treatment-period central-laboratory assessments of FIX activity by the aPTT assay for a subject (during the Post-treatment period).

The structure of the analysis:

- Let i denote the ordinal number of a subject's uncontaminated during-the-treatment-period central-laboratory assessment of FIX activity.
- Let A_i denote a subject's ith uncontaminated during-the-treatment period central-laboratory aPTT FIX activity.
- Let t_i denote the date of the FIX activity-assessment sample.
- Let M_i be the mean FIX activity of consecutive assessments:
  - $M_i = (A_i + A_{i+1})/2$
- Let B_i denote the number of bleeds occurring for a subject in the time interval encompassing dates after t_i but before or on t_{i+1} (i.e., for dates in the time interval $t_i <$ date $<= t_{i+1}$).
  - The exception is that any bleeds occurring on date t_1 will also be counted in B_1,
  - – i.e., a bleed occurring on the date of the first (applicable) sample will be counted in the first-time interval.

- Let D_i denote the duration of the time interval:
  - $D_i = t_{i+1} - t_i$ days for i>=2, and
  - $D_i = t_2 – t_1 + 1$ days for i=1 (i.e., for the first-time interval).
- All triples of B_i, M_i, and D_i – across all subject-specific time intervals and across all subjects – will be employed in a generalized additive model, negative-binomial-regressing B_i on M_i, with an offset equal to $\ln(D_i)$, where ln is the natural logarithm and where D_i is converted to units of years.
- This analysis will treat all triples of B_i, M_i, and D_i across all time intervals and subjects equally and will thus not explicitly account for intrasubject correlation.

Time sub-intervals that are contaminated by exogenous FIX use (by the 5 half-life rule) will be removed from the time at risk (D_i).

This analysis will be used to graph the estimated ABR and its 95% CI as a function of mean FIX activity. The analysis will be based on the FAS population. .

7.6.6 Rate of Traumatic Bleeding Events Following Stable FIX Expression (6 Months) Post-treatment Follow-up Compared to the Lead-in Period

The same negative binomial GEE detailed in Section 7.5.2.1 for the primary analysis (total bleeds ABR) will be repeated with traumatic bleeding events as the response. Estimated ABR for the lead-in and post-treatment (beginning at Month 7) periods along with the associated rate ratio will be provided.

The estimated rate ratio (between Post-treatment and the Lead-in) will be tested using the following hypotheses:

H_o: rate ratio (Post-treatment)/(lead-in) = 1 (no effect of treatment)
H_1: rate ratio (Post-treatment)/(lead-in) < 1.

The hypothesis that *(Post-treatment)/(lead-in) = 1* (i.e., no difference between the 2 treatment periods) will be tested and a one-sided p-value <= 0.025 will be regarded statistically significant. The one-sided p-value and two-sided 95% CIs for the rate ratio will be obtained and presented in a table. The treatments will be compared for superiority. The main population will be the FAS.

The number of traumatic bleeding events and the time-at-risk of traumatic bleeding events will be listed by period for each subject.

7.6.7 Subgroup Analyses

- Subgroup analyses will be carried out for the following endpoints (the subgroups are mentioned a bit farther below in this SAP):

- Endogenous FIX activity at Month 18
- Annualized consumption of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding replacement for invasive procedures, compared to the Lead-in period
- Annualized infusion rate of FIX replacement therapy during the 52 weeks following stable FIX expression (7-18 months) Post-treatment follow-up, excluding replacement for invasive procedures, compared to the Lead-in period
- ABR comparison between CSL222 (during the 52 weeks following stable FIX expression [7-18 months] Post-treatment follow-up) and FIX prophylaxis (during the Lead-in period)
- Comparison of the percentage of subjects with FIX activity <12% of normal between the Lead-in period and after treatment with CSL222 over the 52 weeks following stable FIX expression (7-18 months)
- Proportion of subjects remaining free of previous prescribed continuous routine prophylaxis during the 52 weeks following stable FIX expression (7-18 months) post-treatment follow-up.

The subgroup analyses for the aforementioned endpoints will be carried out for the following subgroups:

- Age categories: <40 years, 40 to <60 years, >= 60 years
- Race and/or Ethnicity subgroups (with categories to be specified later because the racial/ethnic frequencies are not well known in advance)
- Zero bleeds versus >=1 bleed in Lead-in
  - Because this subgrouping is defined using information from the Lead-in period, the analysis will provide descriptive statistics only and will provide those descriptive statistics for only the Post-treatment period.
- Presence or absence of target joints at Screening
- Baseline NAB titer categories: positive titer (>= LOD) versus negative titer (<LOD), where LOD denotes limit of detection.
- Subjects with baseline NAB titer < 3000.
- HIV-negative vs. controlled HIV positive (CD4+ count >200 /µL) at Baseline
- History of Hepatitis B or C at Baseline

- Baseline liver pathology, according to Baseline FibroScan™ or equivalent shear wave elastography, magnetic resonance elastography result:
  - Degree of fibrosis [≥9Kpa versus <9Kpa]
  - Degree of steatosis [Controlled Attenuation Parameter (CAP) score ≥S2 (≥260 dB/m) versus <S2 (<260 dB/m)] versus Missing.

For any subgroup that has fewer than 10 subjects (N < 10), descriptive statistics will be provided, and model-based statistics will not be conducted due to possible convergence and estimability issues.

In addition, the effect of ALT elevation (defined from the AE data using the preferred term of "Alanine aminotransferase increased" that occurred within 6 months Post-treatment), the effect of corticosteroid use for ALT elevations, the effect of pre-existing NABs to AAV5, and the effect of baseline FIX activity on FIX activity (by one-stage aPTT assay), and FIX protein using the FAS and PP populations will be summarized in tables. Mean ± SD of FIX activity and FIX protein by ALT elevation as well as corticosteroid use for ALT elevations using the FAS and PP populations will be plotted in figures.

7.6.8 Sub-Study Efficacy and Quality of Life Endpoint Analyses

Sub-study endpoints consist of:

- PROBE questionnaire sub-study summary scores
- Musculoskeletal ultrasound sub-study results.

Analysis will be based on the FAS population for the set of subjects participating in the respective sub-study. Any subject with at least one assessment of the sub-study endpoint will be considered to be participating in the respective sub-study.

7.6.8.1 PROBE Questionnaire Summary Scores

PROBE Summary Scores and individual item responses will be summarized descriptively by treatment and visit. Summary Scores and individual item responses also will be listed. A higher score indicates better health. The PROBE Summary Score ranges from 0 to 1.

7.6.8.2 Musculoskeletal Ultrasound Results

Detailed information about MSKUS result analyses will be described in a separate SAP.

7.7 Safety Analyses

All safety analyses will be based on the safety population.

The safety endpoints to be analyzed are:

- TEAEs
- Changes in abdominal ultrasound
- Anti-AAV5 antibodies (total [IgM and IgG], neutralizing antibodies)
- AAV5 capsid-specific T cells
- Anti-FIX antibodies
- FIX inhibitors and recovery
- Hematology and serum chemistry parameters
- ALT and AST levels and corticosteroid use for ALT and AST increases
- Vector DNA in blood and semen
- Inflammatory markers: IL-1β, IL-2, IL-6, IFNγ, MCP-1
- AFP.

7.7.1 Adverse Events

An adverse event is considered to be treatment-emergent for the CSL222 treatment (i.e., a TEAE) if the event occurs after the administration of the IMP, or if the AE worsened during the study after the dose of study drug (intensity and/or severity changed to a worsened grade). An adverse event that begins on the same date as the IMP administration is treatment-emergent if the AE begins after the time of dose or if the time of AE onset is unknown. Additionally, if an AE has an onset date during Post-treatment period and has an outcome of death, that death will be considered to be treatment-emergent. Furthermore, if the AE could possibly be treatment-emergent, based on the missing or incomplete date, then the AE will be regarded as treatment-emergent. A treatment-emergent adverse event can be described as having had "incidence" during the treatment period.

An adverse event will be counted as having had "incidence" during the Lead-in period if it occurs during the Lead-in period, or if the AE worsened during the Lead-in period (intensity and/or severity changed to a worsened grade). Additionally, if an AE has an onset date during the lead-in period and has an outcome of death, that death will be counted as having incidence during the Lead-in period. Furthermore, if the AE could have had incidence during the Lead-in period, based on the missing or incomplete date, then the AE will be regarded as having incidence during the Lead-in period.

An adverse event incidence table for the safety populations will be created displaying the number of subjects (and percentage) experiencing an incident event and the number of incident events for: any AEs, AEs of special notification, serious AEs, related AEs, serious and related AEs, AEs

leading to early treatment discontinuation (i.e., to a partial dose), mild/moderate/severe AEs, and deaths.

The following AE incidence summary tables will be presented by decreasing frequency of occurrence based on SOC and preferred term:

1. AEs for the Lead-in and Post-treatment safety populations
2. Serious AEs for the Lead-in and Post-treatment safety populations
3. Related TEAEs for the Post-treatment safety population
4. Related serious TEAEs for the Post-treatment safety population
5. TEAEs leading to treatment discontinuation for the Post-treatment safety population (treatment discontinuation means receiving only a partial dose)
6. Serious TEAEs leading to treatment discontinuation for the Post-treatment safety population (treatment discontinuation means receiving only a partial dose)
7. Fatal AEs
8. TEAEs by highest severity for the Post-treatment safety population
9. Related TEAEs by highest severity for the Post-treatment safety population
10. Incidence of TEAEs for Special Notification for the Post-treatment safety population
11. Incidence of non-serious TEAEs occurring in at least 5% of subjects in the Post-treatment period
12. The incidence of TEAEs occurring in at least 10% of subjects in the Post-treatment period.

All incident AEs will be tabulated by SOC and preferred terms within each SOC according to the Medical Dictionary for Regulatory Activities (MedDRA) terminology list. The version of the MedDRA that is current at the time of database lock will be used to code verbatim terms for AEs for final analysis of the data. A glossary of MedDRA preferred terms used for adverse events reported in the study along with the associated Investigator's verbatim term will be provided. No hypothesis tests will be performed.

The summary tables will be accompanied by individual subject listings of all AEs, including pre-treatment AEs and information on actual AE description, date/time of start and end of AE, preferred term (MedDRA), SOC (MedDRA), severity, relationship/causality, type of AE, action taken, seriousness and outcome. Pre-existing AEs will be flagged. Pre-existing AEs are not considered to be treatment emergent, except in case of worsening during/after trial treatment (to be collected as a separate AE). Separate listings will be created for AEs for special notification,

deaths, and SAEs. All adverse events, whether treatment-emergent or not, will be included in the listings. A listing of any reported deaths during Lead-in and Post-treatment periods will be provided and will include the number of days since IMP administration.

The following will be done for events with irregular onset dates**.** All AEs will be included in the data listings regardless of the completeness of the onset dates. Any partial dates will be used in order to determine whether an AE is Lead-in-incident or treatment-emergent using the rules in Appendix 1; however, imputed dates will not be provided in the data listings.

7.7.1.1 Adverse Events of Special Notification

Table 5 contains (S)AEs that qualify for special notification as they are seen as safety issues of particular concern for Advanced Therapy Medicinal Product (ENTR/F/2/SF/dn D (2009) 35810. Brussels, 03/12/2009) and gene therapy medicinal products (EMA/CHMP/GTWP/60436/2007).

Table 5: Adverse Events of Special Notification

AEs related to the IMP administration procedure
Suspected or confirmed cases of opportunistic or serious infections that in the investigator's opinion might be related to the IMP
Unexpected reactions (e.g., hypersensitivity, immunological, toxic or other as consequence of a change in the construction or function of the viral vector [e.g., generation of replication competent virus])
AEs related to product failure (including lack of efficacy)
AEs related to mandatory concomitant medication (e.g., immunosuppression)
AEs related to medical devices which form part of the product or are used for application of the product
Development of any new/recurrent cancer

These AEs should be reported and followed in the same manner as SAEs. Note that the AEs may be serious or non-serious by definition (please see the protocol for more details). AEs of special notification are designated as such on the eCRF and therefore do not need to be derived.

7.7.1.2 Severity of Adverse Event

If an AE changes severity over time, the severity of maximum severity (i.e., intensity) will be reported.

7.7.1.3 Relationship Between IMP and Adverse Event

Please refer to the protocol for the definitions of related to IMP, probably related to IMP, possibly related to IMP, and not related to IMP.

7.7.2 Changes in Abdominal Ultrasound

To monitor subjects for liver fibrosis and potential occurrences of liver malignancies, abdominal ultrasounds will be performed. These ultrasounds will occur at the final Lead-in visit at the latest (to establish baseline status), at Post-treatment Month 12, and then annually thereafter.

A shift table will be used to summarize normal and abnormal results at Month 12 and the subsequent follow-up visits relative to the results obtained at baseline. Baseline abdominal ultrasound is the most recent assessment prior to the dose of study medication.

All abdominal ultrasound data will be listed.

7.7.3 Anti-AAV5 Antibodies, Anti-FIX Antibodies, and FIX Inhibitors

Total IgG and IgM antibodies against the vector capsid is evaluated by the enzyme-linked immunosorbent assay (ELISA), anti-AAV5 neutralizing antibodies are assessed with the luciferase assay. Antibodies against FIX will be evaluated by ELISA and will be reported as IgG, IgM. The results from the total IgG and IgM antibodies against the vector capsid, neutralizing antibodies against the vector capsid and non-inhibitory FIX antibodies will be tabulated by visit using descriptive statistics (for the titer) of n, mean (SE), SD, Q1, median, Q3, and max titer at each visit. The titer of FIX inhibitors will be reported in Bethesda Units and the sub-class of immunoglobulin of the inhibitor will be displayed as IgG, IgM or others. These results will be displayed at each visit. All data will be listed.

Occurrences of “suffering from FIX inhibitors” will be flagged in the FIX inhibitor listing.

Measurement of FIX recovery (maximum concentration [C_{max}]) and incremental recovery measured as increase in activity per unit infused (IU/ml per U/kg) at 30 min after infusion of a dose of FIX will be performed at baseline Visit L-Final. Additionally, measurement of FIX recovery and incremental recovery should be done at suspicion of FIX inhibitor as judged by the investigator.

All data will be listed.

7.7.4 AAV5 Capsid-specific T Cells

The AAV5 capsid-specific T cells testing (ELISPOT) will be summarized by visit using descriptive statistics of n, mean (SE), SD, Q1, median, Q3, minimum, and maximum.

The data will also be listed.

7.7.5 Clinical Laboratory Measurements

The lab parameters collected include the following:

Table 6: Safety Lab Parameters

Hematology	
Hemoglobin	White blood cells with differential count
Hematocrit	CD4+ count
Platelet count	
Red blood cells	
Serum Chemistry	
Sodium serum electrolytes	Alkaline phosphatase (ALP)
Potassium serum electrolytes	C-Reactive Protein
Creatinine	Albumin
Gamma-glutamyltransferase	Total Bilirubin
AST	Glucose (non-fasting)
ALT	
Coagulation	
aPTT	
PT (or International Normalized Ratio [INR])	
Serology	
HIV 1/2 antibody differentiation	Hepatitis B extracellular antigen (HBeAG) *
HIV 1/2 screen	Hepatitis B virus (HBV) DNA
Hepatitis B surface antigen (HBsAG)	Hepatitis C virus (HCV) RNA
Alpha-fetoprotein	
AFP	
Local Laboratory	
AST	
ALT	

* This parameter was removed with Protocol Amendment 3.

A Clinically Significant Laboratory Abnormality as identified by the investigator after the study drug is administered will be recorded as an Adverse Event and tabulated as an AE in the AE analysis. Abnormalities occurring prior to the IMP administration will be noted in medical history and presented in a data listing.

All laboratory data will be stored in the database with the units in which they were originally reported. Laboratory data not reported in International System of Units (SI units; Système International d'Unités) will be converted to SI units before data analysis.

Individual clinical laboratory variables for hematology, serum chemistry, coagulation, serology, and local laboratory will be provided in listings. Comments for laboratory testing will be listed. For listings, laboratory values will be flagged as low or high based on the reference ranges provided by the central laboratory.

If there are multiple laboratory values for the same parameter for a visit, the last value will be chosen for analysis.

Summary statistics (n, mean, Q1, median, Q3, standard deviation, minimum, and maximum) for the baseline assessment and change from baseline at each post-baseline visit for scheduled lab assessments of continuous laboratory variables will be tabulated for Post-treatment safety population. Data from unscheduled visits or early discontinuation visits will not be used for the by-visit summaries (unless they have been assigned to a scheduled visit according to the Time Windows for Statistical Analysis). Data from both scheduled and unscheduled visits (or early discontinuation visits) will be listed.

Shift tables will be produced using the categories defined by the Common Terminology Criteria for Adverse Events (CTCAE) grades for the Post-treatment safety population for hematology and serum chemistry. For these shift tables, the subject's pre-IMP grade will be cross-tabulated by the subject's maximum Post-treatment follow-up; also, the subject's maximum post-IMP grade during Post-treatment follow-up will be tabulated for all baseline grades combined. Percentages of subjects in each maximum post-IMP grade will be calculated for each pre-dose grade for the treatment and also for all baseline grades combined. Laboratory abnormal values on-treatment will be flagged as High or Low values based on laboratory reference ranges provided by LabCorp Laboratories (found in Appendix 3). These flags along with the reference ranges will be provided in the laboratory data listings.

Potentially Clinically Significant Laboratory Values Above/Below a Clinically Relevant Threshold on-treatment, based on CTCAE and other criteria, will be identified based on the thresholds in the table below.

Table 7: Potentially Clinically Significant (PCS) Laboratory Parameter Criteria

Central Laboratory	Post-Baseline Criteria
Serum Chemistry	
Sodium serum electrolytes	NA
Potassium serum Electrolytes	<3.0 mmol/L >6.0 mmol/L
Creatinine	>2 x ULN
Gamma-Glutamyltransferase	NA
AST	>2 x Baseline
ALT	>2 x Baseline
ALP	>2 x ULN
CRP	NA
Albumin	NA
Total bilirubin	>2 x ULN
Glucose (non-fasting)	NA
Hematology	
Hemoglobin	<8.0 g/dL (<80 g/L) Increase of >40 g/L to a value above the ULN
Hematocrit	NA
Platelet count	<50 x 10^9/L >999 x 10^9/L
Red blood cells	NA
White blood cells with differential count	<2 x 10^9/L >35 x 10^9/L
CD4+ count	≤200/µL
Coagulation	
aPTT	NA
PT (or INR)	NA
Serology	
HIV viral load	>200 copies/mL
HBsAg	NA
HBeAG	NA

Hepatitis B Virus DNA (HBV DNA)	NA
Hepatitis C Virus RNA (HCV RNA)	NA
Alpha-fetoprotein	
AFP	NA
Local Laboratory	
AST	>3 x ULN
ALT	>3 x ULN

NA: Not Applicable

Clinically significant laboratory values will be tabulated for the Lead-in safety population and the Post-treatment safety population. For all laboratory data for the parameter identified as potentially clinically significant for a subject will be listed. Low platelet counts are counted as being clinically significant only if they occur >= 4 weeks after IMP administration.

On the listings, the reference range and flag indicating if the measurement in question is outside the reference range will be provided.

7.7.6 ALT Levels, AST Levels and Corticosteroid Use for ALT and AST Increases

Summary statistics (n, mean, Q1, median, Q3, standard deviation, minimum, and maximum) for the baseline assessment and change from baseline at each post-baseline visit for ALT levels, AST levels and corticosteroid use will be tabulated and listed. Data from unscheduled visits or early discontinuation visits will not be used for the by-visit summaries (unless they have been assigned to a scheduled visit according to the Time Windows for Statistical Analysis). Data from both scheduled and unscheduled visits (or early discontinuation visits) will be listed.

Plots of individual subject profiles of ALT and AST levels over time will also be displayed. Corticosteroid use will be indicated on the plots.

7.7.7 Vector Genome Detection

The number of days until vector DNA can no longer be detected in semen and blood will be tabulated. The number of days is calculated using the date of collection of the third consecutive negative sample for each matrix.

All data will be listed.

Time to first shedding negative will be defined for each type of matrix and each subject as the Post-treatment time point where a negative result is measured for the first time in a consecutive order of 3 or more time points with a negative result. A negative result is defined as a result of either ‘0’ or ‘LOD’ (limit of detection). The time to first shedding negative will be flagged on the above-mentioned listings.

The time to first shedding negative for the Post-treatment period will also be summarized using a Kaplan-Meier curve. The censoring time will be truncated at the data cut-off date, the time of completion of the study, or time of early withdrawal from the study, whichever is earlier. For the 5-year analysis (and 5-year CSR), there will be no data cut-off date.

7.7.8 Inflammatory Markers

Blood samples will be taken to assess IL-1β, IL-2, IL-6, IFNγ and MCP-1 (monocyte chemotactic protein-1) using ELISA.

The test results will be summarized by visit using descriptive statistics and will include a change from baseline calculation for each post-baseline measurement. All data will also be listed. Inflammatory markers will not be summarized or presented following the interim 6-month data cut as the full set of inflammatory marker data will not be available as part of the 6-month database lock.

7.7.9 Alpha-fetoprotein

Alpha-fetoprotein results will be summarized by period and visit. They will also be listed.

7.7.10 Physical Examination (Including Height and Weight)

A physical examination will be performed at Screening (Visit S), L-Final, Visit D (pre-IMP), during the Post-treatment follow-up at visits F1, F2, F4, F6, F12, F13, F15, F17, F19, and F-Final, and during the long-term follow-up at visits LTF1, LTF2, LTF3, LTF4, LTF6, and LTF8. Height will be measured only at Screening and weight will be measured only at Screening and Visit L-Final.

Height (without shoes) will be measured and recorded, rounded to the nearest centimeter. Body weight (without overcoat and shoes) will be measured and recorded, rounded to the nearest kilogram.

The physical examination will include general appearance and bedside examination of the following body systems: Lymph nodes, eyes and ears, mouth and throat, lungs, abdomen, extremities, musculoskeletal system, neurological system, cardiovascular system and skin.

Abnormal physical examination findings will be reported as adverse events.

Abnormalities (e.g., scar at the left side at knee following total knee replacement, or arthropathy of left ankle due to hemophilia B) identified at Screening will be documented in the subject's source documents and on the medical history eCRF. Changes after the Screening Visit will be captured as AEs on the AE eCRF page, as deemed clinically significant in the opinion of the investigator. These abnormalities are to be followed until they reached "final outcome" (please refer to the protocol).

7.7.11 Vital Signs

Blood pressure, pulse, and body temperature will be measured at Screening (Visit S), Visit L-Final, at pre-IMP and post-IMP (3 hours) on Visit D and at all visits during the post-treatment period. Before measurement of blood pressure and pulse, the subject should rest for at least 5 minutes. For the individual subject, all measurements should be performed while the subject is in the same position (i.e., sitting or lying) throughout the trial.

A summary of baseline weight, height, and BMI will be presented by treatment period for the FAS, PP, and safety populations in the demographics table.

Vital signs values will be listed.

8. INTERIM 6 MONTH ANALYSIS

A partial database lock and data extraction was performed once the last subject had achieved 6 months after CSL222 therapy.

The first secondary efficacy endpoint, endogenous FIX activity, was analyzed. This endpoint/analysis was included in (added to) the 18-month-data-cut CSR.

FIX activity was summarized (and listed) by visit, overall and by subject, over the 6-month period since administration of CSL222. By-subject plots of FIX activity over time were overlaid with plots of exogenous FIX consumption (time of administration) and with the times of occurrence of bleeding events over the first 6 months subsequent to CSL222 administration.

The ratio of FIX activity (%) to FIX protein (%) was tabulated. A table was also provided to summarize the FIX activity (%) by subjects with or without pre-existing neutralizing antibodies to FIX. A scatter plot of FIX activity (%) by baseline titer of neutralizing antibodies to AAV5 was also presented to show the correlation of baseline titer of neutralizing antibodies to AAV5 and FIX activity at Month 6.

Bleeding episodes were tabulated by the following bleed types: all bleeds, spontaneous, traumatic, unknown, and medical/dental/other. The estimated ABR will be tabulated descriptively by treatment.

The FIX replacement during the Post-treatment period and the actual exogenous FIX use was tabulated overall and was listed by subject.

The incidence of AEs, SAEs, AEs in descending frequency, related AEs, and related SAEs was tabulated.

Additional figures showing ALT and AST levels (U/L) and corticosteroid use for ALT and AST elevations over time, T-cell (AAV5-capsid) ELISPOT ((SFC)/million PBMCs) over time were produced. Subject disposition and demographic data were listed for all subjects screened and all subjects treated. Baseline characteristics were listed and included the following: hemophilia B

history, joint status at Screening, bleeding history in the year prior to Screening, history of previous FIX replacement therapy use, prior medication/therapy, medical and surgical history, and FIX gene sequencing.

Additional efficacy endpoints were listed and include the following: FIX protein concentration (%).

Safety listings included the following: AEs, SAEs, AEs of special notification, PCS laboratory values, vital signs, other laboratory values, and vector shedding.

9. INTERIM 12 MONTH ANALYSIS

A partial database lock and data extraction was performed once the last subject had achieved 12 months after CSL222 therapy.

The second secondary efficacy endpoint, endogenous FIX activity, was analyzed. This endpoint/analysis was included in (added to) the 18-month-data-cut CSR.

FIX activity was summarized (and listed) by visit, overall and by subject, over the 12-month period since administration of CSL222. By-subject plots of FIX activity over time were overlaid with plots of exogenous FIX consumption (time of administration) and with the times of occurrence of bleeding events over the first 12 months subsequent to CSL222 administration.

The ratio of FIX activity (%) to FIX protein (%) was tabulated. A table was also provided to summarize the FIX activity (%) by subjects with or without pre-existing neutralizing antibodies to FIX. A scatter plot of FIX activity (%) by baseline titer of neutralizing antibodies to AAV5 was also presented to show the correlation of baseline titer of neutralizing antibodies to AAV5 and FIX activity over months 6-12.

Bleeding episodes were tabulated by the following bleed types: all bleeds, spontaneous, traumatic, unknown, and medical/dental/other. The estimated ABR was tabulated descriptively by treatment.

The FIX replacement during the Post-treatment period and the actual exogenous FIX use was tabulated overall and was listed by subject.

The incidence of AEs, SAEs, AEs in descending frequency, related AEs, and related SAEs was tabulated.

Additional figures showing ALT and AST levels (U/L) and corticosteroid use for ALT and AST elevations over time, T-cell (AAV5-capsid) ELISPOT ((SFC)/million PBMCs) over time were produced. Subject disposition and demographic data were listed for all subjects screened and all subjects treated. Baseline characteristics were listed and included the following: hemophilia B history, joint status at Screening, bleeding history in the year prior to Screening, history of previous FIX replacement therapy use, prior medication/therapy, medical and surgical history, and FIX gene sequencing.

Additional efficacy endpoints were listed and included the following: FIX protein concentration (%).

Safety listings included the following: AEs, SAEs, AEs of special notification, PCS laboratory values, vital signs, other laboratory values, and vector shedding.

10. CSR

After 52 weeks following stable FIX expression (18 months Post-treatment), the database was locked, and all available efficacy and safety data collected between Screening and 52 weeks following stable FIX expression (18 months Post-treatment follow-up time) were analyzed and reported in a full CSR. All of the efficacy endpoints were analyzed.

Data up to each analysis time point were considered locked and were not changed (with the exception of ending dates for continuing events and treatments).

Table, listing, and figure shells for the Final CSR were provided in a separate document.

FIX activity was summarized (and listed) by visit, overall and by subject, over the 18-month period since administration of CSL222. By-subject plots of FIX activity over time were overlaid with plots of exogenous FIX consumption (time of administration) and with the times of occurrence of bleeding events over the 52 weeks following stable FIX expression (7-18 months) subsequent to CSL222 administration.

Descriptive statistics was provided for the estimated unadjusted ABR during time periods subsequent to stable FIX expression (6 months) after the dose of CSL222 and during the Lead-in period. The unadjusted ABR was the number of bleeds divided by the person-time at risk during a given time period. For the 52 weeks following stable FIX expression (7-18 months) post-treatment, bleeds and person time on or after Day 1 and prior to stable FIX expression (Month 6) Post-treatment was not included in the calculation.

Bleeding events were listed by subject. The number of bleeding events and the time-at-risk of bleeding events were listed by period for each subject.

The ratio of FIX activity (%) to FIX protein (%) was tabulated. A table was also provided to summarize the FIX activity (%) by subjects with or without pre-existing neutralizing antibodies to FIX. A scatter plot of FIX activity (%) by baseline titer of neutralizing antibodies to AAV5 was also presented to show the correlation of baseline titer of neutralizing antibodies to AAV5 and FIX activity over the 52 weeks following stable FIX expression (7-18 months).

11. CSR ADDENDUM

At the end of the trial, all efficacy and safety data from the long-term follow-up will be analyzed and reported in a CSR addendum. The data will be analyzed cumulatively so that efficacy and safety analyses will cover the entire 5-year Post-treatment period. All of the efficacy endpoints will be analyzed mirroring the methodologies described in Sections 7.5 and 7.6.

Analyses of the secondary FIX activity endpoint will be conducted also for Year 2, Year 3, Year 4, and Year 5 as exploratory analyses.

Analyses of the primary and secondary ABR endpoints will be conducted also for Years 0-2, Years 0-3, Years 0-4, and Years 0-5 as exploratory analyses. Analyses of the annualized consumption and infusion rate of FIX replacement therapy will be conducted also for Years 0-2, Years 0-3, Years 0-4, and Years 0-5 as exploratory analyses.

Analyses of the other secondary efficacy endpoints will be conducted also for Year 2, Year 3, Year 4, and Year 5 as exploratory analyses.

Table, listing, and figure shells for all output including those to be included in the CSR Addendum will be provided in a supplemental document.

12. CHANGES FROM METHODS PLANNED IN THE PROTOCOL

- The full analysis set has become the primary population for the ABR non-inferiority analysis, while the per-protocol population has been relegated to a sensitivity analysis. The reason is that the Food and Drug Administration (FDA) statistical team requested this.

- The contamination period due to exposure to exogenous FIX has been changed to the 5 half-life rule (from a 10-day rule) to allow greater accuracy.

- ABR has now become the sole primary endpoint. The reason is that the FDA statistical and clinical teams requested this.

- Have changed the data cut for the main CSR to be at Month 18 Post-treatment. The reason is that the FDA asked for the efficacy analysis to pertain to the year after a stable FIX activity level is reached.

13. STATISTICAL SOFTWARE

Data processing, statistical Screening, descriptive reporting and analysis of the efficacy and safety data will be performed using SAS (Version 9.4 or higher).

14. REFERENCES

Bretz F, Maurer W, and Maca J “Graphical Approaches to Multiple Testing.” Clinical Trial Biostatistics and Biopharmaceutical Applications, edited by Young W and Chen D, CRC Press, 2015, 349-394.

Detailed guideline on good clinical practice specific to advanced therapy medicinal products, Brussels, 03/12/2009, ENTR/F/2/SF/dn D(2009) 35810.

EQ-5D: https://eq-5dpublications.euroqol.org/details?id=0_78345

EuroQol Group. EQ-5D. Accessed at www.euroqol.org/about-eq-5d/how-to-use-eq-5d.html (November 2014).

Farrar JT, Dworkin RH, Max MB. Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: Making clinical trial data more understandable. J Pain Symptom Manage 2006; 31:369-377.

Guideline on Clinical Follow-Up of Patients Administered with Gene Therapy Medicinal Products, EMA/CHMP/GTWP/60436/2007.

Guilbaud O. Simultaneous confidence regions corresponding to Holm's stepdown procedure and other closed testing procedures. Biometrical Journal, 2008; 50:678-692.

Kessler C, Mariani G. Clinical manifestations and therapy of the haemophilias. In: Colman R, Hirsh J, Marder V, Clowes A, George J, eds. Hemostasis and Thrombosis: Basic Principles and Clinical Practice. 5th ed. Philadelphia: Lippincott-Raven. 2006; P 887-904

Paul J, Venugopal R, et al. Measurement of controlled attenuation parameter: A surrogate marker of hepatic steatosis in patients of nonalcoholic fatty liver disease on lifestyle modification – A prospective follow-up study. Arq. Gastroenterol. 55(1): São Paulo Mar./Apr. 2018.

Pickard S, Law E, Jiang R, Pullenayegum E, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. Value in Health, 2019; 22(8):931–941.

SAS Institute Inc. 2008. SAS/STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Sasso M, Beaugrand M, de Ledinghen V, et al. Controlled attenuation parameter (CAP): a novel VCTE(TM) guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: Preliminary study and validation in a cohort of patients with chronic liver disease from various causes. Ultrasound Med Biol. 2010; 36:1825-35.

Simioni P, Tormene D, Tognin G, Gavasso S, Bulato C, Iacobelli NP, Finn JD, Spiezia L, Radu C, Arruda VR. X-Linked Thrombophilia with a Mutant Factor IX (Factor IX Padua). NEJM 2009.

Statistical Analysis Plan for Study CSL222_3001, Version 1.0, 29 Jan 2019.

Statistical Analysis Plan for Study CSL222_3001, Version 4.0, 10 Jun 2021.

Statistical Analysis Plan for Study CSL222_3001, Version 5.0, 15 Apr 2022.

Strassburger K and Bretz F. Compatible simultaneous lower confidence bounds for the holm procedure and other bonferroni based closed tests. Statistics in Medicine, 2008; 27: 4914-4927.

van Reenan M, Janssen B. EuroQol Group EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument, April 2015

World Federation of Haemophilia. World Federation of Haemophilia report on the annual global survey 2017. dated October 2018. http://www1.wfh.org/publications/files/pdf-1714.pdf date accessed: 23 November 2018

APPENDIX 1: DATA HANDLING RULES

Programming of the tables, listings and figures will be performed using SAS Version 9.4 or a more recent version. The following table presents the algorithms to be used in SAS to calculate the derived variables, including rules for handling other missing data or partial dates, or irregular/unexpected data issues.

Category	Description	Data Handling Rule
1. Age (years)	Age (years)	Age = Year of informed consent – Year of birth
2. Medical History	Medical History Begin Date of Condition	Begin date of condition will be imputed for all subjects as the 1st of the month for the purpose of computing the onset day.
3. Surgical History	Surgical History Date of Surgery	Date of surgery will be imputed for all subjects as the 1st of the month for the purpose of computing the onset day.
4. Treatment Date	date/time of first study treatment	The date and time (24 hr. clock) of the dose of IMP (study treatment) will be taken from the Dosing eCRF. It is not necessary to define a first treatment date for the Lead-in period since the Lead-in treatment is not qualitatively different from pre-study therapies.
5. Last Visit Date	Date of Last Visit	Date of last visit according to the Visit eCRF.
6. Last Study Participation Date (STDM variable, typically named RFPENDTC)	Last Study Participation Date (SDTM variable, RFPENDTC), where SDTM denotes Study Data Tabulation Model	Last study participation date is defined as last known date of contact, which would be the later of the following dates: last visit date, date of last contact if lost-to-follow-up, date of telephone follow-up, or death date.

Category	Description	Data Handling Rule
7. Study Day Definitions	**Study Day** for assessment/event that occurs on or after the beginning of the period.	For the Post-treatment period, Study Day = Date of assessment/event – date of IMP administration + 1. For the Lead-in period (or for overall study day), Study Day is the Date of assessment/event – date of L1 Visit + 1.
	Study Day for assessments/events on days prior to the period	For the Post-treatment period, Study Day = Date of assessment/event – date of IMP administration. For the Lead-in period, Study Day is the Date of assessment/event – date of L1 Visit.
	Dose Day	**Dose Day** in the study is defined as the study day of the trial drug administration (Study Day 1 for the Post-treatment period).
	Last Study Day	**For subjects who did not receive the dose of trial drug**, Last Study Day is defined as (the later of the last visit date and the date of last contact for subjects lost-to-follow-up from the Study Completion/Early Discontinuation CRF) – Date of Screening Visit + 1. **For subjects who received the dose of trial drug,** Last Study Day is defined as (the later of the last visit date and the date of last contact for subjects lost-to-follow-up from the Study Completion/Early Discontinuation CRF) – date of IMP administration + 1.
	Days Since IMP drug administration for event (e.g., Death)	**Days Since IMP drug administration** is defined as date of event – date of IMP drug administration.
8. Duration of event	The duration of any event	The duration of any event is defined as (stop date – start date + 1).

Category	Description	Data Handling Rule
9. Distance between Event	Distance between FIX activity measurement and most recent FIX replacement therapy administration	Date of FIX activity measurement – Date Preceding FIX Replacement Therapy Administration) + 1 The date and time of the FIX activity measurement in question and the FIX replacement therapy administrations respectively are used to find the latest FIX replacement therapy administration preceding the FIX activity measurement in question. In case the dates of the FIX activity measurement in question and a FIX replacement therapy administration are the same and no time is indicated, it is assumed that the FIX replacement therapy administration precedes the FIX activity measurement in question, and the above defined distance therefore becomes equal to 1.
10. Multiple assessments for the same visit	Vital Sign and Laboratory assessments	• All data will be listed in data listings. • The last of multiple valid assessments within a post-baseline study time window will be used for summaries. • If there are multiple laboratory values for the same parameter at post-baseline pre-dose of a visit, the last value will be chosen for analysis.
11. Special Lab Value Handling for Safety Lab values	Lab values with a prefix such as '>', '<', '+' and 'Less than' etc.…	• '>': use the available original value +0.001 in the analyses. • '<': use the available original value –0.001 in the analyses. • '+': use the available original value without the prefix in the analyses. • '>=': use the available original value in the analyses.

Category	Description	Data Handling Rule
		• '<=': use the available original value in the analyses.
12. Prior and concomitant medication	Prior, and Lead-in concomitant, and Post-treatment concomitant medication	**1.** Prior medication/treatment: is any medication/therapy (including herbal treatments, vitamins, non-pharmacological treatment such as psychotherapy as appropriate) received will be considered prior if the start date of the medication/therapy is missing or the medication/therapy start date is before Visit L1 for the Lead-in period. **2.** A medication/therapy will be identified as a "Post-treatment concomitant" medication/therapy if it is being continued by the subject at the date of CSL222 dosing or is any new medication/therapy received during the Post-treatment period. A medication with end date that is the same as the CSL222 dosing date will not be considered to be "Post-treatment concomitant". A medication/therapy will be identified as a "Lead-in" concomitant medication/therapy if it is being continued by the subject at the date of the L1 Visit or is any new medication/therapy received during the Lead-in period prior to the date of CSL222 dosing. The distinction will be made between Lead-in concomitant medications and Post-treatment concomitant medications. **3.** Any medication/therapy that cannot be identified as Prior, Lead-in Concomitant, or Post-treatment Concomitant will be considered as being in each of the possible categories depending on available information. The designation of concomitant medication will be done in a manner that is specific to either the Lead-in period or the Post-treatment

Category	Description	Data Handling Rule
		period. Given that the study treatment is permanent, there cannot be a medication category subsequent to "concomitant" with respect to the Post-treatment period.
13. Adverse event	Missing severity	For the AE summary by severity, an AE with missing severity will be deemed as Severe.
	Missing relationship to study drug	For AE summary by relationship, an AE with a missing relationship to study drug will be deemed as related.
	Treatment-emergent adverse event	An adverse event is considered treatment-emergent for the Post-treatment period if an event occurs (or if there was a worsening [intensity and/or severity changed to worsened grades]) on or after the date of dosing with CSL222. A treatment-emergent adverse can be described as having incidence during the post-treatment period. An adverse event is considered to have had incidence during the Lead-in period if an event occurs (or if there was a worsening [intensity and/or severity changed to worsened grades]) on or after the Visit L1 date and before the date of dosing with CSL222. Prior to the Visit L1 date, adverse events are considered to be a part of the medical history. A death is considered to be treatment-emergent for the Post-treatment period if any of the adverse events that led to the death occurred on or after the date of administration of the IMP. A death is considered to have had incidence during the Lead-in period if any of the adverse events that led to the death occurred on or after the Visit L1 date and before the date of administration of IMP.

Category	Description	Data Handling Rule
		If the AE start date is partial/missing, then • If AE start date is completely missing, then the AE is considered as both treatment-emergent during the post-treatment period and to have had incidence during the Lead-in period. • If both AE start month and day are missing and AE start year is the same or after the IMP dosing year, then the AE is considered as treatment-emergent for the post-treatment period. If both AE start month and day are missing and AE start year is the same or after the L1 Visit year and on or before the IMP dosing year, then the AE is considered as having had incidence during the Lead-in period. • If AE start day is missing and AE start year and month are the same or after the IMP dosing year and month, then the AE is considered as treatment-emergent for the Post-treatment period. If AE start day is missing and AE start "year and month" are the same or after the L1 Visit "year and month" and on or before the IMP dosing "year and month", then the AE is considered as having had incidence during the Lead-in period. Missing/incomplete (partial) AE start and end dates will not be imputed for data listings.
14. Hard coding	Hard coding for data analysis	Hard Coding is not allowed during data analysis unless agreed to in writing by CSL.
15. QoL/PRO data	Bleed event	Assessments within two weeks of a bleed event will not be included in any analysis.
16. Listing outputs	Data excluded	All data not used for efficacy analysis will be flagged in listings.

Category	Description	Data Handling Rule
17. Contamination due to exogenous FIX (infusion) use	Contamination of FIX activity or protein assessment (or in some cases bleeding assessment) due to exogenous FIX (infusion) use	The date/time (where available) – rather than just date – for the time of the exogenous FIX infusion and the time of the blood draw for FIX activity (or protein) assessment will be used for the determination of contamination. The use of date/time (instead of just date) should be applied for the 5-half-life contamination rule. If only the date – but not the time – of the exogenous FIX infusion is known, then the contamination period will (conservatively) be the time period beginning on midnight at the beginning of that day and ending at the time which is 24 hours plus five half-lives later. If only the date but not the time of the FIX activity assessment is known and if any point in time on that date overlaps with the contamination period, then the activity assessment will be deemed contaminated. As alluded to in the SAP text section about secondary efficacy for FIX activity, the "5 half-life" contamination period is actually being applied to the 12-month-data-cut analysis and the 18-month-data-cut analysis (refer to the 18-month SAP for Study CSL222_3001, Version 4.0, 10 Jun 2021), and later analyses.

APPENDIX 2: ANALYSIS DATASET SPECIFICATIONS

Analysis datasets will be built to gain efficiency and ensure consistency in data analyses and presentation for this trial. The specifications for each analysis data set will be prepared separately and will not be a part of this SAP.

APPENDIX 3: CENTRAL LABORATORY REFERENCE RANGES FOR USE IN FLAGGING ABNORMAL VALUES

This appendix is provided as an attachment to this document.

APPENDIX 4: EQ-5D-5L US VALUE SET

This appendix is provided as an attachment to this document.

APPENDIX 5: SAS CODE FOR STATISTICAL ANALYSES

The prototype SAS Code will be in a separate document.

DocuSign

Certificate Of Completion

PPD

Status: Completed

Subject: Complete with Docusign: CSL222_3001 SAP Text v6.0_20241004.docx

Source Envelope:

Document Pages: 88 | Signatures: 2 | Envelope Originator:

Certificate Pages: 5 | Initials: 0

PPD

AutoNav: Enabled

,

EnvelopeId Stamping: Enabled

,, , ,

Time Zone: (UTC-08:00) Pacific Time (US & Canada)

PPD

IP Address: 108.24.139.61

Record Tracking

Status: Original
10/8/2024 1:03:42 PM

Holder: PPD
PPD

Location: DocuSign

Signer Events	Signature	Timestamp
PPD Security Level: Email, Account Authentication (Required)	PPD Signature Adoption: Pre-selected Style Signature ID: PPD Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I have reviewed this document	Sent: 10/8/2024 1:09:50 PM Viewed: 10/9/2024 8:22:55 AM Signed: 10/9/2024 8:27:17 AM
Electronic Record and Signature Disclosure: Accepted: 9/29/2021 11:37:01 AM ID: PPD		
PPD Security Level: Email, Account Authentication (Required)	PPD Signature Adoption: Pre-selected Style Signature ID: PPD Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 10/8/2024 1:09:49 PM Viewed: 10/10/2024 6:05:09 AM Signed: 10/10/2024 6:06:09 AM
Electronic Record and Signature Disclosure: Accepted: 10/1/2020 6:47:33 AM ID: PPD		

In Person Signer Events	Signature	Timestamp

Editor Delivery Events	Status	Timestamp

Agent Delivery Events	Status	Timestamp

Intermediary Delivery Events	Status	Timestamp

Certified Delivery Events	Status	Timestamp

Carbon Copy Events	Status	Timestamp

Witness Events	Signature	Timestamp

Notary Events	Signature	Timestamp

Envelope Summary Events	Status	Timestamps
Envelope Sent	Hashed/Encrypted	10/8/2024 1:09:50 PM
Certified Delivered	Security Checked	10/10/2024 6:05:09 AM
Signing Complete	Security Checked	10/10/2024 6:06:09 AM
Completed	Security Checked	10/10/2024 6:06:09 AM

Payment Events	Status	Timestamps

Electronic Record and Signature Disclosure

ELECTRONIC RECORD AND SIGNATURE DISCLOSURE

From time to time, CSL QA Release (Part 11) (we, us or Company) may be required by law to provide to you certain written notices or disclosures. Described below are the terms and conditions for providing to you such notices and disclosures electronically through the DocuSign system. Please read the information below carefully and thoroughly, and if you can access this information electronically to your satisfaction and agree to this Electronic Record and Signature Disclosure (ERSD), please confirm your agreement by selecting the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

Getting paper copies

At any time, you may request from us a paper copy of any record provided or made available electronically to you by us. You will have the ability to download and print documents we send to you through the DocuSign system during and immediately after the signing session and, if you elect to create a DocuSign account, you may access the documents for a limited period of time (usually 30 days) after such documents are first sent to you. After such time, if you wish for us to send you paper copies of any such documents from our office to you, you will be charged a $0.00 per-page fee. You may request delivery of such paper copies from us by following the procedure described below.

Withdrawing your consent

If you decide to receive notices and disclosures from us electronically, you may at any time change your mind and tell us that thereafter you want to receive required notices and disclosures only in paper format. How you must inform us of your decision to receive future notices and disclosure in paper format and withdraw your consent to receive notices and disclosures electronically is described below.

Consequences of changing your mind

If you elect to receive required notices and disclosures only in paper format, it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices or disclosures. Further, you will no longer be able to use the DocuSign system to receive required notices and consents electronically from us or to sign electronically documents from us.

All notices and disclosures will be sent to you electronically

Unless you tell us otherwise in accordance with the procedures described herein, we will provide electronically to you through the DocuSign system all required notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you during the course of our relationship with you. To reduce the chance of you inadvertently not receiving any notice or disclosure, we prefer to provide all of the required notices and disclosures to you by the same method and to the same address that you have given us. Thus, you can receive all the disclosures and notices electronically or in paper format through the paper mail delivery system. If you do not agree with this process, please let us know as described below. Please also see the paragraph immediately above that describes the consequences of your electing not to receive delivery of the notices and disclosures electronically from us.

How to contact CSL QA Release (Part 11):

You may contact us to let us know of your changes as to how we may contact you electronically, to request paper copies of certain information from us, and to withdraw your prior consent to receive notices and disclosures electronically as follows:
To contact us by email send messages to: silvia.hausmann@cslbehring.com

To advise CSL QA Release (Part 11) of your new email address

To let us know of a change in your email address where we should send notices and disclosures electronically to you, you must send an email message to us at silvia.hausmann@cslbehring.com and in the body of such request you must state: your previous email address, your new email address. We do not require any other information from you to change your email address.

If you created a DocuSign account, you may update it with your new email address through your account preferences.

To request paper copies from CSL QA Release (Part 11)

To request delivery from us of paper copies of the notices and disclosures previously provided by us to you electronically, you must send us an email to silvia.hausmann@cslbehring.com and in the body of such request you must state your email address, full name, mailing address, and telephone number. We will bill you for any fees at that time, if any.

To withdraw your consent with CSL QA Release (Part 11)

To inform us that you no longer wish to receive future notices and disclosures in electronic format you may:

i. decline to sign a document from within your signing session, and on the subsequent page, select the check-box indicating you wish to withdraw your consent, or you may;

ii. send us an email to silvia.hausmann@cslbehring.com and in the body of such request you must state your email, full name, mailing address, and telephone number. We do not need any other information from you to withdraw consent.. The consequences of your withdrawing consent for online documents will be that transactions may take a longer time to process..

Required hardware and software

The minimum system requirements for using the DocuSign system may change over time. The current system requirements are found here: https://support.docusign.com/guides/signer-guide-signing-system-requirements.

Acknowledging your access and consent to receive and sign documents electronically

To confirm to us that you can access this information electronically, which will be similar to other electronic notices and disclosures that we will provide to you, please confirm that you have read this ERSD, and (i) that you are able to print on paper or electronically save this ERSD for your future reference and access; or (ii) that you are able to email this ERSD to an email address where you will be able to print on paper or save it for your future reference and access. Further, if you consent to receiving notices and disclosures exclusively in electronic format as described herein, then select the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

By selecting the check-box next to 'I agree to use electronic records and signatures', you confirm that:

- You can access and read this Electronic Record and Signature Disclosure; and
- You can print on paper this Electronic Record and Signature Disclosure, or save or send this Electronic Record and Disclosure to a location where you can print it, for future reference and access; and
- Until or unless you notify CSL QA Release (Part 11) as described above, you consent to receive exclusively through electronic means all notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you by CSL QA Release (Part 11) during the course of your relationship with CSL QA Release (Part 11).