Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02141074
Sponsor trial ID:	NN7999-3895
Official title of study:	An Open-label Single-arm Multicentre Non-controlled Phase 3 a Trial Investigating Safety and Efficacy of Nonacog Beta Pegol (N9-GP) in Prophylaxis and Treatment of Bleeding Episodes in Previously Untreated Patients With Haemophilia B (FIX Activity Below or Equal to 2 Percent)
Document date*	15 July 2019

*Document date refers to the date on which the document was most recently updated.
Note: The date in the header of Page 2 is the date of compilation of the documents and not of an update to content.

16.1.9 Documentation of statistical methods

List of contents

Redacted statistical analysis plan
Includes redaction of personal identifiable information only.

.

[redacted]

Statistical Analysis Plan

Safety and Efficacy of nonacog beta pegol (N9-GP) in Previously Untreated Patients with Haemophilia B

Protocol Number: NN7999-3895

Sponsor: Novo Nordisk

Address: [redacted]

SAP Version: 2.0
SAP Date: 15 Jul 2019
Protocol Version: 5.30 Date: 25 June 2018

[redacted]

1. Note

[redacted] has prepared a Statistical Analysis Plan (SAP) for the Sponsor to review and sign-off for study NN7999-3895. Analyses will be conducted after this document has been finalized and officially signed. Anything in italics is taken directly from the protocol. For more details, please refer to the study protocol.

2. Version History

Version Number	Date	Reason for Revision
1.0	22 April 2019	Initial Version
2.0	15 July 2019	Revised based on sponsor requests

[redacted]

3. Signature Page for SAP Approval

The following signatures indicate the approval of the statistical analysis plan for the study NN7999-3895.

[redacted]

Name (print): [redacted]

Position: [redacted]

Novo Nordisk

Name (print): [redacted]

Position: [redacted]

Date (DD-MMM-YYYY): 16-Jul-2019

[redacted]

4. Table of Contents

[redacted]

5. Abbreviations

Abbreviation	Description
ABAS-3	Adaptive Behavior Assessment System – Third Edition
BASC-3	Behavior Assessment System for Children – Third Edition
BRIEF2	Behavior Rating Inventory of Executive Function – Second Edition
BRIEF-A	Behavior Rating Inventory of Executive Function – Adult Edition
BRIEF-P	Behavior Rating Inventory of Executive Function – Preschool Edition
CI	Confidence Interval
DET	Detection Test
ED	Exposure Day (Definition: One exposure day is defined as each day when a patient is administered coagulation factor IX/blood components for any reason regardless number of doses)
FAS	Full Analysis Set
HNAS	Hemophilia Normative Analysis Set
IDN	Identification Test
IQR	Interquartile range (25th, 75th percentiles)
N9-GP	nonacog beta pegol, glycopegylated recombinant coagulation factor IX
ONB	One Back Test
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SD	Standard Deviation
SDH	Structured Development History
TLF	Tables, Listings, and Figures
WASI-II	Wechsler Abbreviated Scale of Intelligence – Second Edition
WPPSI-IV	Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition
WSD	Within-Subject Deviation

[redacted]

6. Introduction

This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the NN7999-3895 Study Protocol related to neurocognitive assessments and includes detailed procedures for executing the statistical analysis of the neurocognitive assessment data.

The SAP is finalized and signed prior to database lock. For operational efficiency, an earlier time is usually targeted. If needed, revisions to the approved SAP may be made prior to database lock. Revisions will be version controlled.

Any changes from the analyses planned in the SAP will be justified in the statistical report.

Prior to database lock, a dry run analysis will occur.

7. Study Objectives Related to Neurocognitive Assessments

To evaluate neurodevelopmental, neurocognitive and neurobehavioral outcomes on N9-GP exposed patients.

8. Study Design

8.1. General Description

Please see Sections 5, 8.1, and 8.2 of the study protocol for a general description of the study design.

8.2. Scheduled Visits

Table 1: Visit Schedule for Neurocognitive Assessments

Visit Number	Scheduled Interval	Analysis Visit
1, 20, 23, 26, 28-50	• BRIEF: Every 6 months (all countries) • [redacted] Battery: Every 6 months (all countries) • SDH: Every 6 months (all countries) • HH: Every 6 months (all countries) • Bayley-III / WPPSI-IV / WASI-II: Every 1 year (English-speaking countries only) • ABAS-3: Every 1 year (English-speaking countries only) • BASC: Every 1 year (English-speaking countries only)	Yes
End of Trial		Yes
Unscheduled Visit		No

Study visits will be renamed such that the first visit is “Initial Assessment”, the next is “Assessment 2”, the following is “Assessment 3”, and so on, to account for the fact that not all subjects will be starting their neurocognitive assessments

at the same visit. If patients start on pre-prophylaxis and switch to prophylaxis at Visit 20, the assessments will be displayed under “Prophylaxis” in the tables.

The patient’s age when the Structured Developmental History (SDH) is completed will be used as the Visit Age for each visit.

9. Analysis Set and Population

9.1. Unblinding Procedure

The unblinding procedure was not described in protocol for study NN7999-3895 because an open-label design was used. Once all data discrepancies within the database are resolved with the clinical research units, the database is locked.

9.2. Analysis Set

9.2.1. Full Analysis Set (FAS)

All main descriptions and analysis of efficacy will be based on the Full Analysis Set (FAS), as defined in ICH E9 guidelines. The FAS includes all patients exposed to N9-GP.

9.3. Sample Size and Analysis Population

60 patients are planned to be screened and 40 patients are expected to complete the trial.

For details, please see protocol Section 16.1.

9.4. Handling of missing data

No imputations will be performed in the event of missing data due to dropouts or omitted visits. All incomplete subject profiles will be retained in the analysis. In view of issues of reliability, all analyses will be conducted with completion failures removed.

10. Analysis Endpoints, Initial Assessment Definition, and Outcome Measures

10.1. Analysis Endpoints

The analysis endpoints are z-scores and change in z-scores for each neurocognitive assessment (see Table 2 for more details) against eTHINK population and US normative population (international normative population for [redacted] computerized tests), as well as responses on the Structural Development History and Hemophilia History.

10.2. Initial Assessment Definition

[redacted]

Initial Assessment data is the first assessment that has cognitive data for each subject. This first set of cognitive data may occur at different visits depending on the subject.

10.3. Definition of Findings – Initial and Follow-Up

Based on the results from the eTHINK trial (all age groups pooled) the mean and standard deviation (SD) will be calculated for composite scores of all tests (see Table 2 for primary outcome measures). Each patient's individual results from each test (initial and follow-up assessments) from this trial will be compared to the mean and SD from the respective test from the eTHINK trial. If the patient's results are greater than 1 SD above the eTHINK mean (or less than -1 SD below the eTHINK mean), this will be treated as potential increase (or decrease) of the respective neurocognitive function and will be discussed with experts at the external expert review panel (EERP). If the absolute value of change from the eTHINK mean is ≤ 1 SD, the change will be categorized as "normal variation" rather than signal. If less than three different test results from the same visit are outside the defined threshold of 1 SD, the result will also be categorized as "normal variation", but still discussed with experts during the EERP.

10.4. Outcome Measures Related to Neurocognitive Assessment Analysis

The cognitive tests administered include computerized tests from the [redacted] Battery and traditional neurocognitive instruments.

Bayley Scales of Infant and Toddler Development – Third Edition (Bayley-III; Neurodevelopment)

Bayley Scales of Infant and Toddler Development – Third Edition (Bayley-III) is a revision of the Second Edition (BSID-II), but while the prior provided mental, motor and behavioral scales, the psychologist-administered Bayley-III provides cognitive, language, motor, social-emotional, and adaptive behavioral scales. To avoid duplication in domains, the adaptive behavioral scale of Bayley-III will not be performed as it is identical to the Adaptive Behavioral Assessment System, 2nd edition (ABAS-2), and ABAS-3 is already being used in this study.

- Bayley III (ages 1:0 to 3:6 years), excluding ABAS-2 domain test

Primary scales of intelligence have continued to be an integral part of assessment of neurocognitive function. Assessment of intelligence across multiple domains is different in the pre-school setting from that of older children and adults.

Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV; General Intelligence)

The Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV) is the latest edition of the individually administered early childhood intelligence test with origins dating back to the original 1967 Wechsler Preschool and Primary Scale of Intelligence. The WPPSI-IV is divided into two distinct age bands (2 years 6 months to 3 years 11 months and 4 years 0 months to developmental changes during the age range covered. Only the latter will be used in this study due 7 years 7 months) corresponding to different subtest batteries due to significant cognitive ability and to overlap with Bayley-III in infants. Six of the WPPSI-IV subtests will be completed in this study: 3 subtests to align to WASI-II (Block Design, Matrix Reasoning, Similarities) and 5 partially overlapping subtests required to generate a composite or full scale IQ (Matrix Reasoning, Information, Similarities, Picture Memory, Bug Search); Matrix Reasoning also provides a proxy for non-verbal IQ.

- WPPSI-IV 4:0-7:7 (ages 4:0-6:11 years)
  - Block Design
  - Matrix Reasoning
  - Similarities
  - Information

- Picture Memory
- Bug Search

Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II; General Intelligence)

The Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II) is an abbreviated measure of intelligence designed for individuals 6 to 90 years of age. Primarily used in clinical, psychoeducational, and research settings, the WASI-II was developed to quickly and accurately estimate intelligence when administration of a full battery is not feasible or necessary. The WASI-II is classified as a Level C measure, and can be administered and interpreted by individuals with a doctorate degree in psychology or a related discipline. Administration time is approximately 30 min for the four-subtest form.

- WASI-II (ages 7-21)

Behavior Assessment System for Children (BASC; Behavior/Emotion)

Behavioral assessment is multidimensional, including adaptive and problem behavior to identify emotional and behavioral problems. It typically includes a combination of assessment in various or school setting, and through self-report of older children and adolescents to identify their thoughts settings, through parents to assess the home or community, through teachers to assess the preschool and feelings. The Behavioral Assessment System for Children (BASC) was developed in 1992 as a multidimensional assessment including Structured Developmental History (SDH), Parent Rating Scale (PRS), Teacher Rating Scale (TRS), Self-Report of Personality (SRP) and Student Observation System (SOS), which can be used separately or in combination. The current third edition, BASC-3, provides a comprehensive set of rating scales that measure areas important for IDEA and DSM-5 classifications. In this cross-sectional study, the parent rating scale (PRS) and self-report of personality (SRP) will be utilized.

- BASC3 Parent response scale Preschool (PRS-P 2-5y) (ages 24 months-5 years)
- BASC3 Parent response scale Child (PRS-C 6-11y) (ages 6-11 years)
- BASC3 Parent response scale Adolescent (PRS-A 12-21y) (ages 12-21)
- BASC3 Self-Reported Personality Child (SRP-C 8-11y) (ages 8-11)
- BASC3 Self-Reported Personality Adolescent (SRP-A 12-21y) (ages 12-17)
- BASC3 Self-Reported Personality College (SRP-College 18-25y) (age 18-21)

Adaptive Behavior System – Third Edition (ABAS-3; Adaptive Behavior)

Adaptive behavior refers to “the level of everyday performance of tasks that is required for a person to fulfill typical roles in society, including maintaining independence and meeting cultural expectations regarding personal and social responsibility”. Measurements of adaptive behavior have been utilized by school psychologists and other mental health professionals as part of assessment/diagnosis, evaluation of behavioral conditions, and one of five domains that are a basis for eligibility for special education, is recommended in guidelines by the American Association on Intellectual and Developmental Disabilities6 and by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Measures of adaptive behavior are impaired in autism spectrum disorders8 and attention deficit hyperactivity disorder. Adaptive behavior scales typically use semi-structured interview forms and are typically completed by parents, teachers, and other caregivers familiar with the child or adolescent and their behavior.

Evidence based reviews suggest that adaptive behavioral assessments should be part of a comprehensive assessment of children and adolescents. In this study, the Adaptive Behavior Assessment System – Third Edition (ABAS-3) scale focused on home and community setting will be used.

- ABAS-3 Parent 0-5y (ages 12 months to 5 years)
- ABAS-3 Parent 5-21y (ages 6-21)

[redacted]

- ABAS-3 Adult 16-89y (ages 16-21)

Behavior Rating Inventory of Executive Function (BRIEF; Executive Function)
Self-regulation skills have pervasive and long-lasting implications for the development of children's academic abilities and psychopathology, and is most associated with the construct of executive function. Executive Function is characterized by three intercorrelated factors: inhibitory control, working memory, and shifting. Inhibitory control (IC) is defined as the ability to inhibit a prepotent response in favor of a subordinate response. Working memory (WM) is defined as the ability to maintain, update, and manipulate information within memory. Shifting (SH) is defined as the ability to switch attention between mental sets or tasks, or the ability to engage and disengage with specific aspects within tasks. Rating scales involving either self-report or parent and teacher reports of EF skills in everyday situations have been developed. The Behavior Rating Inventory of Executive Function (BRIEF) is a parent- or teacher-rating scale of everyday EF behaviors in school-age children. The second edition BRIEF2 was designed and modified to be used in conjunction with performance-based measures to allow for a comprehensive understanding of children's EF skills. Modified from BRIEF to address preschool issues, Behavior Rating Inventory of Executive Function – Preschool version (BRIEF-P), is intended to be completed by teachers or parents of 2- to 6-year old children, with ratings based on the child's observed behaviors within the relevant home or school environments. Behavior Rating Inventory of Executive Function – Adult version (BRIEF-A) is modified from BRIEF to address issues in adults aged 18-90 years old. In this study, parent ratings will be captured for ages 2-18 (BRIEF-P and BRIEF2) and self-report of adolescents and adults (ages 11-21, BRIEF2 and BRIEF-A).

- BRIEF-P (Preschool) Parent 2:0-5:11y (ages 24 months to 5 years)
- BRIEF2 Parent 5-18y (ages 6-18 years)
- BRIEF2 Self Report 11-18y (ages 11-18 years)
- BRIEF-A Self Report 18-90y (ages 18-21)

[redacted] Custom Battery (DET, IDN, ONB; Attention/Processing Speed)
Clinical assessment of attention usually depends on three sources of information: (1) psychometric tests designed to measure other cognitive functions, (2) specific neuropsychological tests of attention; and direct behavioral observation and measurement. Assessment often requires the clinician to obtain information about performance under different conditions. Measures of attention include attention span, divided attention, switching, response intention and planning, sustained performance and vigilance, and information-processing speed.

Information processing speed along with working memory have been identified as measures of cognitive impairment. While traditionally clinicians assessed with the Paced Auditory Serial Addition Test (PASAT), over the past decade the advent of computerized tests of information processing have been relied on to assess processing speed.
The [redacted] computerized cognitive test battery proposed for this study includes an assessment of processing speed, attention/vigilance, and working memory.

- [redacted] Custom Battery (Peds platform, ages 4-9; Adult platform, ages 10-21)
  - Detection test (processing speed, psychomotor function)
  - Identification test (attention/vigilance)
  - One Back Test (working memory)

Note: Ages 4-5 years old will be administered with pediatric Detection and Identification. Ages 6-9 years old will be administered with pediatric Detection, Identification, and One Back. Ages 10-21 will be administered with adult Detection, Identification, and One Back.

[redacted]

Although each of the [redacted] computerized cognitive tests yields multiple outcome measures, research by [redacted] has identified a set of measures that are optimal for the detection of cognitive change in clinical trials at both the group and individual level (Falleti et al., 2006; Maruff et al., 2009; Bland & Altman, 1996a; Bland & Altman, 1996b).

For each [redacted] computerized cognitive test, a single primary outcome measure was selected prior to data analysis from each test in the battery to minimize experiment-wise error rates. Each primary outcome measure was selected because it has been shown to be optimal for the detection of change because:

a) It is drawn from a data distribution that contains only a small probability of floor or ceiling effects and no restriction in the range of possible performance values (Falleti et al., 2006; Bland & Altman, 1996a; Bland & Altman, 1996b).
b) It is drawn from a normal distribution or a distribution which can be corrected to normal through the use of appropriate mathematical transformation (e.g., logarithmic base 10, or arcsine) (Falleti et al., 2006; Bland & Altman, 1996a; Bland & Altman, 1996b).

Table 2 below summarizes the outcome measures for the neurocognitive assessments with the instruments from which they were derived.

Table 2: Neurocognitive Instruments Administered in Study NN7999-3895 and their Primary, Secondary, and Exploratory Outcome Measures

Primary Outcome Measure	Secondary Outcome Measure	Exploratory Outcomes
BAYLEY-III[1]		
Cognition Composite **Communication Composite** **Motor Composite** **Social-Emotional Composite**	Receptive communication Expressive communication Fine Motor Gross Motor	
WPPSI-IV[1]		
Full-Scale IQ **Verbal Comprehension Index**	Block Design subtest Matrix Reasoning subtest Similarities subtest Information subtest Picture Memory subtest Bug Search subtest	
WASI-II[1]		
Full-Scale IQ **Verbal Comprehension Index** **Perceptual Reasoning Index**	Vocabulary subtest Similarities subtest Matrix Reasoning subtest Block Design subtest	
BRIEF-P[2]		
Global Executive Composite Index **Inhibitory Self-Control Index** **Flexibility Index** **Emergent Metacognition Index**	Inhibit Scale Emotional Control Scale Shift Scale Working Memory Scale Plan/Organize Scale	
BRIEF-2 Parent Report[2]		

[redacted]

Global Executive Composite Index **Behavioral Regulation Index** **Emotion Regulation Index** **Cognitive Regulation Index**	Inhibit Scale Self-Monitor Scale Emotional Control Scale Shift Scale Initiate Scale Working Memory Scale Plan/Organization Scale Organization of Materials Scale Task Monitor Scale	
BRIEF-2 Self Report[2]		
Global Executive Composite Index **Behavioral Regulation Index** **Emotion Regulation Index** **Cognitive Regulation Index**	Inhibit Scale Self-Monitor Scale Emotional Control Scale Shift Scale Working Memory Scale Plan/Organization Scale Task Completion Scale	
BRIEF-A[2]		
Global Executive Composite Index **Behavioral Regulation Index** **Metacognitive Index**	Inhibit Scale Self-Monitor Scale Emotional Control Scale Shift Scale Initiate Scale Working Memory Scale Plan/Organization Scale Organization of Materials Scale Task Monitor Scale	
BASC3 Parent Rating Scale Preschool (PRS-P)		
Externalizing Problems Composite[2] **Internalizing Problems Composite**[2] ***Behavioral Symptoms Index***[2]	Hyperactivity Domain[2] Aggression Domain[2] Anxiety Domain[2] Depression Domain[2] Somatization Domain[2] Attention Problems Domain[2]	Atypicality Domain[2] Withdrawal Domain[2] Adaptive Skills Composite[1] Adaptability Domain[1] Social Skills Domain[1] Activities of Daily Living Domain [1] Functional Communication Domain[1]
BASC3 Parent Rating Scale Child (PRS-C)		
Externalizing Problems Composite[2] **Internalizing Problems Composite**[2] ***Behavioral Symptoms Index***[2]	Hyperactivity Domain[2] Aggression Domain[2] Conduct Problems Domain[2] Anxiety Domain[2] Depression Domain[2] Somatization Domain[2] Attention Problems Domain[2]	Atypicality Domain[2] Withdrawal Domain[2] Adaptive Skills Composite[1] Adaptability Domain[1] Social Skills Domain[1] Activities of Daily Living Domain[1] Functional Communication Domain[1] Leadership Domain[1]
BASC3 Parent Rating Scale Adolescent (PRS-A)		
Externalizing Problems Composite[2] **Internalizing Problems Composite**[2] ***Behavioral Symptoms Index***[2]	Hyperactivity Domain[2] Aggression Domain[2] Conduct Problems Domain[2] Anxiety Domain[2] Depression Domain[2] Somatization Domain[2] Attention Problems Domain[2]	Atypicality Domain[2] Withdrawal Domain[2] Adaptive Skills Composite[1] Adaptability Domain[1] Social Skills Domain[1] Activities of Daily Living Domain[1] Functional Communication Domain[1] Leadership Domain[1]

[redacted]

BASC3 Self Rating Scale Child (SRP-C)		
School Problems Composite[2] **Internalizing Problems Composite**[2] **Inattention/Hyperactivity Composite**[2] **Emotional Symptoms Index**[2]	Attitude to School Domain[2] Attitude to Teachers Domain[2] Locus of Control Domain[2] Social Stress Domain[2] Anxiety Domain[2] Depression Domain[2] Sense of Inadequacy Domain[2]	Atypicality Domain[2] Personal Adjustment Composite[1] Relations with Parents Domain[1] Interpersonal Relationships Domain[1] Self-Esteem Domain[1] Self-Reliance Domain[1]
BASC3 Self Rating Scale Adolescent (SRP-A)		
School Problems Composite[2] **Internalizing Problems Composite**[2] **Inattention/Hyperactivity Composite**[2] **Emotional Symptoms Index**[2]	Attitude to School Domain[2] Attitude to Teachers Domain[2] Sensation Seeking Domain[2] Locus of Control Domain[2] Social Stress Domain[2] Anxiety Domain[2] Depression Domain[2] Somatization Domain[2] Sense of Inadequacy Domain[2] Hyperactivity Domain[2] Attention Problems Domain[2]	Atypicality Domain[2] Personal Adjustment Composite[1] Relations with Parents Domain[1] Interpersonal Relationships Domain[1] Self-Esteem Domain[1] Self-Reliance Domain[1]
BASC3 Self Rating Scale College (SRP-COL)		
Internalizing Problems Composite[2] **Inattention/Hyperactivity Composite**[2] **Emotional Symptoms Index**[2]	Locus of Control Domain[2] Social Stress Domain[2] Anxiety Domain[2] Depression Domain[2] Sense of Inadequacy Domain[2] Somatization Domain[2] Attention Problems Domain[2] Hyperactivity Domain[2] Sensation Seeking Domain[2] Alcohol Abuse Domain[2] School Maladjustment Domain[2]	Atypicality Domain[2] Personal Adjustment Composite[1] Relations with Parents Domain[1] Interpersonal Relationships Domain[1] Self-Esteem Domain[1] Self-Reliance Domain[1]
ABAS Parent 0-5 year old[1]		
Conceptual Reasoning Domain **Social Interactions Domain** **Practical Functioning Domain** ***Adaptive Behavior Domain Score***	Communication Skill Area Functional Pre-Academics Skill Area Self-Direction Skill Area Leisure Skill Area Social Skill Area Community Use Skill Area Home Living Skill Area Health and Safety Skill Area Self-Care Skill Area Motor Skill Area	
ABAS Parent 5-21 year old[1]		
Conceptual Reasoning Domain **Social Interactions Domain** **Practical Functioning Domain** ***Adaptive Behavior Domain Score***	Communication Skill Area Functional Academics Skill Area Self-Direction Skill Area Leisure Skill Area Social Skill Area Community Use Skill Area Home Living Skill Area Health and Safety Skill Area Self-Care Skill Area	

[redacted]

ABAS Adult 16-89 year old[1]		
Conceptual Reasoning Domain **Social Interactions Domain** **Practical Functioning Domain** ***Adaptive Behavior Domain Score***	Communication Skill Area Functional Academics Skill Area Self-Direction Skill Area Leisure Skill Area Social Skill Area Community Use Skill Area Home Living Skill Area Health and Safety Skill Area Self-Care Skill Area	
	[redacted] **Pediatrics 4-5 year old**	
Detection[2] **Identification[2]** **Attention Domain Composite[1]**		
	[redacted] **Pediatric 6-9 year old**	
Detection[2] **Identification[2]** **One Back (speed)[2]** **One Back (accuracy)[1]** **Attention Domain Composite[1]**		
	[redacted] **Adult 10-21 year old**	
Detection[2] **Identification[2]** **One Back (speed)[2]** **One Back (accuracy)[1]** **Attention Domain Composite[1]**		

[1] Higher scores indicate better performance (multiplicand = +1).
[2] Lower scores indicate better performance (multiplicand = -1).

10.5. Other Outcome Measures

The following composite scores will be computed and analyzed:

- Attention Domain Composite
  - Peds Age 4-5: DET, IDN speed scores must exist
  - Peds Age 6-9: DET, IDN, ONB speed scores; at least two of three scores must exist
  - Ages 10-21: DET, IDN, ONB speed scores; at least two of three scores must exist

The composite scores will be calculated by:

1) Standardizing each of the test scores included in the cognitive domains against age-matched US normative population (international normative population for [redacted] computerized tests) data for that test (Section 10.6.1 specifies the procedure for the standardization of the tests for the purposes of creating the Attention Domain Composite score).
2) Computing the average of the available standardized test scores if the required number of scores available is met (as described above).

The Attention Domain Composite score is normalized based on age and therefore serves as the age-adjusted z-score in summary tables without any further calculations.

[redacted]

10.6. Derived Variables

10.6.1. Standardized Scores: Comparison of Individual Scores to Age-Matched Normative Data for [redacted] Computerized Cognitive Tests (Age-Adjusted Z-Scores)

A comparison of individual scores to age-matched normative data will also be conducted. The process will be as follows:

- Performance on each neurocognitive outcome measure will be standardized relative to age-matched normative data (i.e., the score will be converted to a z-score by subtracting the age matched mean from the normative sample and dividing by the age-matched standard deviation (SD) from the same normative sample)
- The multiplicand equals 1 for outcome measures for which a higher score is indicative of better cognitive performance and -1 for outcome measures where a lower score is indicative of better cognitive performance. Please see Table 2 for details regarding the multiplicand for each outcome measure.
- The z-score will be calculated as follows:

$$z - Score\ (z_{ijt}) = \frac{(x_{ijt} - \bar{x}_t)}{SD_t} * Multiplicand$$

Where:

t = is the outcome measure indicator

i = indexes subject *i*

j = indexes the *jth* assessment for subject *i*

x = cognitive score

$\bar{x}_t$ = mean performance score of the age-matched normative sample for outcome measure t

SD_t = Standard Deviation of the age-matched normative sample for outcome measure t

10.6.1.1. Age-Matched Normative Data for [redacted] Computerized Cognitive Tests

Note: These tables are stored in XLS format in nD ID: 004758297.

Table 3. Detection

Age (Years)	Mean	SD
4	2.82934	0.12057
5	2.77951	0.11345
6	2.74635	0.09264
7	2.66173	0.09901
8	2.63129	0.13746
9	2.58581	0.13506
10	2.56241	0.08007
11	2.53830	0.07896
12	2.52938	0.07682
13	2.51597	0.07535

[redacted]

14	2.51298	0.08101
15	2.50910	0.07674
16	2.50538	0.07644
17	2.50197	0.07528
18-34	2.45444	0.07337

Table 4. Identification

Age (Years)	Mean	SD
4	3.01615	0.07030
5	2.94433	0.11321
6	2.94156	0.06801
7	2.87557	0.07695
8	2.84589	0.11548
9	2.80141	0.09222
10	2.76101	0.07096
11	2.73223	0.07068
12	2.71368	0.06719
13	2.69591	0.06782
14	2.68636	0.06954
15	2.67744	0.06687
16	2.67182	0.06561
17	2.66634	0.06456
18-34	2.64200	0.06351

Table 5. One Back – Speed

Age (Years)	Mean	SD
6	3.17567	0.06515
7	3.07410	0.11496
8	3.03745	0.11140
9	3.00543	0.11960
10	2.96038	0.09147
11	2.92315	0.08742
12	2.90079	0.08970
13	2.88112	0.08832
14	2.86992	0.08875
15	2.85673	0.08837
16	2.84739	0.08920
17	2.83990	0.09017
18-34	2.78579	0.09418

Table 6. One Back – Accuracy

Age (Years)	Mean	SD
6	1.14048	0.23689

[redacted]

7	1.13669	0.20779
8	1.16183	0.23773
9	1.30178	0.18364
10	1.30448	0.14716
11	1.31396	0.14210
12	1.32934	0.14177
13	1.32993	0.13466
14	1.33324	0.13896
15	1.34055	0.13843
16	1.34830	0.13575
17	1.35444	0.13472
18-34	1.39617	0.12998

10.6.2. Conversion of Scale Outcome Measures into Age-Adjusted Z-Scores

All primary, secondary, and exploratory outcome measures for scales will need to be converted to z-scores. Outcome measures will be available in one of three forms: standard score, T-score, or scaled score. The Table 7 below defined the general properties of each score form, along with the conversion formula for changing these into z-scores.

Table 7: Scale Outcome Measure Conversion to Z-Scores

Score Name	Mean	SD	Z-Score Conversion Formula
Standard Score	100	15	(Standard Score – 100) / 15 * multiplicand
T-Score	50	10	(T-Score – 50) / 10 * multiplicand
Scaled Score	10	3	(Scaled Score – 10) / 3 * multiplicand

10.6.3. Standardized Scores: Comparison of Individual Scores to eTHINK Hemophilia Normative Data (Hemophilia-Adjusted Z-Scores)

A comparison of individual scores to eTHINK hemophilia normative data will also be conducted. The process will be as follows:

- Performance on each neurocognitive outcome measure will be standardized relative to eTHINK hemophilia normative data (i.e., the age-adjusted z-score will be converted to a hemophilia-adjusted z-score by subtracting the mean from the eTHINK hemophilia normative sample and dividing by the eTHINK hemophilia normative standard deviation (SD) from the same normative sample)
- The z-score will be calculated as follows:

$$z_{ijt} = \frac{\left(x_{ijt} - \bar{x}_t\right)}{SD_t}$$

Where:

$\bar{z}_{ijt}$ = cognitive hemophilia-adjusted z-score

t = is the outcome measure indicator

i = indexes subject *i*

[redacted]

j = indexes the *jth* assessment for subject *i*

x = cognitive age-adjusted z-score

$\bar{x}_t$ = mean performance score of the eTHINK hemophilia normative sample for outcome measure t

SD_t = Standard Deviation of the eTHINK hemophilia normative sample for outcome measure t

Tables similar to the above Tables 3-6 will be generated based on the eTHINK study and re-used here.

11. Data Quality Assurance

[redacted] Computerized Tests

Data from the [redacted] computerized test battery will be collected on iPads at site(s) and uploaded to the [redacted] database for processing. [redacted] data management staff will query any data discrepancies. Queries will be confirmed and resolved with the sponsor.

Traditional Neurocognitive Instruments

Before raters could administer the neurocognitive instruments to subjects, they were required to become certified by completing online training requirements (i.e., didactic training videos and for some scales (Bayley-III, WPPSI-IV, and WASI-II) demonstrations showing proper administration and scoring of the scales). A neuropsychologist at [redacted] then evaluated the completed source document and audio recording from the practice administration. The neuropsychologist issued corrective feedback, if necessary. During the study, [redacted] monitored the first Bayley-III, WPPSI-IV, and WSI-II administrations for each rater to ensure quality. For these administrations, raters submitted the source document of their administration. A neuropsychologist at [redacted] then evaluated the completed source document and issued corrective feedback, if necessary.

11.1. [redacted] Computerized Test Completion Criteria

For each of the [redacted] computerized tests, subjects must provide sufficient responses to allow computation of reliable performance measures. For the majority of [redacted] computerized tests, the term “sufficient” has been defined as a Test Completion criterion. The number of trials required for Test Completion is unique to each test. They do not vary for different patient groups or study samples.

The completion criteria set forth a priori for each test were as follows:

- DET: The number of responses provided by the subject is ≥ 100% of the desired number of trials (responses ≥ 35)
- IDN: The number of responses provided by the subject is ≥ 100% of the desired number of trials (responses ≥ 30)
- ONB: The number of responses provided by the subject is ≥ 100% of the desired number of trials (responses ≥ 31)

[redacted]

12. Statistical Methodology

12.1. Analysis Overview

All statistical analyses and summary information will be generated according to this Statistical Analysis Plan (SAP). Data collected in this study will be presented using standardized scores, comparing cognitive outcome measures against age-matched norms, hemophilia norms, the most recent prior assessment, and the Initial Assessment.

12.2. Analyses

Cognitive outcome measures from the neurocognitive instruments will have a series of z-scores calculated. This includes age-adjusted US normative (international normative for [redacted] computerized tests) z-scores (see Section 10.6.1), age-adjusted hemophilia z-scores (see Section 10.6.3), z-score change from prior (for both age- and hemophilia-adjusted z-scores), and z-score change from Initial Assessment (for both hemophilia- and age-adjusted z-scores). The above z-scores will be tabulated descriptively (n, mean, standard deviation, median, interquartile range (first quartile, third quartile), min, max).

Individual domains/scores by visit will be plotted per patient for each instrument.

A subject listing including scores, z-scores (age- and hemophilia-adjusted), change from prior z-scores (age- and hemophilia-adjusted) and change from initial assessment z-scores (age- and hemophilia-adjusted) on each neurocognitive instrument and outcome measure will be generated. All z-scores will be adjusted so that positive scores reflect improvement.

All analyses will be done in SAS (v9.4).

See Appendix 1 for an example of the tables, listing, and figures (TLF).

[redacted]

References

Bland, J. M., & Altman, D. G. (1996a). Statistics notes: Measurement error. *British Medical Journal*, 312(7047), 1654.

Bland, J. M., & Altman, D. G. (1996b). Statistics notes: Measurement error and correlation coefficients. *British Medical Journal*, 313(7048), 41-42.

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1095-1112.