Alexion Pharmaceuticals, Inc.

STATISTICAL ANALYSIS PLAN
Primary Evaluation Period
for US/Japan Submission

PROTOCOL NUMBER: WTX101-301

A PHASE 3, RANDOMIZED, RATER-BLINDED, MULTI-CENTER STUDY TO EVALUATE THE EFFICACY AND SAFETY OF ALXN1840 ADMINISTERED FOR 48 WEEKS VERSUS STANDARD OF CARE IN PATIENTS WITH WILSON DISEASE AGED 12 YEARS AND OLDER WITH AN EXTENSION PERIOD OF UP TO 60 MONTHS

Authors: [redacted]

Date: 06-Aug-2021

Version: 6.0 Final

1. APPROVAL SIGNATURES

06-Aug-2021 | 10:11:29 EDT

[redacted] Date dd Mmm yyyy

06-Aug-2021 | 11:02:30 EDT

[redacted] Date dd Mmm yyyy

06-Aug-2021 | 10:18:26 EDT

[redacted] Date dd Mmm yyyy

2. TABLE OF CONTENTS, LIST OF TABLES, AND LIST OF FIGURES

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

3. LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

The following abbreviations and acronyms in Table 1 are used in this Statistical Analysis Plan (SAP).

Table 1: Abbreviations and acronyms

Abbreviation or acronym	Explanation
ADL	Activities of Daily Living
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AUEC	area under the effect-time curve
BMI	body mass index
BPRS-24	Brief Psychiatric Rating Scale-24
BSA	body surface area
CCR	color congruent ratio
CGI	Clinical Global Impression
CGI-I	Clinical Global Impression-Improvement Scale
CGI-S	Clinical Global Impression-Severity Scale
CI	confidence interval
CIR	color incongruent ratio
cNCC	calculated non-ceruloplasmin-bound copper
$cNCC_{corrected}$	calculated non-ceruloplasmin-bound copper corrected for copper bound in tetrathiomolybdate-copper-albumin complexes
COVID-19	Coronavirus Disease 2019
Cp	ceruloplasmin
CpC	ceruloplasmin-bound copper
CP_{ni}	conditional power for non-inferiority
CP_s	conditional power for superiority
CRF	case report form
CTCAE	Common Terminology Criteria for Adverse Events
Cu	copper
CV	coefficient of variation
DNA	deoxyribonucleic acid
dNCC	directly measured non-ceruloplasmin-bound copper
dNCC $AUEC_{0-48W}$	area under the effect-time curve of directly measured non-ceruloplasmin-bound copper from 0 to 48 weeks
ECG	electrocardiogram
eDISH	evaluation of drug induced serious hepatotoxicity
EOS	end of study
EQ-5D	EuroQOL-5 Dimensions
EQ-5D-5L	EuroQOL-5 Dimensions-5 Levels
EQ VAS	EuroQol Visual Analogue Scale
ET	Early Termination
FA	Full Analysis
FIB-4	fibrosis-4
GFR	glomerular filtration rate
GGT	gamma-glutamyltransferase

Table 1: Abbreviations and acronyms

Abbreviation or acronym	Explanation
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
INR	international normalized ratio
LBC	labile bound copper
LLN	lower limit of normal
LLOQ	lower limit of quantification
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-Stage Liver Disease
MMRM	mixed model for repeated measures
Mo	molybdenum
NCC	non-ceruloplasmin-bound copper
NCI	National Cancer Institute
PD	pharmacodynamic
PK	pharmacokinetic
PP	Per Protocol
PTAE	pretreatment adverse event
PUF	plasma ultrafiltrate
PY	patient-years
QoL	Quality of Life
QTc	corrected QT interval
SAE	serious adverse event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SCr	serum creatinine
SE	standard error
SoC	standard of care
SSR	sample size re-estimation
TEAE	treatment-emergent adverse event
TPC	tripartite complex
TSQM-9	Treatment Satisfaction Questionnaire for Medication-9
ULN	upper limit of normal
UWDRS	Unified Wilson's Disease Rating Scale
VAS	Visual Analogue Scale
WD	Wilson Disease
WHO	World Health Organization
9-HPT	Nine-Hole Peg Test

4. DESCRIPTION OF THE PROTOCOL

This SAP relates to Alexion Pharmaceuticals Inc. Protocol WTX101-301 (A Phase 3, Randomized, Rater-Blinded, Multi-Center Study to Evaluate the Efficacy and Safety of ALXN1840 Administered for 48 Weeks Versus Standard of Care in Patients with Wilson Disease Aged 12 and Older with an Extension Period of up to 60 Months), Amendment 2 dated 25 March 2021.

This is a randomized, rater-blinded, multicenter study assessing the efficacy and safety of ALXN1840 versus standard of care (SoC). In the Primary Evaluation Period, efficacy and safety were assessed for an individualized ALXN1840 dosing regimen compared with SoC administered for 48 weeks in patients with Wilson Disease (WD) who are aged 12 years and older (18 years and older in Germany).

Patients meeting all inclusion and no exclusion criteria were enrolled into the study and studied as outpatients. Eligible patients with WD were enrolled into 1 of 2 cohorts, in a ratio of 3:1 (Cohort 1: Cohort 2).

- Cohort 1: Patients who had received SoC therapy (i.e., chelation therapy with penicillamine or trientine, zinc therapy, or a combination of both chelation and zinc therapy) for > 28 days prior to first dose
- Cohort 2: Patients who were treatment naïve or who had received SoC therapy for ≤ 28 days prior to first dose

Approximately 180 patients were enrolled and randomized by cohort in a 2:1 ratio to treatment with ALXN1840 or SoC (either as continued therapy in Cohort 1 or as continued or initial therapy in Cohort 2) to obtain data from approximately 150 evaluable patients (100 ALXN1840, 50 SoC) for the primary analysis. Treatments were assigned randomly, stratified by cohort.

Patients who were randomized to receive ALXN1840 were required to withhold treatment with SoC for ≥ 48 hours immediately prior to first study assessment on Day 1. Patients who were randomized to ALXN1840 received ALXN1840 as delayed-release tablets for oral administration at doses ranging from 15 mg every other day (QOD) to 60 mg once daily (QD). Efficacy and safety assessments were performed at scheduled visits, while adverse events (AEs) and concomitant medications were monitored continuously throughout the study. Patients randomized to SoC initiated treatment or continued treatment on their current regimen where possible, without compromising the safety of individual patients.

The Primary Evaluation Period consisted of an up to 28-day Screening Period, a 1-day Enrollment Visit, a 48-week Treatment Period, and a Follow-up Visit 4 weeks after the last dose for patients who do not elect to continue in the Extension Period.

Patients in Study WTX101-301 who completed the 48-week Treatment Period and patients who completed participation in Study WTX101-201 will be offered the opportunity to participate in an up to 60-month Extension Period within Study WTX101-301 to evaluate the long-term safety and efficacy of ALXN1840.

This SAP describes the analytical plan for the randomized Primary Evaluation Period. A separate SAP will be generated for the non-randomized exploratory open label Extension Period.

Refer to the Study WTX101-301 protocol for additional details.

4.1. Changes from Analyses Specified in the Protocol

- LBC responder rate at 48 weeks will not be analyzed.
- Physical examination results will not be tabulated independently. Abnormalities will be reported as part of Medical History or as adverse events.

4.2. Changes from Analyses Specified in the Previous Version of the SAP

More details of the changes can be found in Section 9.2.

5. DEFINITIONS

5.1. Efficacy

5.1.1. Primary Endpoint(s)

The primary efficacy analysis will be conducted on the data from the Primary Evaluation Period for all patients. The primary endpoint is the daily mean area under the effect-time curve (AUEC) of directly measured non-ceruloplasmin-bound copper (dNCC) from 0 to 48 weeks (dNCC $AUEC_{0\text{-}48W}$). dNCC is the directly quantified copper not bound to ceruloplasmin, obtained by inductively coupled plasma mass spectrometry after immunocapture and removal of ceruloplasmin. The mean of dNCC $AUEC_{0\text{-}48W}$ will be compared between ALXN1840 and SoC.

Based on results obtained from the completed Phase 1 and Phase 2 studies, it was determined that an integrative measure (AUEC) of non-ceruloplasmin-bound copper mobilization and elimination over time is a more appropriate surrogate endpoint to assess the treatment effect on tissue copper overload. In contrast, point measurements of blood species of copper (eg, total plasma copper, labile bound copper, and calculated non-ceruloplasmin-bound copper [cNCC]), do not reflect the underlying burden of copper overload in tissues, and, therefore, do not provide meaningful data to direct therapeutic decisions.

5.1.2. Secondary Endpoints

The key secondary endpoints are:

- Change from baseline in the Unified Wilson's Disease Rating Scale (UWDRS) Part II total score.
- Change from baseline in UWDRS Part III Functional Subscale score
- Change from baseline in UWDRS Part III individual functional items: arising from a chair, gait, handwriting, and speech.

The non-key secondary endpoints are:

- Change from baseline in UWDRS Part III total score
- Clinical Global Impression-Improvement Scale (CGI-I)
- Change from baseline in Clinical Global Impression-Severity Scale (CGI-S)
- Change from baseline in Model for End-Stage Liver Disease (MELD) score
- Absolute change and percentage change from baseline (Day 1) to 48 weeks in cNCC in plasma. For ALXN1840-treated patients, the cNCC in plasma will be corrected for the amount of copper bound to the ALXN1840 tripartite complex (TPC) ($cNCC_{corrected}$)
- cNCC/$cNCC_{corrected}$ responder rate at 48 weeks.

Descriptions and algorithms for some of the secondary endpoints are described in Section 5.1.2.

5.1.2.1. Unified Wilson Disease Rating Scale (Parts I, II, and III)

The UWDRS (Czlonkowska, 2007; Leinweber, 2008) is a clinical rating scale designed to evaluate the neurological manifestations of WD that generally can be divided into 3 movement disorder syndromes: dystonic, ataxic, and Parkinsonian syndrome. The UWDRS comprises 3 parts: UWDRS Part I (level of consciousness, item 1), UWDRS Part II (a patient-reported review of daily activity items [disability], items 2 to 11 [10 items in total]), and UWDRS Part III (a detailed neurological examination, items 12 to 34 [23 items in total]). The UWDRS Part I and III will be assessed by a neurologist who is blinded to the treatment randomization, while UWDRS Part II may be reported to a non-blinded member of the study team, by the patient, family member or caregiver.

The UWDRS will be assessed at Day 1, Week 4, Week 12, Week 24, Week 36, and Week 48/ET. The UWDRS may also be assessed at the Screening Visit and EOS for patients who enrolled under the original protocol prior to the implementation of Protocol Amendment 1. Refer to Appendix 9.4 for further details on calculating the UWDRS scores.

The UWDRS has not been formally evaluated in adolescents. However, the components are not fundamentally different between adults and adolescents. Patients aged 12 years and older are expected to be able to comply with UWDRS assessments (Rohay, 2020).

5.1.2.2. Clinical Global Impression-Improvement Scale and the Clinical Global Impression-Severity Scale

The clinical global impression (CGI) rating scales are commonly used measures of symptom severity, treatment response and the efficacy of treatments in treatment studies of adult and pediatric patients. Although originally developed for patients with mental disorders (Guy, 1976) the CGI rating scales have been adapted to use generalized language (Busner, 2007), which supports their application to any clinically assessed disease setting.

The CGI-S is a 7-point scale that requires the clinician to rate the severity of the patient's illness at the time of the assessment, relative to the clinician's past experience with patients who have the same diagnosis. Considering total clinical experience, a patient is assessed on severity of illness at the time of rating as: 1, normal, not at all ill; 2, borderline ill; 3, mildly ill; 4, moderately ill; 5, markedly ill; 6, severely ill; or 7, extremely ill. The CGI-S will be performed at Day 1, Week 4, Week 12, Week 24, Week 36, and Week 48/ET. The CGI-S may also be assessed at Week 8, Week 18 and EOS for patients who enrolled under the original protocol prior to the implementation of Protocol Amendment 1.

The CGI-I is a 7-point scale that requires the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of the intervention and rated as: 1, very much improved; 2, much improved; 3, minimally improved; 4, no change; 5, minimally worse; 6, much worse; or 7, very much worse. The CGI-I will be performed at the same visits as CGI-S with the exception of Day 1.

5.1.2.3. Model for End-Stage Liver Disease Score

The MELD (Alcorn, 2015) is a scoring system for assessing the severity of chronic liver disease in patients 12 years and older. The MELD score (range 6-40, with higher values indicating more advanced disease) uses the patient's values for serum bilirubin, serum creatinine, and the

international normalized ratio (INR). The initial MELD score, $MELD_{(i)}$ is calculated according to the following formula:

$$MELD_{(i)} = 3.78\times\ln[\text{serum bilirubin (mg/dL)}] + 11.2\times\ln[\text{INR}] + 9.57\times\ln[\text{serum creatinine (mg/dL)}] + 6.43$$

Creatinine, bilirubin, and INR values less than 1.0 are set to 1.0 and creatinine values greater than 4.0 are set to 4.0 when calculating $MELD_{(i)}$. Additionally, creatinine, bilirubin, and INR are rounded to the 10th decimal place prior to performing the calculation. The initial MELD score is then rounded to the nearest integer. The maximum MELD score is 40.

A modification to the MELD score exists for MELD scores greater than 11 (Alcorn, 2015) and this will be applied to remain up-to-date with the current guidance. Thus, if $MELD_{(i)}$ is greater than 11, MELD is recalculated as follows:

$$MELD = MELD_{(i)} + 1.32*(137\text{-sodium [mmol/L]}) - 0.033*MELD_{(i)}*(137\text{-sodium [mmol/L]})$$

Sodium values less than 125 mmol/L will be set to 125 and values greater than 137 mmol/L will be set to 137.

The MELD score will be calculated each time the appropriate clinical chemistry parameters are obtained at the Screening Visit, Day 1, Week 4, Week 6, Week 8, Week 12, Week 18, Week 24, Week 36, Week 48/Early Termination (ET), and End of Study (EOS). If any of the parameters required to compute MELD score are unavailable at any given visit, the MELD score will not be derived for that patient visit.

5.1.2.4. cNCC/cNCC$_{corrected}$, cNCC/cNCC$_{corrected}$ Responder

cNCC will be calculated as follows (EASL, 2012; Roberts et al., 2008):

$$cNCC[\mu mol/L] = \frac{Plasma\ Total\ Cu\ [\mu g/L] - (3.15 * ceruloplasmin\ [mg/L])}{63.5\ [\mu g/\mu mol]}$$

For ALXN1840-treated patients, the cNCC in plasma will be corrected for the amount of copper bound to the ALXN1840 TPC using the square root-based cNCC correction method ($cNCC_{corrected}$) as determined based on data from the WTX101-201 study as follows (Plitz, 2017):

$$cNCC_{corrected} = \left(\sqrt{cNCC} - 0.993\sqrt{Mo}\right)^2$$

In the calculation of cNCC and $cNCC_{corrected}$ the following rules apply:

- For plasma total copper concentration values < lower limit of quantification (LLOQ), cNCC will be considered missing;
- Serum ceruloplasmin concentration values < LLOQ are set to 0;
- Plasma total molybdenum concentration values < LLOQ are set to 0;
- In cases where cNCC calculation produces a negative result, cNCC will be considered missing and $cNCC_{corrected}$ will not be derived;
- $cNCC_{corrected}$ will be set to 0 when $0.993\sqrt{Mo} > \sqrt{cNCC}$.

cNCC/cNCC$_{corrected}$ responder is defined as patients who achieved or maintained normalized cNCC/cNCC$_{corrected}$ concentration (0.8-2.3 µM) (Brewer 2009) within (at or before) 48 weeks or reached a reduction of at least 25% in cNCC/cNCC$_{corrected}$ within 48 weeks. Thus, a patient will be considered a cNCC/cNCC$_{corrected}$ responder if they met at least one of the following criteria:

- Achieved normalized cNCC/cNCC$_{corrected}$ concentration for two consecutive measurements within 48 weeks, for patients who had elevated cNCC concentrations at baseline;
- Maintained normalized cNCC/cNCC$_{corrected}$ concentration within 48 weeks, for patients who had normal cNCC concentrations at baseline;
- Reached a reduction of at least 25% in cNCC/cNCC$_{corrected}$ for two consecutive measurements within 48 weeks.

Nonresponder will be defined as patients who did not meet the responder criteria.

5.1.3. Exploratory Endpoints

The exploratory endpoints are:

- Individualized assessment of each patient's 3 most troublesome symptoms;
- Change from baseline in the Fibrosis-4 (FIB-4) Index and by transient elastography;
- Change from baseline in Modified Nazer Score;
- Change from baseline in Brief Psychiatric Rating Scale-24 (BPRS-24);
- Change from baseline in Euro Quality of Life (QoL) 5 Dimensions (EQ-5D);
- Evaluation of Treatment Satisfaction Questionnaire for Medication-9 (TSQM-9);
- Change from baseline in the timed 25 Foot Walk Test;
- Change from baseline in the 9-Hole Peg Test (9-HPT);
- Change from baseline in the nonverbal Stroop Interference Test;
- Change from baseline in the Digit Span Test;
- Evaluation of plasma total copper, plasma ultrafiltrate (PUF)-copper, dNCC, labile bound copper (LBC), ceruloplasmin (Cp), and ceruloplasmin-bound Cu (CpC) concentration-time profiles;
- Daily mean AUEC of plasma dNCC (from 0 to 24, and 24 to 48 weeks);
- Daily mean AUEC of LBC from 0 to 48 weeks;
- Daily mean AUEC of plasma total copper from 0 to 48 weeks;
- Absolute and percentage change from baseline (Day 1) to 48 weeks in dNCC;
- Absolute and percentage change from baseline (Day 1) to 48 weeks in LBC;
- Time to first confirmed increase in plasma dNCC, and total copper concentration;

- Time to minimum and maximum concentration (TE_{max}) and the maximum concentration of (E_{max}):
  - − Plasma dNCC
  - − Plasma total copper
  - − Plasma LBC
  - − Ratio plasma dNCC:total copper
  - − Ratio plasma LBC:total copper
  - − Urinary molybdenum
  - − Ratio urinary molybdenum:urinary copper
  - − Ratio urinary molybdenum:dosed molybdenum
  - − Plasma Cp
  - − Plasma CpC
  - − Ratio plasma Cp:total copper
  - − Ratio plasma CpC:total copper
  - − Ratio plasma CpC:Cp;
- Time for return to predose baseline for the following parameters:
  - − Plasma dNCC
  - − Plasma total copper
  - − Plasma LBC
  - − Ratio plasma dNCC:total copper
  - − Ratio plasma LBC:total copper
  - − Urinary molybdenum
  - − Ratio urinary molybdenum:urinary copper
  - − Ratio urinary molybdenum:dosed molybdenum
  - − Plasma Cp concentration
  - − Plasma CpC concentration
  - − Ratio plasma Cp:total copper
  - − Ratio plasma CpC:total copper
  - − Ratio plasma CpC:Cp;
- Change from baseline of total molybdenum and PUF-molybdenum (Mo) in plasma;
- Change from baseline of 24-hour urinary copper and urinary molybdenum;

- Change from baseline in the UWDRS Part I score and UWDRS total score;
- Change from baseline in the UWDRS Part II-FDA suggested scoring;
- Change from baseline in the UWDRS Part III Functional Subscale-FDA suggested scoring
- Change from baseline in the UWDRS Part III Individual Items-FDA suggested scoring: arising from a chair, gait, handwriting, and speech
- Change from baseline in the UWDRS Part II ADL subscale score
- UWDRS Part II responder rate at 48 weeks (proportion of patients who achieve clinically meaningful improvement)
- UWDRS Part III responder rate at 48 weeks (proportion of patients who achieve clinically meaningful improvement)

Descriptions and algorithms for the exploratory efficacy endpoints in the Primary Evaluation Period are described in Sections 5.1.3.1 to Section 5.1.3.13.

5.1.3.1. Three Most Troublesome Symptoms

At randomization each patient will record their 3 most troublesome symptoms. These symptoms will be recorded in the case report form (CRF) as well as the impact these symptoms have on the patient's activities of daily living (ADL). The 3 most troublesome symptoms will be assessed at Day 1; they will be subsequently re-presented to the patient at Week 12, Week 24, and Week 48/ET when the patient will be asked if their symptoms have improved, remained the same or worsened relative to Day 1. The influence of patients with missing data at either Week 12, Week 24 or Week 48 will be explored by assuming responses of (i) 'remained the same' and (ii) 'worsened'.

5.1.3.2. Fibrosis-4 Index/Transient Elastography

The FIB-4 Index (Vallet-Pichard, 2007) is a formula used to predict liver fibrosis based on standard biochemical values (alanine aminotransferase [ALT], aspartate aminotransferase [AST], and platelet count) and age. The FIB-4 Index will be calculated by a Central Laboratory. The formula is as follows:

$$[\text{Age (years)} * \text{AST (U/L)}] / [\text{Platelets } (10^9/\text{L}) * \text{sqrt}(\text{ALT } \{\text{U/L}\})]$$

Transient elastography is a non-invasive imaging method that evaluates the degree of liver fibrosis or fatty deposits in the liver, by determining the speed of sound waves through the liver utilizing a sonogram.

The FIB-4 Index will be assessed at Day 1, Week 4, Week 6, Week 8, Week 12, Week 18, Week 24, Week 36, Week 48/ET, and EOS. Transient elastography will be performed at Day 1 and Week 48/ET. If any of the parameters required to compute the FIB-4 Index are unavailable at any given visit, FIB-4 Index will be set to missing for that patient visit.

5.1.3.3. Modified Nazer Score

The modified Nazer Score (Dhawan, 2005) is an assessment of liver status and consists of a composite of 5 laboratory parameters: bilirubin, AST, INR, leukocytes (white cell count), and albumin. The score has a total range of 0 to 20, and lower values indicate a healthier liver status. The modified Nazer Score will be determined at the Screening Visit, Day 1, Week 4, Week 6, Week 8, Week 12, Week 18, Week 24, Week 36, Week 48/ET, and EOS. If any of the 5 parameters required to compute the modified Nazer Score are unavailable at any given visit, the modified Nazer Score will not be derived at that patient visit. The score for an individual analyte (bilirubin, AST, INR, leukocytes, and albumin) should be derived from Table 2 , and then all 5 scores will be added to obtain the final score.

Table 2: Modified Nazer Score

Score	Bilirubin (µmol/L)	AST (IU/L)	INR	Leukocytes (10^9/L)	Albumin (g/L)
0	0-100	0-100	0-1.29	0-6.7	> 44
1	101-150	101-150	1.3-1.6	6.8-8.3	34-44
2	151-200	151-300	1.7-1.9	8.4-10.3	25-33
3	201-300	301-400	2.0-2.4	10.4-15.3	21-24
4	> 300	> 400	> 2.4	> 15.3	< 21

Abbreviations: AST = aspartate aminotransferase; INR = international normalized ratio

5.1.3.4. Brief Psychiatric Rating Scale-24

The BPRS-24 (Ventura, 1993) is a 24-item instrument that allows the rater to measure psychopathology severity. The presence and severity of psychiatric symptoms are rated on a Likert scale ranging from 1 (not present) to 7 (extremely severe). The BPRS-24 can be performed by a qualified person (eg, neurologist, psychiatrist, psychologist, licensed mental health practitioner, social worker, etc.) who has completed the training required to administer the instrument.

The BPRS-24 will be obtained at Day 1, Week 24, and Week 48/ET. The BPRS-24 may also be assessed at the Screening Visit and EOS for patients who enrolled under the original protocol prior to the implementation of Protocol Amendment 1. The total score across the 24 items will be used as the endpoint for analysis. A minimum of 20/24 items are required to be completed at any given visit; the total score will then be taken for the non-missing items and scaled up to the 24-item score for the purposes of data summary and analysis. If fewer than 20/24 items are completed at any given visit, the BPRS-24 total score will be set to missing for that patient visit.

5.1.3.5. EuroQoL-5 Dimensions

The EQ-5D (EuroQol Group, 2015) consists of 2 different assessments: the EQ-5D-5L Descriptive System and the EuroQol Visual Analogue Scale (EQ VAS). The descriptive system comprises measures of health-related QoL state and consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels of severity: no problems, slight problems, moderate problems, severe problems, or extreme problems. For the scoring in the EQ-5D-5L Descriptive System, the respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number expressing the

level selected for that dimension. The digits for 5 dimensions can be combined in a 5-digit number describing the respondent's health state. The EQ VAS records the patient's self-rated health on a vertical visual analogue scale. Together, this can be used as a quantitative measure of health outcome that reflects the patient's own judgment.

The EQ-5D will be administered at Day 1, Week 4, Week 12, Week 24, Week 36, and Week 48/ET. The EQ-5D may also be assessed at Week 8 and Week 18 for patients who enrolled under the original protocol prior to the implementation of Protocol Amendment 1.

5.1.3.6. Treatment Satisfaction Questionnaire for Medication

The TSQM-9 is used to assess the overall level of satisfaction or dissatisfaction with medication patients are taking (Atkinson, 2004; Atkinson, 2005; Bharmal, 2009).

The TSQM-9 has 9 questions making up 3 components: effectiveness score, convenience score, and global satisfaction score. Details on deriving the component scores are given below including how to handle a single missing question within a component. If more than one question is missing for a component, the component score cannot be derived.

Effectiveness score is comprised of Questions 1 to 3, each of them ranging from 1 to 7. The score is calculated as follows: (Q1+Q2+Q3-3)/18*100. If one of the questions is missing, the score is calculated as follows (including the missing item): (Q1+Q2+Q3-2)/12*100.

Convenience score is comprised of Questions 4 to 6, each ranging from 1 to 7. The score is calculated as follows: (Q4+Q5+Q6-3)/18*100. If one of the questions is missing, the score is calculated as follows (including the missing item): (Q4+Q5+Q6-2)/12*100.

Global satisfaction is comprised of Questions 7 to 9, where Questions 7 and 8 range from 1 to 5 and Question 9 ranges from 1 to 7. The score is calculated as follows: (Q7+Q8+Q9-3)/14*100. If either Question 7 or 8 is missing, the score is calculated as follows (including the missing item): (Q7+Q8+Q9-2)/10*100. If Question 9 is missing, the score is calculated as follows: (Q7+Q8-2)/8*100.

The TSQM-9 will be administered at Week 4, Week 12, Week 24, Week 36, and Week 48/ET. The TSQM-9 may also be assessed at Week 8 and Week 18 for patients who enrolled under the original protocol prior to the implementation of Protocol Amendment 1.

5.1.3.7. Timed 25F Walk Test

The Timed 25F Walk Test (Hobart, 2013; Coleman, 2012) is a quantitative mobility and leg function performance test based on a timed 25-foot walk. Scoring for the Timed 25F Walk Test is the average of the 2 trials in seconds captured in the CRF. The walking times will then be transformed and analyzed as walking speed (more normally distributed) in feet per second by dividing 25 feet by the time in seconds required to complete the walk. Patients may use assistive devices when doing this task. The test will be performed at Day 1, Week 12, Week 24, and Week 48/ET. If one of the trials is missing at any given visit, the time taken for the complete trial will be used as the test result.

5.1.3.8. Nine-Hole Peg Test

The 9-HPT (Mathiowetz et al, 1985) is a brief, standardized, quantitative test of upper extremity function. Both the dominant and non-dominant hands are tested twice. The patient is seated at a table with a small, shallow container holding 9 pegs and a wood or plastic block containing 9 empty holes. On a start command when a stopwatch is started, the patient picks up the 9 pegs, one at a time as quickly as possible, puts them in the 9 holes, and, once they are in the holes, removes them again as quickly as possible one at a time, replacing them into the shallow container. Two consecutive trials with the dominant hand are immediately followed by 2 consecutive trials with the non-dominant hand. The result for the 9-HPT is the average of the 4 trials in minutes. At least 1 complete test per hand is required to provide an overall result for the 9-HPT for that patient visit; if both tests are missing for either hand, then the overall result for that patient visit will be set to missing. The 9-HPT will be performed at Day 1, Week 12, Week 24, and Week 48, and ET.

5.1.3.9. Nonverbal Stroop Interference Test

In psychology, the Stroop effect (Koch, 2012) is a demonstration of interference in the reaction time of a task. There will not be any verbal communication during this test; the test will be taught with nonverbal directions, using gestures and demonstrations. The test will be performed at Day 1, Week 12, Week 24, and Week 48/ET. Patients are presented with 32 color congruent cards and 72 color incongruent cards and are tasked with categorizing these cards correctly according to the nonverbal instruction provided. The number of cards correctly and incorrectly categorized and the time taken to complete the test are recorded onto the record form. The endpoint for analysis is the Stroop Interference Effect, which is calculated for each patient using the color congruent ratio (CCR) and color incongruent ratio (CIR) as follows:

$$CCR = \frac{\text{Number of seconds to complete all color congruent cards}}{\text{Number of correct color congruent cards} - \text{Number of incorrect color congruent cards}};$$

$$CIR = \frac{\text{Number of seconds to complete all color incongruent cards}}{\text{Number of correct color incongruent cards} - \text{Number of incorrect color incongruent cards}};$$

Stroop Interference Effect = CIR – CCR.

Based on age of the patient, the Stroop Effect Scaled Score is then derived from the Stroop Interference Effect score to have a mean of 50 and standard deviation of 10. Scores between 35 and 65 are considered normal. Higher scores indicate worse performance generally.

5.1.3.10. Digit Span Test

The complete Digit Span Test (Wechsler, 2008) is measured for forward and reverse-order (backward) recall of digit sequences and digit span sequencing. Digit sequences are presented beginning with a length of 2 digits and 2 trials are presented at increasing list length. Testing ceases when the patient fails to accurately report either trial at 1 sequence length or when the maximal list length is reached (9 digits, 8 backwards). The Digit Span Test will be performed at

Day 1, Week 12, Week 24, and Week 48/ET. Patients who cannot speak clearly enough to perform the test will be assigned the lowest possible outcome score.

5.1.3.11. Plasma Total Molybdenum, Plasma Ultrafiltrate Molybdenum, Plasma Total Copper, Plasma Ultrafiltrate Copper, dNCC, LBC, Ceruloplasmin, and Ceruloplasmin-bound Copper

Plasma concentrations of total molybdenum, PUF-Mo, plasma total copper, PUF-Cu, dNCC, LBC, Cp, and CpC will be measured at the Screening Visit, Day 1, Week 4, Week 6, Week 8, Week 12, Week 18, Week 24, Week 36, and Week 48/ET.

5.1.3.12. Time to First Confirmed Increase, Time to Minimum and Maximum, and Time for Return to Predose Baseline

Time to first confirmed increase is defined as weeks from baseline to the second timepoint of the 2 consecutive timepoints above baseline. Time to minimum is defined as weeks from baseline to the timepoint with minimum postbaseline values within the Primary Evaluation Period. Time to maximum is defined as weeks from baseline to the timepoint with maximum postbaseline values within the Primary Evaluation Period. Time for return to predose baseline is defined as weeks from baseline to the first timepoint with postbaseline values less or equal to the predose baseline. The maximum observed concentration values will also be reported.

5.1.3.13. 24-Hour Urine Copper and Molybdenum

Twenty-four hour urine copper and molybdenum will be measured at Day 1, Week 4, Week 12, Week 24, Week 36, and Week 48/ET.

5.2. Safety

The safety endpoints in this study are the safety and tolerability of individualized dosing of ALXN1840 over time. Safety will be assessed by analysis of AEs. AE data collected will include onset, duration, seriousness, intensity, and relatedness to study drugs. Deaths and other serious adverse events (SAEs) will also be evaluated and will be collected on a separate CRF. Changes in physical examinations, vital signs (resting heart rate, semi-supine systolic and diastolic blood pressure, respiratory rate, temperature, and weight), and laboratory values will also be evaluated and assessed. Time points for recording of safety parameters is provided in Table 4 of the Protocol.

5.2.1. Adverse Events

The occurrence of AEs at each visit will be recorded on designated CRF pages. Each AE is to be characterized (ie, verbatim term) and information provided regarding its seriousness, start and stop dates, severity, outcome, and causal relationship with the study drug. The safety evaluation will include an assessment of all AEs, SAEs, AE severity, and AE causality. AE severity will be evaluated using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 (published 27 Nov 2017) (NCI, 2017). AE causality will be evaluated by Investigators to be either not related or related. Further details are given in Protocol Section 8.3, Protocol Appendix 3, and in Appendix 9.4.

5.2.2. Vital Signs

Vital sign measurements will be taken after the patient has been resting for at least 5 minutes, and will include systolic and diastolic blood pressure (mmHg), heart rate (beats/minute), respiratory rate (breaths/minute), height, weight, and temperature (°C or °F). The timepoints for recording of vital signs is provided in Table 4 of the Protocol.

5.2.3. Laboratory Assessments

Clinical laboratory measures include chemistry, hematology, coagulation, copper and molybdenum analyses testing, and urinalysis (with microscopy). For females of childbearing potential, a serum or urine pregnancy test (ie, beta-human chorionic gonadotropin) will be performed. For postmenopausal females, follicle-stimulating hormone will be performed. Additionally, deoxyribonucleic acid (DNA) data will be available for some patients as it is optional. The specific laboratory assessments are provided in Appendix 2 of the Protocol.

5.2.4. Physical Examinations

Physical examinations will include an assessment of the following: general appearance, respiratory, cardiovascular, abdomen, skin, head and neck (including ears, eyes, nose, and throat), lymph nodes, thyroid, and musculoskeletal (including spine and extremities) systems.

An abbreviated physical examination consists of a body system relevant examination based upon Investigator (or designee) judgment and patient symptoms. At least 1 body system must be checked for an abbreviated examination.

5.2.5. Electrocardiogram

Single 12-lead electrocardiogram (ECG) will be obtained in triplicate using an ECG machine that automatically calculates the heart rate and measures PR, QRS, QT, and corrected QT (QTc) intervals. Patients must be supine for approximately 5 to 10 minutes before ECG collection and remain supine but awake during ECG collection. The Investigator or designee will be responsible for reviewing the ECG to assess whether the ECG is within normal limits and determine the clinical significance of the results.

5.2.6. Other Events of Special Interest

Any new neurological symptom or clinically significant worsening of an ongoing neurological symptom after initiation of study drug (ALXN1840 or SoC) will be designated to be an adverse event of special interest (AESI), whether serious or nonserious. If a patient has an AESI, in addition to assessments deemed clinically relevant by the Investigator, the following additional assessments should be performed to the extent possible to help assess the AE and patient status: UWDRS Part III, nonverbal Stroop Interference Test, Digit Span Test, and the CGI-I and CGI-S.

6. DATA SETS ANALYZED (STUDY POPULATIONS)

6.1. Full Analysis Set

The Full Analysis (FA) Set includes all randomized patients who received at least 1 dose of randomized treatment. Patients will be analyzed as randomized.

6.2. Per Protocol Set

The Per Protocol (PP) Set includes all patients who are randomized and had at least baseline and 48-week efficacy assessments for dNCC in the Primary Evaluation Period. Patients with major protocol deviations that are likely to impact the primary efficacy analysis will be excluded from the PP Set. Major protocol deviations, and the PP Set, will be defined, documented, and agreed within Alexion prior to database lock.

6.3. Safety Set

The safety analysis will be performed on the Safety Analysis Set. This dataset includes all patients who received at least 1 dose of randomized treatment. Patients will be summarized according to the treatment actually received.

6.4. Pharmacokinetic Set

Patients meeting the definition of the FA Set and who have any baseline and postbaseline data for plasma molybdenum will be included in the pharmacokinetic (PK) set.

6.5. Pharmacodynamic and Biomarker Set

Patients meeting the definition of the FA Set and who have any baseline and postbaseline measurable concentration data reported for any plasma copper (pharmacodynamics; PD) and ceruloplasmin (biomarker) will be included in the PD and Biomarker set.

7. STATISTICAL ANALYSIS

For the statistical analyses below, descriptive statistics (n, mean, median, SD, minimum, and maximum) will be provided for each continuous variable, and frequencies and percentages will be provided for each categorical variable. Certain continuous variables may also include 95% confidence intervals (CIs) and certain categorical variables will be summarized as proportions with 95% CIs. All data will be displayed unless otherwise indicated. Analyses will be conducted using Statistical Analysis Software (SAS©) version 9.4 or higher.

In general, analyses will be presented by randomized treatment within cohort and overall where specified. Cohort will be derived using the patient's prior WD medication history. Patients with SoC treatment > 28 days (treatment experienced) prior to first dose will be in Cohort 1 and patients with SoC treatment ≤ 28 days or who are SoC treatment naïve prior to first dose will be in Cohort 2.

7.1. Study Patients

7.1.1. Disposition of Patients

The number and percentage of all patients enrolled, randomized, included in the FA, PP, Safety, PK and PD sets, and those who have come from Study WTX101-201 or WTX101-203 will be summarized. The reasons for exclusion from the analysis sets will also be provided. Frequency counts and percentages of patients excluded prior to randomization will be provided for patients who failed to meet study entry requirements during Screening at the start of the study.

The number and percentage of patients who completed, or prematurely discontinued from the study will be described by randomized treatment group within each cohort, for rollover patients and overall. For patients who discontinued the study, the number and percentage will be summarized by their reason for premature discontinuation and withdrawal of consent. A summary will be provided of patients by region, country, and site. Additionally, a summary of patients who did not meet inclusion or who met exclusion criteria will be provided.

Descriptive statistics of the number of days in the study will be summarized. The date of first and last use of study medication in each period, and the study termination date will be listed. Individual reasons for premature discontinuation and withdrawal of consent from the study will be presented in a listing. All enrolled patients will be listed indicating their analysis set along with the reason for exclusion. A listing of screen failure patients will also be provided. Additionally, a listing of the inclusion/exclusion criteria and a listing of patients and the inclusion criteria they failed to meet and the exclusion criteria they met will be provided.

7.1.2. Protocol Deviations

All protocol violations will be determined and appropriately categorized prior to database lock. The number and percentage of patients with any important/not important protocol violations, as well as the number and percentage of patients with violations within each category, will be presented. A listing will also be provided.

7.1.3. Demographics, Disease Characteristics, and History

All demographic and baseline characteristics information will be summarized on the FA Set. Summary statistics will be presented by randomized treatment within cohort, for rollover patients and overall. Continuous variables will be presented using descriptive statistics, and categorical variables will be presented using frequencies and percentages. Age will be calculated relative to date of informed consent and will be summarized as both a continuous and categorical variable. Time since WD treatment start date (months) will be calculated as months occurring between date of informed consent and the start date from WD treatment history CRF and will be summarized using descriptive statistics. Listings will also be provided.

7.1.3.1. Demographics

The following demographic variables will be summarized:

- Age (years),
- Age in years group (≥12-<18, ≥18), (≥12-<18, ≥18-<65, ≥65), (<25, ≥25)
- Sex (Male, Female),
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Unknown),
- Race (White, Black or African American, Asian, American Indian or Alaska native, Native Hawaiian or other Pacific Islander, Other),
- Country/Region (United States, United Kingdom, Germany, Japan, Rest of World),
- Height (cm),
- Weight (kg),
- Body mass index (kg/m^2)

7.1.3.2. Disease Characteristics

The following disease characteristics will be summarized:

- Time since WD treatment start date (months),
- Cumulative duration of prior WD treatment (months),
- Prior WD therapy: (i) Zinc monotherapy vs Penicillamine or Trientine (± Zinc); (ii) Zinc monotherapy vs Penicillamine (± Zinc) vs Trientine (± Zinc)
- Cirrhosis (Yes, No),
- CGI severity,
- Psychiatric symptom (Yes, No): at least 1 symptom present on BPRS-24
- dNCC,
- Plasma total copper,
- Plasma total molybdenum,
- PUF-Cu,
- PUF-Mo,
- Cp,
- CpC,
- 24-hour urinary copper concentration,
- 24-hour urinary molybdenum concentration
- LBC,
- MELD score,

- Modified Nazer Score,
- UWDRS Part II total score (original score and FDA suggested score)
- UWDRS Part III total score,
- UWDRS Part III Functional Subscale (original score and FDA suggested score),
- UWDRS Part III individual symptoms (original score and FDA suggested score of arising from a chair, gait, speech, and handwriting),
- UWDRS ADL Subgroup at Baseline,
- EQ-5D visual analogue scale (VAS),
- EQ-5D-5L.

In addition, the following baseline laboratory measures will be summarized:

- ALT,
- ALT groups (> 2× upper limit of normal [ULN], ≤ 2 × ULN),
- AST,
- Total bilirubin,
- Total bilirubin groups (>ULN, ≤ ULN),
- Gamma-glutamyl transferase (GGT),
- Platelets,
- Leukocytes,
- Creatinine,
- Total Cholesterol, HDL, LDL,
- Triglycerides,
- Renal Function Status at Baseline:
    - Normalized glomerular filtration rate (GFR) (mL/min/1.73 m^2)
        - G1: GFR > 90 mL/min/1.73 m^2
        - G2: GFR 60 to 89 mL/min/1.73 m^2
        - G3a: GFR 45 to 59 mL/min/1.73 m^2
        - G3b: GFR 30 to 44 mL/min/1.73 m^2
        - G4: GFR 15 to 29 mL/min/1.73 m^2
        - G5: GFR < 15 mL/min/1.73 m^2 or treatment by dialysis
    - Absolute GFR (mL/min)
        - G1: GFR ≥ 90 mL/min
        - G2: GFR 60 to < 90 mL/min
        - G3: GFR 30 to < 60 mL/min
        - G4: GFR < 30 mL/min not requiring dialysis
        - G5: GFR < 15 mL/min requiring dialysis treatment

Renal Function Status at Baseline:

Patients will be classified into 6 different stages of kidney disease based on normalized GFR estimated from the serum concentration of creatinine at baseline. Patients will also be classified based on the absolute GFR (EMA/CHMP/83874, 2014).

Patients will also be classified into 6 different stages of kidney disease based on normalized GFR estimated from the serum concentration of creatinine at baseline.

The normalized GFR can be calculated using Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for participants with Age ≥ 18 years (Levey et al., 2009, Stevens and Levey, 2009 and Frequently Asked Questions, National Kidney Foundation, 2014):

Normalized GFR (mL/min/1.73 m^2)

$= 141 \times \min(S_{Cr}/\kappa, 1)^{\alpha} \times \max(S_{Cr}/\kappa, 1)^{-1.209} \times (0.993)^{Age} \times$(1.018 if female)×(1.159 if African American),

where S_{Cr} = standardized serum creatinine assay is given in milligrams per deciliter (mg/dl), the parameter κ is 0.7 for females and 0.9 for males, the parameter α is -0.329 for females and -0.411 for males, min indicates the minimum of 1 or S_{Cr}/κ, and max indicates the maximum of 1 or S_{Cr}/κ.

For children and adolescent participants (Age <18 years), the equation provided by Schwartz, 2009 ("bedside" formula) provides a better estimate of the normalized GFR (Frequently Asked Questions, National Kidney Foundation, 2014):

Normalized GFR (mL/min/1.73 m^2) = 0.41×[Height (cm)/ S_{Cr} (mg/dl)]

The relationship between normalized GFR and absolute GFR can be obtained using body surface area (BSA) as presented in Frequently Asked Questions, National Kidney Foundation, 2014:

Absolute GFR (mL/min) = Normalized GFR (mL/min/1.73 m^2)×BSA/1.73

where BSA = Body Surface Area.

For adults, the BSA can be estimated using weight and height of the participant as follows (DuBois and DuBois, 1916, National Kidney Disease Education Program, 2009, and Frequently Asked Questions, National Kidney Foundation, 2014):

Age ≥ 18 years: BSA (m^2) = $0.007184 \times (\text{Weight (kg)})^{0.4250} \times (\text{Height (cm)})^{0.7250}$

For young and adolescent participants, the BSA can be determined using formula by Haycock et al., 1978, and also presented in Schwartz, 2009:

Age < 18 years: BSA (m^2) = $0.024265 \times (\text{Weight (kg)})^{0.5378} \times (\text{Height (cm)})^{0.3964}$

Patients with end-stage renal disease on dialysis (Chronic Kidney Disease Stage 5) or creatinine clearance < 30 mL/min are excluded from the study.

7.1.3.3. Medical / Surgical History, and WD History

Medical history will be summarized by counts and percentages and displayed by system organ class and Preferred Term within each system organ class. System Organ Class and Preferred Term will be coded using the most up-to-date version of the Medical Dictionary for Regulatory Activities (MedDRA), version 21.1 or higher, available at the start of the study. This dictionary

will be used throughout the life of the study and will not be updated during study conduct. The number and percentage of patients will be presented for ongoing conditions and previous conditions separately by System Organ Class and Preferred Term. A by-patient listing will also be created.

All details of WD diagnosis will be listed in full. WD treatment history will be summarized by counts and percentages for Cohort 1 and Cohort 2 by prior SoC, where prior SoC is defined as the later of the treatment at the time of Screening or most recent treatment prior to Screening and is categorized as penicillamine, trientine hydrochloride, zinc, or combination.

7.1.4. Prior and Concomitant Medications / Therapies

The World Health Organization (WHO) Drug Dictionary version from March 2018 or later will be used to code the medications. Medications will be summarized by Anatomical Therapeutic Chemical (ATC) level 3 class and generic drug name.

Prior and concomitant medications will be summarized for the FA Set by randomized treatment within treatment cohort, for rollover patients and overall. The number and percentage of patients receiving any concomitant medication will be summarized, as well as the number and percentage receiving any concomitant medication by ATC drug class and generic drug name. Patients reporting use of more than one medication at each level of summarization (any medication received, ATC class, and generic drug name) will be counted only once. ATC class terms will be displayed by descending order of incidence, as well as generic drug names within each ATC class. Prior medications used to treat WD and an additional analysis of all prior medications will be summarized similarly.

Prior medications will be defined as medications that were discontinued prior to the start of study drug. Concomitant medications will be defined as medications that either started prior to first dose of study drug and were continuing at the time of first dose of study drug or started on or after the date of the first dose of study drug. If it cannot be determined whether a medication was stopped prior to the start of study drug dosing due to partial or missing medication start or end dates, it will be considered a concomitant medication.

Prior and concomitant medications will be presented in a by-patient data listing by patient and medication name.

7.2. Efficacy Analyses

In general, all primary and key secondary efficacy endpoints will be assessed in both the FA and PP Sets. The non-key secondary and exploratory efficacy endpoints will only be assessed in the FA Set. The principal analysis for all efficacy endpoints will be performed on the overall population (Cohort 1 and Cohort 2 combined) in the FA set. Corresponding supportive analyses will be performed in Cohort 1 and Cohort 2. All efficacy endpoint data will be summarized descriptively by visit, by randomized treatment (with cohorts combined), and by visit by randomized treatment within each cohort. Supporting listings will also be presented. Least square (LS) means ± standard error (SE), when calculated, will be displayed graphically over time by randomized treatment group for primary and secondary endpoints.

For endpoints which do not have a baseline assessment (eg, CGI-I, TSQM-9 scores), the results at each time point will be summarized.

7.2.1. Primary Analysis

The primary estimand is the difference in daily mean dNCC $AUEC_{0\text{-}48W}$ between ALXN1840 and SoC in patients with WD, regardless of less-than-complete adherence or use of another medication that affects plasma dNCC, with no benefit derived from treatment after death.

Treatment: ALXN1840 and SoC

Population: Patients with WD who meet all inclusion and no exclusion criteria.

Variable/endpoint: daily mean dNCC $AUEC_{0\text{-}48W}$

Intercurrent Events: For patients who die, a composite strategy will be applied reflecting no benefit derived from treatment after death. The baseline value of plasma dNCC will be carried forward from the point of discontinuation to Week 48.

For patients with less-than-complete adherence, including discontinuation of treatment for any reason other than death, the treatment policy strategy will be applied. The missing values of plasma dNCC will be imputed based on the response for SoC patients.

For patients who use another medication that affects plasma dNCC the treatment policy strategy will be applied.

Population Level Summary: Comparison of the mean of dNCC $AUEC_{0\text{-}48W}$ between treatment groups.

Overall Analysis:

The primary endpoint is the daily mean dNCC $AUEC_{0\text{-}48W}$. The AUEC for dNCC concentration will be calculated using the trapezoidal rule.

$$AUEC_{0-t_n} = \sum_{i=0}^{n} \frac{c_{i-1}+c_i}{2} (t_i - t_{i-1}) \text{ , n= number of measured timepoints}$$

The AUEC is then divided by the number of days to yield daily mean dNCC $AUEC_{0\text{-}48W}$. The group means for daily mean dNCC $AUEC_{0\text{-}48W}$ will be compared between ALXN1840 and SoC using an analysis of covariance (ANCOVA) statistical model; treatment arm, baseline dNCC, and cohort, will be included in the model. Tests will be performed at a significance level of 0.05 (2-sided).

Model-based estimates of the difference between randomized treatments (ALXN1840 – SoC) in dNCC $AUEC_{0\text{-}48W}$, along with a 2-sided 95% CI and p-value will be provided. If the lower 2-sided 95% CI exceeds 0 μM, then superiority will be concluded.

Accounting for Sample Size Re-estimation:

In the present study, the interim analysis for sample size re-estimation (SSR) was performed using the data from the first approximately 35% patients evaluable for the primary analysis. The decision was to keep the original sample size, with no adaptation. When no adaptation occurs (ie, interim decision to keep the original planned sample size), then the conventional statistic can be used to determine statistical significance (Wassmer, 2016).

In contrast, if an adaptation had occurred (ie, interim decision to increase the sample size), then the final analysis would have used the weighted statistic proposed by Cui, Hung, and Wang (1999) in which the independent increments of the Z statistics of the 2 stages are combined by prespecified weights that are based on the planned proportion of total number of patients. The planned proportions were 0.35 and 0.65 for the 2 stages, and the independent incremental Z statistics for the 2 stages combined with weights that equal the square root of 0.35 for Stage 1 and square root of 0.65 for Stage 2. Patients assessed for the primary efficacy endpoint prior to the time the interim analysis would make up the Stage 1 sample, and those patients assessed after the interim analysis would make up the Stage 2 sample. Based on the Stage 1 and Stage 2 patient samples, the 2-independent stage-wise test statistics for superiority, Z_1 and Z_2, would be computed. If an adaptation had occurred, the final test statistics would be:

$$Z_{\mathrm{CHW}} = \sqrt{w}\, Z_1 + \sqrt{1-w}\, Z_2$$

where $w = 0.35$, the planned information fraction at the time of the interim analysis. Since no adaptation occurred, the conventional statistics will be used.

7.2.1.1. Handling of Dropouts or Missing Data

For intermediate missing data, interpolation will be used to fill out missing values.

For patients who die during study, their baseline plasma dNCC concentration will be carried forward.

For patients who drop out due to other reasons, multiple imputation will be used to impute missing dNCC concentration assuming data are missing not at random. In this case, responses for both treatment groups will be imputed with multiple imputation based on the response for SoC patients (jump to reference method). Cohort, baseline, and previous visit value will be included in the imputation model. The following steps will be followed:

Step 1: Missing values at each visit will be imputed 10 times via SAS PROC MI, and imputed values will be retained in the imputed datasets.

Step 2: For each imputed dataset, daily mean dNCC $AUEC_{0\text{-}48W}$ will be derived using the trapezoidal rule (Yeh, 1978).

Step 3: ANCOVA analysis will performed for each of the 10 imputed dataset, with LS means (SE) for each treatment arm saved from each of the 10 analyses.

Step 4: 10 sets of analysis results will be combined using Rubin's rules (Rubin, 1987) via SAS PROC MIANALYZE. The treatment differences, CIs and p-values will be estimated by MODELEFFECTS and STDERR statement.

Tipping Point Analysis:

A tipping point analysis will be conducted for the primary endpoint, by splitting the patients into 2 subgroups: "observed" and "missing". "Observed" patients are those with baseline and Week 48 assessment in plasma dNCC concentration, while "missing" refers to all other randomized patients. An efficient 2-dimensional tipping point analysis will be accomplished by performing two 1-dimensional analyses: (a) vary assumptions about the missing outcomes in the ALXN1840 arm, while fixing those about the SoC arm, and (b) vary assumptions about the missing outcomes in the SoC arm, while fixing those in the ALXN1840 arm. The combination of assumed missing data values that maintain the statistically significant conclusion are bound by the two one-way tipping points (Gorst-Rasmussen, 2020).

Step 1: Within the "observed" subgroup, ANCOVA analysis will be performed on the daily mean dNCC $AUEC_{0\text{-}48W}$ in the 2 treatment arms, with cohort and baseline dNCC as covariates. The LS means and SEs for ALXN1840 and SoC, respectively, and the residual error estimate, will be extracted.

a. ALXN1840 varied, SoC fixed

Step 2:

i. For "missing" SoC patients, the daily mean dNCC $AUEC_{0\text{-}48W}$ concentration is generated randomly as a deviate from a normal distribution with mean equal to the LS mean for "observed" SoC patients, and SD equal to the residual error estimate, as obtained in Step 1.

ii. For "missing" ALXN1840 patients, the LS mean for daily mean dNCC $AUEC_{0\text{-}48W}$ in Step 1 will be decreased by a small increment (δ=10%), to reflect a marginally worse $AUEC_{0\text{-}48W}$ value. For these patients, the daily mean dNCC $AUEC_{0\text{-}48W}$ will be generated randomly as a deviate from a normal distribution with mean equal to (100% − δ) of the LS mean for "observed" ALXN1840 patients, and SD equal to the residual error estimate, as obtained in Step 1.

iii. Combine "observed" and "missing" patients into complete dataset. Analyze complete dataset using ANCOVA with cohort and baseline dNCC as covariates, to give LS means and SEs for ALXN1840 and SoC, with treatment effect estimate and associated SE.

Step 3:

i. Step 2 will be repeated 10 times for the same increment δ, giving rise to 10 sets of LS means, treatment effect estimates, and associated SEs.

ii. The 10 treatment effect estimates will be combined via PROC MIANALYZE using Rubin's method.

Step 4: Steps 2 and 3 will be repeated with larger increments (δ = 20%, 30%,…, 90%) to reflect an even-poorer outcome with ALXN1840 treatment.

Step 5: The treatment effect estimates, associated SEs, and CIs generated for all increments (δ = 10%, 20%,…, 90%) will be summarized in a table, and the first point at which superiority fails (if it exists) will be highlighted.

b. ALXN1840 fixed, SoC varied

Step 2:

i. For "missing" ALXN1840 patients, the daily mean dNCC $AUEC_{0\text{-}48W}$ is generated randomly as a deviate from a normal distribution with mean equal to the LS mean for "observed" ALXN1840 patients, and SD equal to the residual error estimate, as obtained in Step 1.

ii. For "missing" SoC patients, the LS mean for daily mean dNCC $AUEC_{0\text{-}48W}$ in Step 1 will be increased by a small increment (δ=10%), to reflect a marginally better $AUEC_{0\text{-}48W}$ value. For these patients, the daily mean dNCC $AUEC_{0\text{-}48W}$ will be generated randomly as a deviate from a normal distribution with mean equal to (100% + δ) of the LS mean for "observed" SoC patients, and SD equal to the residual error estimate, as obtained in Step 1.

iii. Combine "observed" and "missing" patients into complete dataset. Analyze complete dataset using ANCOVA with cohort and baseline dNCC as covariates, to give LS means and SEs for ALXN1840 and SoC, with treatment effect estimate and associated SE.

Step 3:

i. Step 2 will be repeated 10 times for the same increment δ, giving rise to 10 sets of LS means, treatment effect estimates, and associated SEs.

ii. The 10 treatment effect estimates will be combined via PROC MIANALYZE using Rubin's method.

Step 4: Steps 2 and 3 will be repeated with larger increments (δ = 20%, 30%,…, 200%) to reflect an even-better outcome with SoC treatment.

Step 5: The treatment effect estimates, associated SEs, and CIs generated for all increments (δ = 10%, 20%,…, 200%) will be summarized in a table, and the first point at which superiority fails (if it exists) will be highlighted.

7.2.1.2. Subgroup Analysis

For exploratory purposes, the primary and secondary efficacy endpoints will also be evaluated in clinically relevant subgroups. As a general rule, a subgroup is analyzed only if the number of patients is ≥ 3 in each treatment arm.

- Previous treatment for WD (within Cohort 1, 2, and total).
  i. Zinc monotherapy vs penicillamine or trientine (± Zinc)
  ii. Zinc monotherapy vs penicillamine (± Zinc) vs trientine (± Zinc)
- Previous treatment for WD (Cohort 2 only)
  - Completely naïve (ie, zero days prior treatment for WD) vs minimally treated (ie, 1 to ≤ 28 days treatment prior treatment for WD).
- SoC treatment during Primary Evaluation Period:
  i. Zinc monotherapy vs penicillamine or trientine (± Zinc)

ii. Zinc monotherapy vs penicillamine (± Zinc) vs trientine (± Zinc)

- Age in years group (≥ 12-< 18, ≥ 18), (≥ 12-<18, ≥18-< 65, ≥ 65), (<25, ≥ 25)
- Patients who have an incomplete or intolerant response to prior WD treatment. Incomplete and intolerant responders will be combined into a single subgroup.
    - "Incomplete" responder is defined as a patient who meets at least 1 of the following criteria:
        1. UWDRS Part II total score at Baseline: > 0
        2. UWDRS Part III total score at Baseline: > 0
    - "Intolerant" responder is defined as a patient who meet at least 1 of the following criteria:
        1. Wilson Disease treatment history "Reason for discontinuation" which indicates intolerance, as reviewed by the Medical Monitor. This review will be blinded to patient and treatment identifiers, and be completed prior to database lock.
        2. Adverse event observed prior to first dose on Day 1 with "Action taken with SoC" as "Drug withdrawn".
- Symptomatic vs pre-symptomatic. Presence of symptoms will be assessed by the following criteria independently:
    - Cirrhosis at Baseline (Yes/No)
    - UWDRS Part II total score at Baseline: > 0 or = 0
    - UWDRS Part III total score at Baseline: >0 or = 0
- Symptomatic vs pre-symptomatic adolescents (≥ 12 - <18 years). Presence of symptoms will be assessed by the following criteria independently:
    - Cirrhosis at baseline (Yes/No)
    - UWDRS Part II total score at Baseline: > 0 or = 0
    - UWDRS Part III total score at Baseline: >0 or = 0
- Sex (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska native, Native Hawaiian or other Pacific Islander, Other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino, Unknown)
- Country/Region (United States, EU, United Kingdom, Germany, Japan, Rest of World)
- BMI Group : <25 kg/m^2, 25-30 kg/m^2, >30 kg/m^2
- Liver impairment status at entry defined as MELD score above (>) or below (≤) the median MELD score at Baseline
- UWDRS Part II total score at Baseline: > 0 or = 0

- UWDRS Part III total score at Baseline: > 0 or = 0
- UWDRS Part III Functional Subscale: > 0 or = 0
- UWDRS ADL Subgroups at Baseline
  - No reported ADL limitations
  - Functionally independent
  - Functionally dependent
- Cirrhosis at Baseline (Yes/No)
- Psychiatric symptoms at Baseline (Yes/No) - at least 1 symptom present on BPRS-24
- Overall Mean Daily Dose Group (< 15 mg, 15- < 30 mg, ≥ 30 mg)
- Renal Function Status at Baseline
  - G1
  - G2
  - G3a/G3b/G4/G5 combined
- ALT at Baseline: > or ≤ ULN

7.2.1.3. Multicenter Studies

Subgroup analyses will be performed for country/region (ie, United States, EU, United Kingdom, Germany, Japan, Rest of World).

7.2.1.3.1. Japanese Specific Analysis

Subgroup analyses for Japan will be based on Japanese patients, which are defined as patients enrolled in the country of Japan and enrolled in the Primary Evaluation Period. The layout of the tables and figures will be the same as the global tables and figures for Study WTX101-301. The table/figure number will be the same as the original number, however, a "J" will be added to the end. To the end of original title of the table/figure, "(Japanese)" will be added. The analysis will be performed for the following sections:

- Disposition of patients
- Demographics and baseline characteristics
- Primary efficacy endpoints
- Secondary efficacy endpoints
- Exploratory endpoints
- AEs.

A descriptive efficacy analysis will be performed on the Japanese patients in the FA Set. No formal hypothesis tests or p-values on the Japanese patients alone will be conducted or reported, due to the small number of patients in Japan. Instead, summary statistics will be computed and

displayed by treatment within cohorts in the same format as global tables. The subgroup analysis listed in Section 7.2.1.2 will not be needed for Japanese specific analysis.

The safety analysis will only include AEs, which will be performed on the Japanese patients in the Safety Set.

Figures will be created based on the Japanese patients, if applicable.

Listings will not be created for Japanese specific analysis.

7.2.1.4. Hypothesis Testing and Significance Level

The efficacy assessments will be performed for the following primary and key secondary endpoints:

- Endpoint 1 (primary): Daily mean dNCC $AUEC_{0\text{-}48W}$
- Endpoint 2: Change from baseline in the UWDRS Part II
- Endpoint 3: Change from baseline in UWDRS Part III Functional Subscale
- Endpoint 4: Change from baseline in UWDRS Part III Arising from a chair
- Endpoint 5: Change from baseline in UWDRS Part III Gait
- Endpoint 6: Change from baseline in UWDRS Part III Handwriting
- Endpoint 7: Change from baseline in UWDRS Part III Speech

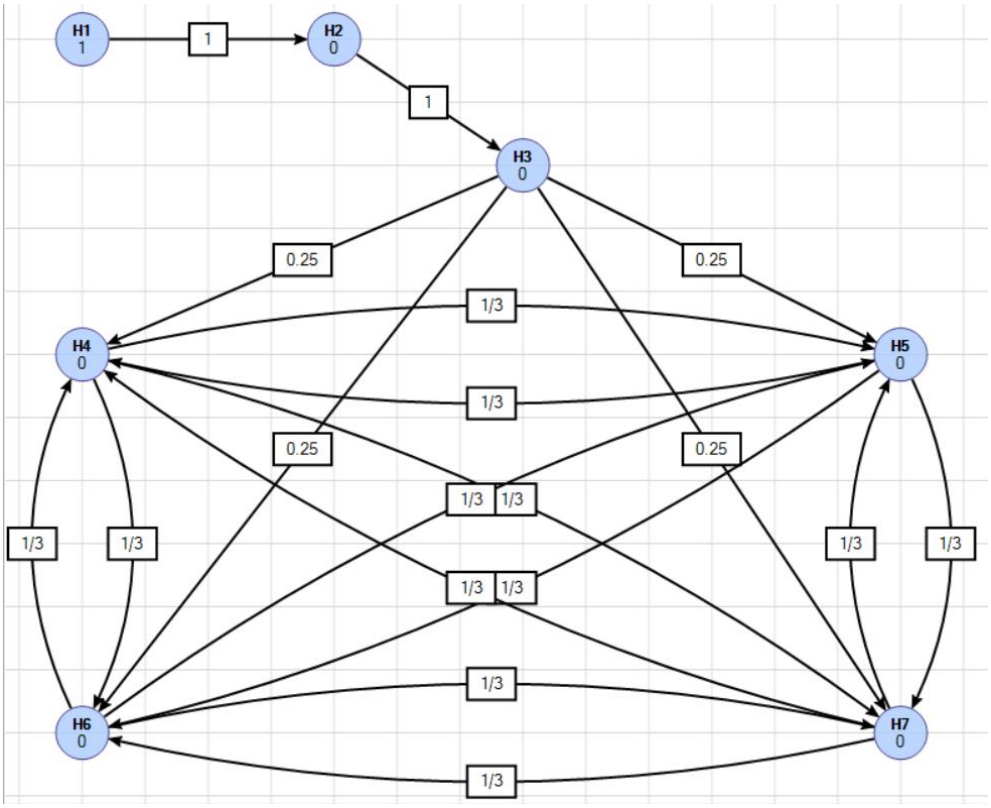
The corresponding null hypotheses will be labeled as follows:

- H_1: Null hypotheses of no effect for Endpoint 1 (dNCC $AUEC_{0\text{-}48W}$).
- H_2: Null hypotheses of no effect for Endpoint 2 (UWDRS Part II).
- H_3: Null hypotheses of no effect for Endpoint 3 (UWDRS Part III Functional Subscale).
- H_4: Null hypotheses of no effect for Endpoint 4 (UWDRS Part III Arising from a chair).
- H_5: Null hypotheses of no effect for Endpoint 5 (UWDRS Part III Gait).
- H_6: Null hypotheses of no effect for Endpoint 6 (UWDRS Part III Handwriting).
- H_7: Null hypotheses of no effect for Endpoint 7 (UWDRS Part III Speech).

A multiplicity adjustment will be applied to address the following 2 sources of multiplicity: (1) analysis of the primary/key secondary endpoints, and (2) SSR rule at the interim analysis. The proposed multiplicity adjustment guarantees strong control of the overall Type I error rate control with respect to these sources of multiplicity at two-sided $\alpha = 0.05$. The multiplicity adjustment relies on a prospectively defined multiple testing procedure. This procedure will be applied in conjunction with the combination function approach (Wassmer, 2016) at the interim and final analyses to ensure strong Type I error rate control. The combination function approach will be applied using the modification proposed in Sugitani et.al (Sugitani,2016).

To address multiplicity induced by the first source of multiplicity in this study, namely, multiplicity induced by the analysis of the primary/key secondary endpoints, a multiple testing procedure will be applied. This procedure will be set up using the graphical method (Bretz, 2009). The multiplicity adjustment strategy will be applied using the serial testing algorithm (Dmitrienko, 2013). Figure 1 illustrates the proposed multiple testing strategy.

Figure 1: Proposed Multiple Testing Strategy.



The hypothesis weights are set up as follows:

- All of the weight is initially assigned to the hypothesis H_1, ie, $w_1 = 1$.
- The other hypothesis weights are set to 0, ie, $w_i = 0$, $i = 2, \ldots, 7$.

The values of the transition parameters are defined as:

- $g_{12} = 1$
- $g_{23} = 1$
- $g_{34} = 1/4$, $g_{35} = 1/4$, $g_{36} = 1/4$, $g_{37} = 1/4$,
- $g_{45} = 1/3$, $g_{46} = 1/3$, $g_{47} = 1/3$,
- $g_{56} = 1/3$, $g_{57} = 1/3$, $g_{54} = 1/3$,
- $g_{67} = 1/3$, $g_{64} = 1/3$, $g_{65} = 1/3$.
- $g_{74} = 1/3$, $g_{75} = 1/3$, $g_{76} = 1/3$.

To account for the second source of multiplicity (data-driven decision rule at the interim analysis), the multiple testing procedure will be applied separately to the 2 study stages:

- Stage 1 includes all patients who are included in the interim database.
- Stage 2 includes all patients who are not included in the interim database.

The multiple testing procedure will account for the interim decision rule to support inferences at the final analysis. For example, if a decision is made to increase the sample size in the study, the evidence of treatment effectiveness from the 2 study stages will be pooled with predefined

weights using the modified combination function principle. The standard analysis will be performed at the final analysis if the sample size is not adjusted at the interim analysis.

Let p_1 through p_7 denote the two-sided treatment effect p-values for the null hypotheses of interest computed from the Stage 1 data. Similarly, let q_1 through q_7 denote the two-sided treatment effect p-values for the null hypotheses computed from the Stage 2 data. If an adaptation at the interim analysis occurred, the inferences for the hypotheses H_1 through H_7 would be performed at the final analysis as follows: The combined p-value for the hypothesis H_i is denoted by r_i, $i = 1, \ldots, 7$, and is derived using the weighted inverse-normal combination function, i.e.,

$$c(x, y) = 1 - \Phi\left(\sqrt{w}\Phi^{-1}(1 - x) + \sqrt{1 - w}\Phi^{-1}(1 - y)\right),$$

where $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution, and $w = 0.35$ and $1 - w = 0.65$ are the predefined weights assigned to Stages 1 and 2.

The combined p-value for the hypothesis H_i would be computed as follows:

$r_i = c(p_i, q_i)$.

The resulting combined p-values would be passed to the multiple testing procedure to perform inferences at the final analysis, ie, the adjusted p-values will be computed for the null hypotheses and each null hypothesis will be rejected at the final analysis if its adjusted p-value does not exceed a two-sided $\alpha = 0.05$. The proposed multiplicity adjustment guarantees overall Type I error rate control in the strong sense with respect to both sources of multiplicity in the study. There will be no Type I error adjustment made for non-key and exploratory efficacy endpoints.

Since no adaptation for sample size increase occurred at the interim analysis, this second source of multiplicity will not be relevant to the final analysis.

7.2.1.5. Sensitivity Analyses

Sensitivity analyses for the primary endpoint only will be performed using different imputation methods on the missing data mentioned in the Section 7.2.1.1.

7.2.2. Secondary Analyses

Similar to the primary estimand, the key secondary estimand is the difference in UWDRS score between ALXN1840 and SoC in patients with WD, regardless of less-than-complete adherence or use of another medication that affects plasma dNCC, with no benefit derived from treatment after death. This estimand applies to all key secondary endpoints.

The secondary endpoints will be analyzed via mixed model for repeated measures (MMRM) analysis with the principal contrast being between SoC versus ALXN1840-treated patients at Week 48. Fixed-effect terms will be included for randomized treatment (ALXN1840 or SoC), cohort, visit, baseline by visit interaction, randomized treatment by visit interaction, and baseline level as a covariate. The treatment by visit interaction will remain in the model regardless of significance. An unstructured covariance matrix will be used to model the within-patient error and the Kenward Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures

will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, and autoregressive.

The key secondary endpoints include:

- Change from baseline in the UWDRS Part II total score.
- Change from baseline in the UWDRS Part III Functional Subscale.
- Change from baseline in the UWDRS Part III individual functional items: arising from a chair, gait, speech, and handwriting)

Other non-key secondary endpoints include:

- CGI-I scale
- Change from baseline in CGI-S scale
- Change from baseline in hepatic status assessed by the MELD score;
- Change from baseline in the UWDRS Part III total score.
- Change and percentage change from baseline (Day 1) in cNCC in plasma. For ALXN1840-treated patients, the cNCC in plasma will be corrected for the amount of copper bound to the ALXN1840 TPC ($cNCC_{corrected}$)
- cNCC/$cNCC_{corrected}$ responder rate at 48 weeks

For cNCC/$cNCC_{corrected}$ responder rate at 48 weeks, patients without 48-week values will be considered as nonresponders. These data will be analyzed via logistic regression with terms for randomized treatment, cohort, and baseline cNCC level.

7.2.3. Exploratory Analyses

The following exploratory endpoints will be analyzed via MMRM analysis with the principal contrast being between SoC versus ALXN1840-treated patients at Week 48. The parameters of total Cu, total Mo, Cp, PUF-Cu, PUF-Mo, 24-hour urine Cu, 24-hour urine Mo, and speciation profile parameters may be log transformed prior to analysis:

- Change from baseline in the FIB-4 Index;
- Change from baseline in Modified Nazer Score;
- Change from baseline in BPRS-24;
- Change from baseline in EQ-5D using the UK health states (including the EQ-5D-5L Descriptive System and the EQ VAS);
- Plasma total Cu, PUF-Cu, dNCC, LBC, Cp, and CpC concentration-time profiles in plasma;
- Total Mo and PUF-Mo concentration-time profiles in plasma;
- Change in 24-hour urinary Cu and urinary Mo;
- Change from baseline in the Timed 25F Walk Test;
- Change from baseline in UWDRS Part I and total score;
- Change from baseline in UWDRS Part II - FDA suggested scoring;

- Change from baseline in UWDRS Part III individual items - FDA suggested scoring: arising from chair, gait, speech, and handwriting.
- Change from baseline in UWDRS Part III Functional Subscale-FDA suggested scoring

The following endpoints will be analyzed using ANCOVA model on the change between Week 48 and baseline. For 9-HPT, nonverbal Stroop Interference Test, and Digit Span Test, an ANCOVA model on the change between baseline and Week 12 and between baseline and Week 24 will also be applied. A fixed-effect term will be included for randomized treatment (ALXN1840 or SoC) and the baseline value will be included as a covariate. LS means will be provided for each randomized treatment group at Week 48 together with the difference in LS means and associated 2-sided 95% CI and p-value.

- Change from baseline in transient elastography;
- Change from baseline in the 9-HPT;
- Change from baseline in the nonverbal Stroop Interference Test;
- Change from baseline in the Digit Span Test.

The 3 most troublesome symptoms will be analyzed at Weeks 12, 24 and 48 using cumulative logit modelling using PROC LOGISTIC in SAS.

The odds ratio and 95% 2-sided CI will be displayed, along with the 2-sided p-value. Model-based estimates of the proportion of patients in each randomized treatment group with responses of ‘improved’, ‘remained the same’, or ‘worsened’ will be computed.

The daily mean AUEC of plasma dNCC (from 0 to 24 and 24 to 48 weeks), daily mean AUEC of plasma total copper from 0 to 48 weeks, and daily mean AUEC of LBC from 0 to 48 weeks will be analyzed in the same manner as the primary endpoint, via ANCOVA analysis with the principal contrast being between SoC versus ALXN1840 treated patients at Week 48.

Descriptive summary and Kaplan-Meier Analysis will be applied to time to first confirmed increase in plasma dNCC and total copper concentration, time to minimum and maximum of PK/PD parameters and biomarkers, and time for return to predose baseline for PK/PD parameters and biomarkers.

- Time to first confirmed increase in plasma dNCC and total copper concentration
- Time to minimum and maximum concentration and the maximum concentration of:

– − Plasma total copper concentration

– − Plasma dNCC concentration

– − Plasma LBC concentration

– − Ratio plasma dNCC:total copper concentration

– − Ratio plasma LBC:total copper concentration

– − Urinary molybdenum concentration

– − Ratio urinary molybdenum:copper concentration

- − Ratio urinary molybdenum:dosed molybdenum concentration
- − Plasma Cp concentration
- − Plasma CpC concentration
- − Ratio plasma Cp:total copper concentration
- − Ratio plasma CpC:total copper concentration
- − Ratio plasma CpC:Cp

- Time for return to predose baseline for the following PK/PD parameters:
- − Plasma total copper
- − Plasma dNCC
- − Plasma LBC
- − Ratio plasma dNCC:total copper
- − Ratio plasma LBC:total copper
- − Urinary molybdenum
- − Ratio urinary molybdenum:copper
- − Ratio urinary molybdenum:dosed molybdenum
- − Plasma Cp concentration
- − Plasma CpC concentration
- − Ratio plasma Cp:total copper
- − Ratio plasma CpC:total copper
- − Ratio plasma CpC:Cp

LBC responder rate will be analyzed the same as the cNCC responder rate analysis described in Section 7.2.2. Note, TSQM-9 will be summarized descriptively.

Relationship Between Copper and Key Secondary Endpoints:
Pearson's correlation coefficient will be estimated to provide summaries of the relationship between copper quantification methods and all key secondary endpoints. For plasma total copper $AUEC_{0\text{-}48W}$ and dNCC $AUEC_{0\text{-}48W}$, the key secondary endpoints will be measured as change from baseline to Week 48 (Table 3).

Table 3: Correlations Between Copper and Key Secondary Endpoints (Change from Baseline)

	dNCC $AUEC_{0-48W}$	PTC $AUEC_{0-48W}$
Change from baseline to Week 48 in the UWDRS Part II total score	X	X
Change from baseline to Week 48 in the UWDRS Part III Functional Subscale	X	X
Change from baseline to Week 48 in arising from a chair	X	X
Change from baseline to Week 48 in gait	X	X
Change from baseline to Week 48 in handwriting	X	X
Change from baseline to Week 48 in speech	X	X

Abbreviations: dNCC = directly measured non-ceruloplasmin-bound copper; PTC = plasma total copper

For plasma total copper, dNCC, plasma LBC, cNCC/$cNCC_{corrected}$, the correlations with key secondary endpoints will be calculated for all visits in the following time windows (Table 4):

- Postbaseline period: Week 4 to Week 48
- 24-Week copper mobilization period: Week 4 to Week 24
- Post 24-Week copper maintenance period: Post Week 24 to 48 (> Week 24 - Week 48)

Table 4: Correlations Between Copper and Key Secondary Endpoints (Measured Values)

	PTC	dNCC	Plasma LBC	cNCC/$cNCC_{corrected}$
UWDRS Part II total score	X	X	X	X
UWDRS Part III Functional Subscale	X	X	X	X
Arising from a chair	X	X	X	X
Gait	X	X	X	X
Handwriting	X	X	X	X
Speech	X	X	X	X

Abbreviations: dNCC = directly measured non-ceruloplasmin-bound copper; PTC = plasma total copper; LBC = labile bound copper; cNCC = calculated non-ceruloplasmin-bound copper calculated non-ceruloplasmin-bound $cNCC_{corrected}$ = copper corrected for copper bound in tetrathiomolybdate-copper-albumin complexes

7.2.4. Extension Period Analyses

Extension Period Analyses will be fully described in a separate SAP.

7.2.5. Pharmacokinetic and Pharmacodynamic Analyses

Raw plasma concentration values, including plasma total molybdenum, PUF-Mo, total copper, PUF-Cu, dNCC, LBC, Cp, and CpC as well as urine total molybdenum and copper will be summarized by randomized treatment within treatment cohort and time point using descriptive statistics, to include the geometric mean and coefficient of variation (CV; %). When calculating the geometric mean, values of 0 will be discarded.

Concentration-time profiles in plasma will be displayed in graphical format for each patient and will be plotted by dose. The mean concentration-time profiles including all patients in the PK and PD/biomarker populations will also be plotted. Concentration-time profiles for plasma total molybdenum, PUF-Mo, total copper, PUF-Cu, dNCC, LBC, Cp, and CpC as well as urine total

molybdenum and copper will be displayed in units of ng/mL and µmol/L, calculated by dividing the concentrations by the relative atomic mass (average molecular weight) of copper (63.546) or molybdenum (95.95).

A by-patient listing will also be provided in units of ng/mL and µmol/L.

7.3. Safety Analyses

Safety and tolerability over the course of the study period is a secondary endpoint and will not be subject to formal analysis. Rather, descriptive statistics will be used to summarize the safety data by randomized treatment within cohort and overall. The Safety Set will be used for the evaluation of safety in the Primary Evaluation Period.

7.3.1. Study Duration, Treatment Compliance, and Exposure

Study duration will be summarized for all enrolled patients by treatment within each cohort. Treatment compliance and exposure will be summarized for the Safety Set by treatment within each cohort. Supportive listings will also be provided.

7.3.1.1. Study Duration

Study duration is defined as the time from informed consent to the end of the Primary Evaluation Period (ie, Week 48) or study discontinuation date, whichever occurs first.

7.3.1.2. Treatment Compliance

The treatment compliance will be defined as:

Compliance (%) = (number of dose received)/(total number of dose scheduled) × 100.

Furthermore, the compliance will be summarized in the following ways:

- Compliance (%)
- Compliance (< 80%, ≥ 80%)

Treatment compliance will be summarized using descriptive statistics by visit, by randomized treatment within cohort and overall. Compliance up to Week 48 (Primary Evaluation Period) will be summarized for all patients. If a patient prematurely discontinues during the Primary Evaluation Period, his or her compliance will be based on the period up to the point of discontinuation from the study. Additionally, drug interruptions or missed doses as the result of a physician decision or AE will be factored into the number of doses scheduled, whereas drug interruptions or missed doses for any reason will be factored into calculating the number of doses received. A supporting listing will also be produced.

7.3.1.3. Exposure

The number of patients exposed to the study drug will be summarized in terms of counts for the overall duration of the study, and by periods specified in International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E3 guideline. The number of patients will also be summarized by the overall mean daily dose groups (< 15 mg, ≥ 15 to < 30 mg, ≥ 30 mg) for the overall duration and in each specified time period listed below.

- 1 day or less
- 2 days to 1 week
- 1 week to 1 month
- 1 month to 6 months
- 6 months to 1 year
- > 1 year

The duration (days) of exposure to treatment will be calculated as date of last exposure to treatment – date of first dose + 1.

Patient-years (PY) of exposure will be derived individually for each patient. PY will be defined for each patient as the total time (in years) from first dose to the last available dose of study drug. Dose adjustments and/or interruptions will not be factored into this derivation. Results will be analyzed using descriptive statistics and the frequency and percentage of patients in the dose categories (< 15 mg, 15- < 30 mg, ≥ 30 mg).

The gap in drug exposure will not be included in the calculation of duration of exposureDosing will be described by mean daily dose, final daily dose, minimum daily dose, maximum daily dose, and total accumulative dose.

- Mean daily dose = sum (each dose × each dose frequency × each period)/total treatment period
- Final daily dose is defined as the last dispensed dose in titration studies
- Minimum daily dose is defined as the minimum daily dose over the Treatment Period
- Maximum daily dose is defined as the maximum daily dose over the Treatment Period
- Total accumulative dose is the sum of doses over the total Treatment Period

The dosing regimen was individualized and varied across protocols. The mean daily dose (mg), minimum daily dose (mg) and maximum daily dose (mg) will be summarized using descriptive statistics. All available dosing data will be presented in a listing. Where available, reasons for dose adjustment will also be presented in a listing.

Listings of exposure, drug interruptions and missed doses will also be presented.

7.3.2. Adverse Events

The verbatim terms as reported in the CRF by Investigators to identify AEs will be coded using the MedDRA, Version 20.0 or higher and summarized by primary System Organ Class and Preferred Term.

AE severity will be evaluated using the NCI CTCAE version 5.0 (published 27 Nov 2017).

AE causality is determined by the Investigator using the following assessment categories: unrelated or related.

Treatment-emergent adverse events (TEAEs) are defined as those AEs with onset after the first dose of randomized treatment or existing events that worsened in severity after the first dose of randomized treatment. Events reported with a partial onset date (eg, month and year are reported but the day is missing) will be considered to be treatment-emergent if it cannot be confirmed that

the event onset was prior to the first dose of study drug based on the available date entries. Additionally, for patients who were originally randomized to SoC in the Primary Evaluation Period and switch to ALXN1840 in the Extension Period, any AEs which initiate after the switch or existing events that worsen in severity will be attributed to ALXN1840.

Although no formal statistical comparison of AEs will be done, exposure-adjusted incidence rates in addition to raw cumulative incidence proportions will be provided for all analyses of safety data. The exposure years will be included in the AE summary tables. Estimated treatment differences and confidence intervals for treatment arm comparisons will be reported for all analyses of safety data. Descriptive analysis to compare treatment groups with respect to risk (eg, with a risk difference, along with a confidence interval) will be reported for AESIs.

7.3.2.1. Overall Summary of Adverse Events

An overall summary of TEAEs will be presented, including frequency of patients experiencing the event (n) and relative frequency (n/N*100, where N is the number of patients in the Safety Set). The summary will include categories indicating how many events are TEAEs, treatment-emergent SAEs, and treatment-emergent non-SAEs. Within TEAEs, the following subcategories will also be summarized:

- Severity of TEAEs (Grade 1 through Grade 5)
- Related TEAEs (not related, related)
- TEAEs leading to withdrawal of study drug
- TEAEs leading to death
- AESIs

A summary of events (n) and number of patients with events (n, %) for pretreatment-emergent adverse events (PTAEs) will also be included with its relevant subcategories.

A listing of all TEAEs by treatment cohort and patient will be presented. Separate listings will be produced for SAEs, AEs leading to study drug withdrawal, AEs resulting in death, AEs leading to withdrawal from the study and PTAEs.

7.3.2.2. AEs by System Organ Class and Preferred Term

The number of TEAEs and the number and percentage of patients with events will be presented by System Organ Class and Preferred Term. Patients are counted once in each System Organ Class and Preferred Term. Percentages will be based on the Safety Set. System Organ Class will be listed in descending frequency as will Preferred Terms within each System Organ Class. If needed, terms will also be ordered alphabetically.

Treatment-emergent SAEs, treatment-emergent non-SAEs, TEAEs leading to withdrawal of study drug, TEAEs leading to death, treatment-emergent AESIs, PTAEs, and TEAEs occurring in $\geq 5\%$ or $\geq 10\%$ of patients will be summarized using the same approach.

7.3.2.3. AEs by System Organ Class

The number of TEAEs and the number and percentage of patients with events will be presented by System Organ Class. Patients are counted once in each System Organ Class. Percentages will be based on the total number of treated patients in the treatment cohort.

7.3.2.4. AEs by Preferred Term

The number of TEAEs and the number and percentage of patients with events will be presented by Preferred Term. Patients are counted once in each Preferred Term. Percentages will be based on the total number of treated patients in the treatment cohort.

7.3.2.5. AEs by System Organ Class, Preferred Term, and Relationship

The number of TEAEs and the number and percentage of patients with events will be presented by System Organ Class and Preferred Term as described in Section 7.3.2.2 by relationship (related, not related). If a patient has more than one occurrence of an AE, the strongest relationship to study treatment will be used in the summary table. If relationship to study drug is missing, the AE will be assumed to be related. A similar analysis will be conducted for treatment-emergent SAEs.

The number of related TEAEs and the number and percentage of patients with related TEAEs will be summarized by System Organ Class and Preferred Term, and separately by Preferred Term only. The same analyses will be produced for related treatment-emergent SAEs.

Lastly, the number of TEAEs by System Organ Class, Preferred Term and relationship, without taking into account the highest relationship, will be analyzed. A similar analysis will be conducted for treatment-emergent SAEs.

7.3.2.6. AEs by System Organ Class, Preferred Term, and Severity

The number of TEAEs and the number and percentage of patients with events will be presented by System Organ Class, Preferred Term, and severity. If a patient has more than one occurrence of an AE, the highest severity reported will be used. If severity is missing, the AE will be assumed to be severe. The number of TEAEs by system organ class, Preferred Term, and severity, without taking into account the highest severity, will also be analyzed.

Additionally, a summary of related TEAEs by System Organ Class, Preferred Term, and severity using the highest severity will be presented.

7.3.2.7. Deaths, Other Serious Adverse Events, and Adverse Events of Special Interest

SAEs and AESIs will be reportable from the time the patient signs the informed consent through the EOS Visit or until the Investigator and Alexion Pharmaceuticals determine that follow-up is no longer necessary. SAEs that are suspected to be drug related may be reported even if they occur more than 30 days after the patient is no longer on the study.

Any new neurological symptom or clinically significant worsening of an ongoing neurological symptom after initiation of study drug therapy will be designated to be an AESI, whether serious or nonserious. This includes all AEs in the MedDRA SOC “Nervous system disorders” or any AE where the “Is this an AESI?” check box is checked in the CRF.

AESIs will be assessed by a panel of 3 independent neurologists not participating in the study. The panel will assess the probability that clinically significant worsening or a new clinically significant neurological symptom is related to disease progression or caused by the study drug (ALXN1840 or SoC). They will be blinded to the treatment given to the patient. All available relevant patient information will be provided to this panel to aid in their assessment. The assessment of AESIs by the panel will be independent of and in addition to the usual assessments of the AE, including assessments of severity (intensity) and causality, by both the Principal Investigator and Alexion. A summary and listing of assessments given by the Neurological Adverse Event Panel will be presented.

A separate independent Hepatic Adjudication Panel, comprising experts in hepatology and drug-induced liver injury will be appointed by the Sponsor. The Hepatic Adjudication Panel will review and monitor study data for abnormalities of liver tests and liver function that may impact safety, effectiveness, and study conduct. A summary and listing of assessments given by the Hepatic Adjudication Panel will be presented.

7.3.2.8. Other Significant Adverse Events

Other significant TEAEs encompass those for abnormal liver function tests (a.k.a. hepatic events), hematopoietic cytopenias, and dyslipidemia. These include any new AE, or worsening of an ongoing AE, after initiation of study drug therapy that meet the criteria described below:

- Hepatic events: All AEs in the SMQ for Drug related hepatic disorder (comprehensive search), or any adverse event where the "Is this a Hepatic AE?" check box is ticked.
- Hematopoietic cytopenias: All AEs in the SMQ "Haematopoietic cytopenias".
- Dyslipidemia: All AEs in the SMQ "Dyslipidaemia".

Relationships to duration of study therapy and dose of study therapy may be explored along with examination within relevant subgroups, eg, WD severity at study entry prior, prior therapies received, age, and sex. WD severity at study entry may be assessed by the following:

- Cirrhosis status (yes, no)
- MELD score at baseline above (>) or below (≤) median
- Modified Nazer Score at baseline above (>) or below (≤) median
- Total Bilirubin (μmol/L) > ULN
- Platelets (10^9/L) < lower limit of normal (LLN)
- UWDRS Part II total score at baseline >0 or = 0
- UWDRS Part III total score at baseline >0 or = 0
- CGI-S at baseline above (>) or below (≤) median
- LBC (μmol/L) at baseline above (>) or below (≤) median
- Leukocytes (10^9/L) < LLN
- Neutrophils (10^9/L) < LLN

Prior therapies received may be assessed as follows:

- Cohort 1 will be categorized by prior SoC, where prior SoC is defined as the most recent treatment prior to or at time of Screening (whichever is later), and is categorized as penicillamine, trientine hydrochloride, zinc, or combination.
- Cohort 2 will be categorized into patients who are treatment naïve and those who have received SoC therapy for ≤ 28 days prior to first dose.

7.3.3. Other Safety Analyses

7.3.3.1. Analyses for Laboratory Tests

Actual values and changes from treatment baseline will be summarized descriptively for patients with available data for each laboratory parameter by randomized treatment within treatment cohort, for rollover patients and overall. Missing laboratory data will not be imputed, and only scheduled assessments will be included in by-visit summaries. A summary for "Last Assessment" will be included for the last available postbaseline result for each patient. A summary of "Worst Postbaseline" from all postbaseline data will also be included for the Primary Evaluation Period and for the Extension Period separately. All data, including that which is only collected at Screening, will be included in by-patient data listings. Laboratory measurements will be listed separately by patient, laboratory test, and unit.

Clinical laboratory measurements, including serum chemistry, hematology, urinalysis, and coagulation, will be summarized. Descriptive statistics will be presented for results and changes from baseline at each visit where parameters were scheduled to be collected per the clinical study protocol.

Where applicable, laboratory results will be classified as "low," "normal," or "high" with respect to the parameter-specific reference ranges (ie, below the lower limit of the normal range, within the normal range, or above the upper limit of the normal range). Contingency tables will be presented for each laboratory parameter to summarize the shift from the baseline category to all visits and to the worst postbaseline measurement, defined as the value numerically farthest outside of the normal range across all postbaseline visits through the end of the study.

Summary results will include the count and percentage of patients within each shift category by scheduled visit. Laboratory values outside the normal range will also be summarized and assessed for trends indicating a safety signal. Additionally, a summary and listing of liver enzyme elevation and a listing of DNA results will be presented.

Clinically Significant Laboratory Test Abnormalities

The following lab tests will be graded according the CTCAE V5.0 at ADaM level.

- **Hematology:** Absolute neutrophil count (neutrophil) (NEUT), total leukocytes (Leukocytes) (WBC), hemoglobin (HGB) (only do the grading for Anemia, no grading for Hemoglobin increased), platelet count (PLAT), Lymphocyte (LYM)
- **Coagulation:** International normalized ratio (INR)
- **Chemistry:** Alanine aminotransferase (ALT), albumin (ALB), alkaline phosphatase (ALP), aspartate aminotransferase (AST), total bilirubin (BILI), creatinine

(CREAT), gamma-glutamyltransferase (GGT), glucose (GLUC) (only grading for Hypoglycemia, no grading for Hyperglycemia), creatine kinase (CK), cholesterol (CHOL), and Triglycerides (TRIG).

Laboratory toxicity grade 0 will be derived for the shift table as: any non-missing results outside the range of the CTCAE criteria will be summarized as "Grade 0". Laboratory tests with missing either LLN or ULN or both will not be graded. Abnormal laboratory tests values will be summarized by CTCAE Laboratory Toxicity Grade. Shift from baseline tables of the number and percentage of patients in each of the CTCAE categories will also be presented for each treatment group for each parameter and visit time window (scheduled visits only). An overall shift summary will also be provided comparing baseline to worst postdose toxicity observed across all scheduled visits.

Lipid profile plots will be produced for individual patients, showing total cholesterol, HDL, LDL, and triglycerides, and ALT as a multiple of the respective baseline (BLN) values, and multiple of ULN values.

7.3.3.2. Evaluation of Drug Induced Serious Hepatotoxicity

A Hy's law case refers to an increase in aminotransferase > 3 × the reference ULN, with bilirubin > 2 × ULN. Possible Hy's law cases can be visualized with use of Evaluation of Drug Induced Serious Hepatotoxicity (eDISH) plots, a log-log scatter plot where the x-axis is the peak postbaseline ALT as a multiple of ULN, and the y-axis is the peak postbaseline total bilirubin as a multiple of ULN (Guo, 2009).

The following series of figures, adapted from Tesfaldet et al (Tesfaldet ,2016), will be produced for each treatment arm (ALXN1840 and SoC) separately for the Safety Analysis Set.

1. Distribution of ULN of liver serum enzymes. The ULN reference values used in each test (bilirubin, ALT, AST, ALP) and their respective frequencies in percentiles.
2. Distribution of baseline liver serum enzymes, for each test by cohort.
3. Distribution of baseline liver serum enzymes (× ULN), for each test by cohort.
4. Status of baseline liver serum enzymes as function of ULN by cohort. Baseline status categorized as Normal, > 1 ×, > 1.5 ×, > 2 ×, > 3 ×, > 5 ×, > 10 ×.
5. eDISH plot by treatment arm.
6. eDISH plot by quadrant.
7. Panel of eDISH quadrant shift plots at baseline. Patients are color coded to correspond to the quadrant they belong in the eDISH plot. The data presented in the panel corresponds to each patients' baseline value as a multiple of the ULN.
8. Panel of shift plots by eDISH quadrants. Patients are color coded to correspond to the quadrant they belong in the eDISH plot. The data is presented as multiples of the baseline value rather than the ULN. The shift in peak postbaseline laboratory value is compared to each patients' baseline value.

9. Panel of eDISH shift plots by eDISH quadrants. The panels represent the on-treatment eDISH quadrants whereas the colored symbols represent the eDISH quadrants of the baseline values (not postbaseline) as multiples of ULN.
10. Time course of liver tests as ×baseline value and ×ULN by cohort.

7.3.3.3. Vital Signs

Changes from baseline in vital signs (blood pressure, heart rate, respiratory rate, and temperature) at each visit will be summarized descriptively by randomized treatment within treatment cohort, for rollover patients and overall. Missing vital signs data will not be imputed and only scheduled assessments will be summarized in tables; unscheduled assessments will be presented in by-patient data listings. A summary for "Last Assessment" will be included for the last available postbaseline result for each patient. Summaries of shifts in blood pressure and heart rate from baseline values will also be created to allow detection of clinically relevant changes; these will be shifts of ± 20 mmHg for systolic or diastolic blood pressure and ± 20 beats per minute for resting heart rate. A listing of vital signs will be presented by randomized treatment, treatment cohort, patient, vital sign, and visit.

7.3.3.4. Electrocardiogram

All ECG data will be fully listed and changes from baseline in ECG data (heart rate, PR interval, RR interval, QRS duration, QT interval) will also be summarized descriptively by scheduled visit.

QT intervals will be corrected for heart rate according to Bazett ($QTcB = QT/[RR^{1/2}]$) and Fridericia ($QTcF = QT/[RR^{1/3}]$). At each time point, the number and percentage of patients falling into the following categories according to the ICH E14 Guidelines will be presented:

- QTc actual values: ≤ 450 ms, > 450 to ≤ 480 ms, > 480 to ≤ 500 ms, and > 500 ms
- QTc increases from baseline of > 30 msec and > 60 msec

Changes from baseline in QTc will also be summarized. Further, ECG data will be classified by the Investigator as "normal," "abnormal, not clinically significant," "abnormal, clinically significant" or "indeterminate" at each timepoint assessed. Contingency tables will be presented to summarize the shift from the baseline category to the worst postbaseline value. Summary results will include the count and percentage of patients within each shift category.

7.3.3.5. Relationship Between Copper and Safety Endpoints

Pearson's correlation coefficient will be estimated to provide summaries of the relationship between plasma total Cu, plasma LBC, cNCC, dNCC, 24-Hour urine copper, and plasma total Mo with safety endpoints. The correlations with safety endpoints will be calculated for all visits in the following time windows (Table 5)

- Week 4 only
- Postbaseline period: Week 4 to Week 48
- 24-week Cu mobilization period: Week 4 to Week 24
- Post 24-week Cu maintenance period: Post Week 24 to 48 (> Week 24 - Week 48)

Table 5: Correlations Between Copper and Safety Endpoints

	Plasma Total Copper	dNCC	Plasma LBC	cNCC/ $cNCC_{corrected}$	24-Hour Urine Copper	Plasma Total Molybdenum
MELD	X	X	X	X	X	X
Modified Nazer	X	X	X	X	X	X
CGI-S	X	X	X	X	X	X
CGI-I	X	X	X	X	X	X
ALT	X	X	X	X	X	X
AST	X	X	X	X	X	X
GGT	X	X	X	X	X	X
Neutrophils	X	X	X	X	X	X
eukocytes	X	X	X	X	X	X
Albumin	X	X	X	X	X	X
Hemoglobin	X	X	X	X	X	X

Abbreviations: ALT = alanine aminotransferase; AST = aspartate aminotransferase; CGI-I = Clinical Global Impression-Improvement Scale; CGI -S= Clinical Global Impression-Severity Scale; cNCC = non-ceruloplasmin-bound copper; $cNCC_{corrected}$ = calculated non-ceruloplasmin-bound copper corrected for copper bound in tetrathiomolybdate-copper-albumin complexes; GGT = gamma-glutamyltransferase; LBC = labile bound copper; MELD = Model for End-Stage Liver Disease

8. REFERENCES

Alcorn JB; United Network for Organ Sharing. Letter to: Transplant Professionals. UNOS. 2015;1-23. Available from https://optn.transplant.hrsa.gov/media/1575/policynotice_20151101.pdf

Atkinson MJ, Sinha A, Hass SL, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. Health Qual Life Outcomes. 2004; 2:12.

Atkinson MJ, Kumar R, Cappelleri JC, Hass SL. Hierarchical construct validity of the Treatment Satisfaction Questionnaire for Medication (TSQM Version II) among outpatient pharmacy consumers. Value Health. 2005;8(suppl 1):S9-S24.

Bharmal M, Payne K, Atkinson MJ, Desrosiers MP, Morisky DE, Gemmen E. Validation of an abbreviated Treatment Satisfaction Questionnaire for Medication (TSQM-9) among patients on antihypertensive medications. Health Qual Life Outcomes 2009;7:36.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

Brewer GJ, Askari F, Dick RB, et al. Treatment of Wilson's disease with tetrathiomolybdate: V. Control of free copper by tetrathiomolybdate and a comparison with trientine. Transl Res. 2009;154(2):70-77.

Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28-37.

Coleman CI, Sobieraj DM, Marinucci LN. Minimally important clinical difference of the Timed 25-Foot Walk Test: results from a randomized controlled trial in patients with multiple sclerosis. Curr Med Res Opin. 2012;28(1):49-56.

Czlonkowska A, Tarnacka B, Möller JC, et al. Unified Wilson's Disease Rating Scale - a proposal for the neurological scoring of Wilson's disease patients. Neurol Neurochir Pol. 2007;41(1):1-12.

Dhawan A, Taylor RM, Cheeseman P, De Silva P, Katsiyiannakis L, Mieli-Vergani G. Wilson's disease in children: 37-year experience and revised King's score for liver transplantation. Liver Transpl. 2005;11(4):441-448.

Dmitrienko A, D'Agostino R Sr. Tutorial in Biostatistics: Traditional multiplicity adjustment methods in clinical trials. Stat Med. 2013;32(29):5172-5218.

EASL. European Association for the Study of the Liver (EASL) Clinical Practice Guidelines: Wilson's disease. J Hepatol. 2012;56(3):671-685.

EuroQol Group. EQ-5D-5L User Guide. Version 2.1. Available at: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed November 2018.

Gorst-Rasmussen A, Tarp-Johansen MJ. Multiway Tipping Point Analyses in Longitudinal Clinical Trials with Missing Data. ASA Biopharm Reg Ind Stats Workshop. Poster 05. 25 Sep 2020.

Guideline on the evaluation of the pharmacokinetics of medicinal products in patients with decreased renal function. EMA/CHMP/83874/2014; Committee for Medicinal Products for Human use (CHMP). 17 Dec 2015.

Guo T, Senior J, Gelperin K; US FDA. How a SAS/IntrNet tool was created at the FDA for the detection of potential drug-induced liver injury using data with CDISC standard. Proceedings of the Western Users of SAS Software Annual Conference, 2009; Silver Spring, MD. Available from https://www.lexjansen.com/wuss/2009/cdi/CDI-Guo.pdf.

Guy W (ed). ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration, 1976.

Hobart J, Blight AR, Goodman A, Lynn F, Putzki N. Timed 25-foot walk: direct evidence that improving 20% or greater is clinically meaningful in MS. Neurology. 2013;80(16):1509-1517.

Koch C, Roid G. Examiner's Manual. Nonverbal Stroop Card Sorting Test (NCCST). Wood Dale, IL: Stoelting Company, 2012.

Leinweber B, Möller JC, Scherag A, et al. Evaluation of the Unified Wilson's Disease Rating Scale (UWDRS) in German patients with treated Wilson's disease. Mov Disord. 2008;23(1):54-62.

Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-612.

Mathiowetz V, Weber K, Kashman N, Volland G. Adult Norms for the Nine Hole Peg Test of finger dexterity. The Occupational Therapy Journal of Research. 1985;5(1):24-38. https://doi.org/10.1177/153944928500500102.

Meyer RD, Ratitch B, Wolbers M, et al., Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic. Stat Biopharm Res. 2020;12(4):399-411.

NCI, 2017. National Cancer Institute Division of Cancer Treatment and Diagnosis (DCTD), National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS). Common Terminology Criteria for Adverse Events V5.0 (CTCAE). Published: November 27, 2017. Available online at http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm. The quick reference guide is available online at:

https://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf.

O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics. 1984;40(4):1079-1087.

Roberts EA, Schilsky ML, American Association for Study of Liver Diseases (AASLD). Diagnosis and treatment of Wilson disease: an update. Hepatology. 2008;47(6):2089-2111.

Rohay J. Measurement Properties of, and Endpoints from, the Unified Wilson Disease Rating Scale (UWDRS). IQVIA. 2020.

Rubin DB. Multiple imputation for nonresponse in surveys. New York. John Wiley and Sons; 1987.

Schwartz, GJ and Work, DF, Measurement and estimation of GFR in children and adolescents. Clin J Am Soc Nephrol. 2009;4(11):1832–1843.

Stevens LA, Levey AS. Measured GFR as a confirmatory test for estimated GFR. J Am Soc Nephrol 20(11):2305-2313, 2009.

Sugitani T, Bretz F, Maurer W. A simple and flexible graphical approach for adaptive group-sequential clinical trials. J Biopharm Stat. 2016;26(2):202-216.

Tamhane AC, Dmitrienko A. Analysis of multiple endpoints in clinical trials. In multiple testing problems in pharmaceutical statistics, Dmitrienko A, Tamhane AC, Bretz F (editors). Chapman and Hall/CRC Press: New York. 2009; 131-163.

Tesfaldet et al. Enhanced visualization of treatment emergent liver effects in integrated eDSH & baseline shift plot. FDA/PHUSE US Computational Science Symposium 2016: https://www.lexjansen.com/css-us/2016/PP18_Final.pdf

Plitz T. WTX101: Development and validation of NCCcorrected, a method to measure free plasma copper in patients treated with WTX101 or other Cu lowering agents. Unpublished report, 2017.

Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. Comparison with Liver Biopsy and FibroTest. Hepatology. 2007; 46(1), 32-36.

Ventura J, Lukoff D, Nuechterlein KH, Liberman RP, Green M, Shaner A. Brief Psychiatric Rating Scale (BPRS) expanded version (4.0) scales, anchor points and administration manual. International Journal of Methods in Psychiatric. 1993;3: 227–244.

Wassmer G, Brannath W. Group sequential and confirmatory adaptive designs in clinical trials. New York: Springer, 2016.

Wechsler, D. Wechsler Adult Intelligence Scale–Fourth Edition. Pearson; San Antonio, TX: 2008.

Yeh KC, Kwan KC. A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. J Pharmacokinet Biopharm. 1978;6(1):79-98.

9. APPENDICES

9.1. Protocol Schedule of Activities

Refer to Table 4 in the Protocol.

9.2. Changes from Analyses Specified in the Previous Version of the SAP

The following updates have been made to the SAP version 5.0 approved 15 Apr 2021:

- Responder rate at 48 weeks for UWDRS Part II and Responder rate at 48 weeks for UWDRS Part III added to list of Exploratory Endpoints and Appendix (Section 5.1.3 and Section 9.4.8, respectively).
- The relationship between normalized GFR and absolute GFR has been updated (Section 7.1.3.2)
- Clarification added that presence of psychiatric symptoms at Baseline refers to at least 1 symptom present on BPRS-24 (Section 7.1.3.2 and Section 7.2.1.2).
- Clarification added in tipping point analysis that when ALXN1840 fixed, SoC varied, the δ varies from 10% to 200% (Section 7.2.1.1).
- Added condition that a subgroup is analyzed only if the number of patients is ≥ 3 in each treatment arm. (Section 7.2.1.2).
- Added additional subgroups for Ethnicity (Section 7.2.1.2).
- Added additional subgroup for patients who have an incomplete or intolerant response to prior WD treatment (Section 7.2.1.2).
- Revised subgroups for Renal Function Status at Baseline (Section 7.2.1.2).
- UWDRS Part III Functional Subscale-FDA suggested scoring added to list of Exploratory Endpoints and Exploratory Analyses (Section 5.1.3 and Section 7.2.3, respectively).
- Added analyses for exposure-adjusted incidence rates (Section 7.3.2).
- Clarification that AESIs includes all AEs in the MedDRA SOC "Nervous system disorders" or any AE where the "Is this an AESI?" check box is ticked (Section 7.3.2.7).
- Added definition of significant TEAEs for abnormal liver function, hematopoietic cytopenias, and dyslipidemia (Section 7.3.2.8).
- Added lipid profile plots (Section 7.3.3.1)
- Added 24-Hour urine copper for correlations with safety endpoints, and updated Table 5 (Section 7.3.3.5).

9.3. Sample Size, Power, and Randomization

9.3.1. Sample Size Justifications

The primary endpoint is copper control assessed as the percent change from baseline to 48 weeks in cNCC levels; for ALXN1840-treated patients, the cNCC level will be corrected for the amount of copper bound to the ALXN1840 TPC.

The hypothesis to be tested with regards to the primary endpoint is:

$H_0: \mu_{ALXN1840} - \mu_{SoC} \geq \Delta$ vs $H_1: \mu_{ALXN1840} - \mu_{SoC} < \Delta$

where Δ=15% is the non-inferiority margin, μ_{ALXN1840} and μ_{SoC} represent 48-week mean percent change for ALXN1840 and SoC, respectively.

For patients who have been previously treated with SoC for > 28 days prior to first dose (ie, corresponding to Cohort 1 in the present study), the cNCC data in the preceding Phase 1 Study WTX101-201, 24-week interim lock gives a mean percent change in cNCC level at 24 weeks of 80% with estimated projected standard deviation of 20% to 30% at 48 weeks. Based on this anticipated degree of variability, 90 Cohort 1 patients, randomized by cohort on a 2:1 basis, will provide between 60% and 91% power to rule out a difference of Δ = 15% in mean percent change in cNCC values, ALXN1840 versus SoC with a 1-sided Type I error of 2.5%.

The number of patients in Cohort 2 is based on feasibility only, and there is no supporting power calculation. The intent of Cohort 2 is to provide at least some randomized data in patients who are treatment naïve or have been previously treated with SoC for ≤ 28 days prior to first dose to descriptively assess within arm changes from baseline. No formal statistical comparisons are planned within Cohort 2 patients alone.

In relation to the overall analysis of cNCC percent change from baseline to 48 weeks, data from the WTX101 201, 24-week interim lock gives a mean percent change in cNCC level at 24 weeks of 78%. A total of 120 patients (90 patients in Cohort 1 and 30 patients in Cohort 2) randomized by cohort on a 3:1 basis will provide between 73% and 97% power to rule out a difference of Δ = 15%, in mean percent change in cNCC values, ALXN1840 versus SoC, in the overall patient population, with a 1-sided Type I error of 2.5%.

Study WTX101-201 had a minimum cNCC value of 0.8 for study entry while this study has no minimum requirement. This may result in the entry of patients with negative baseline cNCC values into this study who will be non-evaluable for the primary analysis and/or have a smaller effect size. For these reasons, and to achieve a pooled safety database of at least 100 patients treated with ALXN1840 for at least 1 year pooling Studies WTX101-201 and WTX101-301, these numbers have been inflated by 25% to achieve a total of 150 evaluable patients (approximately 113 patients in Cohort 1 and 37 patients in Cohort 2).

To help ensure that at least 150 evaluable patients complete 48 weeks of treatment, approximately 180 patients (approximately 135 patients in Cohort 1 and approximately 45 patients in Cohort 2) will be enrolled in total, this will also provide additional safety and tolerability data.

The proposed 150 evaluable patients will provide adequate power (approximately 99%) for the new primary endpoint, namely, daily mean dNCC $AUEC_{0\text{-}48W}$. This power calculation uses

previously observed data on Cohort 1 patients. In Study WTX101-201, Cohort 1 patients treated with ALXN1840 showed AUEC of dNCC with a mean (SD) of 3.96 (2.047), while in Study WTX101-203, Cohort 1 patients treated with SoC showed a mean (SD) of 0.85 (0.425). With a significance level of 2.5% using a one-sided unequal variance t-test, 150 patients (100 to ALXN1840 and 50 to SoC) will provide > 99% power to reject the null hypothesis of equal means.

Because the study size and power are sensitive to the assumptions used for effect size, variability, and withdrawal rates, an SSR will be performed. Details of the SSR are provided in Section 9.3.2.

9.3.2. Sample Size Re-estimation

The interim analysis for SSR will be performed using the data from only the first approximately 31% of enrolled patients. There will be no evaluation of the interim data for early stopping for futility or efficacy. The sample size will not be reduced from the planned enrollment of 180 designed to achieve 150 evaluable patients and the maximum total number of patients enrolled may be increased to achieve up to 225 evaluable patients.

The SSR will be based on a conditional power for non-inferiority (CPni) using a promising zone with boundaries of ≥ 30% and < 80%. The CPni will be calculated using the results obtained at this interim analysis and will assume population mean percent change in cNCC equals the observed sample mean percent change in cNCC at the time of the SSR. If the CPni is below 30%, or equal to or above 80% (ie, outside the promising zone) then no increase in sample size will be made. If the CPni is within the promising zone, then the sample size will be increased by the amount estimated needed to achieve a CPni of 80%. The decision rule is illustrated in the Protocol Section 9.2.8 Table 11. The SSR will be repeated based on conditional power for superiority using the same boundaries and decision rules described for non-inferiority.

9.3.3. Randomization

After meeting all inclusion and none of the exclusion criteria, patients will be randomized, stratified by cohort, via an interactive voice/web response system in a 2:1 ratio to treatment with ALXN1840 or SoC. Randomization may be required on or before Day -2 to allow for SoC to be withheld for ≥ 48 hours for patients randomized to receive ALXN1840.

9.4. Technical Specifications for Derived Variables

The following derived data will be calculated prior to analysis. The subsections below give more details.

9.4.1. Adverse Events

The analysis of AEs is described in detail in Section 7.3.2.

TEAEs are events with start dates and start times on or after the date and time of the first study drug dose or existing events that worsened in severity after the first dose of randomized treatment. If the start date of an AE is partially or completely missing and the end (stop) date and time of the AE does not indicate that it occurred prior to first dose, then the determination of treatment-emergent status will be based on the following:

- If the start year is after the year of the first study drug dose, then the AE is treatment-emergent; else,
- If the start year is the same as the year of the first study drug dose, and
  - the start month is missing, then the AE is treatment-emergent; else if
  - the start month is present and is the same or after the month of the first study drug dose, then the AE is treatment-emergent; else,
- If the start date is completely missing, then the AE is treatment-emergent.

All other AEs are considered PTAEs.

9.4.2. Age

Age will be presented as the number of years between date of birth and the reference date. The following age in Table 6 may be computed, with reference date indicated.

Table 6: Age and Reference Date

Age	Reference Date
• Age at Enrollment	• Date of signing ICF

The following formula should be followed for calculation of age if needed:

Age (year) = FLOOR([reference date – date of birth]/365.25*12)

where FLOOR( ) function returns the integer part of the result.

In cases where only the month and year are provided for a date, the day for the date will be imputed as 15. Missing month will be imputed as June. In cases where the day is observed but the month is missing, the date will be imputed as June 15. In instances when the imputed reference date is earlier than the birth date, the birth date will be used as the reference date.

9.4.3. Analysis Relative Day

Analysis relative day is the day relative to the first dosing day. It will be calculated as: analysis date – first dose date + 1 if analysis date is after the first dose date, or else as: first dose date – analysis date.

9.4.4. Baseline Value

The baseline is defined as the last non-missing value collected on or prior to first dose.

9.4.5. Body Mass Index (BMI)

BMI is derived as follows: weight (kg) / [height (cm) / 100]^2

9.4.6. Change from Baseline

Change from baseline will be calculated as: postbaseline assessment value – baseline assessment value when both values are not missing.

Percent change from baseline is calculated as (change from baseline/baseline result * 100).

If either the baseline or the postbaseline result is missing, the change from baseline and/or percentage change from baseline is set to missing. Additionally, if the baseline is 0, the percentage change will be missing.

9.4.7. Medications and Therapies

Concomitant medications/therapies are any events with administration dates and times on or after the date and time of the first study drug dose. If the start date of a medication or therapy is partially or completely missing and the end (stop) date and time of the medication/therapy does not indicate that it occurred prior to first dose, then the determination of concomitant status will be based on the following:

If the start year is after the year of the first study drug dose, then the medication/therapy is concomitant; else, if the start year is the same as the year of the first study drug dose and the start month is missing, then the medication/therapy is concomitant; else if the start month is present and is the same or after the month of the first study drug dose, then the medication/therapy is concomitant; else, if the start date is completely missing, then the medication/therapy is concomitant.

All other medications/therapies are considered prior medications/therapies and could occur from the 30 days prior to informed consent up through the Screening Period and prior to the first dose.

9.4.8. UWDRS

As described in Section 5.1.2.1, UWDRS has 3 subscores. The algorithms for calculating the subscores are given below. The UWDRS total score will be the sum of the 3 subscores.

UWDRS I: Consciousness; Maximum score of 3

- Set to Question 1, range 0 to 3

UWDRS Part II total score: Disability; Maximum score of 40

- Calculate the sum of Question 2 to Question 11, each question has range 0 to 4
- If all 10 items are populated, the subscore will be the sum calculated above. If 8 or 9 items are populated, prorate the score by dividing by the number of answered items and multiplying by 10. This is done in order to estimate the value the patient could have achieved if they had answered all the questions. If less than 8 items are populated, score won't be derived.
    - Eg, Patient has answered 8 of 10 questions and the total of their answered questions is 23. The pro-rated score will then be calculated as: 23/8*10=28.75

UWDRS Part III total score: Neurological Status; Maximum score of 175

- Calculate the following new scores for Questions 12 to 34:
    - Question 12: range 0 to 4

      Set to Q12A
    - Question 13: range 0 to 6

If Q13=0 then set to 0;

else if Q13=1 then do;

if Q13A>2 then set to Q13A

else set to Q13A+Q13B

end;

- Question 14: range 0 to 1

  Set to Question 14

- Question 15: range 0 to 16

  Set to Q15A+Q15B+Q15C+Q15D

- Question 16: range 0 to 4

  Set to Q16

- Question 17: range 0 to 20

  Set to Q17A+Q17B+Q17C+Q17D+Q17E

- Question 18: range 0 to 8

  Set to Q18A+Q18B

- Question 19: range 0 to 8

  Set to Q19A+Q19B

- Question 20: range 0 to 4

  Set to Q20

- Question 21: range 0 to 16

  Set to Q21A1+Q21A2+Q21B1+Q21B2

- Question 22: range 0 to 8

  Set to Q22A+Q22B

- Question 23: range 0 to 8

  Set to Q23A+Q23B

- Question 24: range 0 to 8

  Set to Q24A+Q24B

- Question 25: range 0 to 4

  Set to Q25

- Question 26: range 0 to 8

  Set to Q26A+Q26B

- Question 27: range 0 to 4

Set to Q27

- Question 28: range 0 to 10

  If Q28=0 then set to 0;
  else if Q28=1 then do;
  if Q28A>2 then set to Q28A
  else set to Q28A+Q28B+QS28C

  end;

- Question 29: range 0 to 10

  If Q29=0 then set to 0;

  else if Q29=1 then do;
  if Q29A1+Q29A2>2 then set to Q29A1+Q29A2
  else set to Q29A1+Q29A2+Q29B+Q29C

  end;

- Question 30: range 0 to 24

  Set to Q30A+Q30B+Q30C+Q30D+Q30E+Q30F

- Question 31: range 0 to 1

  Set to Q31

- Question 32: range 0 to 1

  Set to Q32

- Question 33: range 0 to 1

  Set to Q33

- Question 34: range 0 to 1

  Set to Q34

- Add up the newly calculated scores for Questions 12 to 34. If all 23 items are completely populated, the subscore will be that sum. If 20 to 22 items are fully populated, then calculate the maximum score of the answered items. Next, divide the sum of the responses to Questions 12 to 34 by the maximum score and multiply by 175 to obtain the subscore. This is done in order to estimate the value the patient could have achieved if they had answered all the questions. If less than 20 items are fully populated, then do not derive. If part of a multi-component question is answered, the maximum score of the answered components will be used to derive the subscore. However, partially answered questions will not be considered as fully populated (and therefore will be viewed as a null response) when counting the number of fully answered questions in order to obtain a minimum of 20 fully answered questions. The exception to this latter rule is for Questions 13, 28 and 29. If the > 2 condition is met with populated components, the question is considered to be fully populated regardless if the unused components are null.
  - Eg, Patient has responded to 21 questions and left Questions 21 and 22 null. The total of their responses to the answered questions is 97 and the maximum score of

the answered questions is 151. The pro-rated score will then be calculated as: 97/151*175=112.41721854.

- Eg, Patient has responded to 22 questions and part of Question 17 so they will be considered as answering 22 questions. The total of their response to the answered questions is 111 and the maximum score of the answered questions and answered components (including any partially answered questions) is 167. The pro-rated score will then be calculated as 111/167*175=116.317365.

UWDRS Part II-FDA Suggested Scoring: UWDRS Part II consists of mobility, falling, transfer, salivation, swallowing, feeding, dressing, grooming, taking a bath or shower, and toilet use. These items will be rescaled to 0- 3 scale which 0 is corresponding to original score of 0 and 1; 1 is corresponding to original score of 2; 2 is corresponding to original score of 3; and 3 is corresponding of original score of 4. The total UWDRS Part II-FDA suggested scoring will be the sum of these 10 items.

UWDRS Part III Individual Items-Original Scoring: For items: speech, handwriting, and arising from chair, original scores are 0-4 scale as in the Unified Wilson's Disease Rating Scale. For gait, the original score is 0-10 by summing of Part A, B, and C.

UWDRS Part III Individual Items-Transformed Original Scoring: For gait, four sections (Part A Right, Part A Left, Part B, and Part C) from three sub-questions will be used to generate one single score which ranges from 0 to 4. Score is 0 if response to UWDRS Q29 is 'normal'. The score will be the maximum score of four sections of Q29 if response is 'not normal'.

UWDRS Part III Individual Items - Standardized Transformed Scoring: For gait, the standardized transformed scores are generated by dividing the transformed original scores by 4 and then multiplying by 10 with a range of 0 to 10.

UWDRS Part III Individual Items-FDA Suggested Scoring: UWDRS Part III following items will be rescaled per FDA suggestion:

For items: speech, handwriting, and arising from chair, score will be rescaled to 0-3 scale in which 0 is corresponding to original score of 0, and 1; 1 is corresponding to original score of 2; 2 is corresponding to original score of 3; and 3 is corresponding of original score of 4.

For gait-FDA suggested scoring, the transformed original score will be rescaled to 0-3 scale in which 0 is corresponding to transformed original score of 0, and 1; 1 is corresponding to transformed original score of 2; 2 is corresponding to transformed original score of 3; and 3 is corresponding of transformed original score of 4.

UWDRS Part III Functional Subscale: UWDRS Part III Functional Subscale consists of speech, handwriting, arising from chair, and gait from UWDRS Part III. As described above, the standardized score of the first 3 items ranges from 0 to 10, and standardized transformed score of gait ranges from 0 to 10. The average of these scores will be used to create the Part III Functional Subscale with a range of 0-10.

The subscale score and test statistic for evaluating the significance of the treatment effect will be derived using the global test for multiple endpoints (O'Brien, 1984; Tamhane, 2009). This test relies on pooling the evidence of effectiveness across the 4 items and will be employed for assessing the evidence to support the global alternative hypothesis that ALXN1840 is uniformly

better than SoC in the mean changes in the 4 scores defined above (treated as a single instrument).

The following notation will be used to define the global test for the UWDRS Part III Functional Subscale. Let μ_i and t_i denote the model-based treatment effect, and the test statistic, respectively, for evaluating the treatment difference on the ith item, $i = 1, \ldots, 4$, respectively. The pairwise correlation between the change from baseline in the ith item and jth item will be denoted by ρ_{ij}, $i = 1, \ldots, 4$, $j = 1, \ldots, 4$. The pairwise correlations will be estimated at the time of the final analysis. Lastly, n_1 and n_2 will denote the sample sizes in the 2 study arms. The global test statistic is given by:

$$t = \sum_{i=1}^{4} t_i \Big/ \sqrt{\sum_{i=1}^{4} \sum_{j=1}^{4} \rho_{ij}}.$$

Under the global null hypothesis of no effect on any of the 4 items, the test statistic follows a t distribution with $(1 + 1/16)(n_1 + n_2 - 2)/2$ degrees of freedom. The mean treatment difference for the UWDRS Part III Functional Subscale is given by:

$$\mu = \sum_{i=1}^{4} \mu_i \Big/ \sqrt{\sum_{i=1}^{4} \sum_{j=1}^{4} \rho_{ij}}.$$

UWDRS Part III Functional Subscale-FDA Suggested Scoring: UWDRS Part III following items will be rescaled per FDA suggestion:

For items: speech, handwriting, and arising from chair, score will be rescaled to 0-3 scale in which 0 is corresponding to original score of 0, and 1; 1 is corresponding to original score of 2; 2 is corresponding to original score of 3; and 3 is corresponding of original score of 4.

For gait-FDA suggested scoring, the transformed original score will be rescaled to 0-3 scale in which 0 is corresponding to transformed original score of 0, and 1; 1 is corresponding to transformed original score of 2; 2 is corresponding to transformed original score of 3; and 3 is corresponding of transformed original score of 4.

These items will be standardized to create a value between 0 to 10, and the average of these scores will be used to create the Part III Functional Subscale.

UWDRS ADL Subgroup at Baseline: Patients can be classified into 3 UWDRS ADL subgroups based on their response to UWDRS Part II questions related to ADL: mobility, feeding, dressing, grooming, taking a bath or shower, and toilet use.

- No reported ADL limitation: If a patient with score of 0 for all 6 questions (original score 0-4)

- Functionally independent: if a patient with score of 0 or 1 for all 6 questions (original score 0-4)

- Functionally dependent: if a patient with score of 2 or more for at least 1 of the 6 questions (original score 0-4)

Responder rate at 48 weeks for UWDRS Part II and Part III: The definition of "clinically meaningful change" will be defined in Clinical Outcome Assessment (COA) validation documents. The definition will be formed using both quantitative data collected in studies WTX101-201 and WTX101-301 and qualitative information from WTX101-301 patient/physician interviews. These activities require information from the WTX101-301 primary

analysis period to be final. Consequently, the specific numerical values for "clinically meaningful change" cannot be prospectively stated.

9.4.9. Visit Windowing

In analysis of data summarized by study visit, all data collection will be reassigned a study visit where data is scheduled for collection based on the actual days relative to baseline. All visits will be assigned a target study day; for the determination of target days, weeks will be assumed to have 7 days. Baseline will have a target study day of 1. Thus, Week 4 would have a target day of 4*7+1=29. For each assessment, the postbaseline period will be divided up using the scheduled visit's target days. The lower bound of each visit interval will be evaluated as the midpoint between the target day and the previous visit's target day in the following manner: study day interval lower bound = target study day – ([target study day – last target study day]/2). If more than 1 value is mapped to the same scheduled visit, the closer of those values will be considered for summarization. Visit windows are intended to be contiguous such that all data collected at all postbaseline visits, whether scheduled or unscheduled, will map to 1 of the visits.

The visit displayed on patient data listings will be reflective of the scheduled visit label as reported on the CRF. Study days relative to baseline will be displayed for each visit so it is apparent which visit the data may have been reassigned to in the summaries.

9.5. Additional Details on Statistical Methods

To maintain study integrity and minimize bias in this open label study, draft tables, listings, and figures will be produced with blinded treatment assignments for review prior to database lock for the Primary Evaluation Period.

9.5.1 Analysis Considerations Related to COVID-19

On 11 Mar 2020, Coronavirus Disease 2019 (COVID-19) was declared a pandemic by the WHO. This section summarizes additional analysis considerations to assess the potential impact of COVID-19 (Meyer, 2020). Since enrollment (last patient informed consent) was completed on 27 Feb 2020, ie, before 11 Mar 2020, the pandemic onset should have no effect on baseline patient characteristics. The following additional sensitivity and supplementary analyses will be included to assess the impact of the pandemic disruption on the trial and to address pandemic-related data missingness.

1. The summary of patient disposition will include COVID-19 related discontinuations and withdrawals.
2. A summary of known COVID-19 exposure or diagnosis during the Primary Evaluation Period will be provided by treatment group using the Safety Set.
3. A summary of COVID-19 related important protocol deviations during the Primary Evaluation Period will be provided. A by-patient listing of all protocol deviations will be provided.
4. A summary of the number and percentage of patients who missed a study visit or had a modified study visit during the Primary Evaluation Period along with the reasons

(COVID-19 related/not), will be provided by treatment group and by visit using the FA Set and Safety Set. For patients who had a modified study visit, the method for the different assessments will be summarized.

5. Alternative data collection methods required during the pandemic may introduce additional variability. Two sensitivity analyses will be performed to assess this possibility. Descriptive statistics for the primary and key secondary endpoints by visit (and by randomized treatment and cohort) can be calculated with the visits split into the categories described below. In the first sensitivity analysis, each visit for each patient will be categorized as either "collected as planned", or "modified". In the second sensitivity analysis, each visit for each patient will be categorized as either "prepandemic" (occurring before 11 Mar 2020), or "during pandemic".
6. Missing data on the primary endpoint due to any COVD-19 related reasons will be handled as described in Section 7.2.1.1 "Handling of Dropouts or Missing Data", where sensitivity analyses assuming missing at random, as well as missing not at random, will be performed.