Statistical Analysis Plan

A Multicenter, Multinational, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of Laquinimod (0.5, 1.0 and 1.5 mg/day) as Treatment in Patients with Huntington's Disease

Study Number TV5600-CNS-20007

NCT02215616

Statistical Analysis Plan Approval Date: 19 June 2018

Statistical Analysis Plan

Study TV5600-CNS-20007

A Multicenter, Multinational, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of Laquinimod (0.5, 1.0 and 1.5 mg/day) as Treatment in Patients with Huntington's Disease

(LEGATO-HD - Laquinimod Efficacy and Safety in a GlobAl Trial Of HD)

Phase 2

IND number: 120331; NDA number: not applicable; EudraCT number: 2014-000418-75

Approval Date: 19 June 2018

Sponsor

Teva Branded Pharmaceutical Products R&D, Inc.
41 Moores Road
Frazer, Pennsylvania 19355
United States

Prepared by: [redacted]

STATISTICAL ANALYSIS PLAN APPROVAL

Study No.: TV5600-CNS-20007

Study Title: A Multicenter, Multinational, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of Laquinimod (0.5, 1.0 and 1.5 mg/day) as Treatment in Patients with Huntington's Disease

Statistical Analysis Plan for:

- ☐ **Interim Analysis**
- ☒ **Final Analysis**
- ☐ **Integrated Summary of Efficacy**
- ☐ **Integrated Summary of Safety**

Amendment: not applicable

Author: [redacted]

[redacted] 19 June, 2018

Approver: **Date**

19 June 2018

Approver: **Date**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
βHCG	beta human chorionic gonadotropin
ANCOVA	analysis of covariance
CAB	Cognitive Assessment Battery
CAG	cytosine-adenosine-guanine repeat
CDR-SB	Clinical Dementia Rating - Sum of Boxes
CDT	Carbohydrate deficient transferrin
CIBIC	Clinician's Interview-Based Impression of Change
CIBIS	Clinician's Interview-Based Impression of Severity
CNS	Central nervous system
CrCl	Creatinine Clearance
CRF	case report form (refers to any media used to collect study data [ie, paper or electronic])
CSC	Clinical Supply Chain
C-SSRS	Columbia-Suicide Severity Rating Scale
CT	Computed tomography
CTCAE	Common Terminology for Adverse Events
CYP	Cytochrome P450
DMC	Data Monitoring Committee
DSM–IV	Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition Text Revision
DSMB	Data and Safety Monitoring Board
ECG	electrocardiography, electrocardiogram
ET	Early Termination
EU	European Union
FAS	full analysis set
FDA	US Food and Drug Administration
HADS	Hospital Anxiety and Depression Scale
HCG	human chorionic gonadotropin
HD	Huntington's Disease
HD- CAB	Huntington's Disease Cognitive Assessment Battery
HD-QoL	Huntington's Disease Quality of Life
HIV	Human Immunodeficiency Virus
HVLT-R	Hopkins Verbal Learning Test, revised
IB	Investigator's Brochure
ICH	International Conference on Harmonisation

Abbreviation	Term
IL	Interleukin
IOI	Inter onset intervals
IPI	Inter peak intervals
IRT	interactive response technology
ISCED	International Standard Classification of Education
ITI	Inter tap intervals
ITT	intent-to-treat
LOCF	Last Observation Carried Forward
Max	maximum
mPPT	Modified Physical Performance Test
MedDRA	Medical Dictionary for Regulatory Activities
MI	Myo-inositol
ML	Maximum Likelihood
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MS	Multiple Sclerosis
OTS	One Touch Stockings of Cambridge
PBA-s	Problem Behaviors Assessment-Short form
PET	Positron Emission Tomography
PGx	pharmacogenomic(s)
PK	Pharmacokinetic(s)
PPK	Population Pharmacokinetics
PT	Preferred Term
QC	Quality Control
Qd	once daily
Q-Motor	Quantitative Motor Assessments
QoL	Quality of Life
REML	Restricted Maximum-Likelihood
SAP	Statistical Analysis Plan
SD	standard deviation
SDMT	Symbol Digit Modalities Test
SOC	system organ class
SOP	standard operating procedure
t_{max}	time to maximal plasma drug concentration

Abbreviation	Term
TD	Tap duration
TE	Echo Time
TF	Tapping forces
TR	Repetition Time
TSH	thyroid stimulating hormone
TSPO	Translocator Protein
(UHDRS-)FA	(Unified Huntington's Disease Rating Scale) Functional Assessment
(UHDRS-)TFC	(Unified Huntington's Disease Rating Scale) Total Functional Capacity
(UHDRS-)TMS	(Unified Huntington's Disease Rating Scale) Total Motor Score
ULN	upper limit of the normal range
WBC	white blood cell
WHO	World Health Organization
WHO Drug	World Health Organization (WHO) drug dictionary
WLQ	Work Limitations Questionnaire

INTRODUCTION

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Teva Branded Pharmaceutical Products R&D, Inc. study TV5600-CNS-20007, (A Multicenter, Multinational, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of Laquinimod [0.5, 1.0 and 1.5 mg/day] as Treatment in Patients with Huntington's Disease), and was written in accordance with SOP GBP_RD_702 (Statistical Analysis Plan).

The reader of this SAP is encouraged to read the study protocol for details on the conduct of this study, the operational aspects of clinical assessments, and the timing for completing the participation of a patient in this study.

The following documents were reviewed in preparation of this SAP:

- Clinical Study Protocol TV5600-CNS-20007 issued on 27MAY2014, amendment 1 issued on 10SEP2014, amendment 2 issued on 16FEB2015, amendment 3 issued on 24SEP2015, amendment 4 issued on 16Feb2016
- Case report form (CRF) for TV5600-CNS-20007
- ICH E9 Guidance on Statistical Principles for Clinical Trials
- ICH E3 Structure and Content of Clinical Study Reports (CSRs)
- Multiple Endpoints in Clinical Trials Guidance for Industry by FDA
- E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials

Any other references indicated throughout the SAP are cited from the study protocol, and the reader is requested to search there for the details.

The SAP is intended to be in agreement with the protocol, especially with regards to the primary and all secondary endpoints and their respective analyses. However, the SAP may contain more details or changes regarding these particular points of interest, or other types of analyses (eg other endpoints). When differences exist in descriptions or explanations provided in the study protocol and this SAP, the SAP prevails; the differences will be explained in the Clinical Study Report.

1. STUDY OBJECTIVES

The primary objective of this study is to assess the efficacy of laquinimod as treatment in patients with Huntington's Disease (HD) after 52 weeks using the Unified Huntington's Disease Rating Scale Total Motor Score (UHDRS-TMS or TMS).

The secondary objective of this study is to assess the effect of laquinimod in patients with HD after 52 weeks of treatment on change in brain volume (i.e. atrophy) using magnetic resonance imaging (MRI) measures of caudate volume.

Other efficacy endpoints will also be analyzed in an exploratory manner; these are detailed in Section 2.4.3.

The additional objectives are as follows:

- To evaluate the safety and tolerability of laquinimod (0.5, 1.0 and 1.5 mg/day) doses in patients with HD during 52 weeks of treatment
- To evaluate the pharmacokinetics of laquinimod and its metabolites in patients with HD

Some additional ancillary studies defined in the Protocol are beyond the scope of this SAP and will be reported separately if applicable (see Section 16).

2. STUDY DESIGN

2.1. General Design

This is a multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the efficacy and safety of laquinimod treatment at dosages of 0.5, 1.0, and 1.5 mg/day in adults with HD.

The study will consist of a screening period (2 weeks up to 5 weeks), followed by a 52-week double-blind treatment period and a follow-up visit (one month after end of treatment). Prior to 10 January 2016, a total of 400 patients were planned to be equally randomized in a 1:1:1:1 ratio (100 patients within each treatment arm) to receive laquinimod 0.5, 1.0, 1.5 mg/day, or matching placebo for 52 weeks. A total of 123 patients were randomized prior to 10 January 2016.

As of 10 January 2016, following the decision to endorse the DMC recommendation to discontinue treatment of the laquinimod 1.5 mg dose arm as a proactive safety measure, additional eligible patients who are enrolled will be randomized in a 1:1:1 ratio to receive laquinimod 0.5 mg/day, 1.0 mg/day, or matching placebo for 52 weeks. Thus, approximately 300 patients (100 patients within each study arm), plus the 30 patients who were already randomized to the laquinimod 1.5 mg treatment arm, were planned to be enrolled in the study. See the study schema from 10 January 2016 on Figure 1 for reference.

As of recruitment completion in May 2017, 352 patients, including the 30 patients who were randomized to the laquinimod 1.5 mg treatment arm, were enrolled in the study. After signing the informed consent, including consent to provide a blood sample for genetic analyses including CAG repeat length analysis, patients will be screened for a period of 2 weeks up to 5 weeks in order to determine whether they are eligible to participate in the study. Patients with a legal guardian should be consented according to local requirements.

Study procedures and assessments with their timing are summarized in Table 1.

During the visits, vital sign measurements, physical examinations and ECG will be performed prior to the blood draw for clinical laboratory tests and pharmacokinetic sampling. Further, the efficacy assessments (UHDRS-TMS, Huntington's Disease Cognitive Assessment Battery (HD-CAB), Clinician's Interview-Based Impression of Change (CIBIC)-Plus and Total Functional Capacity (UHDRS-TFC)) will be done prior to the other assessments.

Patients who complete all scheduled visits will have final procedures and assessments performed at the final visit (Visit 8, Month 12). Patients who withdraw from the study before completing the 52-week evaluation period will have Visit 8 procedures and assessments performed at their last visit.

A single blood sample will be collected from all patients at Months 1, 3, 6 and 12 for evaluation of laquinimod and its metabolites.

For PK ancillary study, PK samples were collected from approximately 4 patients per each of the three continuing treatment groups (at selected sites at Month 1), for a total of 13 patients at pre-dose, 15, 30 min and 1, 2, 3, 6 and 24 hours post dose. Additionally, PK samples were collected from 2 patients in the laquinimod 1.5 mg/day treatment group at the time this treatment group was discontinued; no further PK samples will be collected from the laquinimod 1.5 mg/day treatment group.

Figure 1: Overall Study Schema (from to 10 January 2016)

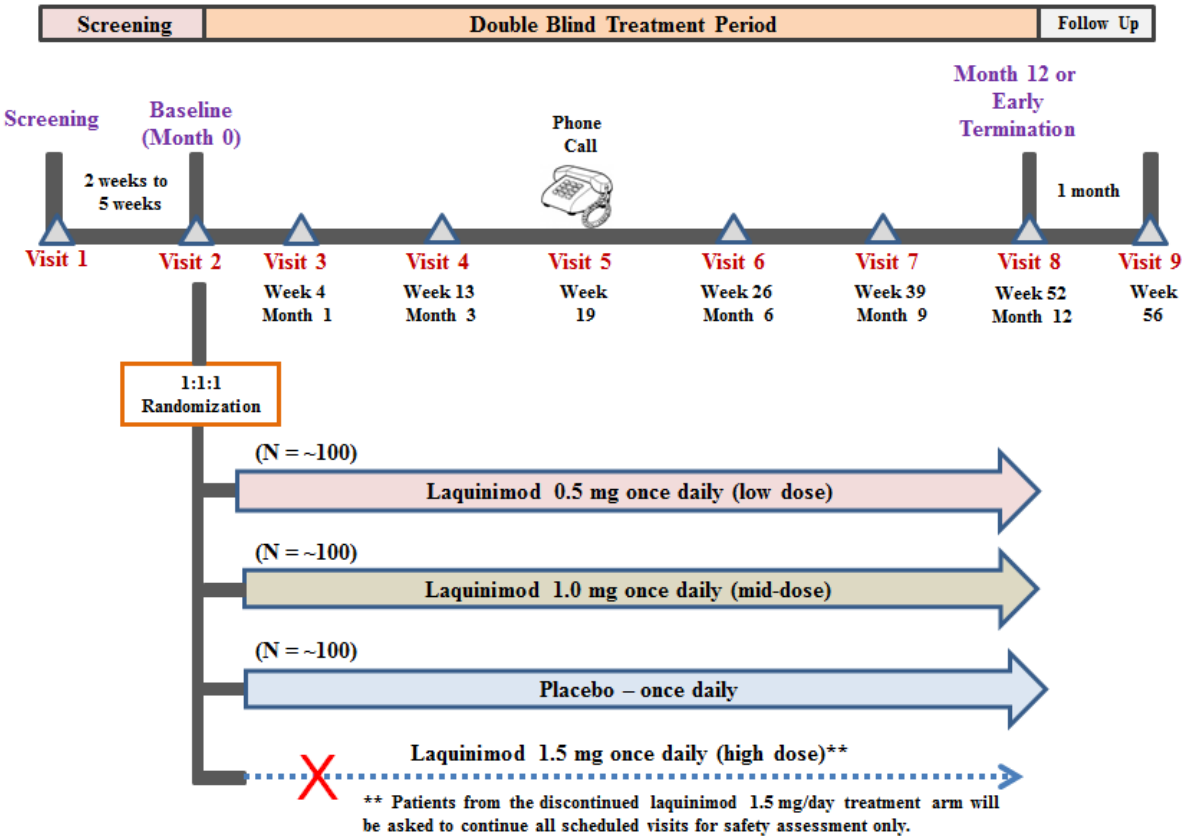


Table 1: TV5600-CNS-20007 (LEGATO-HD) - Study Procedures and Assessments

TV5600-CNS-20007 (LEGATO-HD)	Pre-treatment	Baseline (Dosing)	Double-blind Treatment Period						Follow-Up	Unscheduled Visit
Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8 - Termination Visit[a]	Visit 9	-
Time (Weeks)	Screening	Baseline	Week 4	Week 13	Week 19	Week 26	Week 39	Week 52	Week 56	-
Time (Months)		Month 0	Month 1	Month 3	☎	Month 6	Month 9	Month 12 (or ET)		
Visit Window	2 weeks up to 5 weeks		±3 days	±7 days	±7 days	±7 days	±7 days	±7 days	±5 days	
Informed consent	X[b]									
Inclusion and exclusion criteria	X[c]	X[c]								
Medical and psychiatric history / demographics	X									
Cardiovascular risk factor assessment and management[d e]	X									
ISCED	X									
HD history	X									
Prior medication history	X									
Drug test and/or CDT test[f]	X									
Clinical laboratory tests[g]	X[h]	X	X	X		X	X	X	X	X[i]
Clinical hematology	X	X	X	X		X	X	X	X	
Anemia panel[j]		X	Anemia panel is performed if there is a confirmed hemoglobin decrease of >1 g/dL from baseline							
Lipid profile[g]		X						X		
Thyroid function (T3, T4, TSH)		X				X		X		
Blood collection for soluble biomarkers (cytokines)		X				X		X		X[i]
Blood collection for gene expression analysis		X				X		X		
Blood sample for genomic analysis (including CAG analysis)	X									
Blood sample for TSPO genotype	X									

TV5600-CNS-20007 (LEGATO-HD)	Pre-treatment	Baseline (Dosing)	Double-blind Treatment Period						Follow-Up	Unscheduled Visit
Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8 - Termination Visit[a]	Visit 9	-
Time (Weeks)	Screening	Baseline	Week 4	Week 13	Week 19	Week 26	Week 39	Week 52	Week 56	-
Time (Months)		Month 0	Month 1	Month 3	☎	Month 6	Month 9	Month 12 (or ET)		
Visit Window	2 weeks up to 5 weeks		±3 days	±7 days	±7 days	±7 days	±7 days	±7 days	±5 days	
analysis[k]										
Blood sample for monocyte analysis[k l]		X						X		
Urinalysis	X									
Physical examination, including weight	X[m]	X	X	X		X	X	X	X	
ECG[n]	X	X	X	X		X		X	X	
Vital signs measurements	X	X	X	X		X	X	X	X	X
Serum β-HCG in women of childbearing potential	X[o]	X[o]	X	X		X	X	X	X	
Urine β-HCG in women of child-bearing potential		X[p]	X	X		X	X	X[p]	X	
Ascertaining use of acceptable contraception		X	X	X	X	X	X	X	X	X
Home urine pregnancy test			Starting after Month1 (Visit 3), a urine β-hCG test will be performed at home in women of child-bearing potential every 28 (±2) days.							
Site call to patient for result inquiry			A telephone call will be scheduled to be performed within 72 hours of the urine test date to inquire about the results.							
C-SSRS (baseline screening version)	X									
C-SSRS (since last visit version)		X	X	X		X	X	X	X	
Randomization		X								
MRI scan[q]	X[n]	X[r]						X[s]		
Estimated creatinine clearance calculation[t]	X	X	X	X		X	X	X	X	X[i]
MRS scan[k]		X						X		
PET scan[k u v]		X[p]						X[p]		

TV5600-CNS-20007 (LEGATO-HD)	Pre-treatment	Baseline (Dosing)	Double-blind Treatment Period						Follow-Up	Unscheduled Visit
Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8 - Termination Visit[a]	Visit 9	-
Time (Weeks)	Screening	Baseline	Week 4	Week 13	Week 19	Week 26	Week 39	Week 52	Week 56	-
Time (Months)		Month 0	Month 1	Month 3	☎	Month 6	Month 9	Month 12 (or ET)		
Visit Window	2 weeks up to 5 weeks		±3 days	±7 days	±7 days	±7 days	±7 days	±7 days	±5 days	
UHDRS-TMS	X[w]	X	X	X		X		X		
UHDRS-TFC	X[w]	X				X		X		
UHDRS-FA		X				X		X		
mPPT		X				X		X		
PBA-s[x]		X						X		
CIBIS[x]		X								
CIBIC-plus[x]						X		X		
HD-QoL[y]		X						X		
EQ-5D-5L[z]		X						X		
WLQ[z]		X						X		
Q-Motor Assessments	X	X	X	X		X		X		
HD-CAB[aa]	X[bb]	X	X[cc]	X[cc]		X		X		
CDR-SB[x]		X						X		
HADS		X						X		
PK (drug and metabolites concentration) sampling			X[y]	X		X		X		X[i]
24-hour PK profiling[dd] for drug and metabolites			X							
Review study compliance		X	X	X	X	X	X	X		X
Study Drug Dispensing		X	X	X		X	X			
Study Drug Collection and Reconciliation			X	X		X	X	X		X

TV5600-CNS-20007 (LEGATO-HD)	Pre-treatment	Baseline (Dosing)	Double-blind Treatment Period						Follow-Up	Unscheduled Visit
Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8 - Termination Visit[a]	Visit 9	-
Time (Weeks)	Screening	Baseline	Week 4	Week 13	Week 19	Week 26	Week 39	Week 52	Week 56	-
Time (Months)		Month 0	Month 1	Month 3	☎	Month 6	Month 9	Month 12 (or ET)		
Visit Window	2 weeks up to 5 weeks		±3 days	±7 days	±7 days	±7 days	±7 days	±7 days	±5 days	
Concomitant Medication Inquiry		Concomitant medication will be monitored throughout the treatment and follow-up periods.								
Adverse events inquiry		Adverse events will be monitored throughout the treatment and follow-up periods.								

[a] For patients in the high dose group (1.5 mg/day) who were discontinued, only safety and no efficacy assessments had to be performed at the Early Termination visit. The patients will be asked to continue all scheduled visits for safety assessment only after study drug discontinuation.
[b] Patients with a legal guardian should be consented according to local requirements.
[c] Inclusion/exclusion criteria should be met at screening and reviewed at baseline (Visit 2) before the patient is randomized.
[d] Including smoking history. In addition, an evaluation of cardiovascular risk factors should take place as soon as possible for patients already in the study, following approval of Global Amendment 04.
[e] Assessment of changes in cardiovascular risk and appropriate cardiovascular risk management with appropriate medical follow-up, if clinically indicated, should be performed during the scheduled and unscheduled visits.
[f] When applicable, patients will be screened for drug substances in urine and/or CDT level in blood at screening to confirm abstinence in former (more than 12 months from screening) alcohol and/or drug abusers.
[g] For visits 2 through 9, Patients must have fasted no less than 8 hours prior to the blood draw.
[h] When applicable per local requirements, patients will undergo an HIV test at screening.
[i] Unscheduled urgent safety laboratory samples, pharmacokinetic blood samples, and/or samples for potential biomarker analysis may be collected at the discretion of the investigator or medical monitor at any time to assist with further investigations of cardiovascular events or other clinical event of interest. The samples should be collected as soon as possible in association with the event.
[j] Anemia panel includes B12, blood smear, serum iron, ferritin, total iron binding capacity, folic acid, haptoglobin, IL-1, IL-6, IFN-γ, TNF-α, and hepcidin. Assessed at baseline and also at 1 subsequent time point with B12 if hemoglobin decrease of >1 g/dL from the patient's hemoglobin level at baseline and the decrease is confirmed.
[k] Only at selected sites in a subgroup of patients
[l] Blood sample for monocyte collection and analysis has to be done in accordance with the laboratory manual. For UK patients, the sample can be drawn prior to the baseline or Month 12 visit, as it coincides with the PET scan visits
[m] Height will be assessed during the physical exam at screening (Visit 1)
[n] ECG will be performed in triplicate at baseline (approximately 10±5 minutes apart). All other visits will have a single ECG performed.
o At the screening visit, only serum β-HCG test will be performed. For the baseline visit - urine pregnancy test (beta human chorionic gonadotropin [β-hCG]) result is required for women of child-bearing potential prior to randomization. The urine test should be conducted at the site to allow randomization. A serum β-HCG test should also be used to confirm the urine test; however, the randomization should be performed based on the results of the urine pregnancy test.

[p] An additional urine β-HCG test will be performed at the PET facilities at Imperial College in London at the Baseline visit and at Visit 8 for women of child-bearing potential participating in the PET substudy. Its result will be available on site prior to any procedures involving ionizing radiation and will be considered accordingly.
[q] If anxiolysis is required in order to perform the MRI scan, the scan should be performed at the end of the study visit day.
[r] The baseline MRI can be performed at the screening visit, but no later than 7 days prior to baseline. If performed at the screening visit, as many screening assessements as possible should be conducted prior to the MRI scan in order to assess eligibility.
[s] If a patient terminates within 3 months of the baseline visit, they will not undergo an Early Termination scan. If a patient terminates prior to Month 12, the Early termination scans should be performed as soon as possible, but not more than 7 days after discontinuation of study drug. Month 12 scans should be performed 7 days prior to the Termination Visit. The early termination MRI scan was not to be done for the patients in the discontinued 1.5 mg/day treatment arm .
[t] Estimated creatinine clearance will be calculated at all in clinic study visits. Patients who develop chronic renal disease associated with moderate or severe functional impairment, defined as estimated creatinine clearance (CrCl) <60 mL/min/1.73 m^2, while participating in the study should stop study medication temporarily and the creatinine clearance assessment should be repeated. If the renal impairment is confirmed (estimated CrCl <60 mL/min/1.73 m^2), the patient should stop study medication permanently.
[u] The PET scan can be done at any time prior to the date of the baseline visit, after eligibility of the patient has been confirmed. It can also be done 2 to 5 weeks prior to the Month 12 visit.
[v] Patients taking benzodiazepines should be instructed to stop taking them for 3 days prior to having the PET scan.
[w] UHDRS-TMS>5 and UHDRS-TFC≥8 are required for study eligibility
[x] Where possible, the same person should act as a patient's caregiver/informant throughout the study. If this is not possible, a patient should have no more than 2 caregivers throughout the study. All possible attempts should be made to assure that caregiver will attend the clinical visits in person together with the patient. If the caregiver/informant is not available to attend the clinic visit, the interview can be done over phone.
[y] HD-QoL will be assessed by both caregiver/informant and patient. All possible attempts should be made to assure that caregiver/informant will attend the clinical visits in person together with the patient. If the caregiver/informant is not available to attend the clinic visit, the caregiver/informant form should be omitted.
[z] The WLQ/EQ-5D-5L could be completed by the patients with caregiver/informant assistance if needed.
[aa] Includes Symbol Digit Modalities Test (SDMT), Emotion Recognition, Trail Making Test, Hopkins Verbal Learning Test, revised (HVLT-R), Paced Tapping at 3 Hz, One Touch Stockings of Cambridge (OTS, abbreviated 10 trial version).
[bb] The HD-CAB is performed at screening to reduce practice effects during the treatment phase.
[cc] Reduced battery (only SDMT and Trail Making Test).
[dd] Only at selected sites in a subgroup of patients (approximately 15 patients per each of the three continuing treatment groups, for a total of approximately 45 patients) at Month 1. Patients participating in the 24-hour PK profiling will not have the single PK sample drawn at Visit 3. Additionally, PK samples were collected from 2 patients in the laquinimod 1.5 mg/day treatment group when the treatment was stopped; no further PK samples will be collected from the laquinimod 1.5 mg/day treatment group.

Abbreviations: ET = early termination; CDT = carbohydrate deficient transferrin; ECG = electrocardiogram; C-SSRS = Columbia-Suicide Severity Rating Scale; UHDRS = Unified Huntington's Disease Rating Scale; CIBIS = Clinician's Interview-Based Impression of Severity; CIBIC-Plus = Clinician's Interview-Based Impression of Change plus Caregiver Input; HD-CAB = HD Cognitive Assessment Battery; HD-QoL = Huntington's disease Quality of Life; HIV = human immunodeficiency virus; CAG = cytosine-adenine-guanine; TMS = Total Motor Score; PBA-s = Problem Behaviors Assessment-Short form; SDMT = Symbol Digit Modalities Test.TFC = Total Functional Capacity; Q-Motor = Quantitative motor; CDR-SB = Clinical Dementia Rating – Sum of Boxes; HADS = Hospital Anxiety and Depression Scale

2.2. Randomization and Blinding

This is a randomized, double-blind, placebo-controlled study. Prior to 10 January 2016, eligible patients were randomly assigned to receive treatment with laquinimod at a dosage of 0.5, 1.0 or 1.5 mg qd or a matching placebo in a 1:1:1:1 ratio.

As of 10 January 2016, following a decision to discontinue treatment of the laquinimod 1.5 mg dose arm, future eligible patients were to be randomized in a 1:1:1 ratio to receive laquinimod 0.5 mg/day or 1.0 mg/day or matching placebo for 52 weeks. No change was performed to the original randomization list except that the patient numbers assigned to laquinimod 1.5 mg/day were removed from the list by the interactive response technology (IRT) vendor.

All patients who discontinued the 1.5 mg/day dose have been unblinded. No attempts were made to re-randomize patients whose 1.5 mg treatment was stopped to a lower dose of laquinimod. The remaining ongoing patients retained their originally randomized treatment assignments.

Patients and investigators will remain blinded to treatment assignment during the study.

Patients were randomly assigned to treatment through a qualified randomization service provider (eg, IRT). This system is used to ensure a balance across treatment groups.

The randomization code were generated by the Clinical Supply Chain (CSC) department following specifications from the Biostatistics Department.

In addition, the sponsor's clinical personnel involved in the study are blinded to the study drug identity until the database is locked for analysis and the treatment assignment revealed. However, in case a prioritized sample analysis is needed, bioanalytical personnel may not be blinded .

A statistician not assigned to the study will be responsible for reviewing the randomization code, and the final randomization code will be maintained by the CSC department.

2.3. Data Monitoring Committee

An independent Data Safety Monitoring Board (DSMB) will oversee the study. The DSMB will review unblinded accumulating safety data on a regular basis to ensure the continuing safety of the study patients and study conduct issues. An external unblinded statistician will provide the unblinded data to DSMB.

2.4. Primary, Secondary and other Efficacy Variables and Safety Variables

2.4.1. Primary Efficacy Variable

The primary objective of this study is to assess the efficacy of laquinimod 1.0 mg in patients with HD after 12 months of treatment using the UHDRS-TMS.

2.4.2. Secondary Efficacy Variable

The secondary objective of this study is to assess the effect of laquinimod 1.0 mg in patients with HD after 12 months of treatment on brain atrophy using MRI measures of caudate volume.

2.4.3. Other Efficacy Variables

Analyses of the primary and secondary endpoints on the 0.5 mg laquinimod arm vs. placebo arm:

- Change from baseline in UHDRS-TMS at week 52;
- Percent change from baseline in caudate volume at week 52.

Additional analyses of each dose, 0.5 mg and 1.0 mg laquinimod, in comparison with placebo:

- Change from baseline in UHDRS-TFC at week 52;
- Change from baseline in HD-CAB composite score at week 52;
- CIBIC-Plus global score at week 52 as compared to baseline (rated by an independent rater);
- Change in brain volume (i.e. atrophy) evaluated at week 52/Early Termination (ET) versus baseline as defined by
  - Percent change from baseline in whole brain volume at week 52/ET;
  - Percent change from baseline in white-matter volume at week 52/ET;
  - Absolute change in ventricular volume at week 52/ET;
- Change from baseline in UHDRS Functional Assessment (FA) at week 52/ET (evaluated at baseline,week 26 and 52)
- Change from baseline in Quantitative-Motor (Q-Motor) assessments at week 52/ET (evaluated at baseline and week 4, 13, 26 and 52)
- Change from baseline in the modified physical performance test (mPPT) at week 52/ET (evaluated at baseline and week 26 and 52)
- Change from baseline in the Huntington's Disease Quality of Life (HD QoL) and 5-Level EuroQol Five Dimensions Questionnaire (EQ-5D-5L) at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in the Work Limitations Questionnaire (WLQ) at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in HD-CAB sub-components at week 52/ET (evaluated at baseline and week 26 and 52) - Symbol Digit Modalities Test (SDMT), Emotion Recognition, Trail Making Test, Hopkins Verbal Learning Test, revised (HVLT-R), Paced Tapping at 3 Hz, One Touch Stockings of Cambridge (OTS, abbreviated 10 trial version)
- Change from baseline in the Clinical Dementia Rating - Sum of Boxes (CDR-SB) at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in the Hospital Anxiety and Depression Scale (HADS) at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in the Problem Behaviors Assessment-Short form (PBA-s) at week 52/ET (evaluated at baseline and week 52)

Pharmacokinetics (PK) analysis

- To potentially evaluate the plasma concentrations of laquinimod in patients with HD

2.5. Sample Size and Power Considerations

The study aims to detect potential beneficial effects in deteriorating clinical signs and symptoms. Based on previous studies in patients with HD, the UHDRS-TMS has been shown to be one of the more sensitive clinical measures to detect decline in symptoms of HD. It is estimated that approximately 100 patients per arm will provide a power of 80% to detect a significant effect of an active laquinimod arm compared to placebo, assuming a true mean difference of 2.5 points or more in the change from baseline in UHDRS-TMS, standard deviation (SD) of 6.2 and type I error of 5%.

As the intention is to investigate laquinimod as a treatment with the potential to slow disease progression and prohibit neuronal death in the CNS, the study should also be sized to be able to detect changes in brain atrophy rate after treatment with laquinimod. One of the most sensitive measures to detect brain atrophy over time in patients with HD is change in the caudate volume.

It is estimated that approximately 100 patients per arm will enable a power of 80% to detect a beneficial effect of 0.95 (30% of the estimated decline in placebo) or more in the percent change from baseline in caudate brain atrophy of an active laquinimod arm compared to placebo, assuming SD of 2.36 and type I error of 5%.

2.6. Sequence of Planned Analyses

2.6.1. Planned Interim Analyses

No interim analysis is planned for this study.

2.6.2. Final Analyses and Reporting

All analyses identified in this Statistical Analysis Plan will be performed after the end of study as defined in the study protocol.

This Statistical Analysis Plan and any corresponding amendments will be approved before database lock, in accordance to SOP GBP_RD_702 (Statistical Analysis Plan).

The randomization codes will not be unblinded until this Statistical Analysis Plan has been approved and issued and the database has been locked.

Any results of ancillary studies or exploratory analyses completed to support study analyses, which were not identified in this SAP, will be documented and reported separately as addendums to the CSR.

3. ANALYSIS SETS

3.1. Intent-to-Treat Population

The intent-to-treat (ITT) population will include all randomized patients. In this population, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

3.2. Safety Population

The safety population will include all randomized patients who receive at least 1 dose of study drug. In this population, treatment will be assigned based upon the treatment patients actually receive, regardless of the treatment to which they were randomized.

3.3. Full Analysis Set

The full analysis set (FAS) will include all patients in the ITT population who receive at least 1 dose of study drug and have at least 1 post-baseline TMS assessment. In analyses and summaries based on the FAS analysis set, patients will be included in the treatment group to which they were randomized, regardless of the treatment that was actually received.

3.4. Pharmacokinetic Analysis Set

The Pharmacokinetic analysis set (PAS) will include all patients from the safety analysis set who also have ≥1 plasma concentration measured.

4. GENERAL ISSUES FOR DATA ANALYSIS

4.1. General

Descriptive statistics for continuous variables include n, mean, SD, standard error (SE), median, minimum, and maximum. Descriptive statistics for categorical variables include patient counts and percentages, missing category will be displayed as appropriate.

4.2. Specification of Baseline Values

Baseline is defined as the last observed data before the first dose of study drug, unless otherwise noted.

4.3. Handling Withdrawals and Missing Data

For all variables, only the observed data from the patients will be used in the by visit summary.

For the TMS component of the UHDRS, if responses of up to 25% of items are missing, the missing responses will be replaced by the average of the remaining responses within the TMS component. If eight or more of the items are missing, the missing items will not be replaced and the TMS will be set to missing.

For the TFC component of the UHDRS, if responses of up to 25% of items are missing, the missing responses will be replaced by the average of the remaining responses within the TFC scale. If responses to more than 25% of the items are missing, the missing responses will not be replaced and the TFC will be set to missing.

For the FA component of the UHDRS, if responses to 1 assessment up to 25% of assessments are missing, the missing responses will be replaced by the average of the remaining responses within the FA component. If more than 25% of the assessments are missing, the missing responses will not be replaced and the UHDRS-FA will be set to missing.

For mPPT, item 1 (balance tasks), item 4 (put on and remove a jacket), and item 6 (turn 360 degrees) have to be calculated before calculating the total score. For all 3 items, if any assessment is missing then the missing responses will not be replaced and the item will be set to missing. After calculating items 1, 4, and 6, if responses of up to two items (25% of 9) are missing, the missing responses will be replaced by the average of the remaining responses. If responses to more than two items are missing, the missing responses will not be replaced and the total score will be set to missing.

For PBA-s, the (severity times frequency) scores per item need to be calculated before calculating the total score. If either the severity or frequency value is missing then the item will be equal to the nonmissing value. After calculating the items, if responses of at least one and up to 25% of items are missing, the missing responses will be replaced by the average of the remaining responses. If responses to more than 25% of the items are missing, the missing responses will not be replaced and the total score will be set to missing.

For tests from the Q-Motor assessments, if either the left or right value is missing, then the average value will be equal to the nonmissing value.

If more than 1 out 6 HD-CAB tests is missing, the HD-CAB composite score will be set to missing.

For the four WLQ scales, if one of the item responses is missing (either left blank or includes a "does not apply to my job" response), then the half-scale imputation rule is applied by assigning the valid item score to the missing item. If both of the items within a scale are missing, the WLQ scale score will be left missing.

The possible impact of missing values on the analysis will not be evaluated.

4.4. Study Days and Visits

Study days are numbered relative to the first day of study drug administration. The start of treatment (Day 1) is defined as the date on which a patient takes the first dose of study drug, as recorded on the Case report form (CRF). Days will be numbered relative to treatment start (ie, ..., –2, –1, 1, 2, ...; with day 1 being the first day of study drug administration and day –1 being the day before the first day of study drug administration).

The actual visit day will be verified against the allowed visit window according to Table 1.

For by-visit summaries, if there are multiple assessments at a postbaseline visit day, then the last non-missing assessment at that visit day will be used for the summary, for both scheduled and unscheduled assessments.

For by-visit efficacy summaries and/or statistical models, early termination or unscheduled visit for a specific measure will be assigned to the closest non-missing planned scheduled visit of this measure. See Table 2 for the rules of the assignment.

For safety summaries, unscheduled, ET and Follow Up visits will be mapped using visit windows according to Table 1.

For both safety and efficacy summaries, if there is a scheduled visit assessment in the visit window, it will be preferably used in the analysis. Among the unscheduled and ET visit (and also Follow Up for safety data), the visit with closest date to the planned scheduled visit will be used.

For last safety assessment, the last available non-missing value will be used. Unscheduled, ET and Follow Up visits that could not be mapped will not be displayed in the safety by visit summaries except they will be considered for the endpoint/last assessment visit.

All assessments will be listed.

Table 2: Visit Windows in days for Early Termination and Unscheduled Visits

Assessments	Time Points			
	Visit 3 (Month 1/ Week 4)	**Visit 4 (Month 3/ Week 13)**	**Visit 6 (Month 6/ Week 26)**	**Visit 8 (Month 12/ Week 52)**
UHDRS-TMS Q-Motor Assessments HD-CAB[a]	2-60	61-135	136-270	≥271
UHDRS-TFC UHDRS-FA mPPT CIBIC-plus HD-CAB[b]			2-270	≥271
PBA-s HD-QoL EQ-5D-5L HADS WLQ CDR-SB MRI[c]				>180

[a] SDMT and Trail Making Test

[b] Other than SDMT and Trail Making Test

[c] The MRI measurement after 6 months (180 days) on study and out of the window ±7 days from the day 365 will be annualized assuming linear change with time. The annualized value will be presented on the descriptive tables and used in the analysis. See details in Section 6.3.1

5. STUDY POPULATION

5.1. General

The ITT population will be used for all study population summaries unless otherwise noted. Summaries will be presented by treatment group and for all patients unless otherwise noted.

Due to the decision from 10 January 2016 to discontinue treatment of the laquinimod 1.5 mg dose arm, and the low number of enrolled patients compared to the target at that time, data from the laquinimod 1.5 mg treatment arm will be presented descriptively only, and will not be included in any inferential analyses for efficacy or safety.

5.2. Patient Disposition

Data from patients screened, patients screened but not randomized, patients in the ITT population (ie, randomized patients), patients who were randomized but not treated, patients in the safety, FAS population, patients who complete the study and patients who withdraw from the study, as well as patients who complete the treatment and patients who withdraw from the treatment, will be summarized using descriptive statistics. Data from patients who withdraw from the study as well as patients who withdraw from the treatment will also be summarized by reason for withdrawal using descriptive statistics.

Patients in the high dose group (1.5mg/day), who were requested to discontinue study drug before week 52 but continued to attend scheduled study visits for safety assessments, will be considered to have completed the study, but not treatment (see Section 16.2).

5.3. Demographics and Baseline Characteristics

Patient demographic and baseline characteristics will be examined regarding the comparability of the treatment groups and will be summarized using descriptive statistics using both ITT and FAS analysis sets. For continuous variables, descriptive statistics (number, mean, SD, SE, median, minimum, and maximum) will be provided. For categorical variables, patient counts and percentages will be provided. Categories for missing data will be presented if necessary. P-values for the simultaneous testing (for detecting the difference between at least two treatment groups) of the placebo, laquinimod 0.5 mg and laquinimod 1.0 mg treatment groups will be presented.

For demographics, the continuous variables of patient age, weight, height, and body mass index (BMI) will be summarized using descriptive statistics. The categorical variables of patient sex, race, and ethnicity will be summarized using descriptive statistics for each category.

For baseline characteristics, the number of Cytosine-Adenosine-Guanine (CAG) repeats, efficacy and safety scales scores will be summarized using descriptive statistics.

Months from HD diagnosis and months from onset of first HD symptoms at the study consent day will be presented. In case the reported dates are partial, with a missing day of the month or missing month, full dates will be imputed. If the day is missing, the first day of the month will be imputed. If the month is missing, it will be imputed as January of that year.

Smoker status will be summarized using descriptive statistics.

5.3.1. Baseline MRI Parameters

The baseline MRI parameters will be presented using descriptive statistics using both ITT and FAS analysis sets: caudate volume, total intracranial volume, ventricular volume, whole brain volume and white-matter volume.

The baseline normalized volume will be calculated by dividing the volume for each region by total intracranial volume, and will be presented using descriptive statistics using both ITT and FAS analysis sets. The total intracranial volume is also called pseudo total intracranial volume, since it is unitless.

The baseline volumes for caudate, whole brain and lateral ventricles are outlined by a trained analyst (semi-automatically) and subsequentially do not have associated endpoint Quality Control (QC) grades. However, total intracranial volume and baseline white-matter volume volume are generated completely automatically and do have endpoint QC grades, due to the importance of testing the reliability of the automated pipeline at this stage.

The baseline white-matter volume endpoint QC is different from the longitudinal white-matter volume change endpoint QC, as these are different processes/pipelines. Baseline white-matter volume endpoint QC represents the reliability of the white matter volume numerical result at baseline. The longitudinal change in white-matter volume endpoint QC grade represents the reliability of the volume change measurement over the scanning interval.

Where there is an endpoint QC fail, the data point will not be shown and will be further removed from the statistical analysis.

5.4. Medical History

All medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of medical history abnormalities will be summarized using descriptive statistics by system organ class (SOC) and preferred term. Patients are counted only once in each preferred term and SOC category. Summaries will be presented by treatment group and for all patients.

5.5. Prior Therapy and Medication

Any prior therapy, medication, or procedure a patient has had within 3 months before study drug administration will be recorded on the CRF. Trade name or International Nonproprietary Name (INN), indication, and dosage will be recorded. The sponsor will encode all therapy and medication according to the World Health Organization drug dictionary (WHO Drug).

The incidence of prior therapies and medications will be summarized using descriptive statistics by therapeutic class and preferred term. Patients are counted only once in each therapeutic class category, and only once in each preferred term category. Prior therapies and medications will include all medications taken and therapies administered before the first day of study drug administration.

5.6. Electrocardiography

Electrocardiogram (ECG) findings (normal and abnormal) at baseline will be summarized using descriptive statistics. The worst result of up to three last measurements during the last visit prior to first dose administration will be used for the interpretation.

5.7. Childbearing Potential and Methods of Contraception

For female patients, information related to childbearing potential, and menopause will be collected at all visits including Week 19 telephone visit. The data will be listed.

5.8. Subject Characteristics

An evaluation of cardiovascular risk factors should take place as soon as possible for patients already in the study, following approval of Global Amendment 04. Assessment of changes in cardiovascular risk and appropriate cardiovascular risk management with appropriate medical follow-up, if clinically indicated, should be performed during the scheduled and unscheduled visits.

Liver disease risk factor evaluation will be performed for the patients who develop any chronic liver disease associated with hepatic function impairment.

Data will be listed.

5.9. Study Protocol Violations

Data from patients with any protocol violations (as recorded in protocol violation CRF) during the study will be summarized overall and for each category using descriptive statistics. Of note, medication errors, overdose, misuse, abuse, off-label use, and occupational exposure will be summarized in category "Non-Compliance to study medication" as per study protocol.

6. EFFICACY ANALYSIS

6.1. General

The FAS will be used for efficacy analyses. Summaries will be presented by treatment unless otherwise noted. As noted in Section 5.1, the data from the laquinimod 1.5 mg treatment arm will be presented descriptively only, and will not be included in any inferential analyses for efficacy.

6.2. Primary Efficacy Endpoint and Analysis

6.2.1. Definition

The UHDRS comprises a broad assessment of features associated with HD. It is a research tool which has been developed to provide a uniform assessment of the clinical features and course of HD. The TMS component of UHDRS comprises 31 assessments from the 15 items of the UHDRS. The TMS is calculated as the sum of the 31 motor assessments.

The primary efficacy variable and endpoint for this study is the change from baseline to week 52/ ET (evaluated at baseline and week 4, 13, 26 and 52) in the TMS. Specifically, the change from baseline to week 52/ET in the TMS comparison between laquinimod 1 mg and placebo is prioritized to serve as the primary efficacy endpoint, while the change from baseline to week 52/ET in the TMS comparison between laquinimod 0.5 mg and placebo will be an exploratory endpoint.

A detailed description for all rating scales can be found in the protocol and below.

6.2.2. Primary Efficacy Analysis

The primary analysis of efficacy will be performed using the FAS analysis set. The change from baseline to week 52/ET in the TMS will be compared between laquinimod 1.0 mg and placebo using a Mixed Model Repeated Measures model (MMRM) using SAS® MIXED procedure with REPEATED sub-command. The model will include the following fixed effects: treatment group with 3 levels: placebo, laquinimod 0.5 mg and laquinimod 1 mg; categorical week in trial with 4 levels: week 4, 13, 26 and 52; treatment by week interaction, country, TMS at baseline and week by TMS at baseline interaction. Categorical week in study by baseline TMS interaction term was added to the primary analysis model since it was found statistically significant in the previous HD studies. Subject will be a random effect.

The unstructured covariance matrix for repeated observations within patients will be used. In case that the model does not converge, the Maximum-Likelihood (ML) estimation method will be used instead of the default Restricted ML (REML). If the model still does not converge then a simpler covariance structures with less parameters will be used, according to the following order: Heterogeneous Autoregressive(1) [ARH(1)], Heterogeneous Compound Symmetry (CSH), Autoregressive(1) [AR(1)], and Compound Symmetry (CS). The Kenward-Rodger (KR) method will be used to calculate the denominator degrees of freedom.

The least square (LS) mean and standard error for the laquinimod 0.5 mg , laquinimod 1.0 mg and placebo groups, and the LS mean difference, 95% confidence interval (CI), and p-value for each laquinimod dose versus placebo comparison (laquinimod 0.5 mg vs placebo and laquinimod 1.0 mg vs placebo) will be presented at all visits.

The inferential p-value for the primary endpoint is for the comparison of laquinimod 1.0 mg vs placebo at week 52.

The comparison between laquinimod 0.5 mg and placebo change from baseline to week 52/ET in the TMS will serve as an exploratory endpoint. See Section 6.4.1 for details.

The SAS code for the repeated measures model is as follows:

```
Ods Output LSMeans=LSM Diffs=Diff;
Proc Mixed Method=REML;

Class USUBJID Arm Week country ;

Model TMSChange = Arm | Week  country TMSBaseline Week*TMSBaseline /DDFM=KR;

Repeated Week /Type=Un Subject= USUBJID R Rcorr;

LSMeans  Arm*Week / PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

The hypotheses for testing the primary analysis are:

$$H_o: \mu_a \equiv \mu_p \text{ vs. } H_a: \mu_a \neq \mu_p$$

where μ_a is the mean change from baseline to week 52/ET in the TMS for the laquinimod 1.0 mg arm and μ_p is the mean change from baseline to week 52/ET in the TMS for placebo arm. This two-sided test will be performed at the alpha level of 0.9*0.05=0.045, as supported by the Fallback with Loop-back approach for the familywise Type I error control for the primary and secondary endpoints. Details are provided in Section 7.

The TMS score values and changes from baseline to week 4, 13, 26 and 52 will be summarized using descriptive statistics for all treatment groups.

6.2.3. Sub-Group Analysis

Subgroup analysis will be performed for the primary efficacy endpoint by gender, median baseline TMS / TFC / caudate volume / CAG repeat length and USA vs outside of USA. Data will be analyzed in a manner analogous to the method described in the Section 6.2.2. Subgroup analysis will be performed on the FAS analysis set with fixed effects for subgroup, treatment, week and their interaction.

```
Ods Output LSMeans=LSM Diffs=Diff;
Proc Mixed Method=REML;

Class USUBJID Arm Week country Subgroup;

Model TMSChange = Arm | Week | Subgroup  country TMSBaseline Week*TMSBaseline
/DDFM=KR;

Repeated Week /Type=Un Subject= USUBJID R Rcorr;

LSMeans  Arm*Week*Subgroup /PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

6.3. Secondary Efficacy Endpoint and Analysis

6.3.1. Definition

The percent change from baseline in caudate volume at week 52 / percent caudate atrophy from baseline at week 52 was chosen as the secondary objective for the study.

The percent change in caudate volume at week 52 or percent caudate atrophy at week 52 is calculated as the change in caudate volume since baseline visit divided by the baseline caudate volume and multiplied by 100.

In this study, brain MRI is used to assess the change in volume of certain regions, including the caudate, whole brain, white-matter, and ventricles, over the time period from baseline to the end of study. Participants will undergo 3 Tesla MRI at baseline and week 52 following unaccelerated volumetric T1-weighted acquisition protocols developed during the ADNI study (www.adni-info.org). Change in caudate volume over the scanning interval was shown to provide robust measures of brain-volume change from multi-site data and has been shown to be sensitive to HD-related pathology over a 12-month interval in the multi-site TRACK-HD study.

A more detailed description of the MRI assessments can be found in the protocol.

Brain atrophy in the caudate and whole brain refers to the shrinkage in volume of these structures, so that a decrease in volume will be a positive value, while an increase in volume will be a negative value. In contrast, for white matter and ventricles a decrease in volume will be a negative value and an increase in volume will be a positive value. Since ventricles are fluid filled spaces, an increase in the volume of the ventricles (i.e., ventricle expansion) reflects a loss of volume of the surrounding brain tissues. The SAP uses the terms "atrophy" and "change in volume" interchangeably, referring to the sign convention explained hitherto.

The caudate atrophy since baseline measurement after 6 months (180 days) on study and out of the scanning window ±7 days from the day 365, will be annualized assuming linear change with time. The annualization calculation formula will be as follows: denoting the distance between the baseline and postbaseline MRI scans in days as X, the caudate atrophy will be multiplied by the factor of 365/X. The annualized value will be presented in the descriptive tables and will be used in the analysis.

Each caudate atrophy result, as well as the other MRI measurements, is accompanied by a specific QC result with a response value "Pass" or "Fail". If the QC results in "Fail", the corresponding caudate atrophy result should not be used in the statistical analysis. All the results will be listed.

6.3.2. Analysis

The Percent caudate atrophy from baseline to week 52/ET will be compared between laquinimod 1.0 mg and placebo using an Analysis Of Covariance (ANCOVA) model (SAS® MIXED procedure). The model will include the following fixed effects: treatment group (Arm, 3 levels: placebo, laquinimod 0.5 mg and laquinimod 1 mg), country, and caudate volume at baseline (CVBaseline).

The estimated means at week 52 will be compared between the laquinimod 1.0 mg and the placebo arm, presented with the least square (LS) mean and SE for the laquinimod 1.0 mg and placebo groups, and the LS mean difference, 95% CI, and inferential p-value for the comparison (laquinimod 1.0 mg vs placebo).

Sample SAS code is as follows:

```
Proc Mixed;

Class Arm country;

Model Change=Arm country CVBaseline;

LSMeans Arm /PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

The analysis will be based on the FAS analysis set. A Fallback with Loop-back approach for the Type I error control will be applied. If the primary endpoint is significant at a 2-sided alpha of 0.045 then the secondary endpoint will be tested at the 2-sided alpha level of 0.05. If the primary endpoint fails to reach significance, then the secondary endpoint will be tested at the 2-sided alpha level of 0.005. Details are described in Section 7.

Percent of change in caudate volume (i.e. atrophy) from baseline to week 52/ET will be summarized using descriptive statistics for all treatment groups.

The percent change in caudate volume from baseline to week 52/ET compared between laquinimod 0.5 mg and placebo will serve as an exploratory endpoint. See Section 6.4.2 for details.

6.3.3. Sub-Group Analysis

Subgroup analysis will be performed for the secondary efficacy endpoint by gender, median baseline TMS / TFC / caudate volume / CAG repeat length and USA vs outside of USA. Data will be analyzed in a manner analogous to the method described in the Section 6.3.2. Subgroup analysis will be performed on the FAS analysis set with fixed effects for gender, treatment and their interaction.

```
Proc Mixed;
Class Arm country Subgroup;
Model Change=Arm | Subgroup country CVBaseline;
LSMeans Arm*Subgroup /PDiff CL;
Where Arm in ('Placebo', 'laquinimod 0.5mg', 'laquinimod 1mg');
Run;
```

6.4. Exploratory Efficacy Endpoints Analysis

Exploratory efficacy endpoints will include the analyses of the primary and secondary endpoints applied on the 0.5 mg laquinimod arm vs. placebo arm, as well as additional endpoints, either based on rating scales or using MRI measures. See the list of the exploratory endpoints below.

Analyses of the primary and secondary endpoints on the 0.5 mg laquinimod arm vs placebo arm:

- Change from baseline in TMS at week 52;
- Percent change in caudate volume (i.e. atrophy) from baseline at week 52.

Additional analyses end-points to be compared between each dose of laquinimod, and placebo:

- Change from baseline in TFC at week 52;
- Change from baseline in HD-CAB composite score at week 52;
- CIBIC-Plus global score at week 52 as compared to baseline (rated by an independent rater);
- Brain atrophy evaluated at week 52/ET versus baseline as defined by
  - Percent whole brain volume change (i.e. atrophy) from baseline at week 52/ET;
  - Percent white-matter volume change from baseline at week 52/ET;
  - Absolute ventricular volume change at week 52/ET;
- Change from baseline in FA at week 52/ET (evaluated at baseline,week 26 and 52)
- Change from baseline in Q-Motor assessments at week 52/ET (evaluated at baseline and week 4, 13, 26 and 52)
- Change from baseline in mPPT at week 52/ET (evaluated at baseline and week 26 and 52)
- Change from baseline in HD QoL and EQ-5D-5L at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in WLQ at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in HD-CAB sub-components at week 52/ET (evaluated at baseline and week 26 and 52) - SDMT, Emotion Recognition, Trail Making Test, HVLT-R, Paced Tapping at 3 Hz, OTS (abbreviated 10 trial version)

- Change from baseline in CDR-SB at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in HADS at week 52/ET (evaluated at baseline and week 52)
- Change from baseline in PBA-s at week 52/ET (evaluated at baseline and week 52)

6.4.1. Change From Baseline in Total Motor Score (laquinimod 0.5 mg arm vs placebo) at Week 52

6.4.1.1. Definition

See Section 6.2.1.

6.4.1.2. Analysis

Analysis of change from baseline to Week 52 in UHDRS-TMS for the laquinimod 0.5 mg arm vs placebo will be analysed as described in Section 6.2.2. The hypothesis testing will be done on the mean change difference between the laquinimod 0.5 mg arm and the placebo arm.

6.4.2. Percent Change from Baseline in Caudate Volume at Week 52 (laquinimod 0.5 mg arm vs placebo)

6.4.2.1. Definition

See Section 6.3.1.

6.4.2.2. Analysis

The analysis will be performed as described in Section 6.3.2 with the hypothesis testing done on the mean change difference between the laquinimod 0.5 mg arm and the placebo arm.

6.4.3. Change from Baseline in the Total Functional Capacity at Week 52

6.4.3.1. Definition

The TFC scale of the UHDRS assesses 5 functional domains associated with disability (occupation, finances, domestic chores, activities of daily living, and care level) and is rated from 0-13 (maximum functionality). The TFC is the sum of the 5 functional capacity domains. TFC is evaluated at baseline, week 26 and 52.

6.4.3.2. Analysis

Analysis of change from baseline in the UHDRS-TFC at week 52 will be analyzed in a manner analogous to that described in Section 6.2.2 except that baseline TFC will be included in the model instead of baseline TMS (including the interaction term). Comparisons will be performed between laquinimod 1.0 mg arm and the placebo arm, as well as between laquinimod 0.5 mg arm and the placebo arm.

The SAS code for the repeated measures model is as follows:

```
Ods Output LSMeans=LSM Diffs=Diff;
Proc Mixed Method=REML;

Class USUBJID Arm Week country ;

Model  TFCChange = Arm | Week  country TFCBaseline Week*TFCBaseline /DDFM=KR;

Repeated Week /Type=Un Subject= USUBJID R Rcorr;

LSMeans  Arm*Week /PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

6.4.4. Change from Baseline in the Functional Assessment at Week 52

The FA scale of the UHDRS assesses functionality in 25 tasks of daily living (eg, "Could patient engage in gainful employment in his/her accustomed work?"). Each question is answered with 'yes=1' or 'no=0'. The FA component of the UHDRS is calculated as the sum of the 25 items.

The FA, similarly to TFC assessment, was collected at baseline, weeks 26 and 52. Analysis of change from baseline in the FA at week 52 will be analyzed using the same model as for TFC with the baseline FA replacing baseline TFC, see Section 6.4.3.2 for details. The comparisons will be the same.

6.4.5. Brain Atrophy Evaluated at Week 52/ET Versus Baseline

6.4.5.1. Definition

Brain atrophy will be evaluated at week 52/ET versus baseline as defined by

a. Percent change in whole brain volume (i.e. atrophy) from baseline at week 52/ET;
b. Percent change in white-matter volume from baseline at week 52/ET;
c. Absolute change in ventricular volume from baseline at week 52/ET;

See Section 6.3.1 for the details about the MRI assessments. In particular, the annualization will be performed on the postbaseline MRI measurement after 6 months (180 days) on study, by multiplying the change or the atrophy by the factor of 365/X, where X is the interval between the baseline and postbaseline MRI scans in days. The annualized value will be presented on the descriptive tables and will be used in the analysis.

Each change or the atrophy result is accompanied by a specific QC result with a response value "Pass" or "Fail". If the QC results in "Fail", the corresponding change or the atrophy result should not be used in the statistical analysis and will not be presented in the descriptive analysis.

6.4.5.2. Analysis

Analysis of percent whole brain atrophy, percent white-matter volume change and absolute ventricular volume change from baseline at week 52 will be analogous to that described in Section 6.3.2 with the corresponding baseline MRI measurement (MRIBaseline) as baseline covariate instead of caudate volume. Comparisons will be performed between laquinimod 1.0 mg arm and the placebo arm, as well as between laquinimod 0.5 mg arm and the placebo arm.

The SAS code for the ANCOVA model is as follows:

```
Proc Mixed;

Class Arm country;

Model Change=Arm country MRIBaseline;

LSMeans Arm /PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

6.4.6. Quantitative Motor Assessments (Q-Motor)

Q-Motor assessments will be performed only in those sites that have access to the devices needed to perform the assessments and, where this is the case, only in those patients who are capable of performing the assessments.

Motor deficits can be objectively assessed using different Q-Motor assessments. All Q-Motor assessments are based on the application of pre-calibrated and temperature controlled force transducers and 3-dimensional position sensors with very high sensitivity and test-retest reliability across sessions and sites in a multicenter clinical study. Q-Motor measures thus aim to reduce the limited sensitivity of categorical clinical rating scales, the intra- and inter-rater variability, and placebo effects observed in scales such as TMS. In addition, Q-Motor assessments allow for the objective monitoring of unintended motor side-effects in clinical studies.

Tasks detailed in the sections below have been selected for use in the current study. Data transfer will be performed using a secure web based platform, allowing continuous centralized data monitoring and quality control. Data analysis will be performed blinded and automated.

The Q-Motor assessments are collected at screening, baseline, week 4, 13, 26 and 52.

6.4.6.1. Digitomotography (Speeded Index Finger Tapping)

The patient will place their hand on a hand rest with their index finger positioned above a force-transducer. Recordings will start after practice runs. The patient will be instructed to finger tap as fast as possible between 2 auditory cues. The beginning of a tap is defined as a rise of the force by 0.05 N above maximal baseline level. The tap ends when it drops to 0.05 N before the maximal baseline level is reached again. The duration and variability of tap durations (TD), inter-onset intervals (IOI), inter peak intervals (IPI), and inter tap intervals (ITI) are the exploratory outcome measures for speeded tapping. In addition, variability of peak tapping forces (TF) will be calculated, and the tapping frequency (Freq), ie, the number of taps between the onsets of the first and the last tap divided by the time in between, will be determined. Five trials of 10 seconds duration are performed with each hand.

6.4.6.2. Dysdiadochomotography (Pronation/Supination Hand Tapping)

This task assesses the regularity of hand taps performed when alternating between the palm and dorsal surface of the hand performing a repetitive pronation/supination movement. The force and duration of the hand tapsare recorded and analyzed similarly to the speeded tapping task. A tone

cues the start and end of an assessment. Five trials of 10 seconds duration are performed with each hand.

6.4.6.3. Manumotography and Choreomotography (Grip Force and Chorea Analysis)

This task assesses the coordination of isometric grip forces in the precision grip between the thumb and index finger. Grip forces are assessed during grip initiation, object transport, and in a static holding phase. Patients are instructed to grasp and lift a device equipped with a force transducer and 3-dimensional position sensor in the precision grip between thumb and index finger and hold it stable adjacent to a marker 10-cm high. Grip forces and 3-dimensional position and orientation of the object are recorded. Mean isometric grip forces and grip force variability in the static phase (expressed as coefficient of variation = SD/mean × 100) (GFV-C) are calculated during a 15-second period prior to the second cueing tone.

Five trials of 20 seconds duration are performed with each hand. Chorea is assessed calculating a "position-index" and "orientation-index". Start and end of assessment are signaled by a cueing tone.

6.4.6.4. Pedomotography (Speeded Foot Tapping)

The patient will place a foot on the foot device such that the ball of the foot is positioned above a force-transducer. Recordings will start after practice runs. The patient will be instructed to tap with the foot as fast as possible between 2 auditory cues. The beginning of a tap is defined as a rise of the force by 0.05 N above maximal baseline level. The tap ends when it dropped to 0.05 N before the maximal baseline level is reached again. The duration and variability of TD, IOI, IPI, and ITI are the exploratory outcome measures for speeded foot tapping. In addition, variability of peak TF will be calculated, and the tapping Freq, ie, the number of taps between the onsets of the first and the last tap divided by the time in between, will be determined. Five trials of 10 seconds duration are performed with each foot.

6.4.6.5. Analysis

The Q-Motor assessments will be analyzed in the same way as described in Section 6.4.3.2 except that the matching baseline assessment will be included in the model as the efficacy measure at baseline.

The summary will be done by the left, right and the average of the two extremeties.

6.4.7. Change from Baseline in the Huntington's Disease Cognitive Assessment Battery Composite Score at Week 52 (laquinimod 1mg arm vs placebo)

6.4.7.1. Definition

The following sections describe the tests that will be included in the HD-CAB.

The CAB assessments will be performed only in those sites that have access to the devices needed to perform the assessments and, where this is the case, only in those patients who are capable of performing the assessments.

To avoid variability due to different operating systems, hardware and computer accessories (i.e. mouse), sites will be provided with standardized equipment (tablets) to perform the computerized cognitive assessments.

6.4.7.1.1. Symbol Digit Modalities Test (SDMT)

The SDMT is a paper-and-pencil test of attention, psychomotor speed and working memory. Participants view a 'key' at the top of the page containing symbols paired with numbers. The remainder of the page displays rows of symbols, and the participant has 90 seconds to write in the corresponding number that matches each symbol.

6.4.7.1.2. Emotion Recognition

Recognition of facial expressions of emotions is examined using computerized images of faces depicting 6 basic emotions or a neutral expression. Participants are asked to indicate the emotion expressed in each photograph by selecting from the words fear, disgust, happy, sad, surprise, angry, and neutral (a total of 36 items). The score used is correct identification of negative emotions (out of 24 possible).

6.4.7.1.3. Trail Making Test

Visual attention and task switching are assessed using the Trail Making test, which consists of 25 circles on a standard sheet of paper. For Trails A, participants are required to connect, as quickly as possible, circles containing numbers in ascending numerical order. For Trails B, participants are to connect, as quickly as possible, circles containing numbers and letters, alternating between numbers and letters in ascending order (eg, 1, A, 2, B, 3, C, etc.). Trail A is administered first (to ensure preparedness for Trail B), followed by Trail B. Only the Trail B score is used as an outcome measure in the HD-CAB.

6.4.7.1.4. Hopkins Verbal Learning Test, Revised (HVLT-R)

The HVLT-R is a paper-based instrument that offers a brief assessment of verbal learning and memory (recall). It is easy to administer and score and is well tolerated even by significantly impaired individuals.

Its use has been validated with brain-disordered populations (eg, Alzheimer's disease, HD, amnestic disorders) as a measure of verbal learning and memory. There are six alternate forms available, but 3 of the 6 forms, which have relatively greater equivalence with each other (Forms 4, 5, and 6) will be used, in randomized order. Each form consists of a list of 12 nouns (targets) with 4 words drawn from each of 3 semantic categories. The semantic categories differ across the 6 forms, but the forms are very similar in their psychometric properties. Only the three learning trials (Trials 1-3) and the delayed recall trial (Trial 4) will be administered as part of the HD-CAB. The primary scores that will be examined as part of the HD-CAB is the sum of Trials 1-4; however, the sum of Trials 1-3 and separately Trial 4 will also be examined in exploratory analyses. The HVLT-R has high test-retest reliability, and its construct, concurrent, and discriminant validity have been well established.

6.4.7.1.5. Paced Tapping at 3 Hz

Psychomotor function is assessed in a computerized Paced Tapping test. Participants tap on left and right mouse buttons, alternating between thumbs, at 3.0 Hz. They first listen to a tone presented at the 3.0 Hz rate, and then begin tapping in time with the tone. After 11 taps with the tone, the repetition tone is discontinued, and participants attempt to continue tapping at the same rate until the end of the trial (31 taps later). Four trials are administered.

6.4.7.1.6. One Touch Stocking of Cambridge (OTS)

OTS is a computerized spatial planning task which gives a measure of frontal lobe function. OTS is a variant of the Stockings of Cambridge task, and places greater demands on working memory as the participant has to visualize the solution. As with Stockings of Cambridge, the participant is shown 2 displays containing 3 colored balls. The displays are presented in such a way that they can easily be perceived as stacks of colored balls held in stockings or socks suspended from a beam.

Along the bottom of the screen, there is a row of numbered boxes. The test administrator first demonstrates to the participant how to use the balls in the lower display to copy the pattern in the upper display, and completes 1 demonstration problem, where the solution requires 1 move. The participant must then complete 3 further problems, 1 each of 2 moves, 3 moves, and 4 moves.

Next, the participant is shown further problems, and must work out in their head how many moves the solutions to these problems require, then select the appropriate box at the bottom of the screen to indicate their response.

6.4.7.1.7. HD-CAB Test Details and Calculation Mehthods

HD-CAB is evaluated at baseline, week 26 and 52. The HD-CAB consists of sub-components described in Table 3.

Table 3: HD-CAB Test Details and Calculation Methods

Test	Variable	Score Range		Interpretation of scores	Recode Method
		Min. value	Max. value		
Hopkins Verbal Learning Test-Revised	Total correct items summed across the 3 learning and 1 delayed recall trials	0	48	High scores = Better performance	No Recode
Symbol Digit Modalities Test	Number of correctly coded items	0	110	High scores = Better performance	No Recode
Trail Making Test: Part B	Time (seconds) to complete the task	0	240	High scores = Worse performance	No Recode
Emotion Recognition Test	Number of negative emotions correctly identified	0	24	High scores = Better performance	No Recode
Paced tapping test (3 Hz)	The reciprocal of the standard deviation of the inter-tap intervals that occurred following cessation of the pacing tones over all trials taken (msec). This must be computed from standard deviation, which is provided in the summary files.	0	Infinite	High scores = Better performance	Mutiply the reciprocal of the raw score (rs), which is the standard deviation, eg 1/rs by 1000
One Touch Stockings of Cambridge test	Mean time to reach a correct response (sec), averaged across all 10 trials	0	Infinite	High scores = Worse performance	No recode

The computation of the standardized HD-CAB composite score is described below.

1. Convert the recoded or raw scores of the 6 HD-CAB modules to standardized Z scores:

   Standardized Z score = (Raw/Recoded score - Total Baseline population mean)/Total Baseline population SD

   Total population should include all randomized patients. Note that the standardized scores are to be calculated on a per visit basis, using the raw/recoded score from that visit.

   To make Trail Making Test: Part B and One Touch Stockings of Cambridge consistent with other 4 HD-CAB tests, the Standardized Z score for Trail Making Test: Part B and One Touch Stockings of Cambridge will be multiplied by -1 for calculation of HD-CAB composite score.

2. For each patient and visit, calculate the average of the Z scores across the 6 HD-CAB tests as the HD-CAB composite score. This creates an equal weighting composite score.

6.4.7.2. Analysis

Change from baseline in the HD-CAB components and the composite score at week 52 will be analyzed in a manner analogous to that described in Section 6.4.3.2 except that appropriate baseline measurement will be included in the model instead of baseline TFC. The comparisons are the same.

6.4.8. Clinician's Interview-Based Impression of Change (CIBIC) - PLUS Global Score at Week 52

6.4.8.1. Definition

Global change in HD will be measured using the CIBIS scale at baseline and the CIBIC-Plus scale at subsequent time points. The CIBIC-Plus (version ADCS-CGIC) was developed, validated, and is commonly used in studies of anti-dementia drugs in Alzheimer's disease.

An independent rater whose only role in the study is to conduct these global assessments will evaluate the patient's overall disease severity during the baseline visit (Visit 2) prior to the administration of study drug. This assessment, known as the CIBIS, rates severity of the patient's HD on a 7-point Likert scale (1 = normal, not at all ill to 7 = among the most extremely ill patients).

At each subsequent visit in which the evaluation is performed (Months 6 and 12; Visits 6 and 8), the CIBIC-Plus will be preferentially administered by the same independent rater, but without knowledge of other endpoint assessments or the Adverse Events (AE) experienced by the patient during the study (so as not to confound the rating of CIBIC-Plus as an efficacy measure or to unblind the study). The independent rater is not permitted to discuss the medical condition of the patient with the treating physician. Instead, the independent rater exclusively will consider observations of the patient's cognitive, functional, and behavioral performance obtained through interviewing the patient and the caregiver. The rater then compares those findings to the baseline assessment. The overall impression of change from baseline (CIBIC-Plus) is rated on a 7-point scale: 1 = marked improvement; 2 = moderate improvement; 3 = minimal improvement; 4 = no

change; 5 = minimal worsening; 6 = moderate worsening; 7 = marked worsening; all assessments were relative to baseline. A higher score indicates a worsening of global function.

6.4.8.2. Analysis

The CIBIC-Plus will be analyzed in the same way as described in Section 6.4.3.2 except that the baseline CIBIS (Clinician's Interview-Based Impression of Severity) will be included in the model as the efficacy measure at baseline.

6.4.9. Modified Physcial Performance Test (mPPT)

6.4.9.1. Definition

The mPPT quantifies the patient's performance in physical tasks. It is a standardized 9-item test that measures the patient's performance on functional tasks. Both the speed and accuracy at which the patients complete the items are taken into account during scoring.

The mPPT total score is the sum of the mPPT items scores,

The maximum score of the test is 36, with higher scores indicating better performance.

The mPPT is collected at baseline, weeks 26 and 52/Early Termination.

The following is the testing protocol and scoring method:

Administer the test as outlined below. Patients are given up to two chances to complete each item. Assistive devices are permitted for tasks 6 – 9.

1. Standing Static Balance

- **Feet together**: "Stand still with your feet together as demonstrated for 10 seconds."
- **Semi Tandem**: "Stand with the heel of one foot placed to the side of the 1st toe of the opposite foot for 10 seconds." Subject chooses which foot goes forward.
- **Tandem**: "Stand with the heel of one foot directly in front of the other foot, for 10 seconds. Patient chooses which foot goes forward.

2. Chair Rise: Use a straight back chair with a solid seat that is 16" high. Ask participant to sit on the chair with arms folded across their chest. "Stand up and sit down as **quickly** as possible 5 times, keeping your arms folded across your chest." Stop timing when the participant stands the 5th time.

3. Book Lift: Place a Physician's Desk Reference Book (1988 PDR: 5.5 lbs) or other heavy book on a table in front of the patient. Ask the patient, when given the command "go" to place the book, as **quickly** as they can, on a shelf above shoulder level. Time from the command "go" until when the book is resting on the shelf. Starting position is with their hands at their side.

4. Put on and remove a jacket: If the patient has a jacket or cardigan sweater, ask them to remove it. If not, give the patient a lab coat. Ask the patient, on the command "go" to **quickly** put the coat on completely such that it is straight on their shoulders and then remove the garment completely. Time from the command "go" until the garment has been completely removed. Hint: it is more accurate to time putting on the garment, then pause (pause the stopwatch), then time taking off the garment.

5. Pick up a penny from floor: Place a penny approximately 12 inches from the patient's foot on the dominant side. Ask the patient, on the command "go" to pick up the penny from the floor and stand up.

This is to be done as quickly as they can; yet allowing for safety and comfort. Time from the command "go" until the patient is standing erect with a penny in hand. If dexterity is a problem, a pen or similar lightweight object can be used.

6. Turn 360 degrees: Ask the patient to turn 360 degrees "as quickly as you can, as you feel comfortable and safe". Evaluate using the scale on PPT scoring sheet. Additional data: count the number of steps required.

7. 50-foot walk test: Bring patient to start on a 50 foot walk test course (25 feet out and 25 feet back) and ask the patient, on the command "go" to walk as quickly as they can to the 25-foot mark and back. Time from the command "go" until the starting line is crossed on the way back.

8-9. Stairs: Take vital signs. Bring patient to foot of stairs (nine to 12 steps) and ask patient, on the command "go" to begin climbing up to a total of 4 flights stairs (as quickly as they can, as they feel comfortable and safe) or until they feel tired and wish to stop. Before beginning this task, alert the patient to the possibility of developing chest pain or shortness of breath and inform the patient to tell you if any of these symptoms occur. You will walk with the patient. Time from the command "go" until the patients' first foot reaches the top of the first flight of stairs. Go on to record the number of flights (maximum is four) completed (up and down is one flight). Provide a chair for resting when completed, so vital signs can be taken immediately post.

Table 4: Modified - Physical Performance Test – Scoring Method

No.	Task	Items Score Values
1.	Balance Tasks	**Standing Balance Feet Together** • ≥ 10 sec = 1 • 0 - <10 sec or Unable = 0
		Standing Balance Semi Tandem • ≥ 10 sec = 1 • 0 - <10 sec or Unable = 0
		Standing Balance Tandem • ≥ 10 sec = 2 • 3 – <10 sec = 1 • 0 - <3 sec or Unable = 0
2.	Chair rise	• >0 - 11 sec = 4 • >11 - 14 sec = 3 • >14 - 17 sec = 2 • >17 sec = 1 • 0, Unable = 0

No.	Task	Items Score Values
3.	Lift a book and put it on a shelf	• >0 - 2 sec = 4 • >2 - 4 sec = 3 • >4 - 6 sec = 2 • >6 sec = 1 • 0, Unable = 0
4.	Put on and remove a jacket (sum of time to put and to remove the jacket)	• >0 - 10 sec = 4 • >10 - 15 sec = 3 • >15 – 20 sec = 2 • >20 sec = 1 • 0, Unable = 0
5.	Pick up a penny from floor	• >0 - 2 sec = 4 • >2 - 4 sec = 3 • >4 - 6 sec = 2 • > 6 sec = 1 • 0, Unable = 0
6.	Turn 360 degrees	**Steps** • Continuous = 2 • Discontinuous = 0
		Steadiness • Steady = 2 • Unsteady (grabs, staggers) = 0
7.	50-foot walk test	• >0 - 15 sec = 4 • >15 - 20 sec = 3 • >20 - 25 sec = 2 • >25 sec = 1 • 0, Unable = 0

No.	Task	Items Score Values
8.	Climb one flight of stairs	• >0 - 5 sec = 4 • >5 - 10 sec = 3 • >10 – 15 sec = 2 • >15 sec = 1 • 0, Unable = 0
9.	Climb 4 flights of stairs (Number of flights climbed up and down)	• 4 flights = 4 • 3 flights = 3 • 2 flights = 2 • 1 flights = 1 • Unable = 0
	Total mPPT Score	**Sum of all Item score values [0-36]**

6.4.9.2. Analysis

The mPPT will be analyzed in the same way as described in Section 6.4.3.2 except that the baseline mPPT will be included in the model as the efficacy measure at baseline.

6.4.10. Huntington's Disease Quality of Life (HD-QoL) and 5-Level EuroQol Five Dimensions Questionnaire (EQ-5D-5L)

6.4.10.1. HD-QoL Definition

The HD-QoL questionnaire includes 40 items (ranged 0='Never' to 6='All the Time') and is collected at baseline and Week 52.

The HD-QoL total score is defined as the sum of the 40 items.

The HD-QoL total score range is from 0 to 240.

6.4.10.2. EQ-5D-5L Definition

The EQ-5D-5L consists of the EQ-5D descriptive system and the EQ visual analogue scale. The EQ-5D descriptive system includes 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The values for each of the dimensions are: 1=no problems, 2=slight problems, 3=moderate problems, 4=severe problems and 5=unable to/extreme problems.

The EQ visual analogue scale records the respondent's self-rated health on a vertical, visual analogue scale where the endpoints are labeled '100=Best imaginable health state' and '0=Worst imaginable health state'.

6.4.10.3. Analysis

The change from baseline in HD-QoL and in EQ-5D-5L domains at endpoint, will be analyzed using an Analysis of Covariance (ANCOVA) Model. The model will include the following fixed effects: treatment, country and baseline HD-QoL or EQ-5D-5L domains scores.

The least square mean and standard error of the least square mean for each treatment group, and the 95% CI for the comparisons (laquinimod dose vs placebo) will be presented.

The SAS code for the ANCOVA model is as follows:

```
Ods Output LSMeans=LSM Diffs=Diff;

Proc Mixed;

Class Arm Country;

Model Change=Arm Country Baseline;

LSMeans Arm /PDiff CL;

Where Arm in ('Placebo', 'laquinimod 0.5mg' , 'laquinimod 1mg');

Run;
```

6.4.11. Work Limitations Questionnaire (WLQ)

HD imposes a substantial burden on patients in the form of impaired ability to perform productive activity such as paid employment, domestic household activity, or schooling. The WLQ captures a multidimensional look at work productivity. It is designed to measure productivity among workers who are employed, but may be performing at less than full capacity. It measures the degree to which health problems interfere with specific aspects of job performance and the productivity impact of these work limitations. The eight-item version (Lerner, et al. revised 2009, unpublished, email communication) will be used to reduce respondent burden, though it is potentially less precise than the full 25-item version (Lerner, et al, 2001).

This version captures the most important subset of the original 25 questions, choosing two items from each of the four full scale dimensions:

- Time Management Scale addresses the difficulty of performing a job easily at the beginning of the workday and starting the job soon after arriving at work (Question 1 – a and b).
- The Physical Tasks Scale covers a person's ability to perform job tasks that involve sitting and standing in one position and performing the same motion repeatedly (Question 2 – a and b).
- The Mental-Interpersonal Tasks Scale assesses the difficulty concentrating on work and a person's ability to interact with people on-the-job (Questions 3 and 4).
- The Output Tasks Scale concerns the person's ability to complete work (Question 5 – a and b).

The responses to each of the WLQ's 8 items describe the amount of time in the prior 2 weeks an employee was limited with regard to performing a specific type of job task. The values of the response for each WLQ item represent, by design, the item labels (to lessen the responder bias) rather than the severity score itself. A severity score for each WLQ item will be defined such that the highest score, 5, corresponds to the highest degree of work limitation, whereas the lowest score, 1, corresponds to no limitation in performing the job tasks.

Thus, for the Time Management Scale, Mental-Interpersonal Tasks Scale and the Output Tasks Scale, the severity score will be defined such that the highest severity score, 5, will correspond to "Difficult all of the time (100%)", while the lowest severity score, 1, will correspond to "Difficult none of the time (0%)".

For the Physical Tasks Scale, the response of "Able all of the time (100%)", will correspond to the severity score of 1, while the response of "Able none of the time (0%)" will correspond to the severity score of 5.

The "Does not apply to my job" category will be set as a missing severity score (Lerner et al, 2009)

For each of the four scales, the average item score will be calculated, after imputing the missing values as specified in Section 4.3, where possible. The average item score is then converted to the scale score by the following formula: WLQ scale score=25*(average item score -1).

The work productivity will be assessed in all employed patients, regardless of their disease severity.

The WLQ could be completed by the patients with caregiver assistance if needed.

The WLQ questionnaire is collected at baseline and Week 52. The endpoint, change from baseline in the WLQ scale, will be analysed as described in Section 6.4.10 for each WLQ scale, except that baseline WLQ scale will be used instead of baseline HD QoL.

6.4.12. CDR-SB

6.4.12.1. Definition

The Clinical Dementia Rating – Sum of Boxes (CDR-SB) is a widely used scale that has demonstrated validity and reliability in the longitudinal assessment of patients with cognitive and functional deficits that do not rise to the level of a diagnosis of overt dementia.

The utilization of CDR-SB scores for staging dementia severity offers several advantages over the global score because the optimal characteristics of both scores can be combined into a single score. First, CDR-SB scores are much simpler to calculate than the global score and they do not require an algorithm for computation, which will ultimately result in fewer calculation errors for those not using the online system. Second, CDR-SB scores can be treated as interval data in statistical analyses, whereas global CDR scores are ordinal by the nature of the algorithm approach to condensing the data. Finally, the most significant advantage to using CDR-SB scores for staging of dementia severity is the increased precision afforded for tracking changes across time.

The CDR is obtained through semistructured interviews of patients and informants, and cognitive functioning is rated in 6 domains of functioning: memory, orientation, judgment and

problem solving, community affairs, home and hobbies, and personal care. Each domain is rated on a 5-point scale of functioning as follows: 0, no impairment; 0.5, questionable impairment; 1, mild impairment; 2, moderate impairment; and 3, severe impairment (personal care is scored on a 4-point scale without a 0.5 rating available). The global CDR score is computed via an algorithm. The CDR-SB score is obtained by summing each of the domain box scores, with scores ranging from 0 to 18.

Where possible, the same person should act as a patient's caregiver/informant throughout the study. If this is not possible, a patient should have no more than 2 caregivers throughout the study. All possible attempts should be made to assure that caregiver/informant will attend the clinical visits in person together with the patient. If the caregiver/informant is not available to attend the clinic visit, the interview can be done over phone.

6.4.12.2. Analysis

The CDR-SB is assessed at baseline and Week 52. The change from baseline in CDR-SB will be analysed as described in Section 6.4.10, except that baseline CDR-SB will be used instead of baseline HD QoL.

6.4.13. Hospital Anxiety and Depression Scale (HADS)

6.4.13.1. Definition

The Hospital Anxiety and Depression Scale, HADS (Zigmond and Snaith, 1983), is a 14 item (each item scored 0-3), self-administered rating scale that consists of two subscales assessing the presence and severity of depression (0–21) and anxiety (0–21), with a global score of 0–42. It was designed to diminish the influence of somatic symptoms and consequently does not include items relating to the physical symptoms of depression and has a medium overall cognitive complexity.

6.4.13.2. Analysis

The HADS scale is assessed at baseline and Week 52. The endpoint, change from baseline, will be analysed as described in Section 6.4.10, except that baseline HADS will be used instead of baseline HD QoL.

6.4.14. Problem Behaviors Assessment-Short Form (PBA-s)

6.4.14.1. Definition

The PBA-s is a brief semi-structured interview covering the most common behavioral and psychiatric manifestations of HD. The interview is not restricted to a single construct, but rather covers several broad symptom domains relevant to HD, comprising 11 items: low mood, suicidal ideation, anxiety, irritability, anger/aggressive behavior, loss of motivation, perseverative thinking or behavior, obsessive-compulsive behaviors, paranoid thinking, hallucinations, behavior suggestive of disorientation. Each symptom is rated for severity on a 5-point scale according to detailed scoring criteria which roughly correspond to the following: 0 = "not at all"; 1 = trivial; 2 = mild; 3 = moderate (disrupting everyday activities) and 4 = severe or intolerable. Each symptom is also scored for frequency on a 5-point scale as follows: 0 = symptom absent;

1 = less than once weekly; 2 = at least once a week; 3 = most days (up to and including some part of every day); and 4 = all day, every day.

Severity and frequency scores are multiplied (after setting all values outside the range of 0-4 to missing) to produce an overall 'PBA-s score' for each symptom domain.

The total PBA-s score is calculated by the sum of all PBA-s scores across symptom domains.

6.4.14.2. Analysis

The PBA-s assessments are collected at baseline and Week 52. The endpoint, change from baseline, will be analysed as described in Section 6.4.10.3, except that baseline PBA-s will be used instead of baseline HD-QoL.

7. MULTIPLE COMPARISONS AND MULTIPLICITY

In order to to maintain the experiment-wise type I error rate of 5% , the Fallback method with the loop-back feature will be used to test the primary and secondary endpoints. The methodology has been recommended in the recent Guidance for Industry on the Multiple Endpoints in Clinical Trials[1] to increase the study power.

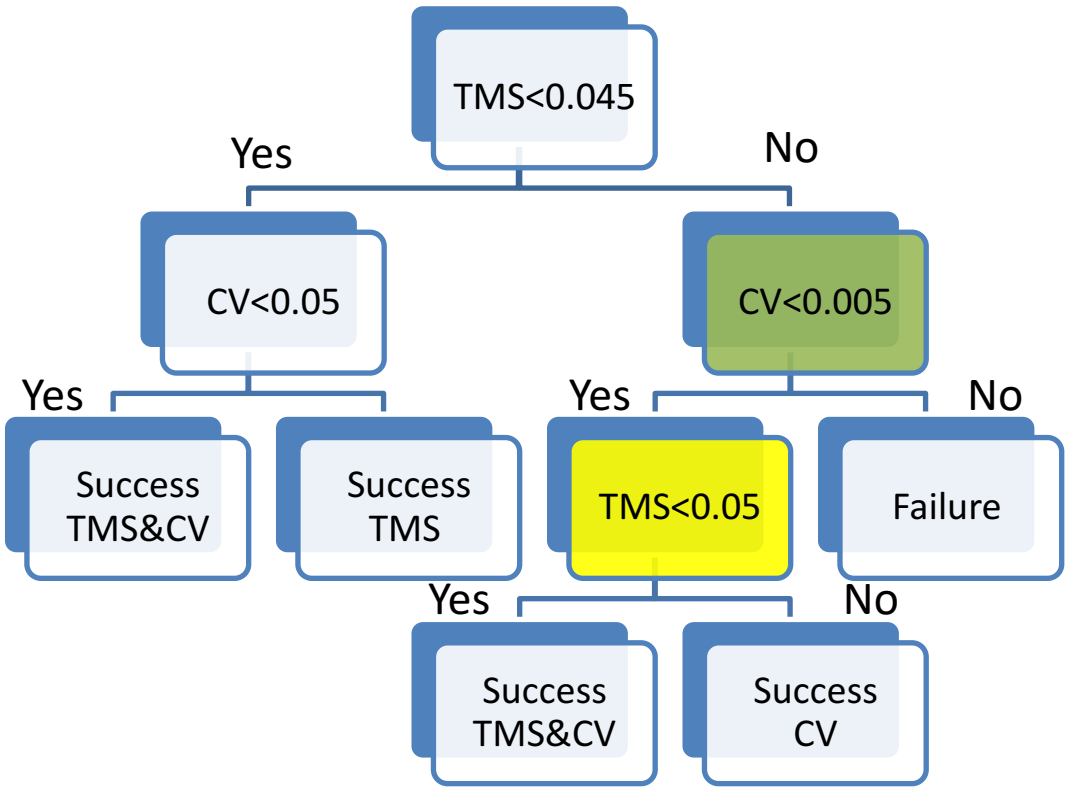
The fallback method permits interpretation of an endpoint with a robust treatment effect using modest amount of alpha retained as a fallback, even if the preceding endpoint is unsuccessful, without inflating the Type I error rate. In this method, the alpha is split between the endpoints of interest using weights reflecting the endpoint clinical importance; sum of weights always equals 1. In our setting, we suggest using weights of 0.9 and 0.1 for the TMS and caudate volume (CV) respectively, yielding alpha of 0.9*0.05=0.045 for TMS, and alpha of 0.1*0.05=0.005 for CV. The hypothesis testing starts with the TMS tested at an alpha of 0.045 and if successful, the CV hypothesis will be tested at the alpha level of 0.05 ("full alpha"). In case the TMS is not successful, the CV hypothesis has the chance of being successful when tested at the alpha of 0.005.

The loop-back feature has been proposed to increase the power of the fallback method giving a "second chance" to an endpoint that was not statistically significant at the initially assigned endpoint-specific alpha, by receiving pass-along or "looped back" alpha from a different endpoint that was successful. Thus, in our previously discussed setting, if the TMS hypothesis fails at the alpha level of 0.045, and the CV hypothesis is successful at the alpha of 0.005, the TMS hypothesis can be retested at the alpha level of 0.05 ("full alpha"). See Figure 2 for the graphical depiction of the hypotheses testing, where the green and the yellow boxes represent the fallback and the "looped back" alpha paths.

This procedure can be viewed as a modification of the Holm procedure for multiple testing, where instead of splitting the alpha equally between the tested hypotheses, it is divided using the prespecified weights.

[1] US Department of Health and Human Services, Center for Drug Evaluation and Research (CDER), Food and Drug Administration. Draft guidance: Multiple endpoints in clinical trials: Guidance for Industry. January 2017.

Figure 2: Fallback with Loopback – Hypotheses Test Flow



Abbreviations: TMS = Total Motor Score; CV = Caudate Volume.
Note:Traditional gate keeping on the left side of the chart, with the the green and the yellow boxes representing the fallback and the "looped back" alpha paths. The TMS endpoint is assigned $\alpha=0.045$ ($w_1=0.9$) and the CV is assigned $\alpha=0.005$ ($w_2=0.1$).

8. SAFETY ANALYSIS

8.1. General

The safety population will be used for all safety analyses, unless otherwise noted. Summaries will be presented by treatment group, unless otherwise stated.

8.2. Duration of Exposure to Study Drug

Duration of treatment (days treated) is the number of days on treatment based on the first and last days of treatment with the study drug (last day of study drug – first day of study drug + 1). Weeks on treatment using the categories ≤2 week, >2 to ≤4 weeks, >4 to ≤6 weeks, >6 to ≤8 weeks, >8 to ≤12 weeks, >12 to ≤16 weeks, >16 to ≤20 weeks, >20 to ≤26 weeks, >26 to ≤39 weeks, >39 to ≤52 weeks or more than 52 weeks will be summarized using descriptive statistics. Duration of treatment (days) will also be summarized using descriptive statistics.

In case of missing end of treatment date, last study drug return date or last date of study drug administration on the Study Drug Administration exposure forms will be used per study team decision.

8.3. Duration of Exposure to Study

Duration of study (days) is the number of days on study based on the difference between the randomization date and study completion date (study completion date – randomization date + 1). Weeks on study using the categories ≤2 week, >2 to ≤4 weeks, >4 to ≤6 weeks, >6 to ≤8 weeks, >8 to ≤12 weeks, >12 to ≤16 weeks, >16 to ≤20 weeks, >20 to ≤26 weeks, >26 to ≤39 weeks, >39 to ≤52 weeks or more than 52 weeks will be summarized using descriptive statistics. Duration of study (days) will also be summarized using descriptive statistics.

In case of missing end of study date or a lost to follow up subject, last date of subject assessments will be used.

8.4. Adverse Events

All adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Each patient will be counted only once in each preferred term or system organ class (SOC) category for the analyses of safety. Summaries will be presented for all treatment emergent adverse events (overall and by severity), adverse events determined by the investigator to be related to study treatment (ie, reasonable possibility) (defined as related or with missing relationship) (overall and by severity), serious adverse events, and adverse events causing withdrawal from the treatment. Summaries will be presented by treatment group.

Multiple records with the same Preferred Term (PT) and AE onset date for the same patient, or with overlapping or consecutive dates, are counted only once selecting the AE with the highest severity and seriousness. If onset date of AEs with the same PT is partially unknown or duration is < 24 hours, then these AEs will be counted as separate AEs, except for the cases where an AE with duration of < 24 hours has the same onset date as another AE with longer duration, then the longer duration AE will be counted.The event rate per 100 years (PY) of patient treatment exposure is calculated as 100*(Number of cases/PY).

Patient listings of all adverse events, serious adverse events, adverse events leading to withdrawal and adverse events leading to death will be presented.

8.5. Deaths

If any patient dies during the study, a listing of deaths will be provided and all relevant information will be discussed in the patient narrative included in the clinical study report.

8.6. Clinical Laboratory Tests

Summary statistics for chemistry and hematology laboratory tests will be presented at baseline, Weeks 4, 13, 26, 39, 52, 56, and last assesment. Laboratory values and changes from baseline to each visit endpoint will be summarized using descriptive statistics.

Shifts (below, within, and above the normal range) from baseline to each visit and endpoint will be summarized using patient counts for chemistry and hematology laboratories.

Summaries of potentially clinically significant abnormal values will include all postbaseline values (including scheduled, unscheduled, and withdrawal visits). The incidence of potentially clinically significant abnormal values will be summarized for laboratory variables using descriptive statistics with the criteria specified in Table 5.

Table 5: Criteria for Potentially Clinically Significant Laboratory Values

Test		Criterion value
Serum chemistry		
Alanine aminotransferase (ALT)		≥3x ULN
Aspartate aminotransferase (AST)		≥3x ULN
Alkaline phosphatase		≥3x ULN
Gamma-glutamyl transpeptidase (GGT)		≥3x ULN
Lactate dehydrogenase (LDH)		≥3x ULN
Blood urea nitrogen (BUN)		≥10.71 mmol/L
Creatinine		≥177 µmol/L
Uric acid	Men	≥625 µmol/L
	Women	≥506 µmol/L
Bilirubin (total)		≥34.2 µmol/L
Hematology		
Hematocrit	Men	<0.37 L/L
	Women	<0.32 L/L
Hemoglobin	Men	≤115 g/L
	Women	≤95 g/L
White blood cell (WBC) counts		≤3 x 10^9/L ≥20 x 10^9/L

Test	Criterion value
Eosinophils	≥10%
Absolute neutrophil counts (ANC)	≤1 x 10^9/L
Platelet counts	≤75 x 10^9/L ≥700 x 10^9/L

In addition, Common Terminology Criteria for Adverse Events grade values for the hemoglobin will be summarized using patient counts of the worst post baseline shift, using the grades shown in Table 6.

Table 6: CTCAE v4.0 Severity Criteria for Hemoglobin

Hematology	Grade 1	Grade 2	Grade 3
Hemoglobin	10.0 g/dL-<LLN	8.0-<10.0 g/dL	<8.0 g/dL

Note: LLN refers to lower limit of normal as provided by the laboratory.

Shifts of liver enzymes (ALT, AST, GGT) from normal levels at baseline to the following abnormal levels any time post-baseline (including unscheduled visits and withdrawal visits) will be presented: >1 and ≤ 3x upper limit of normal range (ULN), >3 and ≤ 5x ULN, >5 and ≤ 8x ULN and > 8x ULN. Similarly, shifts of bilirubin from normal levels at baseline to the following abnormal level>=2xULN will be presented. Patients will be categorized for their highest post-baseline shift at any time during study.

All the laboratory results will be presented in listings.

8.6.1. Anemia Panel

The anemia panel is assessed at baseline and also at one subsequent time point (with B12) if hemoglobin decrease of >1 g/dL from the patient's hemoglobin level at baseline and the decrease is confirmed.

At baseline the following measurements will be taken: B12, blood smear, serum iron, ferritin, total iron binding capacity, folic acid, haptoglobin, IL-1, IL-6, interferon IFN-γ, TNF-α, and hepcidin. They will be summarized using descriptive statistics, and also presented in the listing.

In case of hemoglobin decrease of >1g/dL from the patient's hemoglobin level at baseline, subject will be re-tested to ascertain true decrease, and if confirmed, the actions taken are detailed in the corresponding section in the protocol. Among others, the same anemia panel measurements as at baseline will be performed and will be presented in the listing.

8.6.2. Estimated Creatinine Clearance Calculation

Estimated Creatinine Clearance (CrCl) will be calculated at all visits to monitor renal function in the study in order to identify patients with a potential renal impairment. The Cockcroft-Gault equation is used for the calculation, and the reported units of measure are mL/min.

Each CrCl result is checked whether <60 mL/min. If yes, CrCl assessment should be repeated and the patient should stop study medication. If the renal impairment is confirmed (estimated CrCl <60 mL/min/1.73 m^2), the patient should stop study medication permanently.

The estimated CrCl values and changes from baseline will be presented by visit using descriptive statistics. The individual CrCl measurements will also be presented in the listing.

There will be a separate listing with all the CrCl values for subjects that have at least 1 CrCl result below 60 mL/min.

8.7. Physical Examinations

Physical examinations, including height (to be measured at the screening visit only) and weight, will be performed at weeks 4, 13, 26, 39, 52 and 56. Any physical examination finding that is judged by the investigator as a clinically significant change (worsening) compared with a baseline value will be considered an adverse event, recorded on the CRF, and monitored as described in Section 7.1.2 of the study protocol.

8.8. Vital Signs

Summary statistics for vital signs blood pressure, pulse, temperature and weight will be presented at weeks 4, 13, 26, 39, 52, 56, and at last assesment. Vital signs values and changes from baseline to each visit will be summarized using descriptive statistics.

Summaries of potentially clinically significant abnormal values will include all postbaseline values (including scheduled, unscheduled, and withdrawal visits). The incidence of potentially clinically significant abnormal values will be summarized using descriptive statistics with the criteria specified in Table 7.

Table 7 specifies the criteria for identifying vital signs as potentially clinically significant abnormal values.

Table 7: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign	Criterion value
Pulse	≥120 bpm
	≤50 bpm
Systolic blood pressure	≥180 mm Hg
	≤90 mm Hg
Diastolic blood pressure	≥105 mm Hg
	≤50 mm Hg
Body temperature	≥38.3°C

Only supine measurements will be presented in the summaries. All the data will be listed.

8.9. Electrocardiography

Shifts (normal and abnormal) from baseline to overall result interpretation and each visit and endpoint, meaning at weeks 4, 13, 26, 39, 52, 56, and last assessment, will be summarized using patient counts. For overall result interpretation the worst postbaseline finding for the patient (the abnormal finding if there are both normal and abnormal findings) will be used in the summaries. Summary statistics for ECG variables values will be presented. Actual values and changes from

baseline at weeks 4, 13, 26, 39, 52, 56, and last assesment will be summarized using descriptive statistics.

In general, 3 ECG assessments are performed per baseline visit, and their average will be calculated. If the number of measurement is different, up to three last measurements during the last visit prior to first dose administration will be used both for the average calculation and for choosing the worst result for the interpretation.

The incidence of potentially clinically significant abnormal values for ECG variables will be summarized using descriptive statistics with the criteria specified below.

- QTcF values >450 ms or >480 ms or >500 ms.
- QTcF change from baseline values >30 or >60.
- PR change from baseline ≥25% and value >200.
- QRS change from baseline ≥25% and value >110.
- Heart rate value <60 bpm or >100 bpm.

8.10. Concomitant Medications or Therapies

Concomitant therapies and medications, including medications that are taken on an as needed basis and occasional therapies, will be monitored during the study. Details of prohibited medications may be found in Section 5.3 of the study protocol. All concomitant medications will be coded using the WHO Drug.

The incidence of concomitant therapies and medications will be summarized using descriptive statistics by therapeutic class category and preferred term. Patients are counted only once in each therapeutic class, and only once in each preferred term category. Concomitant therapies and medications will include all medications up to the end of study as defined in the study protocol.

8.11. Columbia-Suicide Severity Rating Scale (C-SSRS)

The Columbia-Suicide Severity Rating Scale Screening/Baseline Version (C-SSRS-BL) is assessed at screening. The Columbia-Suicide Severity Rating Scale Since Last Visit (C-SSRS-SLV) is assessed at baseline, weeks 4, 13, 26, 39, 52, 52, and 56.

Any positive answer to the behavior subcomponents at screening or baseline identifies a patient as with "Suicidal Behavior at Baseline". Similarly, any positive answer to the ideation subcomponents at any of these two visits identified a patient as with "Suicidal Ideation at Baseline". A patient identified with either "Suicidal Behavior at Baseline" or with "Suicidal Ideation at Baseline" is also classified as with "Suicidal Behavior or Ideation at Baseline".

Similarly, any positive answer to the behavior subcomponents in any of the post-dosing visits identifies a patient as with "Suicidal Behavior Post Dosing". Also, any positive answer to the ideation subcomponents in any of the post-randomization visits identified a patient as with "Suicidal Ideation Post Dosing". A patient identified with either "Suicidal Behavior Post Dosing" or with "Suicidal Ideation Post Dosing" is also classified as with "Suicidal Behavior or Ideation Post Dosing"

Frequency counts and percentages of the C-SSRS outcomes: Suicidal Behavior at Baseline, Suicidal Ideation at Baseline, Suicidal Behavior or Ideation at Baseline, Suicidal Behavior Post

Dosing, Suicidal Ideation Post Dosing, Suicidal Behavior or Ideation Post Dosing, and shifts from baseline will be summarized.

9. PHARMACOKINETIC ANALYSIS

A single blood sample will be collected from all patients at Months 1, 3, 6 and 12 for evaluation of laquinimod, see Table 1.

Individual plasma concentrations of laquinimod will be listed by dose and visit. Descriptive statistics summary tables for plasma concentrations by dose and visit will be presented. The pharmacokinetic analysis set (PAS) will be used for the descriptive statistics summary.

Pharmacokinetics of laquinimod and its metabolites, and potentially the effect of various covariates (e.g., demographics and clinical parameters) on laquinimod's pharmacokinetics, may be evaluated in this study using a population pharmacokinetics approach. This analysis will be reported in a separate population pharmacokinetic report.

10. PHARMACODYNAMIC ANALYSIS

The pharmacodynamic analysis will be performed as described in the study protocol and reported separately, if applicable.

11. PHARMACOKINETIC/PHARMACODYNAMIC ANALYSIS

The exploratory pharmacokinetic/pharmacodynamic analysis described in the study protocol is beyond the scope of this SAP and may be reported separately, if applicable.

12. BIOMARKER ANALYSIS

The biomarker analysis will be performed as described in the study protocol and will be reported separately if applicable.

Any collected data will be presented descriptively only and listed as part of the laboratory data.

13. ANCILLARY STUDIES ANALYSIS

The ancillary study objectives (substudies) (see Section 2.2.5. in the protocol) are beyond the scope of this SAP and might be reported separately if applicable.

14. ANCILLARY PK ANALYSIS

PK samples of laquinimod and its metabolites will be collected from approximately 4 patients per each of the 3 continuing treatment groups (at selected sites at Month 1), for a total of approximately 12 patients. Additionally, PK samples were collected from 2 patients in the laquinimod 1.5 mg/day treatment group when the treatment was stopped; no further PK samples will be collected from the laquinimod 1.5 mg/day treatment group.

Samples will be collected at pre-dose, 15, 30 min and 1, 2, 3, 6 and 24 hours post dose. Individual concentrations by treatment and timepoint will be listed.

15. STATISTICAL SOFTWARE

All data listings, summaries, and statistical analyses will be generated using SAS® version 9.3 or later.

16. CHANGES TO ANALYSES SPECIFIED IN THE STUDY PROTOCOL

16.1. Definition of Primary, Secondary and Exploratory Endpoints

The SAP defines that the primary objective of this study is to assess the efficacy of laquinimod 1.0 mg in patients with HD using the UHDRS-TMS after 12 months of treatment. Previously, in the protocol, the treatment groups were not distinguished with regard to analysis of the primary or secondary endpoints.

The SAP defines the secondary objective of this study as determining the effect of laquinimod 1.0 mg in patients with HD after 12 months of treatment on brain atrophy using MRI measures of caudate volume.

In the protocol, several additional secondary endopoints were defined to evaluate the effect of laquinimod in patients with HD after 12 months of treatment, as follows:

- Functional capacity using the TFC scale;
- Change from baseline in HD-CAB total score (sum of the standardized sub-components) at Month 12/ET (evaluated at baseline and Months 6 and 12)
- Clinical global impression using the CIBIC-Plus.

As per the SAP, these endpoints are now considered exploratory.

Note that the HD-CAB total score has been revised to HD-CAB composite score (Stout et al 2014), calculated as the average of the Z scores across the 6 HD-CAB tests. The reason for the change is that the HD-CAB composite score, rather than HD-CAB total score, is described in the literature and is also used as an exploratory endpoint in other HD studies. The change is rather technical in nature, being just a division of the sum by the number of the tests.

In addition, in the definition of the HD-CAB component "Emotion Recognition", the wording about the number of items was changed from "10 stimuli per emotion" to "the total of 36". The change reflects the modification of the original test (Stout et al 2014), as well as the data collected in the study. It was also clarified that only identification of negative emotions (out of 24 possible) was used in the composite score.

In addition, exploratory efficacy endpoints now also include the analyses of the primary and secondary endpoints on the 0.5 mg laquinimod arm vs. placebo arm.

The following exploratory study objective is outside of the scope of the SAP:

- To investigate the relationship between exposure to laquinimod and its metabolites and outcome measures (eg, clinical effect and toxicity parameters).

The following exploratory study objective has been changed as follows:

- Protocol: To evaluate the pharmacokinetics of laquinimod and its metabolites in patients with HD
- SAP: To potentially evaluate the plasma concentrations of laquinimod in patients with HD

16.2. Patient Disposition of Laquinimod 1.5 mg/Day Treatment Group

Patients in the high dose group (1.5mg/day) who were requested to discontinue study drug before week 52, but continued to attend scheduled study visits for safety assessments, will be considered to have completed the study, contrary to the text in Protocol Section 3.11.3.1.5.

16.3. Type 1 Error Control

The Fallback method with the Loop-back feature (see Section 7) will be used to test the primary and secondary endpoints instead of the initially-described hierarchical method to control inflation in type I error rate.

16.4. Country Replaces Site in the Efficacy Analyses

The country, instead of site, will serve as a covariate in the primary efficacy analysis, as well as in the other statistical models that include site as a covariate. The change was done due to the small numbers of patients in each site.

16.5. No Application of Last Observation Carried Forward (LOCF) for Early Terminating Patients

The Last Observation Carried Forward (LOCF) method will not be applied for observations of patients who terminated the study early. Visit Windows for Early Termination and Unscheduled Visits will be implied instead (see Section 4.4 for reference).

16.6. Ancillary Studies Beyond the Scope of This SAP

The ancillary studies in the protocol are beyond the scope of this SAP and might be reported separately if applicable.

16.7. Estimated Creatinine Clearance Calculation Units

The estimated creatinine clearance (CrCl) threshold is 60 mL/min, and not 60 mL/min/1.73 m^2. The difference in the units reflects the method of estimation, and the values in mL/min can be compared with those in mL/min/1.73 m^2 without the need of a conversion factor. The units mL/min appear in the Protocol Section 4.2 in a description of the relevant patient exclusion criteria. In Protocol Section 7.7, CrCl <60 mL/min/1.73 m^2 is defined as the renal impairment criterion.

REFERENCES

Stout JC, Queller S, Baker KN, Cowlishaw S, Sampaio C, Fitzer-Attas C, Borowsky B, and the HD-CAB Investigators. HD-CAB: A Cognitive Assessment Battery for Clinical Trials in Huntington's Disease. Movement Disorders 2014; 29(10):1281-1288.

Walker T, Tullar J, Diamond P, Kohl H, Amick B. Validity and Reliability of the 8-Item Work Limitations Questionnaire. Journal of Occupational Rehabilitation. 2017; 27(4):576-583.

[redacted]

[redacted]

Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Medcial Care. 2001;39(1):72–85.