THE LANCET
Neurology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Furr Stimming E, Claassen DO, Kayson E, et al. Safety and efficacy of valbenazine for the treatment of chorea associated with Huntington's disease (KINECT-HD): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 2023; **22:** 494–505.

APPENDIX

PRESPECIFIED EXPLORATORY EFFICACY ENDPOINTS IN THE FULL ANALYSIS SET

The following were prespecified as exploratory efficacy endpoints for KINECT-HD, as listed in the Statistical Analysis Plan (SAP):

- Changes from the screening period baseline to each postbaseline study visit (Weeks 2 through 12) in the TMC based on site assessments
- Changes from the screening period baseline to maintenance (the average of the Week 10 and Week 12 assessments) in the TMC based on video recording central rater assessments
- CGI-C response statuses at Weeks 2 through 10
- PGI-C response statuses at Weeks 2 through 10
- The change from baseline to Weeks 4, 8, and 10 in the Neuro-QoL Upper Extremity Function T-score
- The change from baseline to Weeks 4, 8, and 10 in the Neuro-QoL Lower Extremity Function T-score
- Clinical Global Impression of Severity (CGI-S) at Weeks 2 through 12
- Patient Global Impression of Severity (PGI-S) at Weeks 2 through 12
- Short Form 36 Health Survey (SF-36) at Week 12
- Huntington Disease Health Index (HD-HI) at Week 10 and Week 12
- EuroQol 5 Dimensions 5 Levels (EQ-5D-5L) at Week 10 and Week 12
- UHDRS scores for motor, behavior, and functional assessment at Weeks 2 through 12
- Other UHDRS scores including TFC and independence scale at Week 12
- Anosognosia Scale (AS) at Week 12

Results for these endpoints are presented below, with brief descriptions of the analytical methods.

UHDRS® Total Maximal Chorea (TMC) Score: Least Squares Mean Changes from Screening and Baseline to Postbaseline Visits (Weeks 2 to 12)

TMC changes over time showed consistently greater chorea improvement with valbenazine than with placebo at all study visits, including the visit at week 2, when participants had only been taking the initial 40 mg dose (see Figure 2B in main article). Least squares mean differences between valbenazine and placebo increased from week 2 (-1.3 [95% CI: -2.2, -0.3]) to week 12 (-3.2 [95% CI: -4.5, -1.9]).

This exploratory endpoint was analysed using the same approach used for the primary endpoint: mixed-effect model repeated measures; screening and baseline period TMC score as a covariate; treatment group, visit, treatment group-by-visit interaction, and baseline-by-visit interaction as fixed effects; and subject as a random effect.

Huntington's Disease Health Index: Mean Scores and Changes from Baseline

Greater decreases from baseline to week 10 and week 12 in HD-HI subscales were noted for valbenazine versus placebo, including mobility, abnormal movements, and hand and arm function. No worsening of overall disease burden was noted in subscales assessing impact, including fatigue, daytime sleepiness, gastrointestinal health (including swallowing function), emotional health, or communication. Higher scores indicate worse disease burden.

	Placebo (n=61)			Valbenazine (n=64)		
	Score at Baseline	**Change at Week 10**	**Change at Week 12**	**Score at Baseline**	**Change at Week 10**	**Change at Week 12**
Subscale scores, mean (SD)						
Mobility	20.7 (20.2)	-4.1 (15.5)	-2.7 (14.4)	24.5 (25.0)	-6.3 (18.3)	-4.2 (19.6)
Abnormal movements	28.5 (24.1)	-9.6 (18.0)	-4.3 (16.1)	27.5 (25.4)	-12.2 (22.9)	-11.2 (19.6)
Hand and arm function	20.1 (19.9)	-1.6 (16.2)	-1.4 (15.0)	24.8 (23.9)	-7.2 (20.9)	-7.2 (19.9)
Emotional health	21.2 (20.7)	-6.5 (15.1)	-2.7 (16.2)	16.6 (18.3)	-4.5 (16.4)	-5.9 (13.8)
Activity participation	16.0 (19.0)	-1.9 (14.3)	-2.1 (15.7)	17.6 (21.4)	1.1 (17.8)	-0.8 (14.4)
Social performance	17.9 (18.5)	-2.1 (16.9)	-0.5 (13.0)	18.0 (20.5)	-1.1 (20.5)	-2.5 (20.7)
Social satisfaction	17.4 (17.8)	-3.5 (13.4)	-0.9 (15.1)	14.3 (18.4)	-4.1 (21.0)	-2.8 (20.4)
Fatigue	17.8 (20.2)	-4.5 (16.9)	1.0 (17.8)	14.7 (20.9)	-0.9 (21.2)	0.1 (17.5)
Pain	9.4 (14.1)	2.7 (15.5)	6.2 (18.3)	9.9 (19.8)	-1.0 (15.5)	1.0 (15.2)
Cognition	26.2 (23.6)	-4.3 (14.7)	-1.8 (18.8)	23.1 (23.4)	-6.3 (22.2)	-3.9 (17.3)
Communication	23.4 (20.6)	-5.3 (15.0)	-4.8 (16.1)	23.7 (20.6)	-4.5 (18.5)	-4.5 (17.4)
GI health/swallowing function	14.9 (20.0)	-0.7 (11.0)	-0.7 (13.8)	12.8 (18.4)	-0.2 (17.4)	-2.0 (16.2)
Daytime sleepiness	16.2 (20.9)	-1.9 (13.4)	-1.3 (20.8)	16.0 (22.5)	1.9 (20.9)	-0.1 (16.6)
Total score, mean (SD)	19.4 (16.2)	-3.5 (10.3)	-1.3 (11.4)	18.7 (17.3)	-3.5 (15.2)	-3.4 (12.2)

Abbreviations: GI, gastrointestinal; SD, standard deviation.

Analysis: Descriptive statistics based on observed values.

Anosognosia Scale: Mean Scores and Changes from Baseline

At baseline and week 12, the mean differences between patient and clinician scores were <6 points, indicating that a majority of KINECT-HD participants did not have anosognosia (table). No clinically meaningful changes from baseline to week 12 in Anosognosia Scale scores (clinician- or patient-rated) were noted in either treatment group.

A post hoc analysis of the baseline data showed a score difference <6 points in 92 (74%) participants (valbenazine, 47 [73%]; placebo, 45 [74%]), indicating absence of anosognosia in these individuals.

	Placebo (n=61)	**Valbenazine (n=64)**
Mean score at baseline (SD)		
Clinician score	-3.5 (3.9)	-4.2 (4.4)
Patient score	-0.4 (7.3)	-1.3 (5.7)
Difference in score (patient minus clinician)[a]	3.1 (5.6)	2.8 (5.7)
Mean score at week 12 (SD)		
Clinician score	-2.8 (3.9)	-4.6 (3.4)
Patient score	0.4 (7.3)	0.2 (7.7)
Difference in score (patient minus clinician)[a]	3.2 (6.4)	4.8 (7.0)
Mean change from baseline to week 12 (SD)		
Clinician change	0.2 (2.1)	-0.7 (3.3)
Patient change	0.3 (5.9)	1.5 (6.1)
Difference in score change (patient minus clinician)	0.1 (6.0)	2.2 (6.3)

[a] Difference of 6 points or greater indicates presence of anosognosia.

Abbreviation: SD, standard deviation.

Analysis: Descriptive statistics based on observed values.

Additional Exploratory Endpoints

Descriptive statistics were used for all of the remaining exploratory endpoints. Results for these endpoints are presented on the next page.

	Direction of favourable effect	Week 2		Week 4		Week 6		Week 8		Week 10		Week 12	
		PBO	VBZ	PBO	VBZ	PBO	VBZ	PBO	VBZ	PBO	VBZ	PBO	VBZ
Mean change from baseline (SD)													
UHDRS® TMC, central video raters[a]	–											-0.9 (0.4)	-3.5 (0.4)
Neuro-QoL UEF T-score	+			-2.3 (6.5)	0.3 (6.4)			-2.9 (6.2)	1.0 (7.1)	-2.7 (6.4)	0.4 (7.2)	*Secondary endpoint*	
Neuro-QoL LEF T-score	+			1.6 (5.3)	1.6 (5.8)			1.1 (5.3)	2.1 (6.8)	1.7 (5.9)	1.4 (7.2)	*Secondary endpoint*	
SF-36 physical health component	+											-0.5 (5.6)	0.2 (6.9)
SF-36 mental health component	+											0.9 (8.1)	1.1 (8.6)
EQ-5D-5L health state index	+									0.002 (0.105)	0.016 (0.178)	-0.011 (0.143)	0.023 (0.141)
EQ-5D-5L visual analog scale	+									4.0 (12.0)	4.0 (15.4)	2.3 (10.8)	1.5 (13.1)
UHDRS® total motor score	–	-3.7 (5.6)	-6.0 (6.7)	-4.3 (5.7)	-8.8 (7.3)	-5.4 (6.7)	-8.7 (8.5)	-5.3 (6.4)	-9.4 (8.0)	-3.9 (6.0)	-8.1 (7.8)	-3.2 (6.4)	-7.5 (7.9)
UHDRS® total behavior frequency	–	-1.5 (3.3)	-0.5 (2.8)	-2.2 (4.1)	-1.3 (3.4)	-2.6 (5.1)	-1.2 (3.4)	-3.4 (5.0)	-1.3 (3.4)	-2.4 (4.0)	-1.5 (3.8)	-2.4 (4.7)	-1.3 (4.1)
UHDRS® total behavior frequency x severity	–	-2.9 (7.6)	-1.2 (6.6)	-4.5 (9.9)	-3.0 (6.8)	-5.3 (13.1)	-2.3 (7.3)	-6.4 (13.1)	-2.3 (7.9)	-4.3 (8.7)	-2.5 (10.2)	-4.4 (9.3)	-1.8 (11.1)
UHDRS® total functional capacity	+											0.1 (1.3)	-0.5 (1.1)
UHDRS® functional assessment	+											0.0 (1.5)	-0.2 (3.7)
UHDRS® independence scale	+											0.6 (5.5)	-0.1 (5.9)
Percentage of participants (n/N)													
CGI-C response status	+	2 (1/60)	9 (6/64)	3 (2/59)	24 (15/63)	14 (8/58)	34 (19/56)	11 (6/54)	46 (26/57)	17 (9/52)	42 (24/57)	*Secondary endpoint*	
PGI-C response status	+	5 (3/61)	23 (15/64)	10 (6/60)	29 (18/63)	16 (9/58)	39 (22/57)	17 (9/54)	43 (25/58)	27 (14/52)	44 (25/57)	*Secondary endpoint*	
CGI-S score ≤3	+	68 (41/60)	69 (44/64)	68 (40/59)	66 (41/62)	67 (39/58)	75 (42/56)	69 (37/54)	70 (40/57)	81 (42/52)	82 (47/57)	66 (35/53)	75 (42/56)
PGI-S score ≤2	+	59 (36/61)	66 (42/64)	67 (40/60)	66 (41/62)	66 (38/58)	74 (42/57)	69 (37/54)	69 (40/58)	73 (38/52)	79 (45/57)	66 (35/53)	71 (40/56)

[a] Least squares mean changes from screening and baseline (average of screening and baseline visits) to maintenance period (average of week 10 and week 12 visits) with standard error.

Blank cell indicates that the assessment was not collected and/or analysed at that timepoint.

Abbreviations: CGI-C, Clinical Global Impression of Change; CGI-S, Clinical Global Impression of Severity; EQ-5D-5L, EuroQol 5 Dimensions 5 Levels; LEF, Lower Extremity Function; Neuro-QoL, Quality of Life in Neurological Disorders; PBO, placebo; PGI-C, Patient Global Impression of Change; PGI-S, Patient Global Impression of Severity; SD, standard deviation; SF-36, Short Form 36 Health Survey; TMC, Total Maximal Chorea; UEF, Upper Extremity Function; UHDRS®, Unified Huntington's Disease Rating Scale; VBZ, valbenazine.

PRESPECIFIED SENSITIVITY ANALYSES IN THE FULL ANALYSIS SET

Three sensitivity analyses were prespecified for the primary efficacy endpoint using the following methods: a multiple imputation analysis assuming data were missing at random: a multiple imputation analysis assuming data were not missing at random; and a tipping point analysis.

Sensitivity analyses for the secondary CGI-C and PGI-C endpoints were conducted by imputing non-responder status to participants who had missing data at week 12. No sensitivity analyses for the secondary Neuro-QoL endpoints were prespecified in the SAP.

Results of the sensitivity analyses were consistent with results for the TMC primary endpoint and the secondary CGI-C and PGI-C response endpoints.

	Placebo (n=61)	**Valbenazine (n=64)**
For primary UHDRS® TMC endpoint		
Multiple imputation for MAR, LS mean change ± SEM (95% CI)[a]	-1.4 ± 0.4 (-2.3, -0.6)	-4.6 ± 0.4 (-5.4, -3.7)
Multiple imputation for MNAR, LS mean change ± SEM (95% CI)[a]	-1.4 ± 0.4 (-2.3, -0.6)	-4.4 ± 0.4 (-5.3, -3.6)
Tipping point analysis[b]	LS mean difference ± SEM (95% CI)	
Delta = 0	-3.0 ± 0.6 (-4.2, -1.8)	
Delta = 8	-1.6 ± 0.9 (-3.4, 0.2)	
For secondary response endpoints, % (n/N)		
CGI-C with non-responder imputation	12 (7/61)	38 (24/64)
PGI-C with non-responder imputation	23 (14/61)	45 (29/64)

[a] Least squares mean changes from screening and baseline (average of screening and baseline visits) to maintenance period (average of week 10 and week 12 visits).

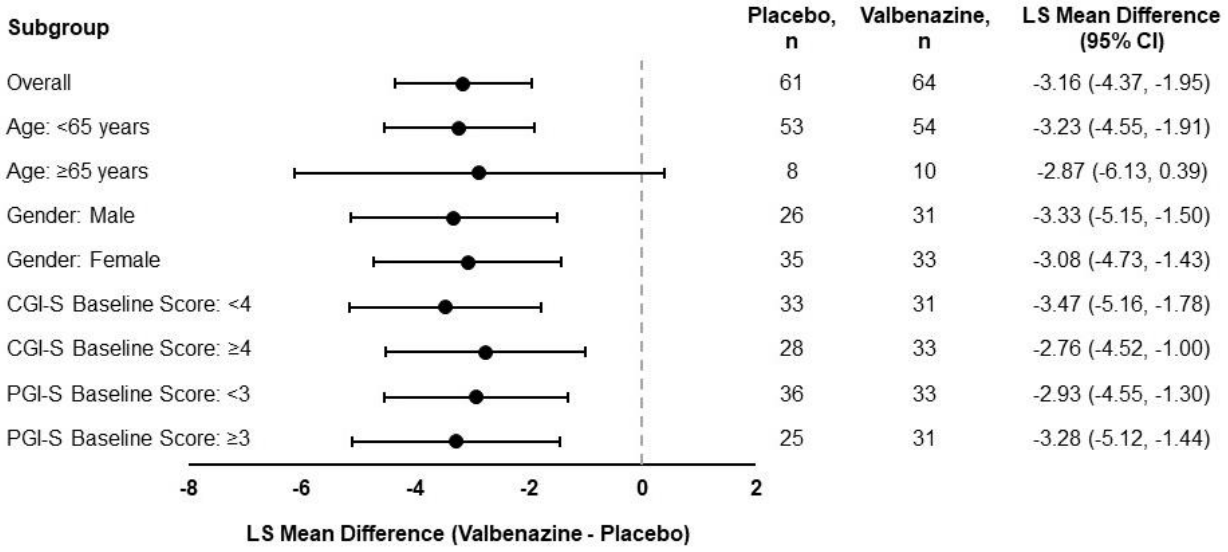
[b] An amount of delta was added to each imputed value in the valbenazine group using the missing imputation method for MAR. The tipping-point assumption was then applied by assuming the trajectories of participants in the valbenazine group were worse by an amount of delta after the first visit with missing data. Successively harsher deltas were imposed on the imputed values in the valbenazine group, starting with a TMC score increase (worsening) by 1. The delta was further increased in steps of 1 (ie, +1, +2, +3, …) until statistical significance was lost (ie, p≥0.05). For the placebo group, the missing imputation using MAR assumption was used. Results are presented as the LS mean difference between treatment groups (valbenazine – placebo) with 95% CI. Delta=8 represents the tipping point (i.e., the smallest delta that resulted in a p-value greater than 0.05).

Abbreviations: CGI-C, Clinical Global Impression of Change; CI, confidence interval; MAR, value missing at random; MNAR, value missing not at random; PGI-C, Patient Global Impression of Change; SEM, standard error of the mean; TMC, Total Maximal Chorea; UEF, Upper Extremity Function; UHDRS®, Unified Huntington's Disease Rating Scale.

PRESPECIFIED SUBGROUP ANALYSES IN THE FULL ANALYSIS SET

The following subgroups have been prespecified for analysis and were used to examine the consistency of effect for the primary efficacy endpoint: age (<65 versus ≥65 years), sex, race, baseline CGI-S categories (score <4 versus ≥4), and baseline PGI-S categories (score <3 versus ≥3).

No difference was observed in the primary efficacy endpoint when evaluated by sex or baseline CGI-S or PGI-S category. The small number of participants who were ≥65 years old precluded meaningful interpretation of analysis by age. Subgroup analysis by race was not performed due to small numbers of participants (15 or fewer) in race categories.



Abbreviations: CGI-S, Clinical Global Impression of Severity; CI, confidence interval; LS, least-squares; PGI-S, Patient Global Impression of Severity.

POST HOC ANALYSES OF WEEK 14 OUTCOMES IN THE FULL ANALYSIS SET

At week 14 (2 weeks after discontinuation of study drug), observed values for the primary TMC endpoint and the secondary Neuro-QoL endpoints returned towards baseline values. For the secondary CGI-C and PGI-C endpoints, which have no baseline value, week 14 results declined relative to week 12 results. Week 14 outcomes were similar between valbenazine and placebo.

	Placebo (n=61)	**Valbenazine (n=64)**
UHDRS® Total Maximal Chorea score, mean (SE)		
For screening and baseline period	12.1 (0.4)	12.2 (0.3)
At week 14	11.3 (0.5)	11.1 (0.5)
CGI-C response status, % (n/N)		
At week 14[a]	9 (5/54)	12 (7/57)
PGI-C response status, % (n/N)		
At week 14[b]	20 (11/54)	28 (16/57)
Neuro-QoL Upper Extremity Function T-score, mean (SE)		
At baseline	47.0 (1.1)	44.5 (1.1)
At week 14	43.7 (1.2)	43.4 (1.3)
Neuro-QoL Lower Extremity Function T-score, mean (SE)		
At baseline	48.4 (1.0)	48.1 (1.1)
At week 14	48.8 (1.1)	48.2 (1.3)

[a] CGI-C response status at week 12: placebo 13% (7/53); valbenazine 43% (24/56).

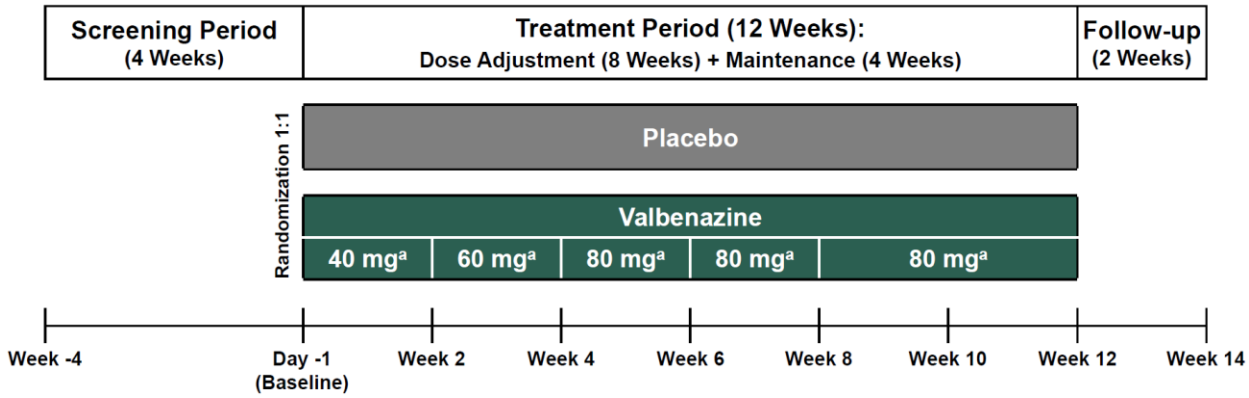
[b] PGI-C response status at week 12: placebo 26% (14/53); valbenazine 53% (29/55).

Abbreviation: CGI-C, Clinical Global Impression of Change; Neuro-QoL, Quality of Life in Neurological Disorders; PGI-C, Patient Global Impression of Change; SE, standard error; UHDRS®, United Huntington's Disease Rating Scale.

Analysis: Descriptive statistics based on observed values.

STUDY DESIGN

KINECT-HD was a randomised, double-blind, placebo-controlled, phase 3 study. It was designed to evaluate the efficacy, safety, and tolerability of once-daily valbenazine in individuals with chorea associated with Huntington disease (HD). The study included a 4-week screening period, 8-week dose-adjustment period, 4-week maintenance period, and a final study visit 2 weeks following the last dose of study drug.



[a] Doses represent maximum daily doses during each 2-week interval in the dose-adjustment period and during the 4-week maintenance period. Early termination could occur at any time; the early termination assessments were the same assessments as those performed at week 12.

STUDY ASSESSMENTS

Unified Huntington's Disease Rating Scale (UHDRS®)

The UHDRS® is a tool developed by the Huntington Study Group (HSG) to assess the clinical features and course of Huntington disease (HD). The UHDRS® has undergone extensive reliability and validity testing and has been used as a major outcome measure in controlled clinical trials (HSG 1996). The full UHDRS® assesses motor symptoms, functioning, cognition, behavior, and independence.

The motor portion consists of 15 items that measure the severity of motor symptoms, with the Total Motor Score (TMS) ranging from 0 to 124 and higher scores indicating more severe motor impairment. Total Maximal Chorea (TMC) is item 12 of the motor assessment and measures chorea in 7 different body regions (face, oral-buccal-lingual region, trunk, right and left upper extremities, right and left lower extremities). The TMC score, which ranges from 0 to 28, is the sum of individual scores from each body region, which are rated from 0 ("absent") to 4 ("marked/prolonged").

Total Functional Capacity (TFC) score ranges from 0 to 13, with higher scores indicating better functioning. In addition, the UHDRS® includes 25 yes/no questions for functional assessment, with the total score ranging from 0 to 25 and higher scores indicating better functioning.

The behavior portion is used to assess the frequency and severity of items such as disruptive/ aggressive behavior and irritable behavior. The total behavior frequency score ranges from 0 to 44, and the total behavior frequency-times-severity score ranges from 0 to 176. Higher scores represent more severe manifestations.

The independence scale is used to indicate the most accurate current level of independence, with a range of 0 to 100 and higher scores indicating better functioning.

A subset of the total motor assessment (items rating retropulsion pull test, finger taps, pronate/supinate hands, rigidity-arms, and bradykinesia-body) was used to assess parkinsonism (safety analysis).

Blinded Central UHDRS® Motor Video Raters

The motor portion of the UHDRS® was video recorded at screening, day -1 (baseline), week 10, and week 12 (or early termination), following a standardized video protocol. The motor and cognition portions of the UHDRS® were not administered at COVID-19–related remote visits. Video recordings were uploaded to a secure, central server and managed by a core laboratory. Access to this dedicated central server was limited and required a user identification and password.

The UHDRS® video recordings were reviewed and scored by blinded, central UHDRS® motor video raters. A triple-blind consensus scoring was conducted by these raters according to scoring guidelines developed by the study sponsor (Neurocrine Biosciences). The sponsor provided the blinded central video raters with digital-secure access to the participants' randomised video files for review and scoring.

The central video raters scored maximal chorea (range, 0 to 4) for each of 7 body regions (face, buccal-oral-lingual, trunk, right and left upper extremities, right and left lower extremities). These raters were blinded to study participants' visits and treatment assignments. Two blinded, central video raters reviewed each video file from beginning to end and had to agree on the TMC score.

Clinical Global Impression of Change (CGI-C)

The CGI-C, which is based on a 7-point scale (range: 1=very much improved to 7=very much worse), was used to rate the overall global improvement of chorea since the initiation of study drug dosing. This scale is a modification of a scale developed by the Psychopharmacology Research Branch of the National Institute of Mental Health to rate a patient's overall improvement in clinical disorder and provides a global evaluation of improvement over time from the clinician's perspective (Guy 1976).

Patient Global Impression of Change (PGI-C)

For the PGI-C, study participants rated the change in their chorea symptoms since initiation of study drug dosing using the same 7-point scale as the CGI-C (range: 1=very much improved to 7=very much worse) (Guy 1976).

Clinical Global Impression of Severity (CGI-S)

The CGI-S was used by clinical investigators to assess overall severity of chorea, using a 7-point scale (range: 1=normal, not at all ill to 7=among the most extremely ill patient) (Guy 1976).

Patient Global Impression of Severity (PGI-S)

The Patient Global Impression of Severity (PGI-S) scale was used by study participants to assess their overall severity of chorea on a 5-point scale (range: 1=none to 5=very severe) (Guy 1976).

Quality of Life in Neurological Disorders (Neuro-QoL)

The Neuro-QoL is a collection of psychometrically sound, clinically relevant, health-related quality of life measurement tools for individuals with neurological conditions. The Neuro-QoL has been demonstrated to be a reliable tool for assessing patient-reported physical functioning measures in patients with HD (Carlozzi 2017). The Lower Extremity Function Short Form and the Upper Extremity Function Short Form each comprise 8 questions about physical abilities, rated from 1 (unable to do) to 5 (without any difficulty).

Short Form 36 Health Survey (SF-36)

The SF-36 is a 36-item, self-administered questionnaire. It measures health on 8 dimensions: vitality, physical functioning, pain, general health perception, physical role limitations, emotional role functioning, social functioning, and mental health (Brazier 1992). Higher scores on all subscales represent better health and functioning. Two component scores are derived from the eight subscales: a physical health component score and a mental health component score.

EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L)

The EQ-5D-5L includes 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each of which is rated using 5 levels (no problems, slight problems, moderate problems, severe problems, and extreme problems) (Herdman 2011). These dimensions are converted into a single country-specific health index score. Using the United States value set, the health state index score ranges from -0.573 to 1.0, with higher scores indicating better health. Subjects also rate their overall health on a vertical visual analog scale, with anchors ranging from 0 (“worst health you can imagine”) to 100 (“best health you can imagine”).

Huntington’s Disease Health Index (HD-HI)

The HD-HI is a disease-specific, patient-reported outcome measure designed to evaluate patient disease burden in therapeutic trials (Brumfield 2022). The HD-HI comprises 13 subscales that each measure a different individual area of HD patient health. Together, the subscales can be utilized to estimate a patient’s overall disease burden. Each question in the instrument was selected based on its high relevance to the HD population, its ability to be consistently understood by patients and clinicians, its content validity, its face

validity, and its potential responsiveness to measure therapeutic benefit of disease progression during clinical trials.

The HD-HI was completed by the study participant (with assistance, as necessary). In standard use, the instrument is handed to a patient who is asked to read the directions and complete the instrument, using a pen, by checking the most appropriate box next to each question.

Upon completion of the HD-HI, 14 scores are generated: the participant receives a score for each of the 13 subscales and a total instrument score, which is a composite of the 13 subscale scores. The score for each subscale and the total instrument ranges from 0 to 100, with 100 representing the highest disease burden.

The Anosognosia Scale

The Anosognosia Scale is an instrument used to screen for anosognosia in daily practice and is specific for HD (Deckel 1996). This scale requires patients and clinicians to rate the patient's ability to perform tasks (relative to other individuals of the same age and education level as the patient) on 8 items using a 5-point scale, from -2 ("very impaired") to +2 ("exceptionally well").

Barnes Akathisia Rating Scale (BARS)

The BARS is a validated 4-item scale used to assess the presence and severity of drug-induced akathisia (Barnes 1989). This scale includes objective items (e.g., observed restlessness) and subjective items (e.g., patient's awareness of restlessness and related distress), and a global assessment of akathisia. Objective akathisia, subjective awareness of restlessness, and subjective distress related to restlessness are rated on a 4-point scale (range: 0 to 3) to give the total score. The global assessment is made on a scale of 0 to 5 (0=absent; 1=questionable; 2=mild akathisia; 3=moderate akathisia; 4=marked akathisia; 5=severe akathisia).

Hospital Anxiety and Depression Scale (HADS)

The HADS is commonly used to determine the levels of anxiety and depression that a person is experiencing. The HADS is a 14-item scale; 7 of the items relate to anxiety and 7 relate to depression (Zigmond 1983). Each item is answered on a 4-point (range: 0 to 3) response category; possible scores range from 0 to 21 for anxiety and 0 to 21 for depression. The HADS has been validated as a measure of depression and anxiety (Barczak et al., 1988).

References

Barnes TR. A rating scale for drug-induced akathisia. *Br J Psychiatry* 1989;154:672-6.

Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.

Brumfield OS, Zizzi CE, Dilek N, et al. The Huntington's Disease Health Index: initial evaluation of a disease-specific patient reported outcome measure. *J Huntingtons Dis* 2022;11:217-26.

Carlozzi NE, Ready RE, Frank S, et al. Patient-reported outcomes in Huntington's disease: quality of life in neurological disorders (Neuro-QoL) and Huntington's disease health-related quality of life (HDQLIFE) physical function measures. *Mov Disord* 2017;32:1096-102.

Deckel AW, Morrison D. Evidence of a neurologically based "denial of illness" in patients with Huntington's disease. *Arch Clin Neuropsychol* 1996;11:295-302.

Guy W, ed. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration; Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.

Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727-36.

Huntington Study Group (HSG). Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996;11:136-42.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361-70.

STUDY METHODS

Inclusion and Exclusion Criteria

To participate in this study, subjects must meet the following criteria:

1. Be a male or female aged 18 to 75 years, inclusive.
2. Diagnosis of motor manifest HD at or before screening.
3. Genetic diagnosis of HD with an expanded CAG repeat (≥37) in huntingtin (HTT) gene at or before Day -1 (baseline).
4. Subjects must be ambulatory, but assist devices are permitted.
5. TMC score ≥8 at screening and Day -1 (baseline).
6. Total Functional Capacity (TFC) score ≥5 at screening. Subjects with a TFC score between 5 and 10 (inclusive) must have a reliable caregiver to ensure study drug administration and attendance at study visits.
7. Subjects of childbearing potential must agree to use contraception consistently while participating in the study until 30 days (females) or 90 days (males) after the last dose of the study drug. A female subject of childbearing potential is defined as a subject who is not surgically sterile (ie, bilateral oophorectomy, hysterectomy, or bilateral tubal ligation for at least 3 months prior to screening) and who has not been postmenopausal for at least 1 year prior to screening. A male subject of childbearing potential is defined as a subject who has not been vasectomized for at least 3 months prior to screening.

Acceptable methods of contraception include the following:

- Condom with spermicide (cream, spray, foam, gel, suppository, or polymer film)
- Diaphragm with spermicide (with or without condom)
- Cervical cap with spermicide (with or without condom)
- Vaginal sponge impregnated with spermicide used with condom
- Intrauterine device (IUD)
- Hormonal contraception taken for at least 3 months prior to screening

The following subjects are not required to use contraception:

- Male and female subjects not of childbearing potential
- Subjects who practice total abstinence from sexual intercourse as the preferred lifestyle (periodic abstinence is not acceptable)
- Female subjects with male partners not of childbearing potential

8. Female subjects of childbearing potential must have a negative serum β-human chorionic gonadotropin (β-hCG) pregnancy test result at screening and a negative urine pregnancy test at Day -1.
9. Have a body mass index (BMI) of 15 to 47 kg/m2 (inclusive) at screening (BMI is defined as the subject's weight in kilograms divided by the square of the subject's height in meters).
10. Subject has voluntarily provided informed consent and has signed an ICF and is willing and able to adhere to the study regimen and study procedures described in the ICF. Subjects must also have been deemed capable of providing consent to study participation using the UBACC prior to signing the ICF.
11. Subject is able to read and understand English.
12. Be willing to provide authorization for access to personal health information in conjunction with US Health Insurance Portability and Accountability Act (HIPAA; US sites only).
13. Subjects participating in the exploratory movement sensor substudy must be willing and able, with the assistance of a reliable caregiver or companion, to comply with all substudy procedures.

Subjects will be excluded from the study if they:

1. Are currently pregnant or breastfeeding.
2. Have clinically manifest dysphagia as defined by a Swallowing Disturbance Questionnaire (SDQ) score ≥11. Subjects with an SDQ score ≥11 may still be eligible per investigator judgement, if they score ≤2 on item 13 (Dysphagia) of the Clinical Rating Scale for Progressive Supranuclear Palsy (CRS-PSP).
3. Have a history or evidence of long QT syndrome, cardiac tachyarrhythmia, left bundle-branch block, atrioventricular (AV) block, uncontrolled bradyarrhythmia, or heart failure.

4. Have an average triplicate electrocardiogram (ECG) corrected QT interval using Fridericia's formula (QTcF) >450 msec (males) or >470 msec (females) or evidence of any significant cardiac abnormality at screening or Day -1 (baseline).
5. Had a medically significant illness within 30 days before Day -1 (baseline), or any history of neuroleptic malignant syndrome.
6. Have a medically significant abnormality, physical examination finding, or any other measurement or observation of clinical significance that may interfere with the objectives of the study observed during screening or Day -1 (baseline).
7. Have an unstable or serious medical or psychiatric illness at screening or Day -1 (baseline).
8. Have an untreated or undertreated psychiatric illness, such as depression. Subjects receiving antidepressant therapy may be enrolled if he/she has been on a stable dose for at least 8 weeks prior to Day -1 (baseline).
9. Have a score ≥11 on the depression subscale of the Hospital Anxiety and Depression Scale (HADS) at screening or Day -1 (baseline).
10. Have a significant risk of suicidal behavior. Subjects with any suicidal behavior or suicidal ideation of type 4 (active suicidal ideation with some intent to act, without specific plan) or type 5 (active suicidal ideation with specific plan and intent) based on the Columbia-Suicide Severity Rating Scale (C-SSRS) in the 3 months prior to screening (using baseline/ screening version) or Day -1 (using Since Last Visit version) will be excluded.
11. Have a positive human immunodeficiency virus antibody (HIV-Ab) test result or hepatitis B surface antigen (HBsAg) test result at screening. Subjects with positive hepatitis C virus antibody (HCV) and confirmatory positive polymerase chain reaction (PCR) reflex test results at screening will be allowed to participate in the study provided that the subject is asymptomatic as assessed by the investigator and does not meet the liver function tests abnormalities for alanine transaminase (ALT), aspartate transaminase (AST), gamma-glutamyl transferase (GGT), and total bilirubin in exclusion criterion 12.
12. Have any of the following laboratory test abnormalities at screening:
    - Serum creatinine >1.5 times the upper limit of normal (ULN).
    - AST ≥2.5 × ULN
    - ALT ≥2.5 × ULN
    - GGT ≥3.0 × ULN
    - Total bilirubin >1.5 mg/dL
13. Have any of the following hematologic abnormalities at screening:
    - Hemoglobin <10 g/dL
    - White blood cell (WBC) count $<3.0 \times 10^3/mm^3$
    - Platelet count $<100,000/mm^3$

14. Have a positive urine drug screen at screening (positive for amphetamines, barbiturates, phencyclidine, benzodiazepines, cocaine, or opiates), except for subjects who have a prescription for benzodiazepines or opiates.
15. History of substance dependence or substance (drug) or alcohol abuse (nicotine and caffeine dependence are not exclusionary), as defined in the Diagnostic and Statistical Manual of Mental Disorders (eg, -IV or -5), within 1 year of screening.
16. Have received any prohibited medication (see Section 9.9.1).
17. Have received gene therapy at any time, or an investigational drug in the context of a clinical study within 30 days or 5 half-lives (if known), whichever is longer, of Day -1 or plan to use such investigational drug (other than the study drug) during this study.
18. Have a blood loss ≥550 mL or have donated blood within 30 days prior to Day -1 (baseline).
19. Have a history of previously established therapy with a VMAT2 inhibitor, in the judgment of the investigator and in consultation with a study medical monitor if needed. Previous exposure to a VMAT2 inhibitor is allowable provided that discontinuation occurred >30 days prior to screening, prior to establishment of a therapeutic response, and was otherwise unrelated to efficacy or tolerability.
20. Subjects participating in the exploratory movement sensor substudy only:
    - Must not have an implanted pacemaker or defibrillator, or other implantable device
    - Must not have known allergies or hypersensitivities to adhesives or hydrogel

Dosing and Blinding

Valbenazine was supplied as orally administered capsules containing 20 or 40 mg doses. Placebo was supplied as capsules that were identical in appearance to valbenazine.

All participants, study investigators, study site personnel, and the study sponsor were blinded to treatment. Participants were identified by a unique subject number and randomised to either valbenazine or placebo using an interactive web response system (IWRS). The randomisation code could only be broken in the following cases: participant was pregnant; participant experienced a serious adverse event that the investigator felt could not be adequately treated without knowing the study drug; or for regulatory reporting requirements. Members of the independent Data Safety Monitoring Board (DSMB) and designated DSMB support individuals were unblinded.

If the investigator determined that a dose reduction was needed, maintenance of the blind was managed identically in each treatment group. A study drug kit containing a lower dose of valbenazine or identically marked placebo was assigned through the IWRS and dispensed by the investigator.

IMPACT OF COVID-19 PANDEMIC ON OUTCOMES

Twelve participants (6 valbenazine, 6 placebo) had 1 or more major protocol deviations due to the COVID-19 pandemic, including missed visit or assessment (n=11), nonstandard or remote assessment collection (n=10), study treatment interruption (n=8), discontinuation of study treatment and/or study (n=7), and presumed or confirmed COVID-19 diagnosis (n=2). None of these deviations were considered to have affected study results or overall outcome of the study.

In addition, supplementary analyses of the primary and secondary endpoints were conducted in a “COVID-19 impact population” that excluded all individuals who discontinued the study prematurely for reasons related to COVID-19. The results of these analyses were very similar to those found in the full analysis set, indicating that discontinuation due to COVID-19 had minimal overall impact on study outcomes.