Characterizing Treatment Effects of Valbenazine for Tardive Dyskinesia:
Additional Results From the KINECT 3 Study

Christoph U. Correll, MD[a,b,c,*]; Andrew J. Cutler, MD[d]; John M. Kane, MD[a,b]; Joseph P. McEvoy, MD[e]; Grace S. Liang, MD[f]; and Christopher F. O'Brien, MD[f]

ABSTRACT

Background: In the KINECT 3 (NCT02274558; October 2014 to September 2015) study, valbenazine efficacy in tardive dyskinesia (TD) was demonstrated based on mean changes from baseline in the Abnormal Involuntary Movement Scale (AIMS) total score (sum of items 1–7). Data from this study were analyzed further to provide a more clinically meaningful interpretation of the primary AIMS results.

Methods: The study included adults who had a *DSM-IV* diagnosis of schizophrenia, schizoaffective disorder, or any mood disorder and also met *DSM-IV* criteria for neuroleptic-induced TD. Study participants received 6 weeks of double-blind treatment with valbenazine (40 or 80 mg/d) or placebo. Post hoc AIMS analyses, based on available data, included Cohen *d* effect sizes, response analyses with odds ratios (ORs) and numbers needed to treat (NNTs), and shift analyses.

Results: At week 6 (N = 202), medium-to-high effect sizes were found for mean improvements in AIMS total score (40 mg/d, *d* = 0.52; 80 mg/d, *d* = 0.89). For AIMS total score responses of ≥ 10% to ≥ 70% improvement from baseline, statistical significance was found for valbenazine 80 mg/d versus placebo (*P* ≤ .01), with ORs (range, 3.0–10.3) and NNTs (range, 3–9) indicating clinical relevance. For response per AIMS item (score ≤ 1 at week 6), significant differences between valbenazine (both doses or 80 mg/d) and placebo were found in the lips, jaw, tongue, and upper extremities. In participants who had an AIMS item score ≥ 1 at baseline, the percentage with a ≥ 1-point improvement at week 6 (shift) was significantly higher with valbenazine (40 and/or 80 mg/d) versus placebo in all 7 body regions.

Conclusions: Consistent with primary published results for KINECT 3, these supplemental analyses indicate that participants treated with valbenazine (40 or 80 mg/d) had statistically significant and clinically relevant improvements in TD severity both overall and in specific body regions.

Trial Registration: ClinicalTrials.gov identifier: NCT02274558

J Clin Psychiatry 2019;80(1):18m12278

[a]Department of Psychiatry, The Zucker Hillside Hospital, Glen Oaks, New York
[b]Department of Psychiatry and Molecular Medicine, The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, New York
[c]Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, New York
[d]Meridien Research, Tampa, Florida
[e]Department of Psychiatry and Health Behavior, Medical College of Georgia, Augusta University, Augusta, Georgia
[f]Neurocrine Biosciences, Inc, San Diego, California
Corresponding author: Christoph U. Correll, MD, 75-59 263rd St, Glen Oaks, NY 11004 (ccorrell@northwell.edu).

The US approval of once-daily valbenazine for adults with tardive dyskinesia (TD) was based on clinical trials in which the Abnormal Involuntary Movement Scale (AIMS) was the primary efficacy assessment. KINECT 3 (NCT02274558) was a phase 3 study that included a 6-week, double-blind, placebo-controlled (DBPC) period with valbenazine (40 or 80 mg/d); a 42-week valbenazine extension (VE) period (with blinded dose); and a 4-week washout period.[1,2] KINECT 3 met statistical significance for the primary 6-week DBPC endpoint (mean change from baseline in AIMS total score); VE results indicated continued AIMS improvement with long-term treatment.

As originally developed by the National Institute of Mental Health, the AIMS consists of 12 items. Items 1–7 assess the severity of abnormal movements in different body regions (face, lips, jaw, tongue, upper extremities, lower extremities, trunk) and are scored on a 5-point scale (0 = none to 4 = severe).[3] One advancement of the valbenazine clinical program was to define anchors for this scale as follows: 1 = low amplitude, present during some but not most of the exam; 2 = low amplitude and present during most of the exam (or moderate amplitude and present during some of the exam); 3 = moderate amplitude and present during most of the exam; and 4 = maximal amplitude and present during most of the exam. Items 8–12 include 3 questions based on clinician judgment (global severity, patient incapacitation, patient awareness) and 2 questions concerning dental status. The AIMS total score, as used in recent clinical trials, is defined by summing regional severity scores (ie, AIMS items 1–7). Because items 8–12 are subjective (especially 8, 9, and 10), with no conventional scoring, they are not usually included in the total.

The AIMS total score is valuable for evaluating treatment outcomes when the AIMS examination is performed rigorously in clinical trial populations, but its current use in clinical practice may be limited due to inadequate training and variable administration. Also, this total score is not a linear scale, and interpretation may require additional information about the constituent items. An AIMS total score of 4 could represent severe abnormal movements in a single body region or minimal abnormal movements in 4 different regions. Additionally, there is insufficient information about incorporating the impact or burden of TD into an overall assessment of severity. A minimal-to-mild abnormal orofacial

Clinical Points

- In contemporary clinical trials of tardive dyskinesia (TD), such as the KINECT 3 valbenazine study, efficacy is defined as a statistically significant improvement from baseline in the Abnormal Involuntary Movement Scale total score relative to placebo. However, additional analytical approaches such as effect sizes, odds ratios, and numbers needed to treat (NNTs) may provide a wider perspective for clinicians.

- These analytic approaches were applied post hoc to data from KINECT 3, which included a 6-week placebo-controlled period, followed by a 42-week blinded extension with dosing and 4-week valbenazine washout.

- Patients who took once-daily valbenazine (40 and/or 80 mg/d dose groups) had clinically meaningful improvements in TD compared to placebo, including moderate-to-strong treatment effects (Cohen d = 0.52–0.89) as well as ORs of 3–10 and NNTs of 3–9 for treatment response depending on the definition (≥ 10% to ≥ 70% improvement).

movement may be more distressing and perceived as more "severe" in a higher functioning patient than a moderate abnormal limb movement in someone with little awareness of his/her TD. This patient's caregiver or family member may find the movements distressing, but the patient may be unaware of his/her own TD.

In October 2016, a Tardive Dyskinesia Assessment Workshop was convened to discuss the challenges of assessing TD in research and clinical settings.[4] Workshop participants agreed that clinically meaningful AIMS analyses are needed and proposed different possible approaches, including providing effect sizes, odds ratios (ORs), and numbers needed to treat (NNTs). Based on these discussions, AIMS data from the KINECT 3 study were analyzed to illustrate different ways that this scale could be used to evaluate TD treatments.

METHODS

Study Participants

As previously reported,[1,2] KINECT 3 was a randomized, double-blind, placebo-controlled, parallel-group, fixed-dose study, conducted from October 2014 to September 2015 at 63 North American centers. All study participants demonstrated the capacity to consent and provided written consent. For each participating center, the study protocol was approved by an institutional review board. KINECT 3 included adults who met the following criteria: *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (*DSM-IV*) diagnosis of schizophrenia, schizoaffective disorder, or any mood disorder; *DSM-IV* diagnosis of neuroleptic-induced TD for ≥ 3 months prior to screening; and moderate-to-severe TD as qualitatively assessed by an expert reviewer at screening. Participants were required to be psychiatrically stable, and maintenance medications to treat psychiatric and medical conditions were allowed.

Participants who completed the 6-week DBPC phase of KINECT 3 were eligible to enroll in the long-term extension study. Participants who initially received placebo were re-randomized to valbenazine 40 mg/d or 80 mg/d; those who were initially randomized to valbenazine continued at the same dose. All participants, study site investigators, and central AIMS video raters were blinded to treatment in both the DBPC and extension phases. The central raters were also blinded to study visit in both phases.

AIMS Analyses

Post hoc analyses were conducted using available AIMS data from participants who received study treatment. Scoring was based on the consensus of 2 central AIMS video raters (movement disorder specialists) who were blinded to treatment and study visit.

Mean score changes from baseline and treatment effect sizes were analyzed for the AIMS total score and individual item scores. Mean changes from baseline to week 6 (end of DBPC) were analyzed using an analysis of covariance model with baseline AIMS total/item score as a fixed effect and with treatment group and diagnosis as covariates. Treatment effect sizes were estimated using Cohen d calculation. The mean percent change from baseline in AIMS total score was also calculated.

A range of AIMS total score responses was analyzed, based on reductions from baseline of ≥ 10%, ≥ 20%, ≥ 30%, ≥ 40%, ≥ 50%, ≥ 60%, ≥ 70%, ≥ 80%, or ≥ 90% at weeks 6, 48, and 52. Worsening in AIMS total score was also analyzed, based on increases from baseline of ≥ 10%, ≥ 20%, ≥ 30%, ≥ 40%, ≥ 50%, ≥ 60%, ≥ 70%, ≥ 80%, ≥ 90%, or ≥ 100% at week 6. For each AIMS item, response was defined as the percentage of participants who had a score ≤ 1 (none/minimal) at weeks 6, 48, and 52, regardless of the baseline score. Complete response was defined as a score ≤ 1 on all AIMS items at weeks 6, 48, and 52, regardless of the baseline score.

Three sets of category shifts were analyzed for each AIMS item as follows: score ≥ 3 (moderate or severe) at baseline and ≤ 2 (mild to none) at weeks 6, 48, and 52; score ≥ 2 (mild to severe) at baseline and score decrease ≥ 2 points at weeks 6, 48, and 52; score ≥ 1 (minimal to severe) at baseline and score decrease ≥ 1 point at weeks 6, 48, and 52.

For response and shift analyses, statistical significance between valbenazine and placebo at week 6 was analyzed using the Cochran-Mantel-Haenszel test. Logistic regression was used to estimate ORs and 95% confidence intervals (95% CI). NNTs were calculated as the inverse of the difference between valbenazine and placebo response/shift rates.

RESULTS

Participants

Of 205 participants who completed the DBPC period (week 6), 198 entered the VE period, 124 completed long-term treatment (week 48), and 121 reached the final post-withdrawal visit (week 52) (Supplementary Table 1). As previously reported,[1,2] participants' baseline characteristics were generally similar across treatment groups. Overall characteristics of participants in the DBPC period were

Figure 1. Least Squares Mean Changes From Baseline to Week 6 in AIMS Total and Item Scores With Treatment Effect Sizes[a]
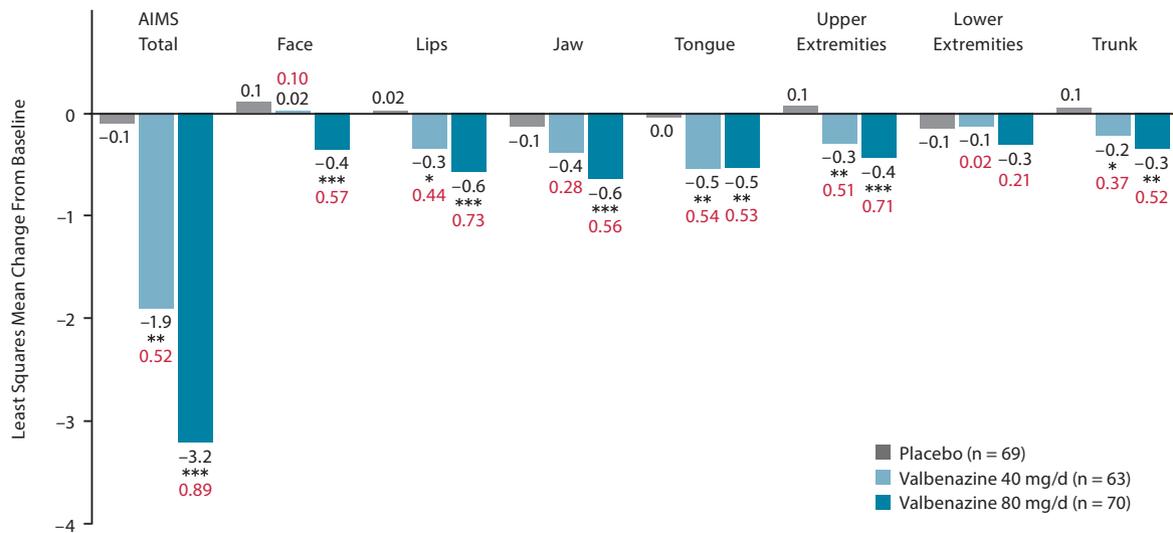
[a]Least squares mean based on the ANCOVA model with AIMS item score, treatment group, and diagnosis as covariates. Cohen d effect sizes are indicated below the least squares mean change values.
*$P < .05$.
**$P < .01$.
***$P ≤ .001$ for valbenazine versus placebo.
Abbreviations: AIMS = Abnormal Involuntary Movement Scale, ANCOVA = analysis of covariance.

as follows: men, 54%; white, 57%; mean age = 56 years; mean body mass index = 28.1 kg/m^2; and mean AIMS total score = 10.0. The majority of participants were diagnosed with schizophrenia/schizoaffective disorder (66%), and most were taking a concomitant antipsychotic (85%). Mean AIMS total and item scores were generally similar across treatment groups (Supplementary Table 2).

Treatment Effect Sizes and Percent Improvements

Mean changes from baseline to week 6 in AIMS total score were significantly greater with valbenazine versus placebo, with placebo-adjusted least squares mean differences (LSMDs) and effect sizes as follows: 40 mg/d (LSMD, −1.8; $d = 0.52$); 80 mg/d (LSMD, −3.1; $d = 0.89$). Mean changes for AIMS item scores were statistically significant for 1 or both valbenazine doses in all body regions except lower extremities (Figure 1). Effect sizes ranged from 0.02 (lower extremities) to 0.54 (tongue) in the 40-mg/d group and from 0.21 (lower extremities) to 0.73 (lips) in the 80-mg/d group.

Calculated as a mean percentage change, AIMS total score improved from baseline to week 6 by 6.6% and 30.0% with valbenazine 40 mg/d and 80 mg/d, respectively, and worsened by 7.7% with placebo. Mean percent changes from baseline in the extension study indicated ongoing improvements with long-term treatment (week 48: 40 mg/d, 23.7%; 80 mg/d, 39.4%) and general loss of effect after washout (week 52: 40 mg/d, 2.3%; 80 mg/d, 2.0%).

Response Threshold Analyses

For AIMS total score responses of ≥ 10% to ≥ 70%, a significantly greater proportion of participants responded with valbenazine 80 mg/d versus placebo (Figure 2). ORs favoring valbenazine over placebo (range, 1.9 to 10.3) were found for all response thresholds (OR for ≥ 90% not estimable due to 0% placebo response rate). NNTs < 10 were found with valbenazine (both doses or 80 mg/d) for responses of ≥ 10% to ≥ 70%. At week 48, at least one-half of participants achieved responses of ≥ 10% to ≥ 50% with valbenazine (both doses or 80 mg/d) (Supplementary Table 3). All response rates decreased from week 48 to week 52. In terms of worsening, significantly fewer valbenazine-treated participants had ≥ 50% increase in AIMS total score from baseline to week 6 (6.0% vs 15.9% for placebo; $P < .05$), with an OR favoring valbenazine of 0.3 and an NNT of 10 for preventing worsening of TD (Supplementary Figure 1).

A significantly higher percentage of participants had an AIMS item response (ie, score ≤ 1 at week 6) with valbenazine (both doses or 80 mg/d) versus placebo in the lips, jaw, tongue, and upper extremities (Figure 3). ORs ≥ 2 favoring valbenazine (both doses or 80 mg/d) were found in the face, lips, jaw, tongue, and upper extremities; NNTs < 10 were found in the same regions. For each AIMS item, at least one-half of participants in both valbenazine dose groups achieved a score ≤ 1 at week 48; at week 52, at least 30% of participants from both dose groups continued to have no or minimal abnormal movements (Supplementary Table 4).

A complete response (score ≤ 1 on all AIMS items at week 6) was found in 24.3% of participants in the 80-mg/d group (Figure 2). There was no statistically significant difference between valbenazine and placebo for complete response, but the OR and NNT for valbenazine 80 mg/d (2.2 and 9, respectively) suggested clinical benefits. Baseline characteristics of the complete responders suggest that the placebo group may have had less severe TD at baseline;

Figure 2. Response Analyses for AIMS Total Score at End of Double-Blind Placebo-Controlled Treatment (Week 6)[a]
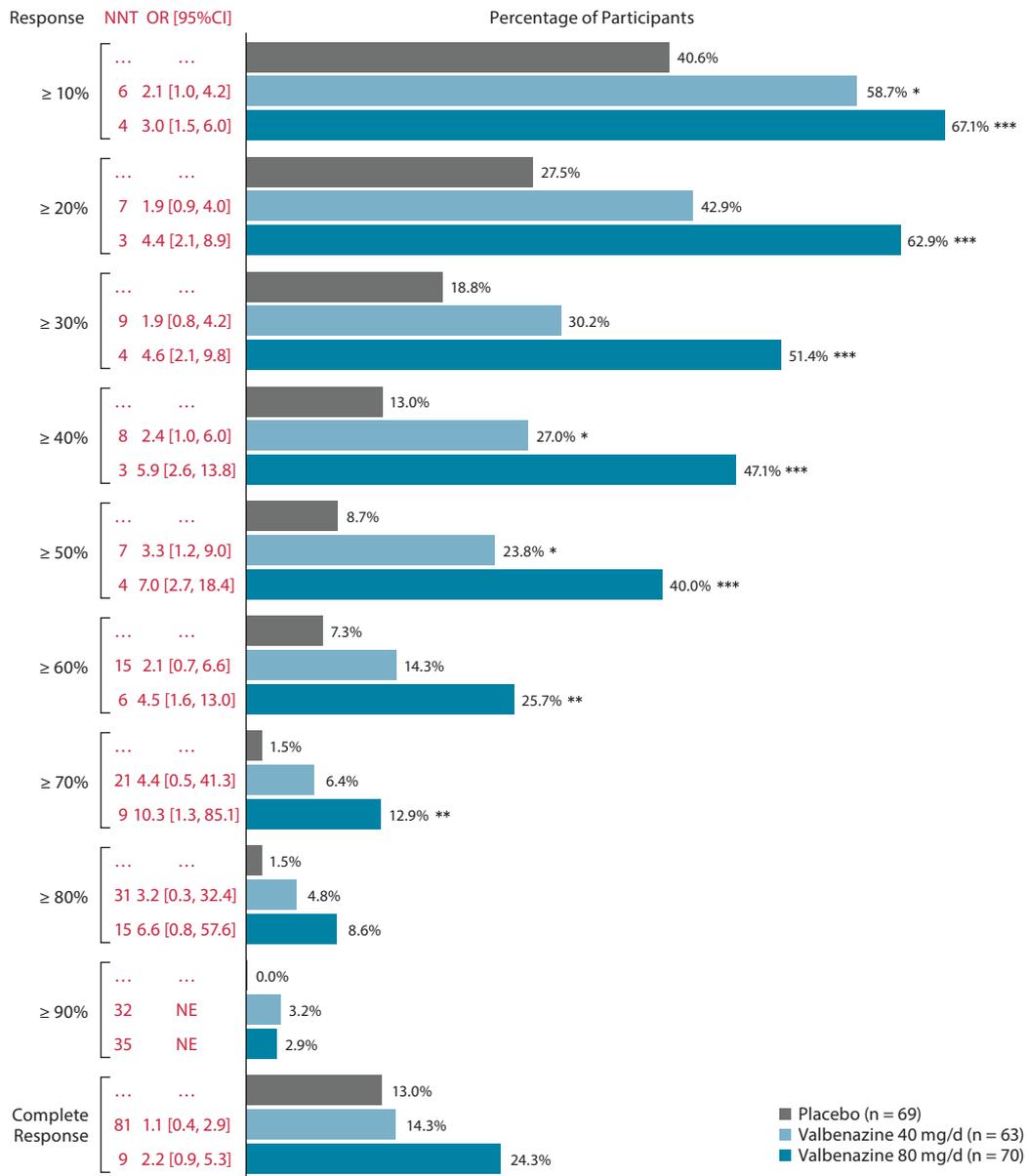
aResponse defined as a percent change from baseline to week 6 in the AIMS total score (sum of items 1–7). Complete response defined as a score ≤ 1 on all AIMS items at week 6. Analyzed using the Cochran-Mantel-Haenszel method with comparison of response rates between treatment groups (valbenazine 40 or 80 mg/d vs placebo) stratified by diagnosis.
*$P < .05$.
**$P ≤ .01$.
***$P < .001$ vs placebo.
Abbreviations: AIMS = Abnormal Involuntary Movement Scale, CI = confidence interval, NE = not estimable, NNT = number needed to treat, OR = odds ratio.

however, the small size of this group (n = 9) should also be noted (Supplementary Table 5). At week 48, complete response was found in 18.3% and 36.5% of participants in the valbenazine 40 mg/d and 80 mg/d groups, respectively; these percentages decreased to 10.0% and 9.8% after washout (Supplementary Table 3).

Shift Analyses

Of participants with available assessments at baseline and week 6, the percentages with AIMS item scores ≥ 1, ≥ 2, and ≥ 3 at baseline are presented in Figure 4. In participants with an item score ≥ 1 at baseline, the percentage with ≥ 1-point improvement at week 6 was significantly higher with valbenazine (40 and/or 80 mg/d) versus placebo in all 7 body regions (Table 1). In participants with an item score ≥ 2 at baseline, the percentage with ≥ 2-point improvement was significantly higher with valbenazine (40 and/or 80 mg/d) versus placebo in the jaw, tongue, upper extremities, and trunk. In participants with an item score ≥ 3 at baseline, a significantly higher percentage improved to a score ≤ 2 at

Figure 3. Response Analysis for AIMS Item Scores at End of Double-Blind, Placebo-Controlled Treatment (Week 6)[a]

AIMS Item	NNT	OR [95%CI]
Face	... / 14 / 7	... / 1.4 [0.7, 2.9] / 2.0 [1.0, 4.3]
Lips	... / 8 / 5	... / 1.7 [0.9, 3.5] / 4.4 [1.2, 4.7]
Jaw	... / ... / 5	... / 0.9 [0.4, 1.7] / 2.4 [1.2, 4.7]
Tongue	... / 8 / 5	... / 1.7 [0.9, 3.4] / 2.4 [1.2, 4.7]
Upper Extremities	... / 5 / 5	... / 2.6 [1.2, 5.4] / 2.7 [1.4, 5.5]
Lower Extremities	... / ... / 52	... / 0.8 [0.4, 1.7] / 1.1 [0.5, 2.2]
Trunk	... / 77 / 17	... / 1.1 [0.5, 2.2] / 1.3 [0.7, 2.7]

Placebo (n = 69); Valbenazine 40 mg/d (n = 63); Valbenazine 80 mg/d (n = 70)
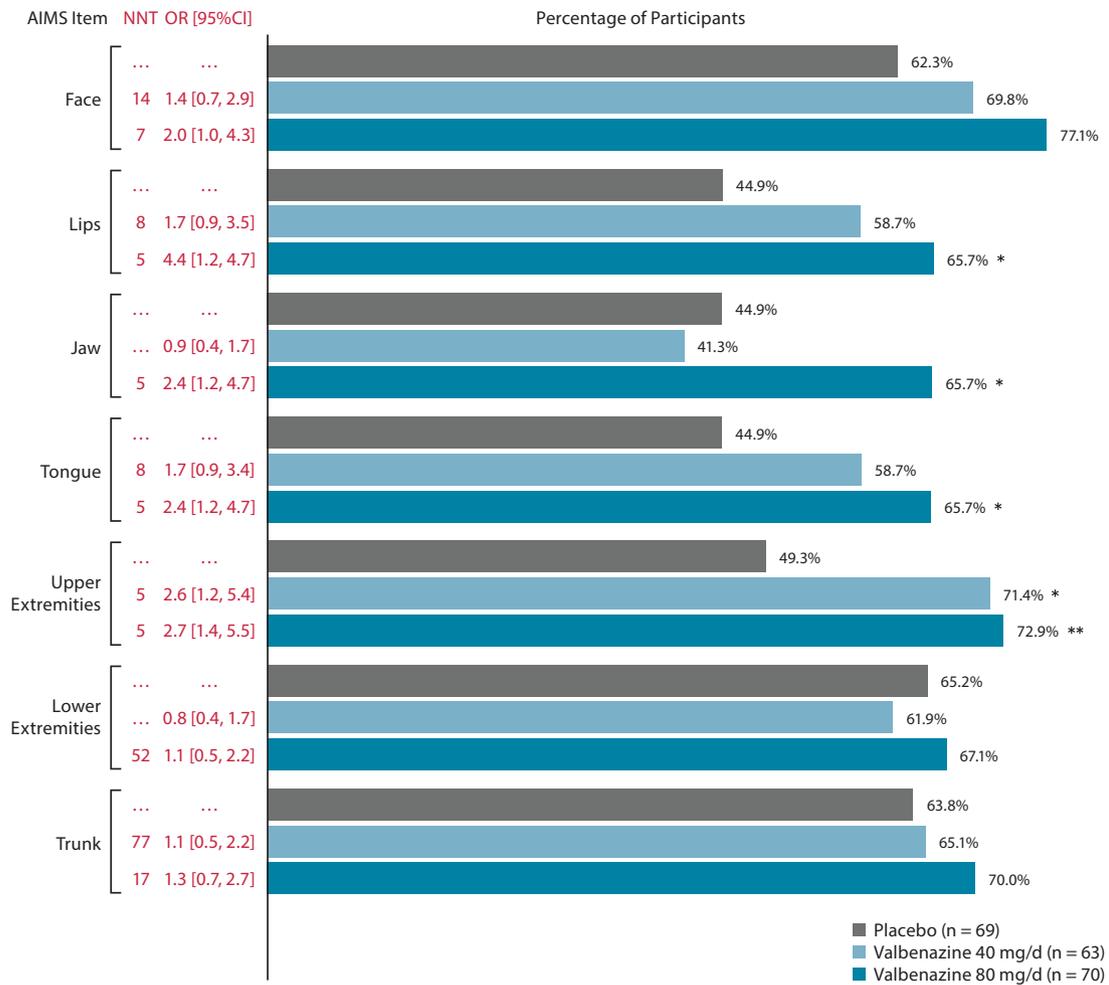
[a]Response defined as an item score ≤ 1 at week 6. Analyzed using the Cochran-Mantel-Haenszel method with comparison of response rates between treatment groups (valbenazine 40 or 80 mg/d vs placebo) stratified by diagnosis.
*P < .05.
**P < .01 vs placebo.
Abbreviations: AIMS = Abnormal Involuntary Movement Scale, CI = confidence interval, NNT = number needed to treat, OR = odds ratio.

Figure 4. Distribution of AIMS Item Scores at Baseline in All Participants With Available Assessments at Baseline and Week 6
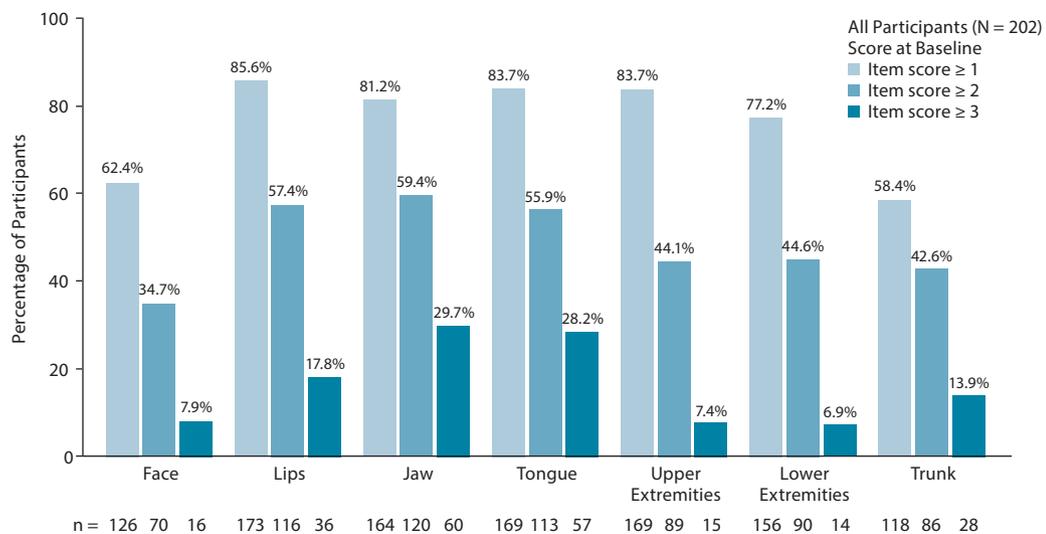
Abbreviation: AIMS = Abnormal Involuntary Movement Scale.

Table 1. Shift Analyses for AIMS Item Scores at End of Double-Blind, Placebo-Controlled Treatment (Week 6)[a]

| | Score ≥ 3 at Baseline Score ≤ 2 at Week 6 | | | Score ≥ 2 at Baseline Score Decrease ≥ 2 at Week 6 | | | Score ≥ 1 at Baseline Score Decrease ≥ 1 at Week 6 | | |
	Placebo	Valbenazine 40 mg/d	Valbenazine 80 mg/d	Placebo	Valbenazine 40 mg/d	Valbenazine 80 mg/d	Placebo	Valbenazine 40 mg/d	Valbenazine 80 mg/d
Face, n/N (%)	1/6 (16.7)	2/4 (50.0)	2/6 (33.3)	1/24 (4.2)	2/22 (9.1)	4/24 (16.7)	11/42 (26.2)	13/33 (39.4)	30/51 (58.8)
P value		.114	.307		.430	.125		.239	.002
OR (95% CI)		NE	6.0 (0.2–162.5)		2.7 (0.2–34.0)	5.7 (0.5–60.7)		1.8 (0.7–4.9)	4.0 (1.7–9.7)
NNT		3	6		21	8		8	4
Lips, n/N (%)	6/13 (46.2)	8/11 (72.7)	12/12 (100)	4/41 (9.8)	6/35 (17.1)	8/40 (20.0)	19/59 (32.2)	22/50 (44.0)	39/64 (60.9)
P value		.134	.002		.336	.208		.181	.001
OR (95% CI)		5.1 (0.6–43.1)	NE		1.9 (0.5–7.1)	2.3 (0.6–8.3)		1.7 (0.8–3.8)	3.3 (1.6–7.1)
NNT		4	2		14	10		9	4
Jaw, n/N (%)	11/24 (45.8)	17/22 (77.3)	12/14 (85.7)	6/41 (14.6)	11/41 (26.8)	14/38 (36.8)	16/55 (29.1)	26/52 (50.0)	35/57 (61.4)
P value		.036	.019		.191	.023		.026	.001
OR (95% CI)		4.0 (1.1–14.4)	6.7 (1.2–35.8)		2.1 (0.7–6.3)	3.3 (1.1–9.7)		2.5 (1.1–5.5)	3.9 (1.8–8.6)
NNT		4	3		9	5		5	4
Tongue, n/N (%)	3/19 (15.8)	11/20 (55.0)	11/18 (61.1)	3/40 (7.5)	13/42 (31.0)	7/31 (22.6)	20/54 (37.0)	33/55 (60.0)	32/60 (53.3)
P value		.012	.007		.009	.066		.018	.091
OR (95% CI)		7.0 (1.5–33.9)	8.8 (1.7–46.1)		4.8 (1.3–17.7)	3.8 (0.9–16.6)		2.5 (1.2–5.4)	1.9 (0.9–4.0)
NNT		3	3		5	7		5	7
Upper extremities, n/N (%)	5/7 (71.4)	2/4 (50.0)	4/4 (100)	1/31 (3.2)	3/23 (13.0)	7/35 (20.0)	13/59 (22.0)	19/49 (38.8)	35/61 (57.4)
P value		.541	.327		.154	.035		.060	<.001
OR (95% CI)		0.4 (0.03–6.1)	NE		5.2 (0.5–58.6)	6.8 (0.9–53.1)		2.2 (1.0–5.1)	4.9 (2.2–10.9)
NNT		−4	4		11	6		6	3
Lower extremities, n/N (%)	2/3 (66.7)	2/3 (66.7)	7/8 (87.5)	4/25 (16.0)	5/28 (17.9)	7/37 (18.9)	14/54 (25.9)	16/42 (38.1)	31/60 (51.7)
P value		1.000	.458		.919	.996		.225	.006
OR (95% CI)		1.0 (0.03–29.8)	2.9 (0.2–53.5)		1.1 (0.3–4.5)	1.0 (0.2–4.1)		1.7 (0.7–4.2)	2.9 (1.3–6.4)
NNT		NE	5		54	35		9	4
Trunk, n/N (%)	2/7 (28.6)	8/10 (80.0)	6/11 (54.6)	1/22 (4.6)	7/27 (25.9)	9/37 (24.3)	8/38 (21.1)	16/36 (44.4)	26/44 (59.1)
P value		.042	.310		.046	.050		.035	<.001
OR (95% CI)		11.4 (1.0–136.8)	2.5 (0.4–15.9)		7.8 (0.8–73.2)	7.3 (0.8–66.0)		2.9 (1.0–7.8)	6.8 (2.3–19.8)
NNT		2	4		5	6		5	3

[a]Analyzed using the Cochran-Mantel-Haenszel method with comparison of shift rates between treatment groups (valbenazine 40 or 80 mg/d vs placebo) stratified by diagnosis. A negative NNT value indicates a lower AIMS response rate with valbenazine vs placebo.

Abbreviations: AIMS = Abnormal Involuntary Movement Scale, CI = confidence interval, N = number of participants who met the score criteria at baseline, n = number of participants who met the shift criteria at week 6, NE = not estimable, NNT = number needed to treat, OR = odds ratio.

week 6 with valbenazine (40 and/or 80 mg/d) versus placebo in the lips, jaw, tongue, and trunk; sample sizes for some regions were notably small (< 10 participants per treatment group).

Ranges for ORs and NNTs across the 3 category shift analyses indicated generally favorable and consistent valbenazine effects in the tongue (OR, 1.9 to 8.8; NNT, 3 to 7), jaw (OR, 2.1 to 6.7; NNT, 3 to 9), and trunk (OR, 2.5 to 11.4; NNT, 2 to 6). Shift analyses at week 48 indicated that participants continued to experience improvements (Supplementary Table 6). Fewer participants met the criteria for shift analyses at week 52 relative to week 48, but some continued to demonstrate ongoing improvements after valbenazine was withdrawn.

DISCUSSION

In recent TD studies, including clinical trials of valbenazine[1,2,5] and deutetrabenazine,[6,7] primary efficacy was based on a mean change in AIMS total score (sum of items 1–7). By necessity, clinical trials need a single measure to define efficacy, and the AIMS total score change has been appropriate for this purpose. However, applying a reported mean score change to clinical practice can be challenging.

Therefore, AIMS data from KINECT 3 were analyzed post hoc to provide clinicians with complementary approaches for interpreting primary clinical trial results.

One interpretative approach is to translate AIMS score changes into a placebo-adjusted outcome (eg, LSMD) or treatment effect (eg, Cohen d) that accounts for placebo effects, number of participants, and SDs. Effect sizes make it easier to compare results from different clinical trials, although differences in study objectives and designs must always be considered. In KINECT 3, the placebo-adjusted AIMS total score change from baseline (40 mg/d, −1.8; 80 mg/d, −3.1) translated into medium-to-high effect sizes[8] (40 mg/d, $d = 0.52$; 80 mg/d, $d = 0.89$) that were dose-related. To assess potential differences across body regions, effect sizes for AIMS item scores were also calculated. A Cohen $d > 0.5$ was found with valbenazine 80 mg/d in all items except for lower extremities, with the largest effect sizes in lips ($d = 0.73$) and upper extremities ($d = 0.71$).

From a practical standpoint, percent changes, ORs, and complete response provide language that clinicians can use to explain study results to their patients. For example, clinicians could state that "on average, patients in clinical trials had approximately 30% improvement after 6 weeks of valbenazine" (percent change), or that "patients who

received valbenazine 80 mg/day were 7 times more likely to have substantial improvement than patients who received no treatment" (OR for AIMS ≥ 50% response), or that "approximately 25% of patients had no or minimal TD symptoms after 6 weeks of valbenazine" (complete response).

In the valbenazine studies, AIMS response was defined a priori as ≥ 50% total score improvement from baseline. However, other studies have used lower thresholds (eg, ≥ 25% or ≥ 30%)[9,10] that might also represent clinically meaningful responses. Response thresholds ranging from ≥ 10% to ≥ 90% were included in the current analysis to illustrate the full range of improvement experienced by KINECT 3 participants, with corresponding ORs and NNTs analyzed for clinical relevance. The ORs indicated that after 6 weeks of treatment, participants receiving valbenazine 80 mg/d had approximately 3 to 7 times greater odds of achieving ≥ 10% to ≥ 50% improvement in AIMS total score than placebo-treated participants. NNTs indicated that 3 or 4 participants required treatment with valbenazine 80 mg/d in order for 1 additional participant to achieve a ≥ 10% to ≥ 50% response. NNT standards have not yet been established for TD, although an NNT < 10 is often considered to be clinically meaningful.[11] For outcomes that did not reach statistical significance (eg, ≥ 20% and ≥ 30% response in the 40-mg/d group), an NNT < 10 may be used to estimate a meaningful improvement and make treatment decisions. For a more comprehensive approach, numbers needed to harm (NNHs) for treatment-emergent adverse events (TEAEs) can also be considered. A previously published analysis of KINECT 3 data showed an NNH of 78 for TEAEs leading to discontinuation (valbenazine, 3.8%; placebo, 2.6%).[12] For the most commonly reported TEAE, somnolence (valbenazine, 5.3%; placebo, 3.9%), the NNH was 74. In conjunction with the NNTs for AIMS total score response, these results suggest that valbenazine is more likely to help than harm.

Since the presentation of TD varies from patient to patient, a response analysis based on each individual AIMS item (ie, score ≤ 1) was conducted to explore whether improvements with valbenazine might differ by body region. The most robust item response with valbenazine 80 mg/d was in the lips (OR = 4.4, NNT = 5), which was comparable to the AIMS ≥ 30% total score response (OR = 4.6, NNT = 4). The next strongest item response was in the upper extremities (OR = 2.7, NNT = 5), followed by jaw and tongue (both OR = 2.4, NNT = 5). Since the orofacial and upper extremity regions are the most visible, they are probably the most meaningful to patients and their caregivers. It should be noted, however, that the AIMS item responses were conducted without consideration of baseline scores. Therefore, a lack of statistical significance in some regions (eg, face) may be attributable to lower baseline scores in this area, which is consistent with the distribution of AIMS item scores at baseline (Figure 4). A more stringent approach would be to define treatment success in terms of "complete response," defined in this analysis as a score ≤ 1

on all 7 AIMS items. Again, it is important to note that this complete response, like the total score responses (≥ 10% to ≥ 90%) and item score responses, does not take the broad range of the baseline scores into account.

One way to account for baseline severity is to conduct shift analyses, which can be defined in any way that captures a potentially clinically meaningful improvement. No shift analysis based on AIMS total score was conducted, but 3 sets of analyses based on AIMS item scores were explored, each in a different subgroup of participants (ie, those with an item score ≥ 3, ≥ 2, and ≥ 1 at baseline). Shift analyses results were generally congruent with AIMS item response results, with compelling ORs and NNTs found in orofacial regions (head, lips, jaw, and tongue) and upper extremities. However, differences in sample sizes need to be considered when interpreting the shifts, particularly with respect to the small number of participants who improved from an item score ≥ 3 at baseline to a score ≤ 2 at week 6. More research is needed to ascertain which (if any) of these shift analyses are generally applicable to patients with TD. In clinical practice, however, it may be possible for health care providers to develop shift criteria on a case-to-case basis, which could then be used to monitor patient progress and treatment response.

Because effect sizes, ORs, and NNTs require a comparator group, this report focused on the 6-week DBPC period of KINECT 3. However, response and shift analyses were also conducted for week 48 (end of VE period) and week 52 (end of washout period). Results from these analyses indicate that TD continued to improve during long-term treatment, particularly in regions that seemed relatively less responsive to valbenazine at week 6 (eg, lower extremities). While analyses based on AIMS total score (ie, mean change from baseline, ≥ 50% response) suggested an overall return to baseline after treatment withdrawal (as reported previously[2]), analyses based on individual AIMS items (ie, item responses, shift analyses) suggested a more complicated scenario. These analyses indicated a greater loss of effect in the 80-mg/d group than in the 40-mg/d group after treatment withdrawal, particularly in orofacial regions. Therefore, the mean worsening in AIMS total score found after treatment withdrawal may have been driven by a subset of participants who had experienced a marked improvement with treatment and a larger, rapid loss of treatment effect in certain regions. More extensive research is needed to better understand the time course of improvement in different body regions, the effects of long-term valbenazine treatment on each region, and the reasons why many participants reverted to baseline levels after treatment withdrawal while some did not.

Another method for interpreting a mean change in AIMS total score is to establish a minimal clinically important difference (MCID) in TD patients using a global rating scale (eg, Clinical Global Impression of Change) as an anchor. Preliminary analyses of data pooled from 3 valbenazine studies indicated an MCID of 2 to 3 points for the AIMS.[13,14] An upcoming publication will provide a more thorough exploration of the AIMS MCID.

The main limitation is that most analyses were conducted post hoc and the KINECT 3 study was not designed to test all of the analyses presented. Additionally, other factors such as functional ability and quality of life need to be considered when evaluating and treating patients with TD. Other AIMS items that were not collected in this study (eg, patient's awareness and distress about abnormal movements) may also need to be considered during clinical evaluation. Moreover, the sample size in some analyzed subgroups was limited, and baseline scores on some AIMS items were low, leaving little room for meaningful improvement. As already discussed, each tested approach has its own limitations; thus, several ways to capture improvement or response should be reported together to provide a more complete picture.

CONCLUSIONS

Results from these supplemental AIMS analyses indicated a consistently stronger effect with valbenazine (40 and/or 80 mg/d) versus placebo, with sustained improvements during double-blind long-term valbenazine treatment. As intended, each analysis provided a slightly different perspective on interpreting the effects of valbenazine on TD. It is also clear that, in addition to the AIMS, validated scales assessing the subjective burden and functional impairment related to TD symptoms (overall and body region) are needed. It is hoped that these analyses will help clinicians interpret clinical trial data and make relevant treatment decisions for their patients with TD.

REFERENCES

1. Hauser RA, Factor SA, Marder SR, et al. KINECT 3: a phase 3 randomized, double-blind, placebo-controlled trial of valbenazine for tardive dyskinesia. *Am J Psychiatry*. 2017;174(5):476–484.
2. Factor SA, Remington G, Comella CL, et al. The effects of valbenazine in participants with tardive dyskinesia: results of the 1-year KINECT 3 extension study. *J Clin Psychiatry*. 2017;78(9):1344–1350.
3. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: US National Institute of Mental Health, Psychopharmacology Research Branch; 1976.
4. Kane JM, Correll CU, Nierenberg AA, et al; Tardive Dyskinesia Assessment Working Group. Revisiting the Abnormal Involuntary Movement Scale: proceedings from the Tardive Dyskinesia Assessment Workshop. *J Clin Psychiatry*. 2018;79(3):17cs11959.
5. O'Brien CF, Jimenez R, Hauser RA, et al. NBI-98854, a selective monoamine transport inhibitor for the treatment of tardive dyskinesia: a randomized, double-blind, placebo-controlled study. *Mov Disord*. 2015;30(12):1681–1687.
6. Anderson KE, Stamler D, Davis MD, et al. Deutetrabenazine for treatment of involuntary movements in patients with tardive dyskinesia (AIM-TD): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Psychiatry*. 2017;4(8):595–604.
7. Fernandez HH, Factor SA, Hauser RA, et al. Randomized controlled trial of deutetrabenazine for tardive dyskinesia: the ARM-TD study. *Neurology*. 2017;88(21):2003–2010.
8. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155–159.
9. Stewart RM, Rollins J, Beckham B, et al. Baclofen in tardive dyskinesia patients maintained on neuroleptics. *Clin Neuropharmacol*. 1982;5(4):365–373.
10. Zhang WF, Tan YL, Zhang XY, et al. Extract of *Ginkgo biloba* treatment for tardive dyskinesia in schizophrenia: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2011;72(5):615–621.
11. Citrome L. Number needed to treat: what it is and what it isn't, and why every clinician should know how to calculate it. *J Clin Psychiatry*. 2011;72(3):412–413.
12. Citrome L. Valbenazine for tardive dyskinesia: a systematic review of the efficacy and safety profile for this newly approved novel medication—what is the number needed to treat, number needed to harm and likelihood to be helped or harmed? *Int J Clin Pract*. 2017;71(7):e12964.
13. Stacy M, Kurlan R, Burke J, et al. An MCID for AIMS dyskinesia total score change in subjects with tardive dyskinesia [abstract]. *Mov Disord*. 2017;32(suppl 2):S158.
14. Sajatovic M, Cutler A, Farahmand K, et al. Estimation of an MCID for AIMS total score change in tardive dyskinesia. Presented at the 30th Annual Psych Congress; New Orleans, LA: September 16–18, 2017. Psych Congress Network site. https://www.psychcongress.com/posters/estimation-mcid-aims-total-score-change-tardive-dyskinesia. Accessed January 5, 2018.

See supplementary material for this article at PSYCHIATRIST.COM.

Supplementary Material

List of Supplementary Material for the article

Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

SUPPLEMENTARY MATERIALS

Table 1. Patient Disposition in the Valbenazine KINECT 3 Trial

Population, n	Placebo	Valbenazine 40 mg/day	Valbenazine 80 mg/day
Double-blind, placebo-controlled period			
Randomized	78	76	80
Discontinued	7	13	9
Adverse event	2	4	2
Noncompliance	2	1	0
Withdrawal of consent	1	5	4
Lost to follow-up	2	1	1
Sponsor/investigator decision	0	2	1
Completed Week 6	71	63	71
Continued to extension period	69[a]	64[b]	65[b]
Valbenazine extension period			
Entered		97	101
Discontinued		36	38
Adverse event		14	17
Withdrawal of consent		9	8
Lost to follow-up		8	6
Noncompliance		3	3
Sponsor/investigator decision		2	2
Death		0	1[c]
Protocol deviation		0	1
Completed Week 48		61	63
Completed Week 52		60	61

[a]Participants in the original placebo group were re-randomized to valbenazine 40 mg/day (n=33) or 80 mg/day (n=36).
[b]Three participants initially randomized to valbenazine 80 mg/day had a dose reduction during the 6-week placebo-controlled period and were included in the 40-mg/day group for the valbenazine extension period.
[c]Attributed to multiple organ failure; judged by the investigator as unrelated to study drug.

Table 2. Mean AIMS Total Score and Item Scores at Baseline (ITT Population)

	Placebo n=76	Valbenazine 40 mg/day n=70	Valbenazine 80 mg/day n=79
Mean total score (SD)	9.9 (4.3)	9.8 (4.1)	10.4 (3.6)
Mean item score (SD)			
Face	1.1 (1.1)	0.9 (1.0)	1.2 (0.9)
Lips	1.6 (1.0)	1.5 (1.0)	1.7 (0.9)
Jaw	1.8 (1.3)	1.8 (1.2)	1.7 (1.1)
Tongue	1.7 (1.2)	1.9 (1.1)	1.6 (1.1)
Upper extremities	1.4 (0.9)	1.2 (0.8)	1.4 (0.8)
Lower extremities	1.2 (0.9)	1.2 (0.9)	1.5 (0.8)
Trunk	1.0 (1.1)	1.2 (1.3)	1.3 (1.1)

AIMS, Abnormal Involuntary Movement Scale; ITT, intent-to-treat; SD, standard deviation.

Table 3. Response Analyses for AIMS Total Score at End of Long-Term Valbenazine Treatment (Week 48) and After 4-Week Treatment Withdrawal (Week 52)

	Week 48		Week 52	
	40 mg/day n=60	80 mg/day n=63	40 mg/day n=60	80 mg/day n=61
Percent improvement from baseline				
≥10%	43 (71.7)	49 (77.8)	26 (43.3)	28 (45.9)
≥20%	37 (61.7)	45 (71.4)	23 (38.3)	25 (41.0)
≥30%	30 (50.0)	40 (63.5)	17 (28.3)	20 (32.8)
≥40%	22 (36.7)	35 (55.6)	15 (25.0)	17 (27.9)
≥50%	17 (28.3)	33 (52.4)	9 (15.0)	12 (19.7)
≥60%	12 (20.0)	26 (41.3)	8 (13.3)	8 (13.1)
≥70%	9 (15.0)	17 (27.0)	3 (5.0)	4 (6.6)
≥80%	7 (11.7)	11 (17.5)	0	3 (4.9)
≥90%	4 (6.7)	6 (9.5)	0	2 (3.3)
Complete response[a]	11 (18.3)	23 (36.5)	6 (10.0)	6 (9.8)

[a]Defined as a score ≤1 on all AIMS items at time of assessment.
AIMS, Abnormal Involuntary Movement Scale.

Table 4. Response Analyses for AIMS Item Scores at End of Long-Term Valbenazine Treatment (Week 48) and After Valbenazine Withdrawal (Week 52)

	Week 48		Week 52	
AIMS Item Response[a]	**40 mg/day n=60**	**80 mg/day n=63**	**40 mg/day n=60**	**80 mg/day n=61**
Face	50 (83.3)	54 (85.7)	42 (70.0)	41 (67.2)
Lips	36 (60.0)	45 (71.4)	30 (50.0)	25 (41.0)
Jaw	30 (50.0)	45 (71.4)	20 (33.3)	26 (42.6)
Tongue	36 (60.0)	43 (68.3)	26 (43.3)	19 (31.2)
Upper extremities	46 (76.7)	45 (71.4)	43 (71.7)	33 (54.1)
Lower extremities	47 (78.3)	49 (77.8)	43 (71.7)	42 (68.9)
Trunk	50 (83.3)	49 (77.8)	45 (75.0)	41 (67.2)

[a]Defined as a score ≤1 on AIMS item at time of assessment.
AIMS, Abnormal Involuntary Movement Scale.

Table 5. Baseline Characteristics of Complete Responders at Week 6[a]

	Placebo n=9	**Valbenazine 40 mg/day n=9**	**Valbenazine 80 mg/day n=17**
Age, mean (SD), years	54.6 (9.2)	60.8 (5.7)	56.0 (10.9)
Women, n (%)	7 (77.8)	5 (55.6)	9 (52.9)
Race, n (%)			
White	4 (44.4)	6 (66.7)	11 (64.7)
Black/African-American	5 (55.6)	2 (22.2)	5 (29.4)
Other/multiple	0	1 (11.1)	1 (5.9)
Body mass index, mean (SD), kg/m^2	29.5 (7.2)	28.6 (6.1)	30.4 (6.7)
Age at TD diagnosis, mean (SD), years	48.2 (11.0)	56.0 (4.6)	48.2 (14.0)
Primary psychiatric diagnosis, n (%)			
Schizophrenia/schizoaffective disorder	7 (77.8)	7 (77.8)	12 (70.6)
Mood disorder	2 (22.2)	2 (22.2)	5 (29.4)
Antipsychotic medication use, n (%)			
Atypical only	5 (55.6)	6 (66.7)	10 (58.8)
Typical only	2 (22.2)	2 (22.2)	3 (17.6)
Both	0	1 (11.1)	2 (11.8)
None	2 (22.2)	0	2 (11.8)
Lifetime history of suicidality, n (%)[b]	4 (44.4)	4 (44.4)	6 (35.3)
AIMS total score, mean (SD)	6.4 (2.8)	9.0 (4.7)	8.6 (2.8)

[a]Defined as a score ≤1 on all AIMS items 1-7 at time of assessment.
[b]Based on Columbia-Suicide Severity Rating Scale.
AIMS, Abnormal Involuntary Movement Scale; SD, standard deviation; TD, tardive dyskinesia.

Table 6. Shift Analyses for AIMS Item Scores at End of Long-Term Valbenazine Treatment (Week 48) and After Valbenazine Withdrawal (Week 52)

	Score ≥3 at Baseline Score ≤2 at Week 48/52		Score ≥2 at Baseline Score Decrease ≥2 points at Week 48/52		Score ≥1 at Baseline Score Decrease ≥1 point at Week 48/52	
	40 mg/day	80 mg/day	40 mg/day	80 mg/day	40 mg/day	80 mg/day
Face						
Week 48	3/4 (75.0)	3/6 (50.0)	9/22 (40.9)	14/25 (56.0)	22/34 (64.7)	32/48 (66.7)
Week 52	3/4 (75.0)	3/6 (50.0)	7/22 (31.8)	9/25 (36.0)	16/34 (47.1)	23/46 (50.0)
Lips						
Week 48	9/10 (90.0)	13/14 (92.9)	8/37 (21.6)	18/37 (48.7)	28/49 (57.1)	35/59 (59.3)
Week 52	5/10 (50.0)	9/13 (69.2)	6/37 (16.2)	6/35 (17.1)	20/49 (40.8)	20/57 (35.1)
Jaw						
Week 48	14/21 (66.7)	13/18 (72.2)	10/34 (29.4)	17/37 (46.0)	26/48 (54.2)	34/51 (66.7)
Week 52	10/21 (47.6)	7/17 (41.2)	7/34 (20.6)	5/35 (14.3)	19/48 (39.6)	17/49 (34.7)
Tongue						
Week 48	12/18 (66.7)	15/21 (71.4)	13/37 (35.1)	14/32 (43.8)	31/51 (60.8)	33/53 (62.3)
Week 52	8/18 (44.4)	12/21 (57.1)	7/37 (18.9)	4/31 (12.9)	21/51 (41.2)	16/51 (31.4)
Upper extremities						
Week 48	2/2 (100)	7/8 (87.5)	4/24 (16.7)	16/36 (44.4)	30/48 (62.5)	38/57 (66.7)
Week 52	0/2 (0.0)	6/8 (75.0)	4/24 (16.7)	8/35 (22.9)	21/48 (43.8)	24/55 (43.6)
Lower extremities						
Week 48	3/3 (100)	5/6 (83.3)	6/22 (27.3)	9/35 (25.7)	20/41 (48.8)	39/56 (69.6)
Week 52	3/3 (100)	4/6 (66.7)	4/22 (18.2)	7/33 (21.2)	19/41 (46.3)	27/54 (50.0)
Trunk						
Week 48	6/7 (85.7)	7/10 (70.0)	13/25 (52.0)	15/34 (44.1)	27/37 (73.0)	34/44 (77.3)
Week 52	5/7 (71.4)	8/10 (80.0)	7/25 (28.0)	13/34 (38.2)	24/37 (64.9)	25/42 (59.5)

AIMS, Abnormal Involuntary Movement Scale.

Figure 1. Worsening in AIMS Total Score at End of Double-Blind Placebo-Controlled Treatment (Week 6)
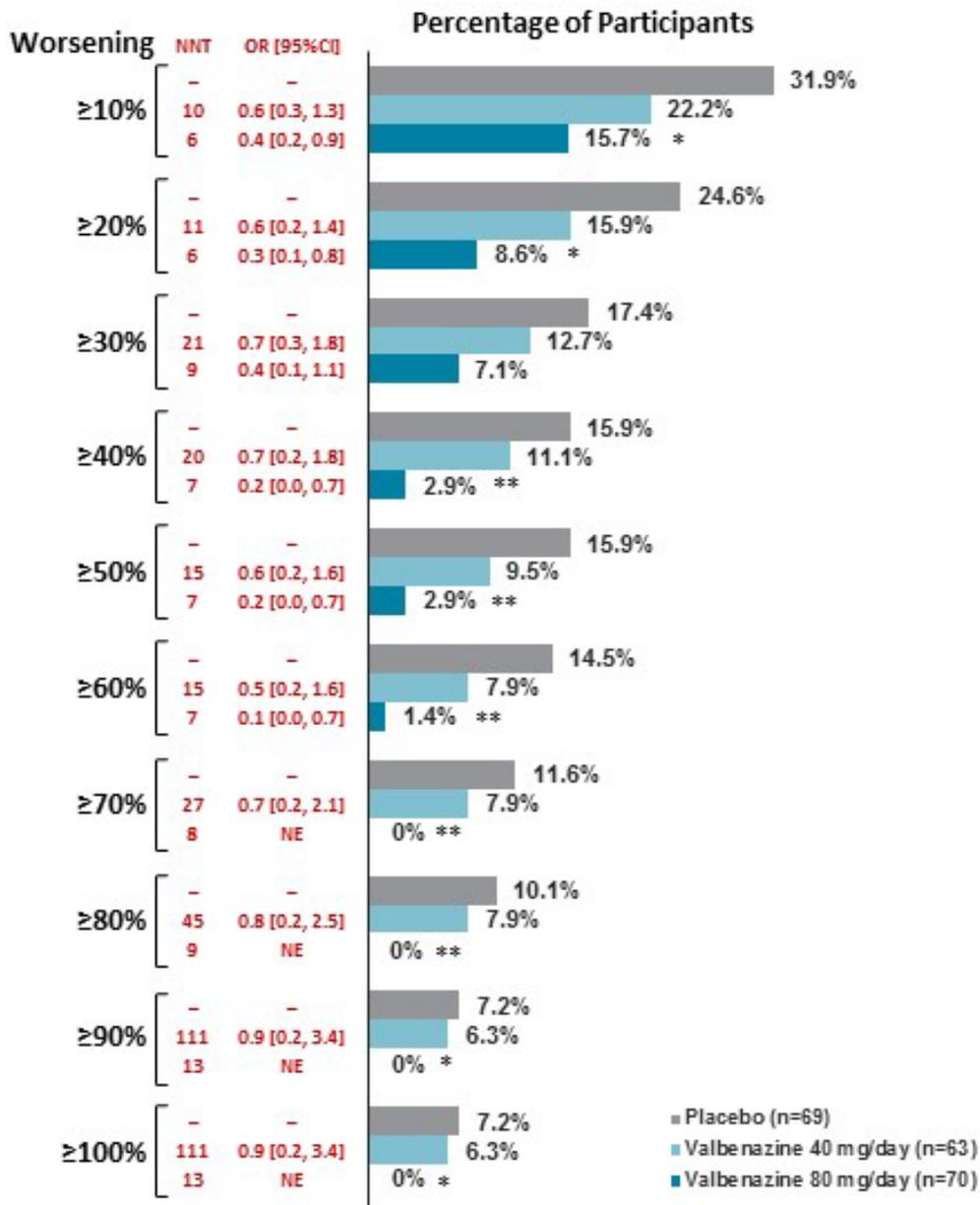
*P<0.05; **P≤0.01; ***P≤0.001 versus placebo.

AIMS, Abnormal Involuntary Movement Scale; NE, not estimable; NNT, number needed to treat for valbenazine to prevent at least 1 patient from worsening; OR, odds ratio.