2019

**Abnormal involuntary movement scale in tardive dyskinesia: Minimal clinically important difference**

M. Stacy

M. Sajatovic

J. M. Kane

*Zucker School of Medicine at Hofstra/Northwell, jkane2@northwell.edu*

A. J. Cutler

G. S. Liang

*See next page for additional authors*

---

Follow this and additional works at: [https://academicworks.medicine.hofstra.edu/articles](https://academicworks.medicine.hofstra.edu/articles)

Part of the Psychiatry Commons

---

**Recommended Citation**

Stacy M, Sajatovic M, Kane JM, Cutler AJ, Liang GS, O'Brien CF, Correll CU. Abnormal involuntary movement scale in tardive dyskinesia: Minimal clinically important difference. 2019 Jan 01; 34(8):Article 5738 [p.]. Available from: [https://academicworks.medicine.hofstra.edu/articles/5738](https://academicworks.medicine.hofstra.edu/articles/5738). Free full text article.

---

This Article is brought to you for free and open access by Donald and Barbara Zucker School of Medicine Academic Works. It has been accepted for inclusion in Journal Articles by an authorized administrator of Donald and Barbara Zucker School of Medicine Academic Works. For more information, please contact academicworks@hofstra.edu.
Authors

This article is available at Donald and Barbara Zucker School of Medicine Academic Works:
https://academicworks.medicine.hofstra.edu/articles/5738
ABSTRACT: Background: A minimal clinically important difference has not been established for the Abnormal Involuntary Movement Scale in patients with tardive dyskinesia. Valbenazine is a vesicular monoamine transporter 2 inhibitor approved for the treatment of tardive dyskinesia in adults. Efficacy in randomized, double-blind, placebo-controlled trials was defined as the change from baseline in Abnormal Involuntary Movement Scale total score (sum of items 1-7).

Objectives: To estimate an minimal clinically important difference for the Abnormal Involuntary Movement Scale using valbenazine trial data and an anchor-based method.

Methods: Data were pooled from three 6-week double-blind, placebo-controlled trials: KINECT (NCT01688037), KINECT 2 (NCT01733121), and KINECT 3 (NCT02274558). Valbenazine doses were pooled for analyses as follows: “low dose,” which includes 40 or 50 mg/day; “high dose,” which includes 75 or 80 mg/day. Mean changes from baseline in Abnormal Involuntary Movement Scale total score were analyzed in all participants (valbenazine- and placebo-treated) with a Clinical Global Impression of Change-Tardive Dyskinesia or Patient Global Impression of Change of score of 1 (very much improved) to 3 (minimally improved).

Results: The least squares mean improvement from baseline to week 6 in Abnormal Involuntary Movement Scale total score was significantly greater with valbenazine (low dose: $-2.4$; high dose: $-3.2$; both, $P < 0.001$) versus placebo ($-0.7$). An minimal clinically important difference of 2 points was estimated based on least squares mean changes in Abnormal Involuntary Movement Scale total score in participants with a Clinical Global Impression of Change-Tardive Dyskinesia score $\leq 3$ at week 6 (mean change: $-2.2$; median change: $-2$) or Patient Global Impression of Change score $\leq 3$ at week 6 (mean change: $-2.0$; median change: $-2$).

Conclusions: Results from an anchor-based method indicate that a 2-point decrease in Abnormal Involuntary Movement Scale total score may be considered clinically important. © 2019 The Authors. Movement Disorders published by Wiley Periodicals, Inc. on behalf of International Parkinson and Movement Disorder Society.

Key Words: AIMS; clinical trial; MCID; tardive dyskinesia; valbenazine
Tardive dyskinesia (TD) is a hyperkinetic movement disorder that is associated with exposure to an antipsychotic or other dopamine receptor blocking agent (DRBA), such as metoclopramide.\textsuperscript{1,2} Despite the development and widespread use of second-generation antipsychotics, TD remains a relevant potential risk of DRBAs.\textsuperscript{3,6} Two medications, valbenazine and deutetrabenazine, are now approved for the treatment of TD in adults. The placebo-controlled clinical trials of these drugs had some differences in study design (e.g., treatment duration, eligibility criteria), but both used the Abnormal Involuntary Movement Scale (AIMS)\textsuperscript{7} to measure the presence, severity, and changes in TD. Results from the valbenazine and deutetrabenazine trials showed that both compounds had measurable and statistically significant benefits as assessed by mean changes in AIMS total score.\textsuperscript{8-12}

Although the mean change in AIMS total score is a current standard for evaluating efficacy in clinical trials, the implications of this outcome for everyday practice are unclear. Along with other analytical approaches (e.g., Cohen’s effect size, number needed to treat [NNT]), one way to estimate clinical relevance of recent TD trial results would be to identify a minimal clinically important difference (MCID) for the AIMS total score. Two approaches are generally used to estimate MCIDs: distribution-based, which relies on a standard deviation (SD) or standard error of the measurement; and anchor-based, which uses an external measure (e.g., 7-point global assessment scale) as an independent criterion for improvement.\textsuperscript{13-15} An MCID for the AIMS has not been established in patients with TD, possibly because of the lack of large, well-controlled, and prospectively designed studies in this population. With the completion of three randomized controlled trials with valbenazine, a data set is now available that includes AIMS results for >350 study participants. Moreover, this data set includes Clinical Global Impression of Change-Tardive Dyskinesia (CGI-TD) and Patient Global Impression of Change (PGIC) results, which may be appropriate anchor scales for estimating an MCID.

Stemming from a Tardive Dyskinesia Assessment Workshop (TD Workshop) that was convened in October 2016,\textsuperscript{16} the results presented in this report are intended to propose a clinically meaningful approach to understanding AIMS results in TD clinical trials. The TD Workshop participants agreed that multiple analytical approaches to interpreting AIMS data should be made available. The MCID estimates proposed in this study are intended to stand alone. However, as discussed in greater detail later, they are also part of a larger initiative by the TD Workshop participants to explore different types of clinically meaningful AIMS analyses.

Materials and Methods

Data and Assessments

Full data sets from three 6-week, randomized, double-blind, placebo-controlled trials (Fig. 1)\textsuperscript{8-10} were made available from the study sponsor (Neurocrine Biosciences, Inc., San Diego, CA) and analyzed post hoc by an independent statistician (Veristat, Inc., Southborough, MA). For efficacy analyses, valbenazine doses were pooled into two groups as follows: (1) “low dose,” which included participants who received 50 mg/day in KINET (including treatment with 100 mg/day for the first 2 weeks), 50 mg/day in KINET 2, or 40 mg/day in KINET 3; and (2) “high dose,” which included participants who received 75 mg/day in KINET 2 or 80 mg/day in KINET 3. Participants who received valbenazine 25 mg/day in KINET 2 were excluded from analyses. All treatments (valbenazine and placebo) were pooled for MCID analyses. Additional information regarding treatment and study participants are summarized in the Supporting Information Appendix.

Outcome measures included the AIMS, CGI-TD, and PGIC. AIMS total score was defined as the sum of items 1 to 7, which focus on severity of abnormal movements in different body regions. Scoring for each of these seven items (range, 0 = none to 4 = severe) was based on the consensus of two central AIMS video raters (movement disorder specialists) who were blinded to treatment and study visit (baseline and weeks 2, 4, and 6). Scoring for the CGI-TD (range, 1 = very much improved to 7 = very much worsened) was based on clinical evaluation by the site investigator. Scoring for the PGIC (range, 1 = very much improved to 7 = very much worsened) was based on self-report by the study participant.

Statistical Analyses

All analyses were conducted in the pooled intent-to-treat (ITT) population, defined as participants who received...
AIMS total score at baseline

Mean (SD) 30.5 (7.6) 31.6 (7.9) 28.9 (6.8)

Concomitant use of antipsychotics, n (%)

Any antipsychotic 130 (82.3) 102 (89.5) 77 (76.2)
Atypical only 102 (78.5) 77 (75.5) 63 (81.8)
Typical only or both 28 (21.5) 25 (24.5) 14 (18.2)

Aims MCID in Tardive Dyskinesia

Results

In the pooled ITT population, baseline characteristics were generally similar across treatment groups (Table 1). Mean improvements from baseline to week 6 in AIMS total score were significantly greater in both valbenazine dose groups than in the placebo group (Fig. 2). Least squares mean differences from baseline were −1.7 and −2.6 in the pooled low- and high-dose valbenazine groups, respectively. The percentage of participants with minimal or better CGI-TD improvement (score ≤3 at week 6) was significantly higher with valbenazine high dose versus placebo (Fig. 3). Both valbenazine doses were found to have a significantly greater percentage of participants meeting the more rigorous response definition of “much improved” or “very much improved” (score ≤2). No statistical significance between valbenazine and placebo was found for either PGIC response analysis (Supporting Information Appendix; Supporting Information Table S1).

MCID Estimation

Based on participants with a CGI-TD score ≤3 at week 6 (Fig. 4A), the estimated MCID for AIMS total score was 2 points. The mean change from baseline (± standard error of the mean [SEM]) in AIMS total score was −2.2 (±0.2) and the median change was −2 (range, −13 to 8). These changes corresponded to a mean percent improvement of 17.2% (±3.5%).

Based on participants with a CGI-TD score ≤2 at week 6 (Fig. 4B), the estimated MCID for AIMS total score was 3 points. The mean change from baseline in AIMS total score was −3.4 (±0.4), corresponding to a percent improvement of 31.0% (±5.3%). The median score change was −3 (range, −13 to 8).

Table 1. Baseline characteristics (pooled ITT population)

<table>
<thead>
<tr>
<th></th>
<th>Placebo (n = 158)</th>
<th>Valbenazine Low Dose* (n = 114)</th>
<th>Valbenazine High Dose** (n = 101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD), years</td>
<td>55.8 (10.1)</td>
<td>54.9 (9.1)</td>
<td>56.2 (10.4)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>89 (56.3)</td>
<td>72 (63.2)</td>
<td>55 (54.5)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>86 (54.4)</td>
<td>64 (56.1)</td>
<td>62 (61.4)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>63 (39.9)</td>
<td>44 (38.6)</td>
<td>36 (35.6)</td>
</tr>
<tr>
<td>Psychiatric diagnosis group, n (%)</td>
<td>116 (73.4)</td>
<td>90 (78.9)</td>
<td>61 (60.4)</td>
</tr>
<tr>
<td>Schizophrenia/schizoaffective disorder</td>
<td>42 (26.6)</td>
<td>24 (21.1)</td>
<td>40 (39.6)</td>
</tr>
<tr>
<td>Mood disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concomitant use of antipsychotics, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any antipsychotic</td>
<td>130 (82.3)</td>
<td>102 (89.5)</td>
<td>77 (76.2)</td>
</tr>
<tr>
<td>Atypical only</td>
<td>102 (78.5)</td>
<td>77 (75.5)</td>
<td>63 (81.8)</td>
</tr>
<tr>
<td>Typical only or both</td>
<td>28 (21.5)</td>
<td>25 (24.5)</td>
<td>14 (18.2)</td>
</tr>
<tr>
<td>BPRS score at screening, mean (SD)</td>
<td>30.5 (7.6)</td>
<td>31.6 (7.9)</td>
<td>28.9 (6.8)</td>
</tr>
<tr>
<td>AIMS total score at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>8.9 (4.4)</td>
<td>9.0 (4.2)</td>
<td>9.5 (3.6)</td>
</tr>
<tr>
<td>Median (minimum, maximum)</td>
<td>8 (1, 26)</td>
<td>9 (0, 20)</td>
<td>9 (3, 20)</td>
</tr>
</tbody>
</table>

*Includes participants who received valbenazine 40 or 50 mg/day.
**Includes participants who received valbenazine 75 or 80 mg/day.
Similar to results for CGI-TD score ≤ 3, analyses based on PGIC score ≤ 3 yielded an MCID estimation of 2 points, with a median 20% improvement from baseline (Supporting Information Appendix; Supporting Information Table S1). Analyses based on PGIC score ≤ 2 also yielded an MCID estimation of 2 points (compared to 3 points for CGI-TD score ≤ 3), with a median 30% total score improvement from baseline.

Discussion

Although the AIMS total score is the current standard for determining efficacy in TD clinical trials, translating this outcome into clinical practice can be challenging.\(^{16}\) To address that challenge, the TD Workshop participants discussed different ways to analyze AIMS data and identified the MCID as one possible approach.\(^{17}\) Based on both clinician- and patient-rated anchors of minimal improvement (CGI-TD and PGIC score ≤ 3 at week 6), mean and median changes in AIMS total score (sum of items 1-7) suggested an MCID of 2 points in adults with TD. Analyses based on more rigorous definitions of global improvement (CGI-TD and PGIC score ≤ 2 at week 6) suggested a clinician-based MCID of 3 points and a patient-based MCID of 2 to 3 points. Clinically, these proposed MCIDs may be useful for interpreting the effects of treatment on TD. However, it may be worth noting that the MCID of 2 points is consistent with the distribution-based approach that uses 0.5 times the baseline SD as a threshold for clinically meaningful change.\(^{18}\) In the pooled data set, the SD of the mean AIMS total score at baseline in all participants was 4.2, which would correspond to an MCID of 2 points.

The current results were consistent with preliminary MCID analyses, which only included CGI-TD anchors.\(^{19,20}\) PGIC anchors were added to the current analyses to address the need for more patient-reported outcomes in TD studies. Given that patients with TD can be unaware of their movements,\(^{21}\) these PGIC-based results should be interpreted with some caution. However, consistent with the CGI-TD results, MCID estimates based on patient-reported improvements suggest that a 2- to 3-point decrease in AIMS total score may be considered clinically meaningful. It should also be noted that both anchor-based methods (CGI-TD and PGIC) included placebo responders to lessen the risk of the MCID being specific to valbenazine treatment. Additional MCID analyses based on data from other TD clinical trials (e.g., deutetrabenazine) would help to further

**FIG. 2.** AIMS total score mean change from baseline to week 6. ***P < 0.001 versus placebo.

**FIG. 3.** CGI-TD response at week 6. *P < 0.05; **P < 0.01; ***P < 0.001 versus placebo. CI, confidence interval.
establish whether an AIMS MCID of 2 to 3 points is applicable to different TD therapies.

An additional goal of the current analysis was to include percentage-based MCIDs for the AIMS total score. Participants with a CGI-TD or PGIC score ≤ 3 at week 6 had approximately 20% improvement from baseline in AIMS total score. Participants with a CGI-TD or PGIC score ≤ 2 had approximately 30% to 40% improvement from baseline in AIMS total score. These percentages are consistent with earlier TD studies that defined response as a ≥ 30% improvement in AIMS total score. They are also consistent with results from the companion piece to this article, which presents a full range of AIMS total score responses (≥ 10% to ≥ 90% improvement from baseline to week 6). They are also consistent with results from the companion piece to this article, which presents a full range of AIMS total score responses (≥ 10% to ≥ 90% improvement from baseline to week 6). They are also consistent with results from the companion piece to this article, which presents a full range of AIMS total score responses (≥ 10% to ≥ 90% improvement from baseline to week 6).

In that analysis, the percentage of participants who achieved a ≥ 20%, ≥ 30%, or ≥ 40% AIMS total score response was significantly higher with valbenazine high dose versus placebo. These results were clinically meaningful, as indicated by ORs for response (OR ≥ 4 for valbenazine vs. placebo) and NNTs (of 3 or 4). In the valbenazine clinical trials, AIMS response was defined a priori as ≥ 50% total score improvement, which is more stringent than the 20% to 40% MCID-based results in the current analysis. Therefore, a greater percentage of patients experienced a clinical benefit in the valbenazine clinical trials than the published ≥ 50% response analyses would imply.

A number of limitations should be noted. First, all analyses were conducted post hoc. None of the valbenazine trials were designed for estimation of an MCID, and the pooled valbenazine dose groups included participants who received slightly different low doses (40 and 50 mg/day) and high doses (75 and 80 mg/day). Second, results of the analyses may not be generalizable to all patients with TD. The trials primarily included psychiatric patients who were exposed to antipsychotic medications, and MCIDs may be different in nonpsychiatric patients who were exposed to an antiemetic (e.g., metoclopramide) or other DRBA. Study participants were also required to be psychiatrically stable, which may not always be true in real-life settings. In addition, participants in the valbenazine studies were required to have moderate or severe TD based on the

FIG. 4. Estimation of AIMS MCID. Based on all participants who met CGI-TD response criteria regardless of treatment (valbenazine or placebo).
The MCID analyses were conducted without considering the AIMS total score at baseline. Nor did they consider AIMS item scores at baseline, which provide more specific information about the location and severity of dyskinetic movements. To address some of these issues, shift analyses based on AIMS item scores were included in the companion piece to this article.17

Limitations of the AIMS itself should also be considered. Given that the AIMS total score is the current “gold standard” for evaluating efficacy in TD clinical trials, determining an MCID based on this measure is a reasonable endeavor. However, the AIMS does not capture the social and functional deficits associated with TD. In addition, one-time or episodic complications related to TD, such as a fall related to gait problems, are not adequately captured by the AIMS. Improvements in these domains must be considered along with dyskinetic movements when determining whether a patient is experiencing clinically meaningful improvements. Methodologies for administering and scoring the AIMS should also be considered. The proposed MCIDs presented in this report are based on AIMS evaluations that were scored by consensus between two central video raters (movement disorder specialists) who were blinded to treatment and study visit. In clinical settings, the AIMS is administered and assessed in real time by a physician or other qualified professional who knows what the patient is taking and how long he or she has been treated. Therefore, an MCID based on clinical trial data, as investigated in this report, should be considered as more of a guideline (rather than an imperative) for everyday practice. Given that the analyses in this report are limited to valbenazine data, they may not be generalizable to all AIMS results, including those that have been reported in other TD clinical trials (e.g., deutetrabenazine). Applying the proposed MCIDs from this report to other TD trials should also be done with caution given that differences in study design (e.g., double-blind vs. open-label, treatment duration, eligibility criteria, and allowance of concomitant medications) may affect treatment outcomes.

Finally, as previously published,15 the limitations of anchor-based methods should be mentioned. First, different anchors may result in different MCIDs, although the current analysis showed consistency between clinician-based (CGI-TD) and patient-based (PGIC) anchors. Moreover, MCIDs from both anchor types were consistent with a commonly used distribution method (i.e., 0.5 times the SD). In addition, anchors can be susceptible to recall bias, and inter-rater agreement was not tested for clinicians or study participants. Despite these various limitations, the TD Workshop participants agreed that the AIMS MCID can be an important advancement for clinicians who treat patients with TD. Taken in conjunction with other types of analyses (e.g., placebo-corrected mean change, effect size, treatment response, and NNT), or even added prospectively to statistical analysis plans, the MCID might help translate trial data into clinically meaningful information. Based on both clinician- and patient-rated anchors, the results of this analysis suggest that a 2-point decrease in AIMS total score may be considered an MCID if minimal improvement is the treatment goal; a 3-point decrease may be the MCID if more robust improvement is desired. Much more research is needed to understand the impact of TD on patients and caregivers, including the benchmarks of physical, functional, and social improvements that constitute a truly meaningful clinical difference. ■

Acknowledgments: The manuscript was critically reviewed by Scott Siegert, PharmD; Khodayar Farahmand, PharmD, and Charles Yoon, PharmD, all of whom are full-time employees of Neurocrine Biosciences, Inc. (San Diego, CA). Medical writing and editorial support were provided by Mildred Bahn, MA, at Prescott Medical Communications Group (Chicago, IL) with support from Neurocrine Biosciences, Inc.

References

13. Doganay Erdogan B, Leung YY, Pohl C, Tennant A, Conaghan PG. Minimal clinically important difference as applied in rheumatology:


Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site.