

# Quantitative and Qualitative Exploration of Meaningful Change on the Vineland Adaptive Behavior Scales (Vineland™-II) in Children and Adolescents with Autism Without Intellectual Disability Following Participation in a Clinical Trial

Susanne Clinch<sup>1</sup>, Stacie Hudgens<sup>2</sup>, Elizabeth Gibbons<sup>3</sup>, Tom Willgoss<sup>1</sup>, Janice Smith<sup>4</sup>, Ela Polek<sup>5</sup>, Claire Burbridge<sup>3</sup>

<sup>1</sup>Patient Centered Outcomes Research, Roche Products Ltd, Welwyn Garden City, Hertfordshire, UK; <sup>2</sup>Quantitative Sciences, Clinical Outcomes Solutions, LLC, Tucson, AZ, USA; <sup>3</sup>Clinical Outcome Assessments, Clinical Outcomes Solutions Ltd, Folkestone, Kent, UK; <sup>4</sup>Global Development, Roche Products Ltd, Welwyn Garden City, Hertfordshire, UK; <sup>5</sup>Quantitative Sciences, Clinical Outcomes Solutions Ltd, Folkestone, Kent, UK

Correspondence: Susanne Clinch, Patient Centered Outcomes Research, Roche Products Ltd, 6 Falcon Way, Shire Park, Welwyn Garden City, Hertfordshire, AL71TW, UK, Email [susanne.clinch@roche.com](mailto:susanne.clinch@roche.com)

**Purpose:** The Vineland™ Adaptive Behavior Scale is often used in autism spectrum disorder (ASD) trials. The Adaptive Behavior Composite Score (VABS-ABC) is the standardized overall score (the average of the Socialization, Communication and Daily Living skills domains), and the standardized 2-Domain Composite Score (VABS-2DC) is a novel outcome measure (average of the Socialization and Communication domains). A within-person meaningful change threshold (MCT) has not been established for the VABS-2DC. This paper presents a quantitative and qualitative interpretation of what constitutes a meaningful change in these scores to individuals with ASD without Intellectual Disability (ID; IQ≥70) and their families, as reported by their study partners (SPs).

**Participants and Methods:** Data were obtained from the aVlation clinical trial in children and adolescents with ASD and associated exit interviews. The intent-to-treat (ITT) clinical trial population included 308 individuals with autism (85.4% male; average age: 12.4 years [standard deviation (SD)=2.97]); 124 in the child cohort (aged 5 to 12 years; average age: 9.4 years [SD=1.86]), and 184 in the adolescent cohort (aged 13 to 17 years; average age: 14.5 years [SD=1.39]). Study partners of 86 trial participants were included in the Exit Interview Population (EIP): participants represented were 83.7% male, average age: 12.3 years [SD=2.98]). Anchor and distribution-based methods were used to estimate within-person change to support a responder definition, to aid interpretation of the clinical trial data; qualitative data were used to contextualize the meaning of changes observed.

**Results:** A within-person MCT range of 4 to 8 points was proposed for both VABS-ABC and VABS-2DC, which was associated with at least a 1-point improvement on 4 different anchors. Evidence for this within-person MCT was further supported by qualitative data, which suggested any change was considered meaningful to the individual with ASD, as reported by their SP, no matter what the magnitude.

**Conclusion:** A change in standardized score of 4 to 8 points constitutes a within-person MCT on both VABS-ABC and novel VABS-2DC in those with ASD and no ID. A change of this, or more, was reported by the SPs in this trial to be meaningful and highly impactful upon the individuals with ASD and their family.

**Keywords:** Autism, Vineland™-II, meaningful change threshold, QOL, MCID

## Introduction

Autism spectrum disorder (ASD) describes a neurodevelopmental condition which appears early on in childhood. While ASD is heterogeneous, it is characterized by difficulties in communication and socialization, with some individuals also experiencing repetitive sensory motor behaviors.<sup>1</sup> Children with ASD may struggle to maintain eye contact, have

difficulty following social conventions or the flow of a conversation, exhibit repetitive behaviors, and often have unusual or overly focused interests.<sup>1,2</sup>

Globally, the prevalence of ASD is estimated to be 7.6 per 1000 based on systematically reviewed epidemiological data,<sup>3</sup> but these estimates may be affected by the lack of accurate screening and diagnostic methods in some parts of the world.<sup>4</sup> Although the Food and Drug Administration (FDA) has approved the use of atypical antipsychotics risperidone and aripiprazole for treating irritability associated with autism; thus far, no medication has proved effective in improving the core challenges faced in ASD related to socialization and communication, as well as daily living skills. There is an unmet need to develop therapies that provide effective, long-term improvements in these core aspects of ASD.

Clinical outcome assessments (COAs) are important tools for capturing severity and impacts of ASD on the individuals and their family. Specifically, patient-reported outcome assessments and observer-reported outcome (ObsRO) assessments can be used to enable a personal, participant perspective to be included within clinical trials and drug evaluation.

The Vineland Adaptive Behavior Scales, Second Edition (Vineland<sup>TM</sup>-II),<sup>5</sup> recently updated to Vineland<sup>TM</sup>-3, is a commonly used COA in ASD used to measure adaptive behavior within the core domains of Socialization, Communication, and Daily Living Skills.<sup>6</sup> The Vineland<sup>TM</sup>-II interview form is a clinician assessment based on a semi-structured interview between an ASD clinician/therapist and a caregiver of the individual with ASD. The standardized Vineland<sup>TM</sup>-II Adaptive Behavior Composite Score (VABS-ABC) is the overall score, which is the average of the 3 core domains, and the standardized 2 Domain Composite Score (VABS-2DC),<sup>7</sup> is a novel outcome assessment, which consists of an average of the Socialization and Communication domains; 2 core symptoms of ASD. The VABS-2DC was the primary endpoint and VABS-ABC a secondary endpoint used in the aVlation clinical trial, which was a Phase 2 multicenter, randomized, double-blind, placebo-controlled trial investigating the efficacy and safety of balovaptan in children and adolescents aged 5 to 17 years with ASD, who had no intellectual disability (ID; IQ $\geq$ 70).<sup>8</sup> In addition, exit interviews were incorporated as part of the trial design to explore qualitatively any changes clinical trial participants (and their families) experienced over the course of the clinical trial and what impact those changes had on their daily lives and health-related quality of life (reported elsewhere).<sup>9</sup>

Research exploring adaptive functioning in ASD using the Vineland<sup>TM</sup>-II has demonstrated that this can vary across individuals ASD and can be related to other factors. One recent study found that those with higher severity levels of ASD had less adaptive functioning, and that this was associated with higher levels of internalizing problems and greater parental stress.<sup>10</sup> Another study showed that IQ was the biggest predictor of adaptive functioning, although a negative relationship with age was also observed.<sup>11</sup>

While work has been conducted, using distribution and anchor-based methods to understand minimal clinically important differences (MCIDs) on the VABS-ABC,<sup>5</sup> which identified a range from 2 to 3.75 point change, it has not been established for the VABS-2DC and there is sparse literature on what a within-person change on the Vineland<sup>TM</sup>-II means in real life for individuals with ASD and their families. This is critical in order to interpret the change reported on this outcome measure.

The meaningful change threshold (MCT) is a term that was coined to describe the threshold of change at which the change becomes meaningful to the individual.<sup>12–14</sup> The methods focus on defining the within-person MCT, which is the difference in an individual's change scores between baseline and a subsequent time point, which individuals perceive as a meaningful improvement or deterioration in their health-related quality of life.<sup>2</sup> This is distinct from previously used terms such as the MCID or the minimal important difference (MID), defined as the minimal change in the score that is meaningful for individuals based on the mean difference between a “no change” group and a “minimally worsened” group; and the minimal detectable change, defined as the minimal change that falls outside the measurement error. These earlier terms reflect estimates that focus on group-level meaningful difference rather than individual-level meaningful change; a distinction that was drawn by Cappelleri and Bushmankin<sup>15</sup> and of particular importance to the FDA.<sup>2,16</sup>

The between-groups MCID and within-person MCT are estimated using blinded clinical trial data that combines the treatment arms. Once unblinding occurs, the within-person MCT can be used as a responder definition in a responder analysis to compare the proportion of individuals who experienced a meaningful change between the treatment and placebo (control) groups, while the MCID can be used to determine whether mean group differences are clinically

meaningful. Several methods can be undertaken to calculate the MCID and within-person MCT that incorporate both anchor- and distribution-based approaches. Both MCID and within-person MCTs are dependent on the context of use, so thresholds calculated can differ between and within indications.<sup>2,16</sup> Furthermore, thresholds can be different for improvement and deterioration.<sup>17</sup>

The use of anchor-based analyses, supplemented by empirical cumulative distribution function (CDF), and supported by qualitative insights, is recommended by the FDA.<sup>16</sup> Anchor-based approaches can focus on an individual- or group-level change and utilize explicit indicators of the meaningful change as assessed usually by single-item global impression questions (global severity and global improvement) that are intuitively interpretable; although items from more complex COA measures can also be used as anchors, providing they meet the required criteria for a suitable anchor.<sup>12,18</sup> A distribution-based method focuses on a group-level change required to exceed intrinsic variability within the population attributed to the measurement error and, as such, are recommended as a complimentary method of exploring meaningful change. In addition to this, qualitative data from interviews with clinical trial participants provide more in-depth information that can contextualize the changes observed, and thus, inform interpretation of meaningful change.<sup>19</sup> Although regulatory agencies highly recommend the use of mixed methods in research, it is still a hugely under-utilized methodology in meaningful change studies.<sup>20</sup>

The objective of this study was thus to address the limited data on a within-person MCT for the Vineland<sup>TM</sup>-II, and an understanding of what such a change means in real-life for individuals with ASD and their families, by exploring meaningful change estimates on the VABS-2DC and VABS-ABC for individuals with ASD and no ID, using a mixture of quantitative (anchor- and distribution-based) and qualitative (thematic analysis of exit interviews) methodology using data from the aVlation ASD clinical trial. This is to help the interpretation of data from this outcome measure in a clinical trial setting.

## Methods

The overall approach of this study was to explore clinical trial instruments and exit interview data to identify those that could be used to derive anchors for estimating MCT on the VABS-ABC and VABS-2DC standardized scores in individuals with ASD with no ID. Data from suitable anchors, and data derived from distribution-based methods, were triangulated to determine MCTs. The qualitative data from the exit interviews was then reviewed to support the agreed MCT and to contextualize the changes observed.

## Clinical Trial Setting

The aVlation clinical trial (NCT02901431)<sup>8</sup> was a randomized, double-blind, 24-week, parallel-group, placebo-controlled Phase 2 trial in individuals with ASD.<sup>8</sup> The primary outcome for this trial was the change from baseline on the standardized VABS-2DC at 24 weeks; secondary outcomes included change from baseline at 24 weeks on standardized Vineland<sup>TM</sup>-II Socialization (VABS-S), Communication (VABS-C), Daily Living Skills (VABS-DLS) domains, and Adaptive Behavior Composite Score (VABS-ABC) (Socialization, Communication and Daily Living domains total score). Although the clinical trial failed to meet the primary endpoint, changes were observed in some individuals.

Ethical approval and informed consent to participate was obtained for all participants for this research as part of the clinical trial. Participants were, in the first instance, randomized in a ratio of 1:1:1 into the 3-arms (placebo, 4 mg adult-equivalent, 10 mg adult-equivalent); however, because of underexposure of the study medication in lower-dose participants in the first weeks of the aVlation trial, the design of the trial was altered to only 2-arms (placebo versus balovaptan 10 mg). The population was stratified by age group and sex, with females limited to 20% of the sample. Study visits occurred at screening, baseline (randomization visit, day 1), every 2 weeks through to week 24, plus follow-up visits at weeks 26 and 30 for participants who did not transition into the open-label extension or upon early withdrawal.

Exit interviews were conducted with the caregivers of a subsample of participants from the aVlation trial within 4 weeks from the Week 24 visit via telephone by a trained qualitative researcher. The caregiver interviewed was the individual who accompanied the individual with ASD during the clinical trial; here, they are referred to as study partners (SPs). The SP was asked to provide informed consent to take part in the interview as part of the clinical trial enrolment.

## Inclusion/Exclusion Criteria

Full inclusion and exclusion criteria for the aVlotion trial can be found in the [Supplemental Material](#). Key inclusion criteria included Clinician Global Impression of Severity (CGI-S) score  $\geq 4$  (moderately ill), Social Responsiveness Scale, second edition *T* score  $\geq 66$ , and IQ  $\geq 70$  as assessed by Wechsler Abbreviated Scale of Intelligence, 2nd Edition. Key exclusion criteria for participants were major changes in psychosocial intervention within 4 weeks prior to screening; unstable or uncontrolled clinically significant psychiatric and/or neurologic disorder; and/or suicidal behavior.

The trial focused on cognitively high functioning individuals with ASD (IQ  $\geq 70$ , ie, those without ID) as these individuals represent a more homogeneous population, facilitating signal detection and it was expected that this population may benefit more from a therapy improving social and communication aspects of ASD. Those with significant social and communication deficits were recruited to ensure room for potential treatment-mediated improvements.

There were no additional inclusion/exclusion criteria for the EIP, beyond meeting the requirements of SP, the participant reaching end of treatment, and consent. All SPs were invited to take part and interviews were scheduled with SPs in the child and adolescent cohorts as clinical trial participants reached the end of trial, until the required number of interviews were conducted. Recruitment was monitoring against target quotas for representation: In line with the clinical trial protocol, the proportion of interviewed caregivers of females within each cohort did not exceed 20% of the sample; it was targeted to obtain 15% of SPs per age group as follows: 5 to 7 years, 8 to 12 years, 13 to 15 years, and 16 to 17 years.

## Clinical Trial Sample

There were 308 individuals with ASD (85.4% male, average age: 12.4 years [SD=2.97]) in the ITT clinical trial population ([Table 1](#)); this included 124 individuals with ASD in the child cohort which had an age range of 5 to 12 years (average age: 9.4 years [SD=1.86]) and 184 individuals with ASD in the adolescent cohort which had an age range of 13 to 17 years (average age: 14.5 years [SD=1.39]). The SP of 86 trial participants were included in the Exit Interview Population (EIP): (participants represented were 83.7% male, average age: 12.3 years [SD=2.98]). Overall, the ITT and EIP populations were very similar in terms of the percentage of male individuals with ASD, average age, ethnicity, race, and age at diagnosis and time from diagnosis. Due to the nature of the recruitment process, it was not recorded how many SPs declined consent.

**Table 1** Comparison of the ITT and EIP Populations

	Populations	
	ITT (n=308)	EIP (n=86)
% Male	85.4	83.7
Average age (SD)	12.4 (2.97)	12.3 (2.98)
<b>Age Ranges</b>		
% in age range 5–7 years	8.1	6.8
% in age range 8–12 years	38.4	33.5
% in age range 13–15 years	25.6	32.1
% in age range 16–17 years	27.9	27.6
<b>Ethnicity</b>		
% Hispanic (Latino)	12.3	14.0
% Not-Hispanic (Latino)	86.4	83.7

(Continued)

**Table I** (Continued).

	Populations	
	ITT (n=308)	EIP (n=86)
% Other	1.2	2.4
<b>Race</b>		
% White	81.2	81.4
% Black	4.5	7.0
% Other Race	14.3	11.6
<b>Age at diagnosis and time from diagnosis</b>		
Age (SD) in years	7.3 (4.24)	6.9 (4.01)
Time from diagnosis (SD) in months	68.1 (51.35)	70.1 (51.25)
<b>Average IQ (SD)</b>	98.4 (16.96)	99.1 (14.82)
<b>Scores at Baseline (SD)</b>		
Vineland-II 2DC	73.5 (10.41)	73.9 (9.78)
Vineland-II ABC	73.4 (10.57)	73.6 (9.69)
Vineland-II Socialization	72.6 (13.20)	71.7 (12.04)
Vineland-II Communication	74.4 (10.82)	76.1 (11.44)
Vineland-II Daily Living Skills	78.6 (14.39)	78.4 (13.81)

**Abbreviations:** ABC, Adaptive Behavior Composite; EIP, exit interview population; ITT, intention-to-treat; SD, standard deviation; 2DC, 2 domain composite.

## Outcome Measures

### VABS-II

A clinician carried out an interview with the child's SP to answer Vineland™-II<sup>21</sup> items; thus, the instrument is a hybrid of a ClinRO and ObsRO. The measure has 11 subdomains that together comprise 4 core adaptive behavior domains: VABS-S (made up of Interpersonal Relationships, Play and Leisure Time, and Coping Skills subdomains), VABS-C (made up of Receptive, Expressive, and Written subdomains), VABS-DLS (including Personal, Domestic, and Community subdomains), and Motor Skills. Domain scores were obtained for the core domains of Socialization, Communication, and Daily Living Skills. The 3 core domain scores (VABS-S, VABS-C, VABS-DLS) were used to calculate the overall VABS-ABC (total) score. The Motor Skills domain was not administered as part of the clinical trial. The Vineland™-II uses standardized algorithm-based scoring that allows for comparison of scores against norm data for a given age group. Standardized VABS-ABC scores range from 20 to 160. For every domain/total score, higher scores indicate better adaptive functioning. All items on the Vineland™-II are scored as 0 (the behavior is never performed or it never occurs without help), 1 (the behavior is sometimes performed without help or reminders) or 2 (the behavior usually occurs without help). Additional to the original core domain and overall scores, a VABS-2DC score was calculated (as a mean score of the Communication and Socialization domains),<sup>22</sup> which was developed and tested in the VANILLA Phase 2 clinical trial of balovaptan.<sup>23</sup> The Vineland™-II<sup>21</sup> was completed at baseline, Week 12 and Week 24 visits in the clinical trial, by the same rater/clinician and caregiver/SP at each visit.

### Instruments Explored to Derive Potential Anchors

The clinical trial instruments that were explored to derive potential anchors for the MCT analysis are as follows. Only a selection of these, those meeting the required criteria (outlined in the anchor-based analyses section), were used.

The CGI-S (prospective anchor) is a single-item ClinRO<sup>24</sup> based on clinician's assessment of the clinical trial participant's overall severity on a 7-point scale from 1 "Normal, not ill at all" to 7 "Among the most severely ill patients." Lower scores indicate less severe ASD symptoms. Data from baseline and week 24 (end of trial [EoT]) visits were utilized in this analysis. Change scores are derived by subtracting baseline severity scores from Week 24 scores. The derived change scores is categorical change on the anchor relative to continuous change on the VABS domain scores. Participants were classified by their change from baseline score on the CGI-S, calculated by subtracting their baseline anchor score from their current anchor score as follows: Marked improvement - 3-point decrease (-3); Improvement - 2-point decrease (-2); Minimal improvement - 1-point decrease (-1); No change - Equal to baseline (0); and Worsening - 1-point increase (1).

The Clinician Global Impression of Improvement (CGI-I; retrospective anchor) is a single-item ClinRO.<sup>25</sup> Assessing clinicians are required to retrospectively judge the clinical trial participant's overall improvement from baseline to the time of the current visit. Assessments are made on a 7-point scale from 1 "Very much improved" to 7 "Very much worse." Lower scores indicate less severe ASD symptoms. Data from week 24 (EoT) visit were utilized in this analysis.

The Caregiver Global Impression of Improvement (CaGI-I; retrospective anchor) is an ObsRO that requires the caregiver to retrospectively assess improvement of the child/adolescent on 4 items (which follow the domains of the VABS -II): overall ASD symptoms, Communication, Socialization, and Daily Living Skills. Responses to each of 4 items are provided on a 7-point rating scale ranging from 1 "Very much improved" to 7 "Very much worse." Lower scores on these items relate to greater improvement in ASD symptoms. Data from the week 24 (EoT) visit were utilized in this analysis.

The Ohio Autism Clinical Impression Scale (OACIS) is a ClinRO that separately measures both global impressions of severity (OACIS-S; prospective anchor) and improvement (OACIS-I; retrospective anchor).<sup>26</sup> Both the OACIS-I and OACIS-S include 4 one-item assessments: Overall autism, Verbal, Non-verbal, and Social Interactions. Each item has 7-response options. For the OACIS-S, responses range from 1 "Normal" to 7 "Among the most severe." For the OACIS-I, responses range from 1 "Very much improved" to 7 "Very much worse." Each of the response options is explained in greater detail on the form in order to aid the clinician's accuracy while making a judgment. Lower scores indicate better functioning. A higher score on the OACIS-S indicates greater severity; a lower score on the OACIS-I indicates greater improvement. Change scores on OACIS-S were calculated by subtracting the baseline score from week 24 scores, with the highest negative change score indicating the highest improvement, 0 - no change; and the highest positive change score indicating the worst deterioration.

### Exit Interview Data Explored to Derive Potential Anchors

Exit interviews were used to qualitatively explore meaningful change from the perspective of the individual with ASD and their family and to derive potential anchors for change. During interviews, SPs - who was the individual who had accompanied the individual with ASD during the course of the clinical trial - were asked to describe what, if any, changes they had noticed in their child/adolescent over the course of the clinical trial, focusing upon the 3 domains of Communication, Socialization, and Daily Living Skills (corresponding to the Vineland<sup>TM</sup>-II domains) and overall ASD. If the caregiver described a change, they were asked to expand upon what the change had meant to them and the individual with autism.<sup>9</sup>

Four study partner perception of meaningful change (SPPMC) ratings were obtained; one for each of the 3 specific domains and one for overall symptoms associated with ASD, for which ratings of change were given. Items were rated on a 7-point scale ranging from 1 "Very much improved" to 7 "Very much worse."

The Independent Clinical Rating of Change (InCRC) was obtained from an independent clinician review of the exit interviews transcripts with SPs using a Transcript Interpretation Rating Guide (TIRG), in which clinicians were asked to provide ratings of change for the 3 domains of Communication, Socialization, and Daily Living Skills (corresponding to the Vineland<sup>TM</sup>-II domains) and overall ASD on the same 7-point scale as the SPPMC. Twenty transcripts were randomly selected (n=10 from each cohort) and these were rated by an independent clinician using the TIRG. Overall, 10 clinicians provided ratings, with each clinician rating 4 of the 20 transcripts. Of these clinicians, 5 were psychiatrists, 2 neurologists, 2 psychologists, and a behavioral analyst, all from the United States.

## Quantitative and Qualitative Analyses

### Anchor-Based Analysis

Within-person MCTs were derived for the Vineland™-II domain scales and composite scores using associated anchors. The MCT analysis was explored in the following hypothesized order of importance: 1) CGI-I and CGI-S (clinical trial anchor); 2) CaGI-I (clinical trial anchor) and SPPMC (exit interview anchor), 3) InCRC (exit interview anchor), and 4) OACIS-S and I (clinical trial anchor).

The process for selection of the anchor and related threshold identification was stepwise, allowing for a more accurate selection of levels of change to be considered as a responder definition. The intent was to select a point on the change scale where individuals who improved are distinctly separated from those who are stable or worsened.

Prior to computing the MCT analysis, correlations between the change scores on the potential anchors and the Vineland™-II domain scores were evaluated. Polyserial correlations were employed due to the ordinal response data collected for anchors and the continuous data elicited on the Vineland™-II. Correlations between the Vineland™-II and anchor-endpoints change score correlations were explored at EoT. Correlation coefficient values  $\geq 0.35$  were considered as acceptable (with values  $\geq 0.40$  considered exceptional for interpretability),<sup>18,27,28</sup> guiding the decision to use an anchor in the MCT derivation.

For each anchor, each individual with ASD was classified into response groups based on their level of change on the anchor between baseline and EoT. For each response group, the mean (SD), median, and 95% confidence interval (CI) between baseline and EoT were calculated.<sup>29,30</sup>

The derivation of the within-person MCT for the Vineland™-II domains began by identifying the lowest improvement category on the anchor, followed by a comparison of the 95% CIs for this category with the “no change” group and the adjacent improvement levels groups, to identify the category with an estimate that did not overlap with the adjacent 95% CI.<sup>18</sup> The 95% CIs were used as this provides a more precise estimate of the threshold for meaningful change than simply considering the mean. The lowest adjacent improvement category was evaluated once the first 2 criteria were met, and an estimate was selected which was above the upper boundary of the lower anchor level 95% CI. In cases where the CIs for anchor categories were overlapping and/or the no-change group demonstrated statistically significant ( $p < 0.05$ ) and large differences in scores between baseline and EoT, the anchor was not used in the analysis, even if anchor-endpoint correlation met the threshold of 0.35. Cumulative distribution function curves were produced for each MCT analysis for the selected anchors. Absolute change from baseline in Vineland™-II domain/total score was expressed on the x-axis, and the cumulative number of participants who expressed a given score was presented on the y-axis.

### Distribution-Based Analysis

A distribution-based approach for defining changes beyond measurement error was used to support the within-person MCT and MCID estimated for the VABS-2DC and VABS-ABC. Specifically, the estimated MCT and MCID must be greater than measurement error. The distribution of the VABS-2DC and VABS-ABC was used to derive the distribution-based MCID estimates for individuals at baseline. This was calculated as 0.5 standard deviation (SD) at baseline and the standard error of measurement (SEM) was calculated using an intraclass correlation coefficient (ICC; correlation of scores between baseline and time 2 (week 12) for those reporting no change from the ITT with available data at each timepoint as the reliability estimate, using the ANOVA methodology<sup>14,21</sup>), where SEM is calculated as  $SEM = SD\sqrt{1 - ICC}$ .<sup>31</sup> 95% CI of SEM was calculated using the formula:  $MDC_{95} = SEM * 1.96 * \sqrt{2}$ .<sup>32,33</sup> Estimates equivalent to 0.2, 0.5, and 0.8 of SDs have also been provided.

While not acceptable as responder definitions in themselves, a distribution-based approach provides boundary information, as any value proposed as a responder definition will need to be at least as large as a distribution-based estimate to rule out the possibility of participants being classified as a responder by chance.

### Qualitative Analysis

The main objective of the interviews was to obtain insights about how SPs conceptualize meaningful change in the individual with ASD and how this meaningful change would demonstrate in the individual with ASD's and SP's daily lives. In addition to deriving the SPPMC, exit interviews were transcribed and subjected to thematic analysis to obtain a qualitative understanding of the impact and meaning of the changes observed. This in-depth analysis was conducted on blinded data and has

been reported elsewhere.<sup>9</sup> However, following the estimation of the within-person MCT on the VABS-2DC and VABS-ABC, the qualitative data (namely the SPPMC and the SP rating of whether the change was meaningful or not in the interviews) was revisited to explore the change reported according to those who met or did not meet the MCT. This analysis was conducted on the exit interview population (EIP) and used to support the interpretation of meaningful change.

## Results

### Anchor-Based Estimations

#### Identifying Appropriate Anchors

The correlation between a total of 22 potential anchors (outlined above) and the Vineland™-II endpoints in the trial was tested. Only 5 overall met the pre-defined requirement of endpoint-anchor correlation of  $\pm 0.35$  or above for the analysis of either the VABS-2DC or VABS-ABC, 4 of which were retrospective and 1 prospective (Table 2). For the VABS-2DC,

**Table 2** Longitudinal Correlations Between Endpoints and Their Hypothesized Anchors\*

Study Measures			VABS-2DC	VABS-ABC	VABS-C	VABS-S	VABS-DLS
Prospective Study Anchors	OACIS-S	Overall	<b>-0.376</b>	-0.339			
		Verbal	-0.272		-0.207		
		Non-verbal	-0.305		-0.216		
		Social interaction	-0.297			-0.302	
	CGI-S	CGI-S	-0.316	-0.342			
Retrospective Study Anchors	OACIS-I	Overall	<b>-0.391</b>	<b>-0.403</b>			
		Verbal	-0.321		-0.216		
		Non-verbal	<b>-0.426</b>		-0.346		
		Social interaction	-0.315			-0.308	
	CaGI-II	Overall	<b>-0.365</b>	<b>-0.353</b>			
		Communication	-0.339		-0.169		
		Socialization	-0.311			-0.331	
		Daily Living Skills					-0.227
	CGI-I	CGI-I	<b>-0.399</b>	<b>-0.397</b>			
Retrospective Exit Interview Anchors	SPPMC	Overall		-0.222			
		Communication	-0.192		-0.072		
		Socialization				-0.264	
		Daily Living Skills	-0.233		-0.181		
	InCRC	Overall		-0.179			
		Communication	-0.415		-0.143		
		Socialization				-0.388	
		Daily Living Skills	-0.287		0.042		

**Notes:** \*Correlations were computed only for the anchor-endpoint combinations which were hypothesized a priori. Correlations in bold are those beyond the cutoff value of  $\pm 0.35$ , warranting further examination of usefulness as an anchor in an MCT analysis. The sample size for the InCRC ratings was low and so not considered for MCT analyses.

**Abbreviations:** CaGI-I, Caregiver Global Impression of Improvement; CGI-I, Clinician Global Impression of Improvement; CGI-S, Clinician Global Impression of Severity; InCRC, Independent Clinical Rating of Change; OACIS-I, Ohio Autism Clinical Impression Scale of Improvement; OACIS-S, Ohio Autism Clinical Impression Scale of Severity; SPPMC, study partner perception of meaningful change; VABS-2DC, Vineland™-II 2 Domain Composite Score; VABS-ABC, Vineland™-II Adaptive Behavior Composite Score; VABS-C, Vineland™-II Communication Domain; VABS-DLS, Vineland™-II Daily Living Skills Domain; VABS-S, Vineland™-II Socialization Domain.

4 met the criterion: OACIS-S Overall, OACIS-I Overall, OACIS-I Non-verbal, and the CaGI-I Overall. For the VABS-ABC, 3 met the criterion: OACIS-I Overall, CaGI-Overall, and CGI-I. The analysis of a within-person MCT on the VABS-2DC and VABS-ABC standardized scores was therefore conducted using data from these anchors only (ITT population). No other anchors derived from the clinical trial or any from the exit interviews met the pre-defined criterion for an anchor-endpoint correlation and so were not used.

### Estimation of a Within-Person MCT for VABS-2DC

For each of the 4 anchors which met the pre-defined anchor-endpoint correlation criterion as above, improvements on the mean change in VABS-2DC scores were observed for each level of categorical improvement on the anchors. However, for 3 of these anchors (CaGI-I-Overall, OACIS-I Non-Verbal, and OACIS-S Overall), the observed within-group mean change in the “no change” anchor group was significant ( $p < 0.05$ ), thus suggesting suitability of these as anchors given anchor categories could be viewed as distinct and non-overlapping. The usefulness of OACIS-S Overall, was rather limited; while the correlation of this anchor met the acceptable criteria for inclusion, the variability observed suggested that all anchor-group categories had overlapping CIs.

Based on the results from these anchors, it was apparent that to arrive at a within-person MCT capable of discriminating between participants experiencing no change with those improving on VABS-2DC, the MCT value would need to be above the 4- to 8-point range (Table 3). This range considers the lower bound of the CI for subjects in the improvement category, where CIs were non-overlapping with the “no change” group as the estimate which is the most conservative and gives the most confidence. When the non-overlapping group represented improvement by 2 points, the upper bound of the CI for the minimal change group was also considered when this did not overlap with the no change CI. For example, for the OACIS-S Overall, the lower bound of the 95% CI for an improvement by 1 point was 3.8, which does not overlap with the upper bound of the 95% CI for the no-change group (which was 3.1). In the OACIS-I Overall, the lower bound of the 95% CI for the much improved group was 7.7, but the upper bound of the CI for the minimally improved group was 4.0, which also did not overlap with the 95% CI for the no-change group. Hence, a change of 4 or more, which is distinct from the no-change group, may be considered a meaningful within-person improvement.

**Table 3** Change in Vineland™-II - 2DC According to Response on Selected Anchors

Anchor	No-Change vs Improvement Groups	N	Mean (SD)	Median	95% CI of Mean
OACIS-S Overall	No Change	127	1.7 (8.09)	1.0	(0.3, 3.1)
	Improvement by 1 point (-1)	72	5.9 (9.00)	4.0	(3.8, 8.0)
	Improvements by 2 points (-2)	22	9.4 (9.52)	8.5	(5.1, 13.6)
OACIS-I Overall	No Change	90	1.3 (8.31)	0.5	(-0.5, 3.0)
	Minimally Improved	84	2.3 (7.65)	2.0	(0.6, 4.0)
	Much Improved	55	10.1 (8.87)	8.0	(7.7, 12.5)
OACIS-I Non-verbal	No Change	138	2.0 (7.63)	1.3	(0.7, 3.3)
	Minimally Improved	65	4.3 (8.20)	3.5	(2.3, 6.4)
	Much Improved	27	11.9 (9.92)	12.0	(8.0, 15.9)
CaGI-I Overall	No Change	62	2.3 (6.89)	1.5	(0.6, 4.1)
	Minimally Improved	91	1.8 (7.28)	1.0	(0.3, 3.3)
	Much Improved	68	8.4 (9.74)	7.3	(6.0, 10.7)

**Abbreviations:** CaGI-I, Caregiver Global Impression of Improvement; OACIS-I, Ohio Autism Clinical Impression Scale of Improvement; OACIS-S, Ohio Autism Clinical Impression Scale of Severity; SD, standard deviation.

For each anchor, the 1-category improvement groups each attained a significant within-group improvement ( $p < 0.05$ ; not shown). Based on the lower CI for the mean of the improvement category (non-overlapping with no-change group) using the anchors OACIS-S Overall, OACIS-I Overall, OACIS-I Non-verbal, and CaGI-I Overall, an improvement around 3.8, 7.7, 8.0, and 6.0 points on the endpoint, respectively, suggested a meaningful improvement, and based on the top level of the CI for the lower category, this was 3.8, 4.0, 6.4 and 4.1 (Table 3). Evaluation across each of the above 4 anchors suggests that a change in the range of 4 to 8 points on the VABS-2DC score results in a meaningful improvement on the VABS-2DC as well as meaningful improvement for each anchor.

### Estimation of a Within-Person MCT for VABS-ABC

For each of the 3 anchors that met the aforementioned  $\geq 0.35$  correlation criterion for the VABS-ABC there was overlap in the “no change” and “minimal change” thresholds, and so the threshold was within the “Much improved” category (Table 4); this increased the threshold to a value beyond what is believed to be a lower value of meaningfulness. Based on these 3 anchors (OACIS-I Overall, CaGI-I-Overall, and CGI-I), an improvement around 7.6, 6.1, and 5.9 points on the VABS-ABC, respectively, if looking at the lower CI for the mean of the improvement category (non-overlapping with no-change group), or 3.6 for all if considering the upper CI of the lower category, suggested a meaningful improvement (Table 4). In sum, a within-person MCT of 4 to 8 points can be considered for this measure.

### Cumulative Distribution Functions

Supportive to this interpretation of the VABS-2DC and VABS-ABC MCT results, a visual inspection of the CDF curves revealed adequate separation between the 1-category improvement and the no-change curves (see an example of the CDF curve for VABS-2DC in Figure 1 and see Figures S1 and S2 in the Supplement for other CDF curves, including those for VABS-ABC). Therefore, utilizing at least a 1-category improvement, even using the upper bound of the 95% CI on the anchor to define the MCT would be appropriate, resulting in correct classification of individuals with autism as responders.

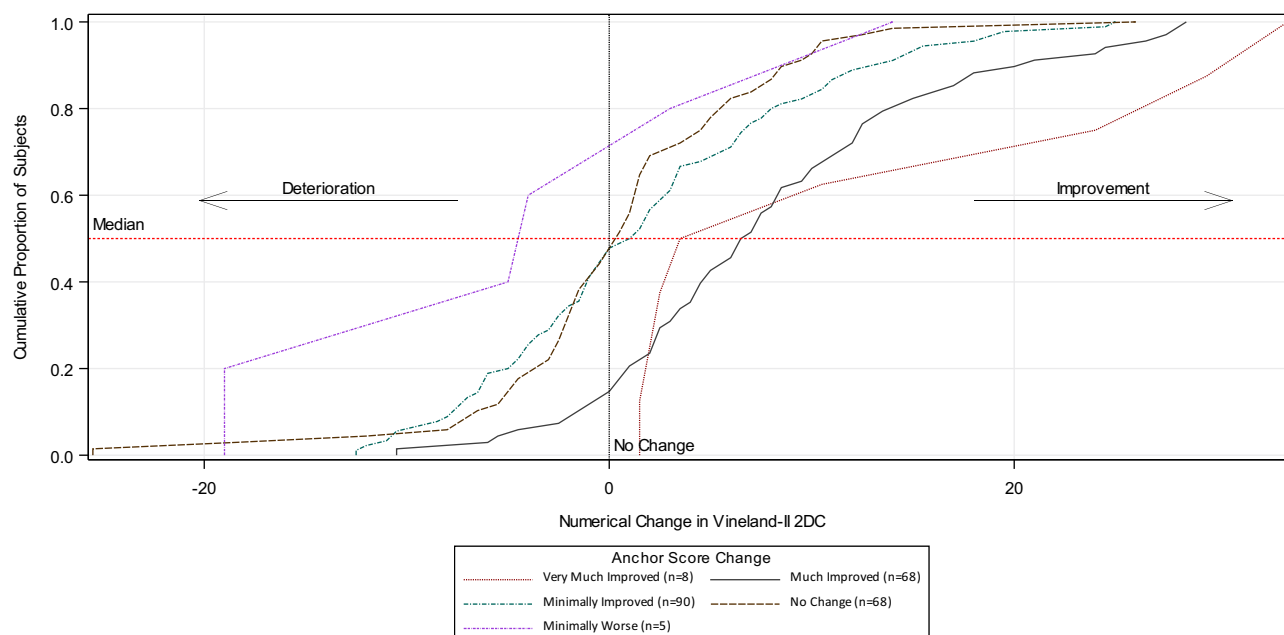
### Distribution-Based Methods

A distribution-based approach for defining changes beyond measurement error was used to support the within-person MCT and MCID estimated for the VABS-2DC and VABS-ABC. Specifically, the estimated MCT and MCID must be greater than measurement error. The distribution of the VABS-2DC and VABS-ABC was used to derive the distribution-based MCID estimates for individuals at baseline; 0.5 SD at baseline was 5.20 and 5.28 (respectively), and the SEM

**Table 4** Change in Vineland™-II -ABC Score According to Response on Selected Anchors

Anchor	No-Change vs Improvement Groups	N	Mean (SD)	Median	95% CI of Mean
OACIS-I Overall	No Change	90	1.3 (8.14)	1.0	(-0.4, 3.0)
	Minimally Improved	84	2.1 (6.97)	2.0	(0.6, 3.6)
	Much Improved	55	9.9 (8.84)	9.0	(7.6, 12.3)
CaGI-I Overall	No Change	62	1.9 (6.68)	1.0	(0.2, 3.6)
	Minimally Improved	91	1.9 (7.09)	1.0	(0.4, 3.4)
	Much Improved	68	8.3 (9.29)	8.0	(6.1, 10.6)
CGI-I	No Change	68	1.0 (7.11)	1.0	(-0.8, 2.7)
	Minimally Improved	90	1.9 (7.93)	1.0	(0.3, 3.6)
	Much Improved	68	7.9 (8.24)	7.5	(5.9, 9.8)

**Abbreviations:** CaGI-I, Caregiver Global Impression of Improvement; OACIS-I, Ohio Autism Clinical Impression Scale of Improvement; SD, standard deviation.



**Figure 1** CDF of change in VABS-2DC from Baseline to Week 24 stratified by CGI-I anchor category.

**Abbreviations:** CDF, cumulative distribution function; CGI-I, Clinician Global Impression of Improvement; VABS-2DC, Vineland™-II 2 Domain Composite Score.

calculated using an ICC was 6.32 and 6.13, respectively (Table 5). Overall, these estimates are in agreement with the anchor-based estimates, suggesting the within-person MCT range above 6 for the VABS-2DC and VABS-ABC.

## Contextualizing Change with Qualitative Data

Qualitative data from the exit interviews were used to confirm the meaning of the individual changes observed. This analysis revealed the range of themes in SPs' narratives. Some examples of themes that emerged and supporting quotes are provided. As all analyses were blinded, the interviews included SPs of those in both placebo and active treatment arms. The [Online Supplement](#) includes more detailed data on thematic analysis.

### 1. Increased willingness to engage, eg:

She used to be one of the kids that would kind of float around the outskirts and would not engage in play or try to join into play without an explicit invitation and with a good awareness of what it is that was expected of her. Now, I'm starting to see her finding little groups and approaching them and participating more openly. [female child, 7 year old]

### 2. Strengthened social skills, eg:

Yeah, this is a kid that never really, he, this kid never really laughed, and now he laughs quite a bit. He, he is starting to understand the jokes and to see the funny in... in things. [male child, 12 year old]

**Table 5** Distribution-Based Method Supporting the Interpretation of Meaningful Change

Domain	N	0.2 SD	0.5 SD	0.8 SD	ICC	SEM	MDC95
Vineland™-II Adaptive Behavior Composite Score	301	2.11	5.28	8.45	0.663	6.130	16.991
Vineland™-II 2 Domain Composite	301	2.08	5.20	8.33	0.630	6.326	17.534

**Abbreviations:** MDC95, minimal detectable change - 95% confidence interval; SEM, standard error of measurement.

### 3. Improved family relationships, eg:

So his-, so I feel like his relationship with his brothers has been the most normal it's ever been. Um, like, especially with his younger brother, he's taken on the role of big brother really well, but because developmentally [Son's Name] is not typical, he does have a lot of common interests with his younger brother, so he's been available to play with him. But yet, with his older brother, like, they like video games, they go bike riding together. I feel like he's been able to be more present in everything that our family does. [male child, 8 year old]

### 4. Improvement in emotional well-being, eg:

... And without realizing what she's saying, she's saying she's happier. If that makes sense? [female child, 10 year old]

### 5. Impact on future, eg:.

if you can't communicate, you can't, you can't be successful in anything. You can't have, uh, functioning relationships, jobs, education, anything. So he's, he's done much better with all of those things. [male child, 10 year old]

### 6. Expression of feelings, eg:

...But, but as far as verbal communication, he is definitely able to tell me things that he never would before. [male child, 7 year old]

### 7. SP improved understanding of the child

...it makes it easier for us to understand what, what it is that he's thinking and feeling, um, there's less guesswork when it comes to his wants and needs, and, um, I think it's just, uh, just a better understanding of him and a calmer environment because of that. [male child, 6 year old]

Very, because. Him expressing to us. I was able to understand when he was upset, didn't like something, and I was able to help him better [male child, 12 year old]

### 8. Improved daily functioning, eg:

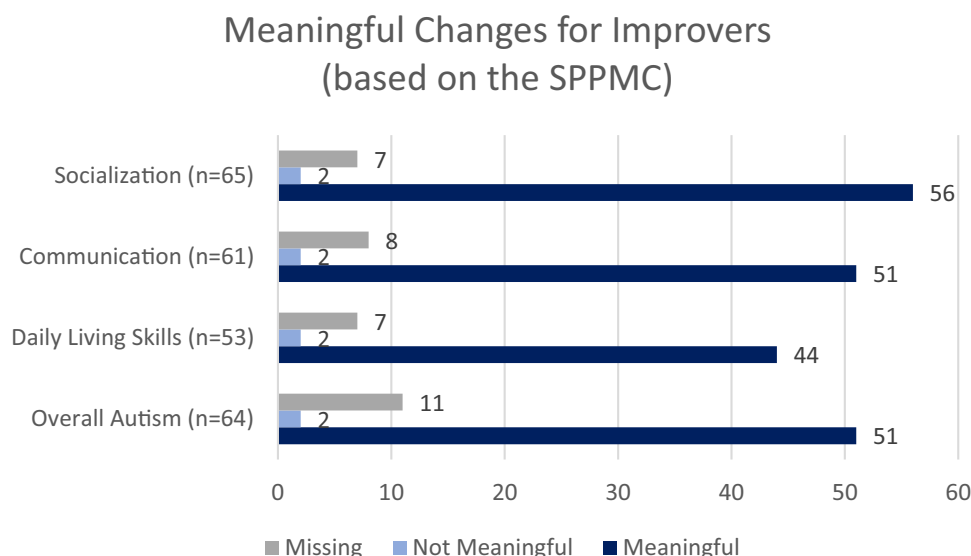
I mean, he might do one or two things. Now he will actually go through, clean his closet, vacuum, um, fold his clothes, put 'em away properly. So, I mean, one day he said, "I-I cleaned my room." And normally, "I cleaned my room", it would be completely a disaster. [male child, 10 year old]

### 9. Increased self-esteem/confidence, eg:

every time he's able to do something, I can, I can see the pride in him and how happy he is and how, you know, excited he is that he's able to do something and he takes and he takes ownership in that. So I know it's a big deal for him. [male child, 10 year old]

During the interviews, SPs discussed a wide range of improvements that were rated from minimally to very much improved on the SPPMC (see emergent themes and the quotes in [Tables S1–S15](#) and [Figures S1–S6](#)). It was clear that the SPs considered almost any change to be meaningful, regardless of the magnitude of the change (as can be seen in [Figure 2](#)).

When comparing the perception of the meaning of the changes reported in the interviews to that captured by change in score on the VABS-2DC in the EIP, the results show that many more SPs reported an improvement on the SPPMC than met the within-person MCT derived from the quantitative analysis. In the EIP,  $n=29$  (33.72%) out of 86 met or exceeded a within-person MCT of 6-point change on the VABS-2DC, and 57 (66.28%) did not. However, the vast majority of EIP participants (whether or not meeting MCT threshold) reported experiencing improvements on the



**Figure 2** Meaning of change reported as an improvement on SPPMC rating in EIP. Y axis: VinelandTM-II domains; X axis: number of participants.  
**Abbreviations:** EIP, exit interview population; MCT, meaningful change threshold; SPPMC, study partner perception of meaningful change.

SPPMC domains (65 out of 86 for Socialization, 61 out of 86 for Communication, and 53 out of 86 for Daily Living Skills), and only 2 SPs considered this improvement not to be meaningful in each domain (Figure 2).

## Discussion

This is the first analysis to explore a within-person MCT on the novel VABS-2DC, used as a primary outcome in aVlation, and the first meaningful change analysis done for VABS-ABC in which a mixed method approach was applied. A range of 4 to 8 points using standardized scores was proposed as the within-person MCT for both VABS-ABC and VABS-2DC, which was found to be associated with at least a 1-point improvement on the 4 different anchors from the clinical trial, aVlation. This could be argued to be a conservative estimate, as within this MCT a change between 4 and less than 7 reflects minimal improvement and a change in score of 7 or more reflects a greater level of improvement (eg, “much improved” on each of the anchors for both endpoints). This finding was supported by the CDF curves showing good separation of curves corresponding to change scores within a given anchor-defined category. The proposed MCT was also in agreement with the results of distribution-based analysis, which found half of an SD for both VABS-ABC and VABS-2DC to be around 5.2 and SEM≈6.

Evidence for an MCT of 4 to 8 points on the VABS-ABC and VABS-2DC was also supported by qualitative approaches including an interview with the SP, although these results suggest that any change was considered meaningful, no matter of the magnitude. In the qualitative interviews, SPs were asked to discuss change that has occurred over the course of the clinical trial, this is anchored to the trial period so that the changes discussed reflect the changes captured on the COAs measures between the administration timepoints in the clinical trial. Such changes could be due to a treatment effect, placebo effect, developmental changes, or another cause. This is appropriate for the evaluation of meaningful change on a COA measure, which does not differ by attribution of the change.

This is a seminal study, both in terms of the methods utilized (drawing upon the data from qualitative and quantitative approaches) and the fact that it used data from a highly controlled clinical trial which involved experienced and trained raters.

The qualitative findings (published elsewhere)<sup>9</sup> confirmed that socialization and communication are the key domains of greatest salience to children and adolescents with ASD and their caregivers/families, supporting the relevance and importance of the VABS-2DC, which is comprised of these 2 domains, as a primary outcome in the aVlation clinical trial. Although the clinical trial failed to meet the primary endpoint, improvements in both of these domains were commonly reported in exit interviews, and the SPs highlighted the positive impact that improvements in these areas had

for the individual with ASD and their family. These reports of the meaning of change at an individual level strongly supported the quantitative estimates of within-person MCT at a 4 to 8-point score change on the VABS-2DC and VABS-ABC; in fact, the qualitative data suggested that these estimates are quite conservative.

Taking the data together, the findings suggest that any change at or above the MCT on the VABS-2DC or VABS-ABC is clearly meaningful to individual with ASD's caregiver and is likely to be meaningful to the individual with ASD. In fact, on an individual level, as any change was found to be meaningful by the caregiver of a child/adolescent with ASD, the qualitative data provided additional confidence that any change in the VABS-2DC at or above the MCT value is truly meaningful for that individual.

The within-person MCT estimated in the current study (4 to 8 points) is higher than the level of change previously estimated as MCID by Chatham et al, for the VABS-ABC standardized score.<sup>5</sup> Data analyses from this study suggested MCID estimates ranged from 2.01 to 3.2 for distribution-based methods, and from 2.42 to 3.75 for sample-size-weighted anchor-based methods. However, when comparing the results of the current study and those of Chatham et al, some key points need to be noted. The data sources from Chatham et al included a variety of datasets from observational, and clinical, prospective epidemiological and community studies; thus, there is likely to be greater variability of data in Chatham et al. They had a much larger sample (pooled data from over 9000 individuals) and identified a range of estimates which then were pooled. Many of the estimates identified by Chatham et al were similar to the estimates based off the current study, and the pooled estimate upper bound (3.75) is highly reflective of the lower bound of 4 found in the current study. In contrast, the current study utilized more conservative methods, a conservative cut-off of 0.35+ to consider an anchor as suitable to be used in anchor-based MCT derivation, and a much smaller sample taken from a homogenous, controlled clinical trial dataset. This is likely to result in higher estimates. Also, the qualitative data in the current study endorsed that almost any change is meaningful to SPs, and therefore they support the lower estimates found by Chatham et al are likely to be meaningful from an individual perspective as well as the more conservative estimates in the current study. Overall, both studies are complimentary and are the only two studies available in this vast research space.

## Limitations

The clinical trial and therefore this study was conducted in those with ASD who were cognitively high functioning ( $IQ \geq 70$ ). Focusing upon a more homogeneous population in which treatment benefit can be observed is appropriate in a clinical trial setting as it facilitates the evaluation of a new therapy. However, this limits the findings of this research to those with ASD and no ID. Furthermore, the population, by necessity, reflects only those diagnosed with ASD. Recent research suggests that this may lead to an underrepresentation of girls with ASD.<sup>34</sup>

Methodologically, a limitation of the analysis was related to the weak anchor-endpoint correlations: out of 39 computed correlations (in total) for 22 (in total) examined anchors only 8 correlations (for 5 anchors) showed correlations with VABS-2DC and VABS-ABC above the threshold of 0.35. This could be due to differences in reporter or conceptualization of a concept across the measures. Clinician ratings were limited to that information observed or shared in a specific interaction, whereas the caregiver may have based their own ratings on more information about daily life that they did not share, and the independent rater was limited only to the interview transcript. Across measures, weaker correlations particularly at the domain level may be because all measures will not capture a concept in the same way and some items within the scales may be conceptually distinct.

Furthermore, while Vineland<sup>TM</sup>-II ratings are based on caregiver-report and the use of caregiver-reported anchors is appropriate, there was no available clinical trial participant self-report endpoint, nor a global impression of change or severity to use as an anchor for MCT derivation. This was in part, due to the age of the individuals enrolled in aVlation and their ability to reliably self report. In addition, the CGI-S is a global measure based on clinician judgement assessing not only symptoms and behavior but also the impact of the symptoms on the individual's ability to function; thus, assessments of symptoms and their impact might have been confounded for this particular measure. Further, one must also acknowledge possible effects of expectancy bias (also known as observer-expectancy effect; referring to the subconscious influence that a researcher can have on the participants of a research study), which could have affected all clinicians' assessments (including all scores on the Vineland<sup>TM</sup>-II), as well as the results obtained in exit interviews

(including the interpretation of the SP's responses to the interview questions). Without knowing expectations of those taking part, or reporting, in the clinical trial or interviews, it is difficult to know what impact such bias may have. However, reflecting the placebo effect and a tendency towards positive reporting, taking part in a clinical trial and discussing the changes that have been observed over that period may bias participants towards looking for changes that have occurred and reporting them to be meaningful. However, both clinical trial and interviews were double-blinded and so there is unlikely to be an expectancy bias for a specific treatment arm, and the interviews were designed to avoid any assumption of change, or that it would be meaningful, on the part of the interviewer.

Finally, due to sample size limitations, the MCT analysis was not computed by age groups. Thus, one has to acknowledge that changes observed in the trial may have been to some degree confounded by the developmental changes experienced by the individuals with ASD due to their age. Similarly, due to the sample size limitations, it was not appropriate to conduct analyses by functional level within the sample of individuals with high functioning ASD and so the consistency of findings across functional level was not explored. Also, given the existing research demonstrating the relationship between adaptive functioning and internalizing problems,<sup>10</sup> the presence of comorbid conditions such as mood disorders, common in ASD, could also be informative when considering meaningful change.

## Methodological Implications

Although the Vineland<sup>TM</sup>-3 has been developed to replace the Vineland<sup>TM</sup>-II, the present study has psychometric implications for the community with for Vineland<sup>TM</sup>-II data, with an interest in comparing the within-person MCT calculated for the Vineland<sup>TM</sup>-II, with future MCTs calculated for Vineland<sup>TM</sup>-3. Mainly, the results highlight an MCT for the novel VABS-2DC composite score, which is consistent with that for the VABS-ABC. The MCT can be utilized in future research to facilitate the interpretation of the VABS-2DC, which the qualitative data from exit interviews with SPs (caregivers), highlighted was meaningful and highly relevant to the individual with ASD without ID and their family. This mixed-methods analysis, utilizing quantitative and qualitative data together, suggest that change on the endpoint that may not meet the MCT but may still be considered meaningful at an individual level.

## Clinical Implications

The Vineland<sup>TM</sup>-II is a well-accepted scale to assess adaptive behavior in ASD in clinical trials, and the novel VABS-2DC assesses the 2 core symptoms in ASD, namely socialization and communication. The current within-person MCT estimates enable the community with access to Vineland<sup>TM</sup>-II data to assess both the statistical and clinical significance of any observed change on VABS-2DC and VABS-ABC in individuals with ASD and no ID. Current findings can be used to inform future ASD trials and facilitate the development of future therapies for the core symptoms of ASD. However, this work should be replicated on Vineland<sup>TM</sup>-3 if used, given the inherent differences in the scales.

## Conclusions

The main finding is that the change in standardized score of 4 to 8 points constitutes the within-person MCT on both VABS-ABC (overall) score and the novel VABS-2DC composite score (comprising Communication and Socialization domains) for individuals with ASD with no ID. The qualitative data (reported elsewhere) confirmed that this change score reflects a change that is considered by caregivers to be meaningful, which could be considered a conservative estimate. Score change within this range or higher was reported by the study partners in this trial to be meaningful and highly impactful upon the individual with ASD and their family.

## Abbreviations

ASD, autism spectrum disorder; CaGI-I, Caregiver Global Impression of Improvement; CDF, cumulative distribution function; CGI-I, Clinician Global Impression of Improvement; CGI-S, Clinician Global Impression of Severity; CI, confidence interval; ClinRO, clinician-reported outcome; COA, clinical outcome assessment; EIP, exit interview population; EoT, end of trial; FDA, Food and Drug Administration; ICC, intraclass correlation coefficient; InCRC, Independent Clinical Rating of Change; ITT, intent-to-treat; MCID, minimal clinically important difference; MCT, meaningful change threshold; MID, minimal important difference; OACIS, Ohio Autism Clinical Impression Scale; OACIS-I, Ohio Autism Clinical Impression

Scale of Improvement; OACIS-S, Ohio Autism Clinical Impression Scale of Severity; ObsRO, observer-reported outcome; SD, standard deviation; SEM, standard error of measurement; SP, study partner; SPGI, Study Partner Global Impression of Improvement; SPPMC, study partner perception of meaningful change; TIRG, Transcript Interpretation Rating Guide; Vineland™-II, Vineland Adaptive Behavior Scales, Second Edition; VABS-2DCM, Vineland™-II 2 Domain Composite Score; VABS-ABC, Vineland™-II Adaptive Behavior Composite Score; VABS-C, Vineland™-II Communication Domain; VABS-DLS, Vineland™-II Daily Living Skills Domain; VABS-S, Vineland™-II Socialization Domain.

## Data Sharing Statement

For up-to-date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: <https://go.roche.com/datasharing>. The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request. Anonymized records for individual participants across more than one data source external to Roche cannot, and should not, be linked due to a potential increase in risk of participant re-identification.

## Ethics Approval and Consent to Participate

For the aVlation study, approval by the Institutional Review Board and Ethics Approval were obtained before study initiation, as appropriate for each site, from an Institutional Review Board or Ethics Committee at each site provided in [Table S16](#). The study was conducted in full accordance with the principles of the Declaration of Helsinki and the International Council for Harmonization E6 guidelines for Good Clinical Practice, or the relevant laws and regulations of the country in which the research was done.

All parent/caregivers gave written informed consent before enrollment which included publication of anonymized responses.

## Acknowledgments

Editorial services were provided by Articulate Science and Clinical Outcomes Solutions. The authors would also like to thank other members of the research team at COS who were involved in the conduct, analysis and/or write up of this study (Ashley Geiger, Bridget Iwamuro, Emily Calderbank, and Jamie Mertoian), Hannah Staunton (employee at Roche) for an extensive review, and, most importantly, those who took part in the research.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This study was funded by Roche Products Ltd.

## Disclosure

SC, TW, JS - employed by Roche Products Ltd. SH and CB - employed by Clinical Outcomes Solutions (which was commissioned by Roche Products Ltd. to conduct the current study). EP and EG - employed by Clinical Outcomes Solutions at the time of the study. The authors report no other conflicts of interest in this work.

## References

1. Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. *Lancet*. 2018;392(10146):508–520. doi:10.1016/S0140-6736(18)31129-2
2. Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007;120(5):1183–1215. doi:10.1542/peds.2007-2361

3. Brugha TS, McManus S, Bankart J, et al. Epidemiology of autism spectrum disorders in adults in the community in England. *Arch Gen Psychiatry*. 2011;68(5):459–465. doi:10.1001/archgenpsychiatry.2011.38
4. Sun X, Allison C, Wei L, et al. Autism prevalence in China is comparable to Western prevalence. *Mol Autism*. 2019;10(1):1–19. doi:10.1186/s13229-018-0246-0
5. Chatham C, Taylor K, Charman T, et al. Adaptive behavior in autism: minimal clinically important differences on the Vineland-II. *Autism Res*. 2018;11(2):270–283. doi:10.1002/aur.1874
6. Sparrow S, Cicchetti D, Balla D. *Vineland Adaptive Behavior Scales*. Third ed. San Antonio: TX Pearson; 2016.
7. Willgoss T, Le Scouiller S, Squassante L, et al. Psychometric Properties of a Novel Vineland-II 2-Domain Composite Score to Assess Social Communication and Social Interaction in ASD. Published in the October 2018 Scientific Proceedings supplement issue of the Journal of the American Academy of Child and Adolescent Psychiatry. Vol. 57; 2018:S231–232.
8. Hollander E, Jacob S, Jou R, et al. Balovaptan vs Placebo for social communication in childhood Autism spectrum disorder: a randomized clinical trial. *JAMA Psychiatry*. 2022. 79(8):760–9
9. Operto FF, Smirni D, Scuoppo C, et al. Neuropsychological profile, emotional/behavioral problems, and parental stress in children with neurodevelopmental disorders. *Brain Sci*. 2021;11(5):584. doi:10.3390/brainsci11050584
10. Operto FF, Pastorino GM, Scuoppo C, et al. Adaptive behavior, emotional/behavioral problems and parental stress in children with autism spectrum disorder. *Front Neurosci*. 2021;15:751465. doi:10.3389/fnins.2021.751465
11. Kanne SM, Gerber AJ, Quirnbach LM, Sparrow SS, Cicchetti DV, Saulnier CA. The role of adaptive behavior in autism spectrum disorders: implications for functional outcome. *J Autism Dev Disord*. 2011;41(8):1007–1018. doi:10.1007/s10803-010-1126-4
12. Kuhlthau K, Payakachat N, Delahaye J, et al. Quality of life for parents of children with autism spectrum disorders. *Res Autism Spectr Disord*. 2014;8(10):1339–1350. doi:10.1016/j.rasd.2014.07.002
13. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales Vineland-II: Survey Forms Manual*. Pearson Minneapolis, MN; 2005.
14. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual Life Res*. 2019;28(4):1029–1033. doi:10.1007/s11136-018-2076-0
15. Cappelleri JC, Bushmakini AG. Interpretation of patient-reported outcomes. *Stat Methods Med Res*. 2014;23(5):460–483. doi:10.1177/0962280213476377
16. Lecavalier L, Leone S, Wiltz J. The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. *J Intellect Disabil Res*. 2006;50(3):172–183. doi:10.1111/j.1365-2788.2005.00732.x
17. Coon CD, Cappelleri JC. Interpreting change in scores on patient-reported outcome instruments. *Ther Innov Regul Sci*. 2016;50(1):22–29. doi:10.1177/2168479015622667
18. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res*. 2018;27(1):33–40. doi:10.1007/s11136-017-1616-3
19. Regnault A, Willgoss T, Barbic S; On behalf of the International Society for Quality of Life Research Mixed Methods Special Interest G. Towards the use of mixed methods inquiry as best practice in health outcomes research. *J Patient-Rep*. 2018;2(1):19. doi:10.1186/s41687-018-0043-8
20. Staunton H, Willgoss T, Nelsen L, et al. An overview of using qualitative techniques to explore and define estimates of clinically important change on clinical outcome assessments. *J Patient-Rep*. 2019;3(1):1–10. doi:10.1186/s41687-019-0100-y
21. Sparrow S, Cicchetti D, Balla D. *Vineland Adaptive Behavior Scales—2nd Edition Manual*. Minneapolis, MN: NCS Pearson Inc.; 2005.
22. Willgoss T, Le Scouiller S, Squassante L, et al. Psychometric properties of a novel vineland-II 2-domain composite score to assess social communication and social interaction in ASD (vol 57, pg S231, 2018). *J Am Acad Child Adolesc Psychiatry*. 2020;59(2):330.
23. Bolognani F, Del Valle Rubido M, Squassante L, et al. A phase 2 clinical trial of a vasopressin V1a receptor antagonist shows improved adaptive behaviors in men with autism spectrum disorder. *Sci Transl Med*. 2019;11(491):eaat7838. doi:10.1126/scitranslmed.aat7838
24. Busner J, Miller DS, Targum SD. A survey of investigator beliefs about including adverse events and comorbidity fluctuations in CNS CGI ratings. *Pharmacopsychiatry*. 2007;25:171–176.
25. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. US Department of Health, Education, and Welfare, Public Health Service; 1976.
26. Butter E, Mulick J. The Ohio autism clinical impressions scale (OACIS). Columbus, OH: Children's Research Institute; 2006.
27. Cappelleri JC, Deal LS, Petrie CD. Reflections on ISPOR's clinician-reported outcomes good measurement practice recommendations. *Value Health*. 2017;20(1):15–17. doi:10.1016/j.jval.2016.12.007
28. Hemphill JF. Interpreting the magnitudes of correlation coefficients. *Am Psychol*. 2003;58(1):78–79. doi:10.1037/0003-066X.58.1.78
29. Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther*. 2011;16(6):563–572. doi:10.1016/j.math.2011.05.006
30. Hudgens S, Floden L, Blackowicz M, et al. Meaningful change in depression symptoms assessed with the Patient Health Questionnaire (PHQ-9) and Montgomery-Åsberg Depression Rating Scale (MADRS) among patients with treatment resistant depression in two, randomized, double-blind, active-controlled trials of esketamine nasal spray combined with a new oral antidepressant. *J Affect Disord*. 2021;281:767–775. doi:10.1016/j.jad.2020.11.066
31. Wyrtwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52(9):861–873. doi:10.1016/S0895-4356(99)00071-2
32. Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *J Rehabil Med*. 2009;41(5):343–346. doi:10.2340/16501977-0337
33. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. USA: Oxford University Press; 2006.
34. Lai M-C, Baron-Cohen S, Buxbaum JD. Understanding autism in the light of sex/gender. *Mol Autism*. 2015;6(1):1–5. doi:10.1186/s13229-015-0021-4

**Patient Related Outcome Measures**

Dovepress

**Publish your work in this journal**

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>