Adaptive Behavior in Autism: Minimal Clinically-Important Differences on the Vineland-II

C. H. Chatham[1], **K.I. Taylor**[1], **T. Charman**[2], **X. Liogier D'ardhuy**[1], **E. Eule**[1], **A. Fedele**[3], **A. Y. Hardan**[4], **E. Loth**[2], **L. Murtagh**[1], **M. del Valle Rubido**[1], **A. San Jose Caceres**[2], **J. Sevigny**[1], **L. Sikich**[5], **L. Snyder**[6], **J. E. Tillmann**[2], **P. E. Ventola**[7], **K. L. Walton-Bowen**[6], **P. P. Wang**[6], **T. Willgoss**[1], and **F. Bolognani**[1]

[1]F. Hoffmann La Roche, Innovation Center Basel, Grenzacherstrasse 170, Hoffmann La Roche, Basel, 4070

[2]Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, UK

[3]Speaks, 1 East 33rd Street, 4th Floor, New York, NY 10016

[4]Psychiatry and Behavioral Sciences, Stanford University, 401 Quarry Road, Stanford, CA 94305-5717

[5]Duke Center for Autism and Brain Development, Pavilion East at Lakeview, 2608 Erwin Rd #300, Durham, NC 27705

[6]Simons Foundation, 160 Fifth Ave, New York, NY 10010, USA

[7]Yale Child Study Center, 230 S Frontage Rd, New Haven, CT 06520, USA

Abstract

Autism Spectrum Disorder (ASD) is associated with persistent impairments in adaptive abilities across multiple domains. These social, personal, and communicative impairments become increasingly pronounced with development, and are present regardless of IQ. The Vineland Adaptive Behavior Scales, Second Edition (Vineland-II) is the most commonly used instrument for quantifying these impairments, but minimal clinically-important differences (MCIDs) on Vineland-II scores have not been rigorously established in ASD. We pooled data from several consortia/registries (EU-AIMS LEAP study, ABIDE-I, ABIDE-II, INFOR, Simons Simplex Collection and Autism Treatment Network [ATN]) and clinical investigations and trials (Stanford, Yale, Roche) resulting in a dataset of over 9,000 individuals with ASD. Two approaches were used to estimate MCIDs: distribution-based methods and anchor-based methods. Distribution-based MCID [d-MCID] estimates included the standard error of the measurement, as well as one-fifth and one-half of the covariate-adjusted standard deviation (both cross-sectionally and longitudinally). Anchor-based MCID [a-MCID] estimates include the slope of linear regression of clinician ratings of severity on the Vineland-II score, the slope of linear regression of clinician

Corresponding author: Christopher Chatham (christopher.chatham@roche.com; tele: +41 79 301 3391; fax: +41 61 687 9037), Grenazacherstrasse 170, 4070 Basel, Switzerland.

ratings of longitudinal improvement category on Vineland-II change, the Vineland-II change score maximally differentiating clinical impressions of minimal vs. no improvement, and equipercentile equating. Across strata, the Vineland-II Adaptive Behavior Composite standardized score MCID estimates range from 2.01–3.2 for distribution-based methods, and from 2.42 to 3.75 for sample-size-weighted anchor-based methods. Lower Vineland-II standardized score MCID estimates were observed for younger and more cognitively-impaired populations. These MCID estimates enable users of Vineland-II to assess both the statistical and clinical significance of any observed change.

Keywords

VABS; intelligence; assessment; treatment; efficacy; intellectual disability; ASD

Introduction

The diagnostic criteria for Autism Spectrum Disorder (ASD) include clinically-significant, functional impairments in social communication and social reciprocity, and the presence of restricted or repetitive behavior and sensory anomalies (DSM-5; APA, 2013), but ASD is also associated with broader impairments in adaptive behaviors that support everyday functioning (Kanne et al, 2011). These adaptive behavior problems span multiple domains including socialization, communication, and daily living skills, and are not fully accounted for by differences in cognitive ability (Charman et al, 2011; Klin et al, 2007). These impairments predict real-world outcomes in ASD, including educational attainment and the likelihood of independent living (Farley et al, 2009; de Bildt, Sytema, Kraijer, Sparrow & Minderaa, 2005). Adaptive behavior impairments in ASD are also associated with both the number of support services received, and the service needs that will go unmet (Taylor & Henninger, 2015). Thus, adaptive behavior is a key target for interventions directed at individual or societal outcomes in ASD (e.g., Veenstra-VanderWeele et al, 2017).

It is not currently possible to fully evaluate the efficacy of interventions directed at adaptive behavior. This is because efficacy claims require a demonstration of not only statistical significance, but also clinical meaningfulness (Jacobson & Truax, 1991; FDA PRO Guidance 2006; 2009; Coon & Cappelleri, 2016). While statistical significance is commonly evaluated, clinical meaningfulness requires that treatment effects exceed a minimally clinically-important difference (MCID). The term MCID was first described in 1989 (Jaeschke, Singer, & Guyatt, 1989) and is now a well-accepted concept to clinicians, regulators and payors to support evaluation of treatments (e.g., Guyatt et al, 2002; Food & Drug Administration, 2009; Medicare Evidence Development & Coverage Advisory Committee, 2006). However, there is currently no consensus on the MCID for adaptive behavior among individuals with ASD.

To remedy this gap, we calculated estimates of the MCID from a large sample of individuals with ASD who were assessed with the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al, 2005) Survey Interview form. The Vineland-II is the most commonly used instrument for quantifying impairments in adaptive behaviors necessary for socialization, communication, and daily functioning (Pugliese et al, 2015; see also Figure

1A). Adaptive behaviors ranging from early developmental milestones to sophisticated demands on language understanding and attention are assessed. The scale can be used from birth to 99 years of age; since the expression of adaptive behavior changes across the lifespan, the score is age-normalized (much like IQ). This instrument characterizes adaptive behaviors at multiple levels of granularity, including an overall composite score as well as domain- and subdomain-level constructs.

The Vineland-II has been used widely in clinical, educational, and research settings, and with populations as diverse as ASD (e.g., Pugliese et al, 2015; Veenstra-VanderWeele et al, 2016; Kanne et al, 2015; Szatmari et al, 2015), Fragile X syndrome (e.g., Fisch et al, 1996), Williams syndrome (Greer et al, 1997), ADHD (Stein et al, 1995), low birth weight (Fjørtoft et al, 2015), Down syndrome (Dykens, Hodapp & Evans, 2006), and other neurodevelopmental disorders. A recent Autism Speaks sponsored working group panel recommended the Vineland-II as suitable (having adequate reliability, validity and responsiveness) to quantify social communication deficits in clinical trials of ASD (Anagnostou et al,. 2015; see also McConachie et al, 2015). The Vineland-II is also increasingly used in randomized trials targeting core ASD symptoms and/or adaptive behavior more broadly (e.g., Chugani et al 2016; Dawson et al 2017; Hardan et al 2015; Scahill et al 2016; Umbricht et al 2016; Veenstra-VanderWeele et al, 2017; Ventola et al 2014), in part due to the relative lack of other reliable measures of change in these behaviors (Anagnostou et al,. 2015). This increasing use may also reflect the relatively broader scope of adaptive behaviors assessed by Vineland-II, relative to the more circumscribed set of core ASD symptoms (Willgoss et al, in press). There are also other widely-used clinical measures that broadly assess adaptive functioning in ASD, including the ABAS-3 and BASC. Both of these measures are completed by caregivers without involvement of a research clinician. They generally yield standard scores which are lower than corresponding Vineland scores in youth with high-functioning autism (e.g., Lopata et al 2013). Unfortunately, neither has been used as a research outcome to evaluate change in individuals with lower IQ's and ASD (Anagnostou et al, 2015) and the BASC has been reported to have highly variable intra-rater reliability (Harrison and Oakland, 2003)."

Despite its widespread use as a comprehensive measure of adaptive behavior in developmental disorders, MCIDs for the Vineland-II are as yet unclear. Previous estimates of clinically-meaningful change in the Vineland-II varied between 2.4 and 6 points (Frye et al, 2016), although this variability could reflect the relatively small, age-restricted sample (n=48 3–14 year-olds) or the use of only a single method of MCID estimation.

Very small changes on Vineland-II scores might be associated with clinically meaningful benefits, given that Vineland-II assesses adaptive behavior in real-world scenarios. Indeed, MCID estimates can (in principle) fall below the measurement error of the instrument itself (e.g., King 2011). If this were the case for the Vineland-II, then *any* statistically-significant change could also be considered clinically meaningful. In addition, the MCID could be dependent on other clinical factors like age or IQ. For example, a given change in Vineland-II scores might be associated with greater clinical benefit early in life, where adaptive behaviors may bootstrap across development. Conversely, MCID estimates might be larger or smaller among those with cognitive impairments: larger if substantive improvements in

adaptive behavior are required before real-world functioning undergoes a meaningful change; or perhaps smaller, given the already significant special support provided to individuals with ASD and comorbid IQ deficits (e.g., Aljunied & Frederickson, 2011).

Methods

Overview

To estimate both the MCID magnitude and its potential dependence on these other clinical factors, we pooled individual patient-level data across a variety of studies, re-derived all Vineland-II standard scores according to US norms, undertook several additional quality control steps, and calculated both distribution-based and anchor-based MCID estimates (Figure 1B). These MCID estimation methods are consistent with recommendations from scientific literature and federal guidance alike (Revicki et al, 2008; King et al, 2011; Food & Drug Administration, 2009; Coon & Cappelleri, 2016; Crosby, Kolotkin & Williams, 2003). At a high-level, distribution-based methods compute the MCID as the change required to exceed some proportion of the intrinsic variability within the affected population; these estimates approximate those achieved through other methods (Norman et al, 2003). By contrast, anchor-based approaches calibrate scores across assessments – mapping from the target assessment to one or more "anchor" assessments where clinical meaningfulness has already been established. For anchor-based estimates, we utilize explicit assessments of clinically-relevant improvement, as quantified by the Clinical Global Impression scale (including its Severity [CGI-S] and Improvement [CGI-I] subforms; Guy, 1976) or, for patients where CGI scores were unavailable, the Ohio Autism Clinical Impressions Scale (OACIS; OSU Research Unit on Pediatric Psychopharmacology, 2005, including its General Level of Autism Severity [OACIS-S] and General Level of Autism Improvement [OACIS-I] subforms). CGI and OACIS assessments were carried out only by clinicians, and not by caregivers or patients themselves (unlike Vineland-II, which was here solely assessed via interview of caregivers by clinicians). Alternative anchors were not evaluated due to lack of standardization across datasets (see also Discussion).

Data Pooling

Individual-level data was pooled from several consortia/registries (EU-AIMS LEAP, ABIDE-I, ABIDE-II, INFOR, Simons Simplex Collection, Autism Speaks Autism Treatment Network), observational studies (Stanford), and clinical trials (Stanford, Yale, Roche). Our pooled data is therefore a large sample of individuals living in Western Europe and North America with ASD. Table 1 describes the studies from which individual-level data was drawn (see also Supplementary Table S1 for further details on each study). All studies used the Survey Interview Form, where adaptive behaviors are assessed in the context of a semi-structured interview with open-ended questions administered by a trained clinician. This form of the test is endorsed as the most reliable and accurate method for obtaining information on adaptive behavior (e.g., Sparrow et al, 2005).

Quality Control

This combined data underwent additional quality control steps, in addition to those already performed on each study independently. First, domain-level scores on the Vineland-II were

recomputed based on Vineland-II raw scores (or v-scale, where raw was unavailable) as a function of the age at assessment. Second, to exclude potentially duplicated visits, we excluded any individuals who had precisely identical standardized scores across all three domains in any pair of follow-up visits. Third, the ranges of all variables were checked (both overall and by dataset) to exclude cases with out-of-range variables (e.g., Vineland-II standard scores outside the range of 20–160, or invalid ages). Fourth, within each dataset, any individuals with recomputed Vineland-II scores falling more than three times the interquartile range above or below the 25th or 75th percentiles (respectively) were excluded as extreme outliers. Overall, 2% of individuals were excluded from the pooled data using these criteria.

Stratification and Covariates

We stratified our combined sample by age (children [<13 years of age], adolescents [13 to <18 years of age] and adults [18 years and above]) and cognitive impairment (as defined by full-scale IQ scores < 70). Within each stratum and in an unstratified analysis (Figure 1B) we controlled for the fixed effects of age, sex, full-scale IQ, and the random effect of dataset where possible. We further adjusted for the effect of age bin, as used in the Vineland-II standard scoring procedure (Sparrow et al, 2005), and its interaction with age. For analysis of longitudinal distribution-based estimates, baseline Vineland-II scores were also used as a covariate. Finally, where regression-derived anchor-based estimates were calculated, potential differences in the administration of the anchor between studies (e.g., OACIS-I/CGI-I and OACIS-S/CGI-S) were captured by allowing the slope between Vineland-II and the anchor to differ as a random effect of dataset. The random effect of dataset is intended to account for differences related to the specific populations enrolled, the differing IQ tests and IQ scoring methods used, and other factors that may increase variability due to pooling individual-level data from many sources. Therefore, the MCID estimates generated were adjusted for the influence of sex, IQ, multiple influences of age and dataset, and (where relevant) baseline Vineland-II scores within each stratum. . See Table 1 for differences between datasets. Each dataset contained a minimum of 10 observations for regression-based analyses.

Distribution-based MCID estimates

For each stratum with more than 30 distinct individuals, three types of distribution-based estimates were computed after adjusting for covariates described above (using R package "lme4"):

1. the standard error of the measurement (SEM; Wyrich, Nienaber & Tierney 1999; Wyrwich, Tierney & Wolinsky, 1999; Wyrwich, Tierney & Wolinsky, 2002; Copay et al, 2007; Farivar, Liu, & Hays, 2004). This measure, distinct from the standard error of the mean, is a lower-limit estimate of scale measurement error for a given individual. It is computed as SD * [SQRT(1-r)], where r is the corrected test-retest reliability estimate reported by the Vineland-II manual.

2. one half the standard deviation (Norman et al, 2003; Le, Doctor, Zoeliner & Feeny, 2013); and

3. one fifth the standard deviation (Beaton 2003; Fayers & Hayes 2014; & Farivar, Liu & Hays 2004; Le, Doctor, Zoeliner & Feeny, 2013).

Anchor-based MCID estimates

For each stratum, anchor-based estimates were computed using the following methods wherever more than 30 individuals could contribute to a given estimate:

1. the slope of a linear regression in which OACIS- /CGI-S category predicted Vineland-II score and the slope of a linear regression in which OACIS- /CGI-I category predicted Vineland-II change score, computed where the Spearman rank correlation of Vineland-II and anchor (OACIS- /CGI-S or OACIS- /CGI-I) exceeded .3 (Revicki, Hays, Cella & Sloan, 2008; see also Fayers & Hays, 2014; Hays, Farivar & Liu, 2005).

2. the Vineland-II change score which maximally (i.e., maximal [sensitivity + specificity]) differentiated OACIS-/CGI-I = 4 ("no change") from OACIS-/CGI-I = 3 ("minimally improved") patients, as determined by the receiver operating characteristics (ROC) curve (Froud & Abel, 2014)

3. equipercentile equating (Leucht et al, 2006; Samara et al, 2014; Kolen & Brennan, 2014), whereby the percentile rank of each subject's Vineland-II change score was used to find corresponding percentile rank on the OACIS-/ CGI-I, thereby identifying the minimal Vineland-II score change that is associated with minimal improvement on the OACIS/CGI (using R-package "equate").

Thus, improvement scores (OACIS-I and CGI-I) were used to anchor longitudinal change in Vineland-II, whereas severity scores (OACIS-S and CGI-S) were used to anchor cross-sectional differences in Vineland-II.

Calculation of Weighted-average MCID estimates

For each stratum individually as well as for the non-stratified analysis, an average distribution-based MCID estimate was computed, weighted by the sample size for each MCID estimate; likewise, an average anchor-based MCID estimate was computed, again weighted by the sample size for each such MCID estimate.

Results

The association of Vineland-II Adaptive Behavior to covariates

The pooled and cleaned final dataset included 9,067 individuals with ASD and a total of 10,326 observations, ranging from 1.16 to 55.7 years of age; 84% of the subjects were males, and 72% of the subjects had IQ 70. Sample sizes stratified by age and the presence of IQ and OACIS/CGI are presented in Table 2.

As expected, individuals with ASD showed substantial adaptive behavior impairments across all domains and at all ages (Figure 2). The most impaired adaptive behavior domain was Socialization, followed by Communication, Daily Living Skills, and Motor Skills, each

with statistically distinct degrees of impairment (Table 3; paired t-test with unequal variances; all p's<4.3e-14). In addition to falling well below the age-normed value of 100 (one-sample t-test; all p's<2.2e-16), all Vineland-II standardized scores also showed a pattern of increased impairment with age (Figure 2). A moderate floor effect was apparent among adults in the Communication and Socialization domains, driven largely by adults with IQ < 70 (see also Figure 3, below). In addition, an apparent inflection point was noted between the ages of 8 and 10 in Communication standardized scores, such that these scores might be seen to improve during early childhood before showing a subsequent decline after age 10. However, this pattern is no longer apparent once adjustments are made for the age bins used in the final stage of Vineland-II scoring (see Supplementary Figure S1 and Supplementary Text S1 for a fuller discussion of this issue), so we refrain from further interpretation.

Full-scale IQ scores were highly variable among those with ASD (with a standard deviation 1.8 times larger than the expected value of 15; Table 3). However, variability in adaptive behavior was not as markedly elevated, remaining relatively close to the expected standard deviation of 15 (Table 3).

Vineland-II and full-scale IQ show a curvilinear association (Figure 3), such that the positive slope relating full-scale IQ and Vineland-II standardized scores becomes shallower at higher IQs (the same pattern is observed regardless of whether verbal, non-verbal, or full-scale measures of IQ are used; Supplementary Figure S2). This attenuation appears least pronounced in the Communication domain, where uniformity among individuals with ASD was found to be lowest (see above).

Sex-related differences in Vineland-II standardized scores were relatively minor. While sex differences in full-scale IQ were 5.6 points on average (95% CI=−7.868–3.33, males: 84.3 [range 7–167]; females: 78.7 [range 18–154] independent unequal variance Welch t(878)= −4.8428 p=1.5e-06), sex differences in Vineland-II scores were always less than half that magnitude, and were statistically significant only for the Daily Living Skills domain (mean difference of 1.8 (95% CI=−3–0.576, males: 75.9 [range 21–123]; females: 74.1 [range 25–136] independent unequal variance Welch t(876.6)= −2.8948 p=0.003; all other p's > .16). Figure 4 depicts these sex-related differences alongside the larger IQ-related differences in adaptive behavior. (See also Supplementary Text S1, Figure S1, and Table S2 for further details on the association of sex and other covariates to the Vineland-II.)

MCID estimates

Table 4 reports stratified, sample size-weighted estimates of both anchor-based and distribution-based MCID estimates for Vineland-II scores. For the Adaptive Behavior Composite score, these weighted anchor-based estimates ranged from 2.44 to 3.76; weighted distribution-based estimates were in good agreement with these weighted anchor-based methods within each stratum, all falling within a half point of the weighted anchor-based estimates. We note that, among adolescents with IQ 70, anchor-based MCIDs could be computed only for the Socialization domain, due to inconsistent associations between other Vineland-II standardized scores and anchors in this stratum.

Good agreement was also observed between the weighted anchor- and distribution-based methods across the Vineland-II domain-level scores. The mean difference between weighted MCID estimates from these two methods was 0.19 points. There was a positive correlation between weighted distribution-based MCID estimates and the corresponding weighted anchor-based MCID estimates across the various strata and Vineland-II standardized scores for which weighted anchor-based MCID estimates could be calculated (n=21, as shown in Table 4; R=.47, p =.029; ICC(2,1) = .42; see also Supplementary Figure S7).

We assessed the robustness of these weighted MCID estimates using sensitivity analysis, including precision-weighting rather than sample size-weighting (Supplementary Table S3), excluding individuals with known syndromic forms of ASD (Fragile X syndrome; Supplementary Table S4), excluding data collected in countries where official non-USA norms are available (Supplementary Table S5), and evaluating alternative age groupings occasionally requested by regulatory authorities (e.g., defining adolescence as starting at 12 years of age and adulthood as starting at age 16 [FDA, 1996; EMA, 2001]; Supplementary Table S6) or utilized in early-life interventional studies (e.g., defining early childhood as age 0 to 4, and middle-to-late childhood as age 5 to age 12; Supplementary Table S7). Across these analyses, the weighted MCID estimates reported in Table 4 were stable, if slightly conservative: the sensitivity analyses revealed slight reductions in weighted MCID estimates on average in the affected strata, although the mean change was small (mean difference: −.15, SD=.59).

These weighted MCID estimates were generally larger for the domain-level standard scores than for the Adaptive Behavior Composite. Weighted MCID estimates for the Socialization and Daily Living Skills domains were larger than the Composite weighted MCID (Wilcoxon Signed Rank Test p=.001 and p=.019, respectively), although this same effect was detected at only a trend level for the Communication domain (Wilcoxon Signed Rank Test p=.08).

The weighted MCID estimates tended to increase with IQ: they were lower in the IQ<70 strata in all 21 of the comparisons enabled by Table 4 (Wilcoxon Signed Rank Test p< 0.0001). Weighted MCID estimates also generally increased with age: they tended to be lower for children than for adolescents (Wilcoxon Signed Rank Test p=.02), and lower for adolescents than adults (Wilcoxon Signed Rank Test p=.04).

These fully-stratified analyses were more consistent than partially-stratified or non-stratified analyses. In non-fully stratified analyses, discrepancies between weighted distribution- and anchor-based MCID estimates could exceed two points. Indeed, anchor- and distribution-weighted averages failed to reliably correlate in analyses that were not fully stratified (R =. 11; Supplementary Figure S7).

The individual, unweighted estimates contributing to the above anchor- and distribution-based weighted-average MCID estimates are presented in Supplementary Figures S3–6 for the Adaptive Behavior Composite, Communication, Socialization, and Daily Living Skills domains (respectively), and are presented in tabular form in Supplementary Table S8.

Discussion

Individuals with ASD increasingly fall behind their peers in adaptive behavior with age. Using the largest sample of adaptive behavior among individuals with ASD to date, we identified the most impaired domain of adaptive behavior as Socialization, followed by Communication, Daily Living Skills, and Motor Skills. We quantify the partial dependency of these impairments on factors like sex, age, and IQ, enabling more precise and powerful assessments of adaptive behavior in ASD (e.g., Lerer et al., 2008).

Pronounced age- and IQ-related shifts are observed in all domains of adaptive behavior among individuals with ASD, with comparatively minor effects of sex (consistent with some prior results; see Howe, Yatchmink, Viscidi & Morrow, 2014). Despite these effects, we find that adaptive behavior impairments are also much more uniform than those observed in full-scale IQ (see Table 3). ASD is considered a spectrum disorder because it is highly heterogeneous (e.g., Beversdorf & MAS Consortium, 2016; Szatmari et al, 2015), yet we find a greater uniformity of adaptive behavior in ASD with persistent impairment across demographic variables like IQ and sex (see Results). This striking uniformity reflects the fact that, even when intellectual ability is intact, adaptive behavior is still impaired – particularly in terms of Socialization and Daily Living Skills (Charman et al, 2011; Klin et al, 2007; Duncan & Bishop, 2013). The curvilinear association between full-scale IQ and adaptive behavior suggests that IQ does not fully explain adaptive behavior impairments among individuals with ASD, particularly at high IQs where the association between IQ and adaptive behavior is attenuated. In summary, given that adaptive behaviors are assessed in real-world, everyday scenarios and represent an area of specific, disproportionate impairment (relative to IQ), the Vineland-II is expected to be an important measure of interventional efficacy in ASD.

To support evaluations of efficacy using the Vineland-II interview form, we used both distribution-based and anchor-based methods to establish estimates of clinically meaningful improvement (as recommended by Coon & Cappelleri, 2016; Crosby, Kolotkin & Williams, 2003; and FDA, 2006, 2009). When stratifying by age and IQ, the sample size-weighted MCIDs produced good agreement both with one another (a half-point or less of discrepancy in the Adaptive Behavior Composite score) and with prior work (Frye et al, 2016's sample of 48 children with ASD using a single method of estimation, the standard error of the measurement). Anchor- and distribution-based weighted MCID estimates were also positively correlated, suggesting both agreement between these methods, as well as systematic variation in the weighted MCID estimates across strata and Vineland-II standardized scores.

One source of systematic variation is the type of standardized score that is used. Weighted MCID estimates were generally larger for any single domain of adaptive behavior as compared to the Adaptive Behavior Composite. This pattern does not mean it is somehow "easier" to achieve a clinically-meaningful effect on the Adaptive Behavior Composite: the criterion for clinical meaningfulness is merely numerically smaller. In other words, a given numerical change on the Adaptive Behavior Composite is more likely to be clinically meaningful than a numerically-identical change that is limited to any given domain (e.g.,

Communication, Socialization, or Daily Living Skills), but numerically-identical changes may not be equally likely across domains. These differing MCID estimates highlight the psychometric independence of the Adaptive Behavior Composite from the domain-level standardized scores that contribute to it (e.g., Sparrow et al 2005). Interventions which do not exceed the MCID on each domain individually could still theoretically exert a clinically-meaningful effect on the Adaptive Behavior Composite, in aggregate.

A second source of systematic variance in Vineland-II weighted MCID estimates relates to the effects of age and IQ, with demonstrably lower MCID estimates for the young and those with cognitive disability. These systematic differences in the MCID estimates do not imply that any given intervention would be more clinically beneficial early in development or among those with cognitive disability, nor do they obviate the need for empirical testing of a given intervention in each population of interest. For example, just as an intervention associated with 5 points change in adults may not yield the same change in children, an intervention associated with clinically meaningful change among adults may or may not also be clinically meaningful among children. Indeed, the dependence of MCID estimates on age and IQ underscores the methodological importance accounting for these variables when designing and assessing interventions influencing adaptive behavior.

Large-scale analysis of pooled individual-level data can attenuate the sampling and ascertainment biases affecting any given study (e.g., Howe, Yatchmink, Viscidi & Morrow, 2014). For example, small or unrepresentative samples may not permit adequate generalization beyond the specific individuals enrolled in that study. On the other hand, pooling can also inflate heterogeneity, for example by mixing distinct populations or distinct methods of diagnosis (see Table 1 and Supplementary Table S1 for details on differences between studies). Here this concern is mitigated by the reduced variance in Vineland-II scores we observe, relative to normalization samples (Sparrow et al, 2015); the robustness of our estimates across various pooling strategies (see Supplementary Tables S2–S6); the statistical modeling of random effects of dataset (see Methods: Stratification and Covariates); and the orderly associations observed between age and Vineland-II scores. Relatedly, our anchors (CGI and OACIS) did not uniformly show robust associations with Vineland-II scores, though again the observed agreement between stratified, weighted distribution-based and anchor-based MCID estimates may lessen this concern. When making these global assessments, clinicians are likely to incorporate features beyond those assessed in the Vineland-II (e.g., eye contact, hand flapping, etc.), including core symptoms of ASD that are not assessed by the Vineland-II. Future work with the Vineland-II should confirm that effects exceeding our weighted MCID estimates are viewed as important by individuals with autism themselves. Future work may also investigate whether the current weighted MCID estimates apply to alternative forms of the Vineland-II (including the Vineland-II Parent/Caregiver Rating Form, Expanded Interview Form, the Teacher Rating Form, and the recently published Vineland-3; Sparrow, Cichetti & Saulnier, 2016), or even to the clinical meaningfulness of biomarkers associated with adaptive behavior (e.g., Lerer et al., 2008).

Interventions exceeding the MCID are often said to be those which would mandate a change in a patient's treatment, (in the absence of other countervailing factors like treatment cost,

potential side effects, etc.). However, interventions falling short of the MCID at the group-level may nonetheless still drive meaningful change in a subset of individuals. For example, suppose a novel intervention for adults increases the Vineland-II Adaptive Behavior Composite standardized score by 2.5 points on average, where half of the participants show an increase of 5 points and the other half show zero treatment-related change. Although the average treatment-related change falls *below* the relevant MCID estimates (e.g., ranging from 2.8–3.8; Table 4), the treatment may still be clinically meaningful for the 50% of subjects showing a treatment-related response. Indeed, the anchor-based MCID estimates we report follow guidelines for defining clinically-meaningful differences *at the individual-level* (as emphasized in the FDA PRO guidance, 2009). Analogously, the MCID estimates reported here may be used to identify subgroups of patients with meaningfully distinct degrees of clinical severity, or to identify biomarkers that index clinically meaningful variation (e.g., Marquand, Wolfers, Mennes, Buitelaar & Beckmann, 2016; Lerer et al., 2008). Vineland MCID estimates may therefore play a broader role, enabling more personalized and biologically-informed treatments for ASD (e.g., Loth, Murphy & Spooren, 2016).

These MCID estimates for adaptive behavior in ASD may be useful for prioritizing the development of promising new interventions. MCID estimates are also acknowledged to play a role in the decision making processes of regulators and payors relating to the potential value of new interventions (see Rennard, 2009 for a specific example in pulmonology, or Hammad & Neyarapally, 2016 for a broader discussion). In those contexts, MCID estimates are not a replacement for directly establishing the meaningfulness (or lack thereof) in any given individual. For example, a sizable proportion of individuals can still experience meaningful improvement even when the MCID is not exceeded on average. Rather, MCID estimates are one of multiple important benchmarks for identifying clinically-meaningful effects.

Conclusions

Our results provide robust, stratified MCID estimates for Vineland-II scores obtained from a large sample of individuals with ASD, and underscore the importance of defining population-specific MCID estimates using multiple methods. Although clinical measures of global functioning were not robustly associated with Vineland-II, stratification by age and IQ yielded good agreement between distribution- and anchor-based methods, supporting the use of age- and IQ-appropriate MCID estimates in evaluating new interventions for efficacy. MCID estimates for the Vineland-II Adaptive Behavior Composite standardized score range from 2–3.8 points for the Adaptive Behavior Composite depending on stratum. Clinical efficacy as measured by the Vineland-II is generally more likely to be seen with early intervention, and among those with impaired intellectual abilities, for a given treatment-related change on the Vineland-II. The developmentally-persistent and relatively uniform (as compared to IQ) profile of adaptive behavior in ASD suggests that adaptive behaviors are a powerful and important dimension of clinically-meaningful variation in this condition. These estimates of the MCID represent an important new benchmark by which new treatments for ASD can be evaluated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aljunied M, Frederickson N. Cognitive indicators of different levels of special educational support needs in autism. Research in Autism Spectrum Disorders. 2011; 5(1):368–376.

Anagnostou E, Jones N, Huerta M, Halladay AK, Wang P, Scahill L, Dawson G. Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. Autism. 2015; 19(5):622–636. [PubMed: 25096930]

Beaton DE. Simple as possible? Or too simple? Possible limits to the universality of the one half standard deviation. Medical Care. 2003; 41(5):593–596. [PubMed: 12719682]

Beversdorf DQ, CONSORTIUM MAS. Phenotyping, Etiological Factors, and Biomarkers: Toward Precision Medicine in Autism Spectrum Disorders. Journal of Developmental and Behavioral Pediatrics. 2016; 37(8):659. [PubMed: 27676697]

Charman T, Pickles A, Simonoff E, Chandler S, Loucas T, Baird G. IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). Psychological Medicine. 2011; 41(03):619–627. [PubMed: 21272389]

Chugani DC, Chugani HT, Wiznitzer M, Parikh S, Evans PA, Hansen RL, Rothermel R. Efficacy of low-dose buspirone for restricted and repetitive behavior in young children with autism spectrum disorder: a randomized trial. The Journal of Pediatrics. 2016; 170:45–53. [PubMed: 26746121]

Coon CD, Cappelleri JC. Interpreting Change in Scores on Patient-Reported Outcome Instruments. Ther. Innov. Regul. Sci. 2016; 50:22–29.

Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine Spine J. Off. J. North Am. Spine Soc. 2007; 7:541–546.

Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J. Clin. Epidemiol. 2003; 56:395–407. [PubMed: 12812812]

Dawson G, Sun JM, Davlantis KS, Murias M, Franz L, Troy J, Kurtzberg J. Autologous cord blood infusions are safe and feasible in young children with autism spectrum disorder: Results of a

single-center phase I open-label trial. Stem Cells Translational Medicine. 2017; 6(5):1332–1339. [PubMed: 28378499]

De Bildt A, Sytema S, Kraijer D, Sparrow S, Minderaa R. Adaptive functioning and behaviour problems in relation to level of education in children and adolescents with intellectual disability. Journal of Intellectual Disability Research. 2005; 49(9):672–681. [PubMed: 16108984]

Duncan AW, Bishop SL. Understanding the gap between cognitive abilities and daily living skills in adolescents with autism spectrum disorders with average intelligence. Autism. 2015; 19(1):64. [PubMed: 24275020]

Dykens E, Hodapp R, Evans D. Profiles and development of adaptive behavior in children with Down syndrome. Down Syndrome Research and Practice. 2006; 9(3):45–50.

European Medicines Agency. [Retrieved Feb 20th 2017] ICH Topic E 11 Clinical Investigation of Medicinal Products in the Pediatric Population. 2001. from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf

Farivar SS, Liu H, Hays RD. Half standard deviation estimate of the minimally important difference in HRQOL scores? Expert Review of Pharmacoeconomics & Outcomes Research. 2004; 4(5):515–523. [PubMed: 19807545]

Farley MA, McMahon WM, Fombonne E, Jenson WR, Miller J, Gardner M, Coon H. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. Autism Research. 2009; 2(2):109–118. [PubMed: 19455645]

Fayers PM, Hays RD. Don't middle your MIDs: regression to the mean shrinks estimates of minimally important differences. Quality of Life Research. 2014; 23(1):1–4.

Fisch GS, Simensen R, Tarleton J, Chalifoux M, Holden JJ, Carpenter N, Maddalena A. Longitudinal study of cognitive abilities and adaptive behavior levels in fragile × males: a prospective multicenter analysis. American Journal of Medical Genetics. 1996; 64(2):356–361. [PubMed: 8844080]

Fjørtoft T, Grunewaldt KH, Løhaugen GCC, Mørkved S, Skranes J, Evensen KAI. Adaptive behavior in 10–11 year old children born preterm with a very low birth weight (VLBW). European Journal of Paediatric Neurology. 2015; 19(2):162–169. [PubMed: 25497593]

Food and Drug Administration. Guidance for industry: The Content and Format for Pediatric Use Supplements. Federal Register. 1996; 61(102):26191–2619.

Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. Federal Register. 2009; 74(235):65132–65133.

Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: The forgotten lesson of Pythagoras. Theoretical considerations and an example application of change in health status. PloS One. 2014; 9(12):e114468. [PubMed: 25474472]

Frye RE, Slattery J, Delhey L, Furgerson B, Strickland T, Tippett M, James SJ. Folinic acid improves verbal communication in children with autism and language impairment: a randomized double-blind placebo-controlled trial. Molecular Psychiatry. 2016; Advance online publication. doi: 10.1038/mp.2016.168

Greer MK, Brown FR, Pai GS, Choudry SH, Klein AJ. Cognitive, adaptive, and behavioral characteristics of Williams syndrome. American Journal of Medical Genetics. 1997; 74(5):521–525. [PubMed: 9342204]

Guy, W. ECDEU Assessment Manual for Psychopharmacology, Revised. National Institute of Mental Health; Rockville, MD: 1976. Clinical Global Impressions; p. 218-222.(DHEW Publ. No. ADM 76-338), (1976)

Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin. Proc. 2002; 77:371–383. [PubMed: 11936935]

Hammad, TA., Neyarapally, GA. Benefit-Risk Assessment Methods in Medical Product Development: Bridging Qualitative and Quantitative Assessments. Chapman and Hall: CRC; 2016. Regulatory and legislative policy and science considerations in the era of patient-centeredness, big data, and value; p. 15-51.

Hardan AY, Gengoux GW, Berquist KL, Libove RA, Ardel CM, Phillips J, Minjarez MB. A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. Journal of Child Psychology and Psychiatry. 2015; 56(8):884–892. [PubMed: 25346345]

Harrison, P., Oakland, T. Adaptive behavior assessment system-(ABAS-II). San Antonio, TX: The Psychological Corporation; 2003.

Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. Journal of Chronic Obstructive Pulmonary Disease. 2005; 2:63–67. [PubMed: 17136964]

Howe YJ, Yatchmink Y, Viscidi EW, Morrow EM. Ascertainment and gender in autism spectrum disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 2014; 53(6):698. [PubMed: 24839890]

Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology. 1991; 59(1):12. [PubMed: 2002127]

Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control. Clin. Trials. 1989; 10:407–415. [PubMed: 2691207]

Kanne SM, Gerber AJ, Quirmbach LM, Sparrow SS, Cicchetti DV, Saulnier CA. The role of adaptive behavior in autism spectrum disorders: Implications for functional outcome. Journal of Autism and Developmental Disorders. 2011; 41(8):1007–1018. [PubMed: 21042872]

King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Review of Pharmacoeconomics & Outcomes Research. 2011; 11(2):171–184. [PubMed: 21476819]

Klin A, Saulnier CA, Sparrow SS, Cicchetti DV, Volkmar FR, Lord C. Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: The Vineland and the ADOS. Journal of Autism and Developmental Disorders. 2007; 37(4):748–759. [PubMed: 17146708]

Kolen, MJ., Brennan, RL. Test Equating, Scaling, and Linking. Springer; New York: 2014.

Le QA, Doctor JN, Zoellner LA, Feeny NC. Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT). Health and Quality of Life Outcomes. 2013; 11(1):59. [PubMed: 23587015]

Lerer E, Levi S, Salomon S, Darvasi A, Yirmiya N, Ebstein RP. Association between the oxytocin receptor (OXTR) gene and autism: relationship to Vineland-II Adaptive Behavior Scales and cognition. Molecular Psychiatry. 2008; 13(10):980–988. [PubMed: 17893705]

Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. Neuropsychopharmacology. 2006; 31(10):2318–2325. [PubMed: 16823384]

Lopata C, Smith RA, Volker MA, Thomeer ML, Lee GK, McDonald CA. Comparison of adaptive behavior measures for children with HFASDs. Autism Research and Treatment. 2013 (2013).

Loth E, Murphy DG, Spooren W. Defining Precision Medicine Approaches to Autism Spectrum Disorders: Concepts and Challenges. Frontiers in Psychiatry. 2016; 7

Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2016; 1(5):433–447. [PubMed: 27642641]

McConachie H, Parr JR, Glod M, Hanratty J, Livingstone N, Oono IP, Garland D. Systematic review of tools to measure outcomes for young children with autism spectrum disorder. Health Technology Assessment. 2015 No. 19.41.

Medicare Evidence Development & Coverage Advisory Committee. [last accessed Feb 20th, 2017] Meeting Minutes for Spinal Fusion for the Treatment of Low Back Pain Secondary to Lumbar Degenerative Disc Disease. 2006. http://www.cms.gov/medicare-coverage-database/details/medcac-meeting-details.aspx?MEDCACId=37

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical Care. 2003; 41(5):582–592. [PubMed: 12719681]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pugliese CE, Anthony L, Strang JF, Dudley K, Wallace GL, Kenworthy L. Increasing adaptive behavior skill deficits from childhood to adolescence in autism spectrum disorder: Role of executive function. Journal of Autism and Developmental Disorders. 2015; 45(6):1579–1587. [PubMed: 25398602]

Rennard SI. Minimal clinically important difference, clinical perspective: an opinion. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2005; 2(1):51–55. [PubMed: 17136962]

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J. Clin. Epidemiol. 2008; 61:102–109. [PubMed: 18177782]

OSU Research Unit on Pediatric Psychopharmacology. OSU Autism Rating Scale. Columbus, OH: 2005.

Samara MT, Engel RR, Millier A, Kandenwein J, Toumi M, Leucht S. Equipercentile linking of scales measuring functioning and symptoms: examining the GAF, SOFAS, CGI-S, and PANSS. European Neuropsychopharmacology. 2014; 24(11):1767–1772. [PubMed: 25219937]

Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures. Pharmacoeconomics. 1999; 15(2):141–155. [PubMed: 10351188]

Scahill L, Bearss K, Lecavalier L, Smith T, Swiezy N, Aman MG, Levato L. Effect of parent training on adaptive behavior in children with autism spectrum disorder and disruptive behavior: results of a randomized trial. Journal of the American Academy of Child & Adolescent Psychiatry. 2016; 55(7):602–609. [PubMed: 27343887]

Sparrow, SS., Cicchetti, DV., Balla, DA. Vineland-II Adaptive Behavior Scales, Second Edition, Survey Forms Manual. PsychCorp; 2005.

Sparrow, SS., Cichetti, DV., Saulnier, CA. Vineland Adaptive Behavior Scales, Third Edition (Vineland-3) Manual. PsychCorp; 2016.

Stein MA, Szumowski E, Blondis TA, Roizen NJ. Adaptive skills dysfunction in ADD and ADHD children. Journal of Child Psychology and Psychiatry. 1995; 36(4):663–670. [PubMed: 7650089]

Szatmari P, Georgiades S, Duku E, Bennett TA, Bryson S, Fombonne E, Volden J. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. JAMA Psychiatry. 2015; 72(3):276–283. [PubMed: 25629657]

Taylor JL, Henninger NA. Frequency and correlates of service access among youth with autism transitioning to adulthood. Journal of Autism and Developmental Disorders. 2015; 45(1):179–191. [PubMed: 25081594]

Umbricht D, del Valle Rubido M, Hollander E, McCracken JT, Shic F, Scahill L, Grundschober C. A Single Dose, Randomized, Controlled Proof-Of-Mechanism Study of a Novel Vasopressin 1a Receptor Antagonist (RG7713) in High-Functioning Adults with Autism Spectrum Disorder. Neuropsychopharmacology. 2016; 1:10.

Ventola P, Friedman HE, Anderson LC, Wolf JM, Oosting D, Foss-Feig J, Pelphrey KA. Improvements in social and adaptive functioning following short-duration PRT program: a clinical replication. Journal of Autism and Developmental Disorders. 2014; 44(11):2862–2870. [PubMed: 24915928]

Veenstra-VanderWeele J, Cook EH, King BH, Zarevics P, Cherubini M, Walton-Bowen K, Carpenter RL. Arbaclofen in Children and Adolescents with Autism Spectrum Disorder: A Randomized, Controlled, Phase 2 Trial. Neuropsychopharmacology. 2017

McDougall F, Willgoss T, Hwang S, Bolognani F, Murtagh L, Anagnostou E, Rofail R. Development of a patient-centered conceptual model of the impact of living with autism spectrum disorder. Autism: International Journal of Research and Practice. (in press).

Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J. Man. Manip. Ther. 2012; 20:160–166. [PubMed: 23904756]

Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Medical Care. 1999; 37(5):469–478. [PubMed: 10335749]

Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. Journal of Clinical Epidemiology. 1999; 52(9):861–873. [PubMed: 10529027]

Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. Quality of Life Research. 2002; 11(1):1–7. [PubMed: 12003051]

Lay Summary: The Vineland Adaptive Behavior Scales (2nd edition; Vineland-II) is the most widely-used scale for assessing day-to-day "adaptive" skills. Yet, it is unknown how much Vineland-II scores must change for those changes to be regarded as clinically significant. We pooled data from over 9000 individuals with ASD to show that changes of 2 to 3.75 points on the Vineland-II Composite score represent the "minimal clinically-important difference". These estimates will help evaluate the benefits of potential new treatments for ASD.
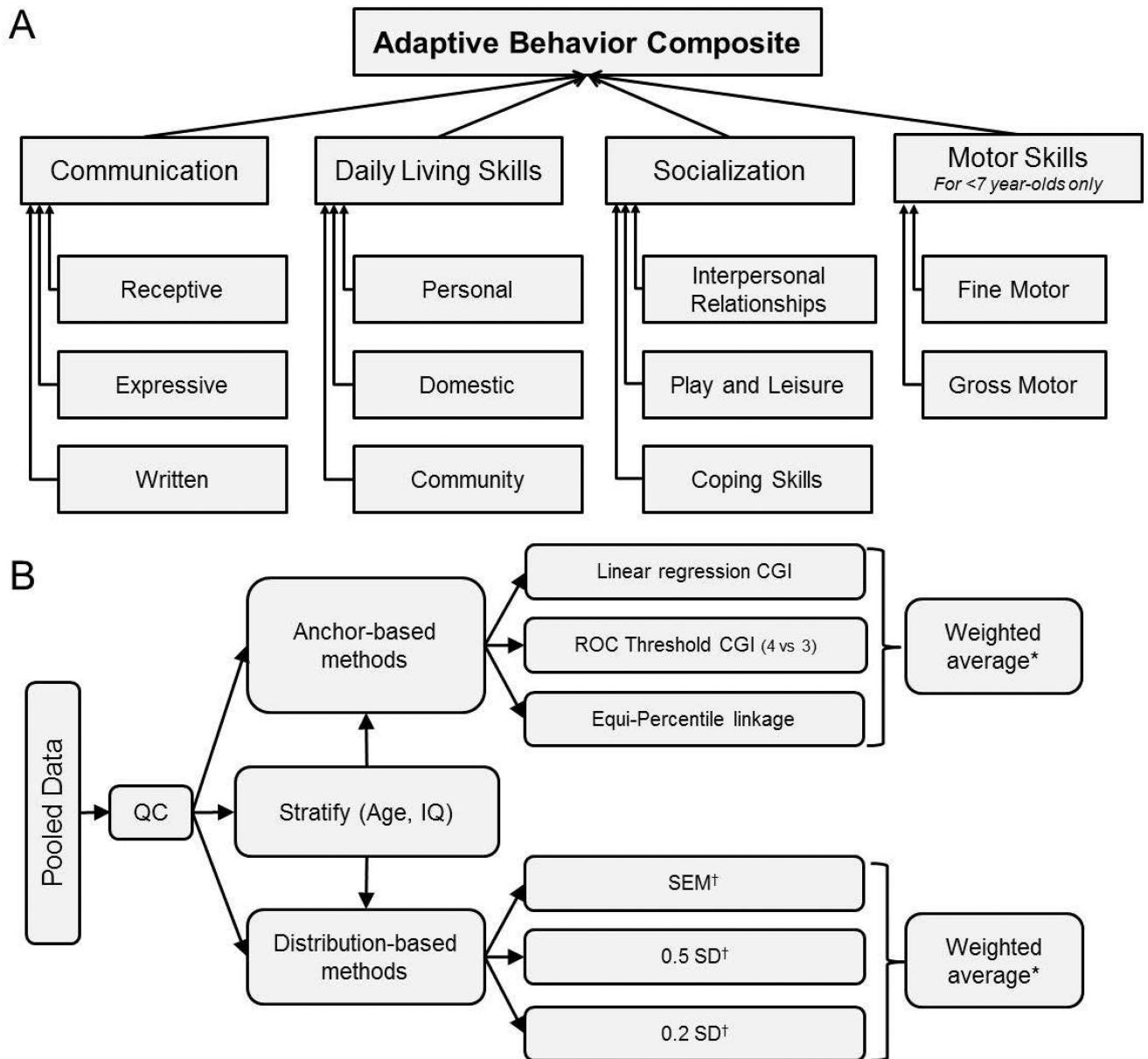
Figure 1.
Structure of the Vineland-II and of the analysis strategy for estimating the MCID. A. The Vineland-II assesses 11 subdomains of adaptive behavior grouped into four domains; sums of these scores are standardized into Communication, Daily Living Skills, Socialization and Motor Skills domain standard scores. Sums of the domain standard scores are then standardized into the Adaptive Behavior Composite score. B. Individual-level data was pooled across multiple consortia, trials and registries, and subjected to several quality control steps. Next, both anchor-based and distribution-based techniques for estimating the MCID were performed on the entire pooled dataset, as well as stratified subsets of the pooled dataset. *Sample-size weighting is used, such that the number of individuals contributing to each anchor- or distribution-based estimate is directly proportional to the

influence of that estimate on the weighted-average; †Both cross-sectional and longitudinal assessments were performed.

Figure 2.
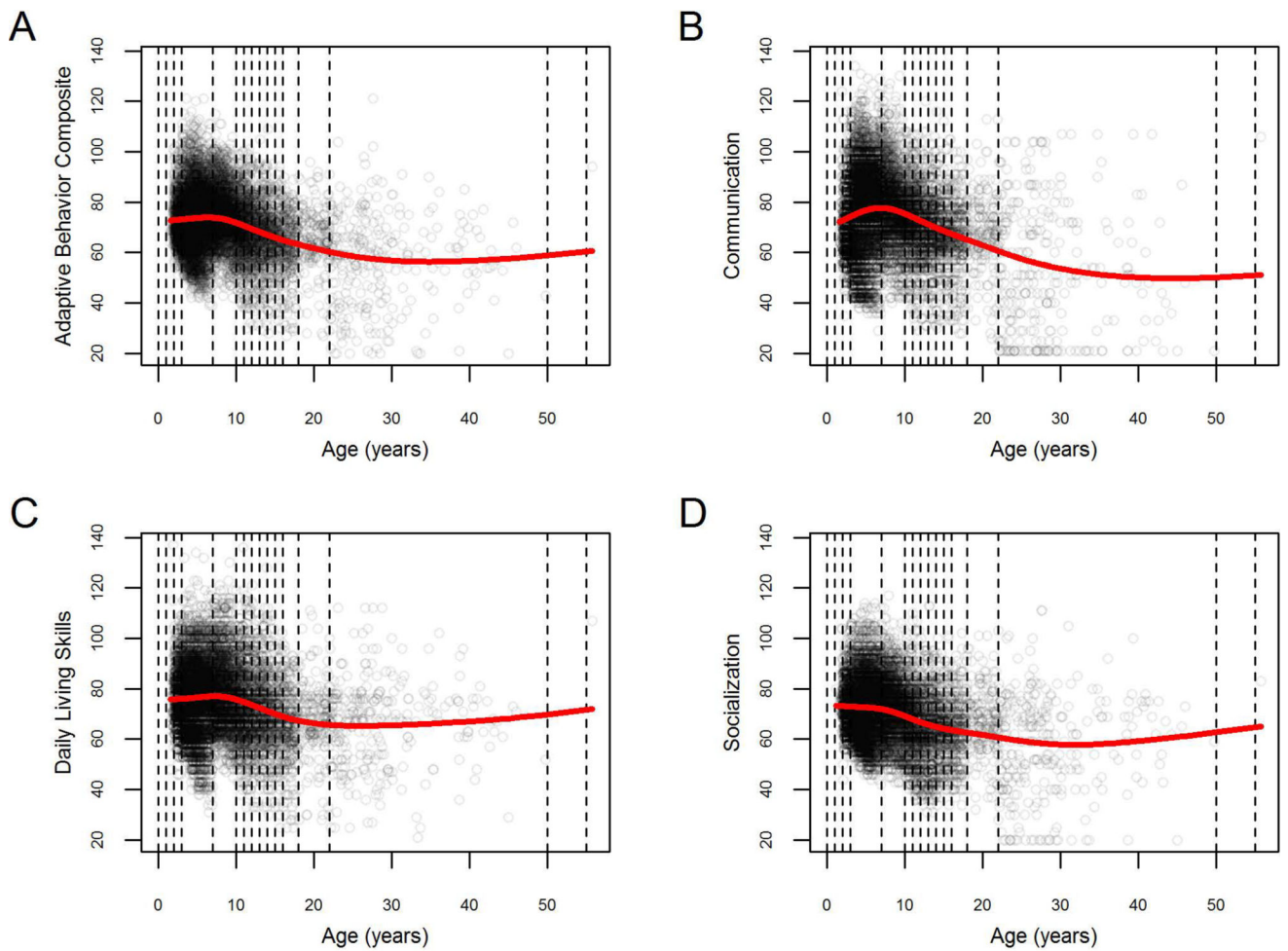depicts the Adaptive Behavior Composite and domain-level standard scores for Communication, Daily Living Skills and Socialization as a function of age in the combined sample. Points represent individual patients; vertical dotted lines demarcate the age bins of the normalization tables for Vineland-II standard scores (per Vineland-II user manual; Sparrow et al,. 2005); red line represents a smoothing spline.
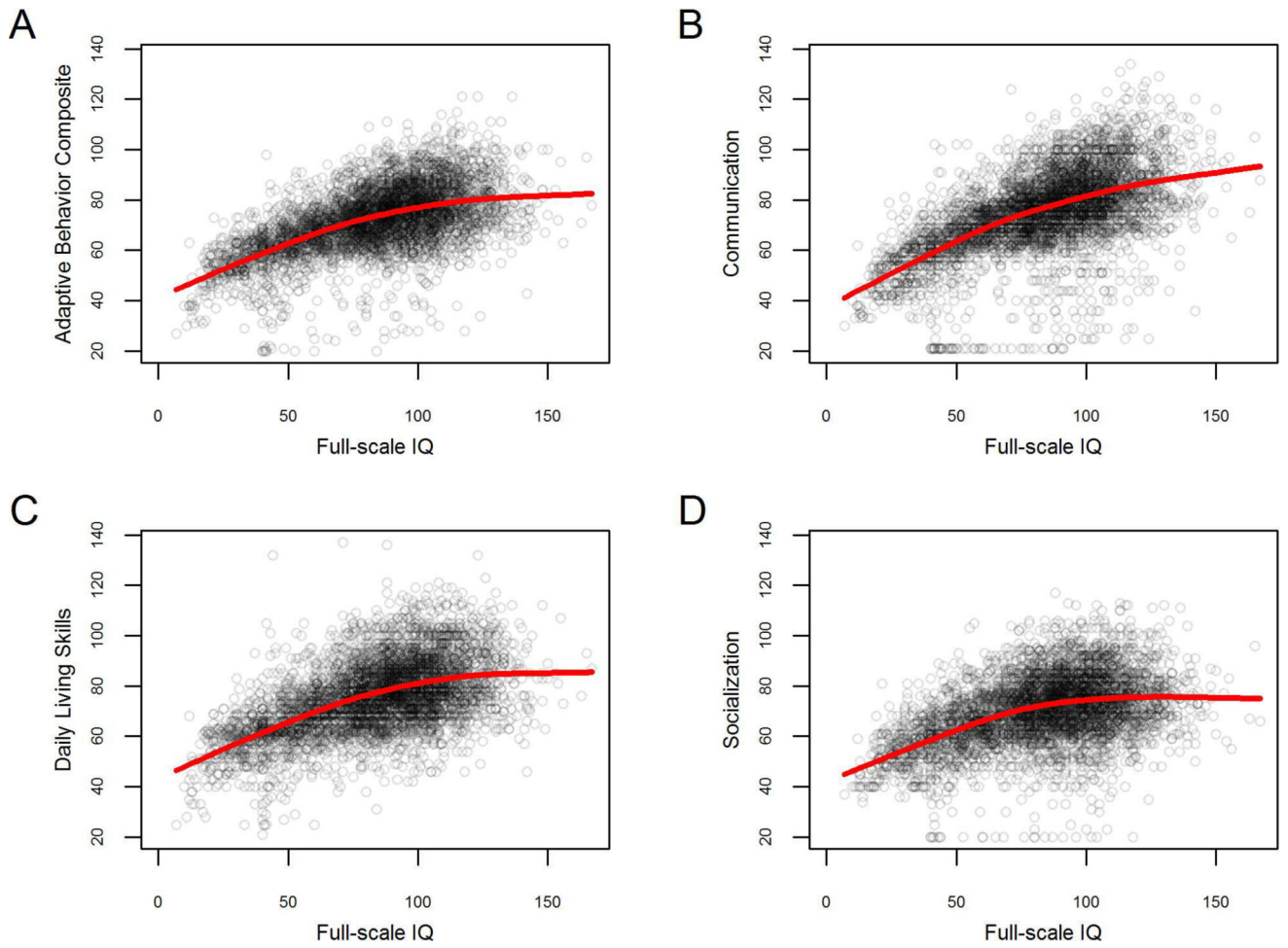
Figure 3.
A-D, depicting Vineland-II standard scores as a function of IQ (x-axis), demonstrating the importance of stratification by IQ; red line represents a smoothing spline.
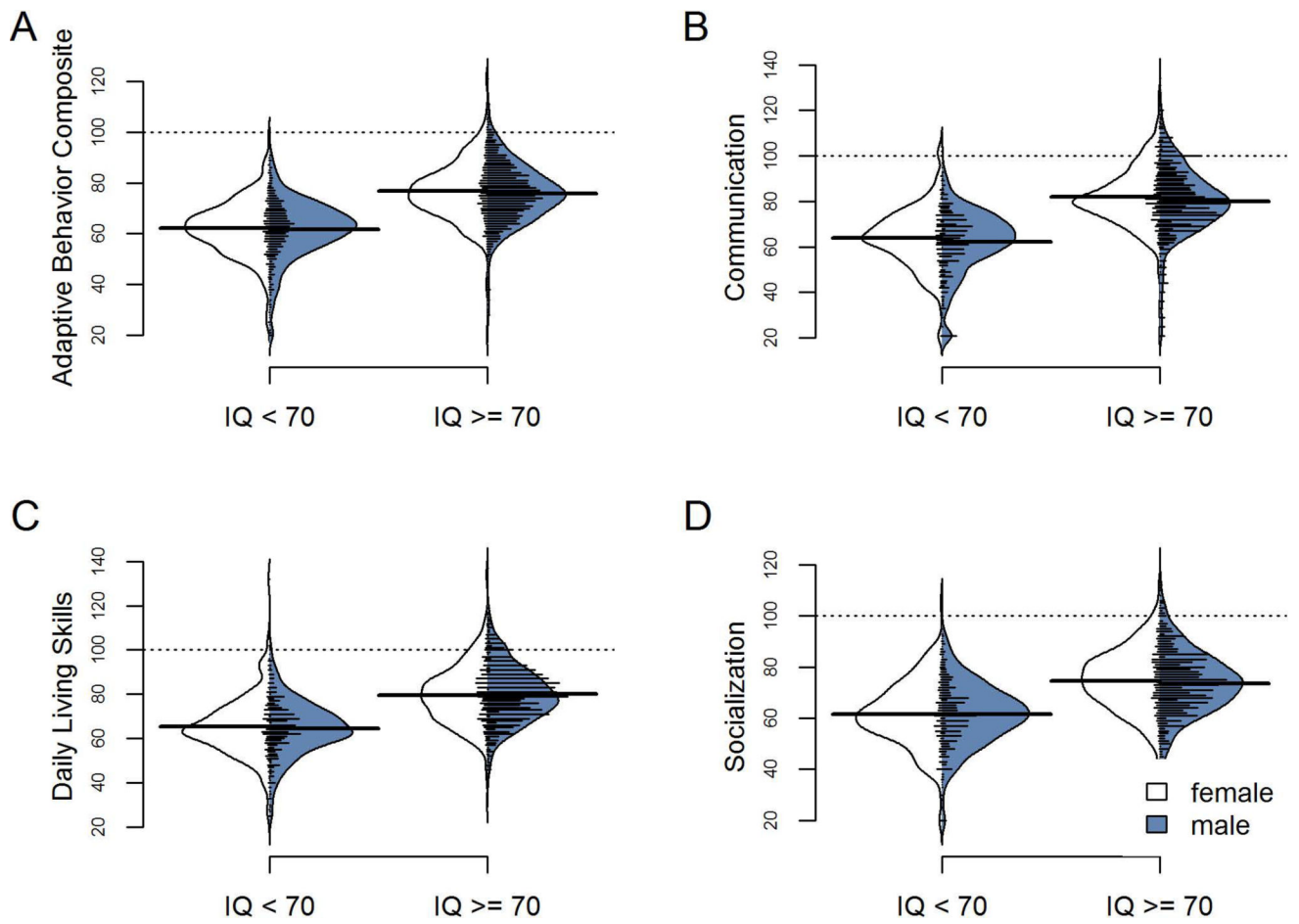
Figure 4.
depicts the distributions of Vineland-II standard scores as a function of sex and IQ; solid horizontal lines reflect sex-specific averages; the dotted line reflects the typically-developing mean of 100 in each Vineland-II standard score. Sex has minimal influence on the observed distributions, whereas the presence of cognitive impairment has a strong effect on the mean score.

Table 1

Descriptives for pooled individual-level data.

Dataset	N w/ Vineland	% male (n)	Data Type*	% w/follow-up	mean Age (range)	mean IQ (range)	% w/ IQ (n)	% w/follow-up & IQ (n)
AS ATN	5312	83% (4384)	L & C	7% (368)	6.18 (1.16 – 17.76)	84.49 (30 – 150)	16% (856)	2% (107)
SSC	2747	86% (2373)	C	n/a	9.03 (4 – 18)	81.17 (7 – 167)	100% (2743)	n/a
BP28420	207	100% (207)	L & C	90% (186)	25.46 (18.19 – 45.57)	98.27 (70 – 143)	98% (203)	90% (186)
LEAP	206	70% (145)	C	n/a	17.36 (6.24 – 30.15)	98.16 (40 – 148)	99% (204)	n/a
Fragxis	104	80% (83)	L & C	99% (103)	23.68 (14.08 – 49.73)	51.83 (38.71 – 81.69)	91% (95)	90% (94)
INFOR	81	86% (70)	L & C	86% (70)	17.28 (6.07 – 55.74)	93.02 (40 – 143)	47% (38)	36% (29)
ABIDEII	74	89% (66)	C	n/a	8.88 (5.13 – 34.76)	103.76 (67 – 143)	99% (73)	n/a
ABIDEI	68	85% (58)	C	n/a	14.91 (7.13 – 39.1)	108.35 (78 – 148)	100% (68)	n/a
Ox Simons/Stanford	62	76% (47)	C	n/a	8.14 (3.08 – 13)	84.93 (40 – 140)	92% (57)	n/a
Stanford PRT-G	48	79% (38)	L & C	79% (38)	4.19 (2.3 – 6.9)	52.68 (19.37 – 83.12)	90% (43)	75% (36)
Stanford PRT-P	42	88% (37)	L & C	93% (39)	3.87 (2.17 – 5.92)	46.73 (20.45 – 99.11)	88% (37)	81% (34)
YALE-ACE	42	100% (42)	C	n/a	13.19 (8 – 18)	101.58 (68 – 167)	100% (42)	n/a
YALE_PRT	28	57% (16)	L & C	100% (28)	6.14 (4.5 – 9)	97.5 (44 – 127)	100% (28)	100% (28)
BP27801	19	100% (19)	C	n/a	24.02 (18.64 – 40.86)	99.95 (78 – 136)	100% (19)	n/a
BP28421	19	100% (19)	C	n/a	25.41 (19.39 – 39.76)	102.26 (71 – 130)	100% (19)	n/a
PNO_Stanford	8	88% (7)	L & C	25% (2)	21.81 (18.08 – 33.33)	46 (40 – 52)	12% (1)	0% (0)
Total	9067	84% (7611)	L & C	9% (834)	8.22 (1.16 – 55.74)	83.56 (7 – 167)	50% (4526)	6% (514)

*
L=longitudinal, C=Cross-sectional

Table 2

Sample size for anchor- and distribution-based analyses, by age and presence of IQ scores.

	Anchor based	Distribution based
# individuals (all ages)	421 (OACIS-/CGI-I), 627 (OACIS-/CGI-S)	9067
# individuals (all ages) w/ IQ	365 (OACIS-/CGI-I), 527 (OACIS-/CGI-S)	4526
# children (<13 years) w/ IQ	91 (OACIS-/CGI-I), 160 (OACIS-/CGI-S)	3398
# adolescents (13–18 years) w/ IQ	38 (OACIS-/CGI-I), 49 (OACIS-/CGI-S)	695
# adults (18+ years) w/ IQ	236 (OACIS-/CGI-I), 318 (OACIS-/CGI-S)	433

Table 3

Descriptive statistics of Vineland-II standardized scores and full-scale IQ in ASD.

Score Type	Mean (95% CI)	SD (95% CI)	Skew (95% CI[*])	Excess Kurtosis (95% CI[*])
Adaptive Behavior Composite	71.58 (71.32–71.84)	12.63 (12.45–12.82)	−0.12 (−0.19–0.05)	0.83 (0.67–0.99)
Socialization	70.2 (69.94–70.46)	12.76 (12.58–12.95)	−0.07 (−0.15–0)	0.62 (0.46–0.79)
Daily Living Skills	75.05 (74.76–75.35)	14.39 (14.18–14.6)	0.08 (0.02–0.14)	0.37 (0.23–0.5)
Communication	74.19 (73.86–74.53)	16.32 (16.09–16.56)	−0.15 (−0.21–0.09)	0.4 (0.3–0.5)
Motor Skills	80.65 (80.29–81.01)	13.6 (13.35–13.86)	0.35 (0.28–0.43)	0.34 (0.15–0.51)
Full-scale IQ	83.56 (82.77–84.34)	27.01 (26.47–27.58)	−0.34 (−0.39–0.29)	−0.32 (−0.42–0.23)

[*] 95% CIs obtained by bootstrapping.

Table 4

Sample size-weighted MCID estimates for the Adaptive Behavior Composite score and each domain standard score, stratified by age and IQ.

	Sample Size-Weighted Anchor-based MCID estimates		Sample Size-Weighted Distribution-Based MCID estimates	
	IQ <70	IQ ≥70	IQ <70	IQ ≥70
Adaptive Behavior Composite				
Children (0–13)	2.5	2.9	2.0	2.7
Adolescents (13–18)	2.4	n/a*	2.5	2.8
Adults (18+)	2.8	3.8	3.0	3.4
Communication Domain				
Children (0–13)	2.0	3.1	2.7	3.2
Adolescents (13–18)	3.2	n/a*	2.8	3.3
Adults (18+)	3.1	3.3	3.8	5.4
Socialization Domain				
Children (0–13)	2.6	3.1	2.9	3.7
Adolescents (13–18)	2.9	5.5	3.2	4.0
Adults (18+)	3.0	5.5	4.8	4.8
Daily Living Skills Domain				
Children (0–13)	2.6	3.7	2.9	3.6
Adolescents (13–18)	1.7	n/a*	3.5	4.1
Adults (18+)	3.0	5.1	3.3	3.3

*
Among studies including adolescents with IQ ≥70, anchors met the inclusion criterion of a correlation above .3 only with Socialization standard scores (see Methods for further details on the correlation criterion). Consequently, anchor-based MCIDs for Daily Living Skills, Communication, and the Adaptive Behavior composite are not available (n/a) for this stratum.