

Title: Vineland-3 growth scale value scores as an outcome in studies of intellectual disability: A simulation study

Authors: Cristan Farmer¹ and Audrey Thurm¹

Introduction: The use of standard scores, which represent the comparison of an individual's level of ability to that of their chronological age-peers, carries several limitations when used in studies of intellectual disability. Clinically relevant improvements in ability may still yield a worsening standard score, if the degree of change is not commensurate with that exhibited by same-age peers. Additionally, the reliability of standard scores is compromised at the extremes of the distribution, which means that minor differences in raw scores can yield large shifts in standard score. Item-response theory (IRT)-based person ability scores are a direct index of ability, and are designed to evaluate within-person change. However, ability scores are unfamiliar to many users and have unknown distributional properties in the population. In this poster, we explore the relative merits and limitations of various score types for use as an outcome measure in clinical trials involving participants with intellectual disability.

Method: We simulated datasets from randomized controlled trials, varying the inclusion criteria (age range and level of impairment) and treatment effect size. The outcome measure of interest is the Vineland Adaptive Behavior Scales-Third Edition, which produces norm-referenced V-scale scores and ability (growth scale values; GSV) scores at the subdomain level. For each factorial combination of conditions (age, level of impairment, and effect size on the V-scale), 10,000 datasets comprised of randomly sampled age and V-scale scores are constructed. The resulting pre- and post-test V-scale scores are back-transformed into GSV scores. Finally, the effects of treatment are evaluated using an ANCOVA model of post- intervention score controlling for pre- intervention score. The standardized mean difference for the GSV is compared to that of the V-scale, which is itself a manipulated value.

Results: The mean effect size estimate (with 95% CI) over the 10,000 datasets, for each set of conditions, will be plotted for each type of score (V-scale and GSV). If the GSV confers greater power to detect a small effect, the effect size for GSV will be larger and/or have narrower confidence intervals than that of the V-scale score.

Discussion: In theory, we expect that ability scores offer some statistical and methodological advantages over norm-referenced scores, when used in studies of intellectual disability (Farmer et al., In Press). However, datasets sufficiently large to evaluate this claim do not exist. The results of this simulation study will help researchers determine whether or under which conditions ability scores may be used as outcomes in clinical trials of intellectual disability.

References: Kaat, A., Thurm, A., Anselm, I., Akshoomoff, N., Bennett, A., Berry, L., ... & Cecil, K. M. (In press). Person ability scores as an alternative to norm-referenced scores as outcome measures in studies of neurodevelopmental disorders. *American Journal on Intellectual and Developmental Disabilities*.

¹Neurodevelopmental and Behavioral Phenotyping Service, Intramural Research Program, National Institute of Mental Health