Precision Matters: An analysis of how various scores behave when measuring change over time; factors that inform score selection for the best results.

Submitter Lynsey Psimas

Affiliation Pearson

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

Poster PDF for download ISCTM Poster DRAFT_LP_PW_2024 (1).pptx

Methodological Issue Being Addressed We will explore the impact of selecting Raw Scores, Standard Scores, Age Equivalents, and Growth Scale Values to identify the most precise metric to measure change. Through simulated hypothetical results, we will highlight the importance of carefully selecting metrics for data analysis and demonstrate the effect on interpretation.

Introduction Choosing the appropriate score to measure change over time in research is challenging, yet critical, to the success of a clinical trial. Selecting the incorrect score type can result in misleading results and faulty conclusions.

We will explore the impact of selecting theses various scores to identify the most precise metric to measure change. Specifically, our analysis will highlight the impact of score selection on study outcomes including changes in raw scores/age equivalents as equal GSV's increase, floor effects, conditional SEM vs. group SEM, and out of level testing.

This analysis illustrates the use and interpretation of several types of scores commonly used in clinical trials. Depending on the purpose, population, and design of a study, Growth Scale Values may provide greater statistical power as a study outcome because they are designed to measure change and because they are equally sensitive to differences at low and average levels of ability.

Methods Through simulated case studies and figures, we will demonstrate the implication of score selection. We have illustrated various scores including age equivalents, raw scores, standard scores, growth scale values, and the implications of using these scores to measure change over time.

For the purposes of illustration, data were drawn from two clinical outcome measures (COAs), the Vineland Adaptive Behavior Scale, Third Edition (Vineland-3) (2016) and the Bayley Scales of Infant and Toddler Development, Fourth Edition (2019). By analyzing score types for these two COAs, we are able to illustrate the importance of carefully selecting the metric used.

Data analysis highlight the impact of various scores on several hypothetical scenarios including changes in raw scores/age equivalents as equal GSV's increase, floor effects, conditional SEM vs.

group SEM, and out of level testing.

Results This poster illustrates an analysis of the impact of various scores frequently used with endpoint measures in clinical trials including Raw Scores, Standard scores, Age equivalents, and Growth Scale Values:

1) Raw scores are widely used in research. While useful, raw scores do not include interval scaling and come with a single group-based SEM.

2) Standard Scores are affected by population change and can be misleading when measuring change that is slower than typical. Norm-referenced scores often decrease over time among people with neurodevelopmental disorders who exhibit slower-than-average increases in ability. Further, the reliability of norm-referenced scores is lower at the tails of the distribution, resulting in floor effects and increased measurement error for people with neurodevelopmental disorders (Farmer et al., 2020).

3) Age Equivalents are often used to track change but have limitations. For example, age equivalents lack of SEMs to determine statistically significant difference. In addition, there are inaccuracies at very low levels, where children in clinical trials often fall. Finally, Age Equivalents are bound to the normative age range, making out of level testing impossible.

4) GSV's demonstrate a viable alternative of adjunctive to norm referenced scores. Growth Scale Values contain interval scaling and known conditional SEM’s, allowing researchers to measure outside of the developmental age range, unlike age equivalents and standard scores. In addition, raw scores and age equivalents can demonstrate wildly different amounts of change across ability levels. For this reason, when a clinical trial includes children with varied developmental levels, averaging the change in raw scores or age equivalents is problematic because you are averaging children with higher and lower amounts of change. Since GSV's offer equal interval scales, this problem is avoided.

Conclusion In research studies, absolute scores, such as raw scores, have advantages over normative scores.

Age equivalent (AE) scores have been the standard metric used in this type of research. While AEs are familiar and interpretable to clinicians and parents, they are imprecise due to lack of standard deviations, standard errors of measurement, and equal intervals between scores. Raw scores have unequal intervals and are not comparable between ages or ability levels. Norm-referenced scores, such as standard scores, do not yield information about gain, stability, or loss of skills, they are not suitable for natural history studies or clinical trials.

Alternatively, Growth Scale Values (GSV’s) offer an accurate measure of within-person change.

GSV’s circumvent inaccurate AEs, are not limited by age, and can be used for out of level testing. GSVs have interval properties and each GSV values has a known SEM. GSVs are recommended to measure change in cognitive and adaptive behavior in natural history studies and in clinical trials for patients with low, flat, or declining performance on outcome measures.

Co-Authors

* Presenting Author

First Name	Last Name	Affiliation
Lynsey *	Psimas *	Pearson
Paul	Williams, PsyD	Pearson

Keywords

Keywords
Growth Scale Values
COA
endpoint
outcome measures
ability score

Guidelines I have read and understand the Poster Guidelines

Disclosures I am an employee of Pearson Clinical Assessments, where I earn a salary for my work.

Related Tables and Supporting Materials Farmeretal.2020.pdf