Objectives

- Assessment of neurocognitive development during the first 1000 days after birth is important, particularly in children in low- and middle-income countries (LMIC).
- Various instruments are used for these assessments, mostly based on a defined set of tasks for the child to perform.
- Tasks typically are scored as a set of ordered categories (e.g., pass/fail).
- Development score (D-score) may integrate data collected using different scales and across different populations.
- The purpose of this work was:
  - To evaluate the assumptions underlying the D-score using data from an LMIC population, and
  - To assess whether the D-score can be used for between-population comparisons.

Methods

Data:
- 2 studies in high-income countries (HIC): ~2000 (Study 1) and ~500 children (Study 2).
- 1 study in an LMIC: ~1900 children.
- All 3 studies: birth to age 2 y.
- Instruments for assessing neurocognitive development differed between studies.
  - The instrument used in the HIC studies included 56 items ("full-set").
  - The instrument used in the LMIC study included 40 items.
- The two instruments had 35 items in common ("matched-set").

Statistical methods:
- A child’s D-score was connected to observed longitudinal outcomes through Rasch model (item-response theory model) (Figure 2).
- Specifically, the probability of a positive response (i) to each item (X) was modeled as a function of the difference between a child’s ability (θ) and an item-level difficulty (τ): Pr(Xi = 1 | θi, τj).
- The item-level difficulty values (τ) were previously estimated using data from one of the HIC studies.
- A child’s D-score at each age was derived as the expected posterior estimate of θ conditional on the item response (X) and a weakly informative, age-dependent prior distribution.

Rasch model assumptions:
- (1) Invariance to the set of items used.
- (2) Common item-level difficulty across populations.
- (3) Items vary only in difficulty (parallel curves).
- Assumption of parameter invariance evaluated by comparing estimated D-scores obtained using the full-set and matched-set of items in the HIC studies.
- Discrimination plots made to compare item difficulty and item discrimination across studies (assumptions 2 & 3).
- Longitudinal D-scores (raw and standardized) were compared between study populations.

Results

- Comparison of the D-score calculated using the full- and matched-set of items in HIC studies (Figure 3).
- High correlations both overall and by age indicated that the D-score may be invariant to set of items.
- Thus, calculation of the D-score using the matched set of items in the LMIC study could proceed.

- Discrimination plots (Figure 1): Item-level difficulty similar across these HIC and LMIC populations for most items (overlapping curves).
- Some items appear to be more difficult in LMIC populations (e.g., says mom/dad).
- Items in LMIC population appear to differ only by difficulty (parallel curves in Figure 4).

Conclusions

- D-score shows promise for facilitating comparisons across populations.
- D-score has not been clearly validated for this purpose.
- Discrimination plots are a useful tool for evaluating assumptions of the Rasch and other IRT models.
- D-score was invariant to choice of items, but item-level difficulties may depend on the population and/or instrument used.
- Additional work is needed to further evaluate D-score, including comparisons using additional LMIC and HIC populations and neurocognitive development instruments.

References


Acknowledgments

Children, families, and investigators who contributed data to this investigation.