HEALTHY BIRTH, GROWTH & DEVELOPMENT

A Case Study in Comparing Cognitive Development Across Populations

Jonathan L. French,[1] Amy Racine-Poon,[2] Stef van Buuren,[3] Jonas Häggström,[4] and the Lifecycle, Auxology, & Neurocognitive Development Surge Team[5]

[1]Metrum Research Group; [2]Novartis, Inc.; [3]TNO; [4]Holy Diver Consulting; [5]Sponsored by the Bill & Melinda Gates Foundation, Healthy Birth, Growth, and Development Initiative

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.[1,2]**
- **The purpose of this work was:**

(1) To evaluate the assumptions underlying the D-score using data from an LMIC population, and

(2) To assess whether the D-score can be used for between-population comparisons.

Figure 1: Discrimination Plots for a Sample of Items

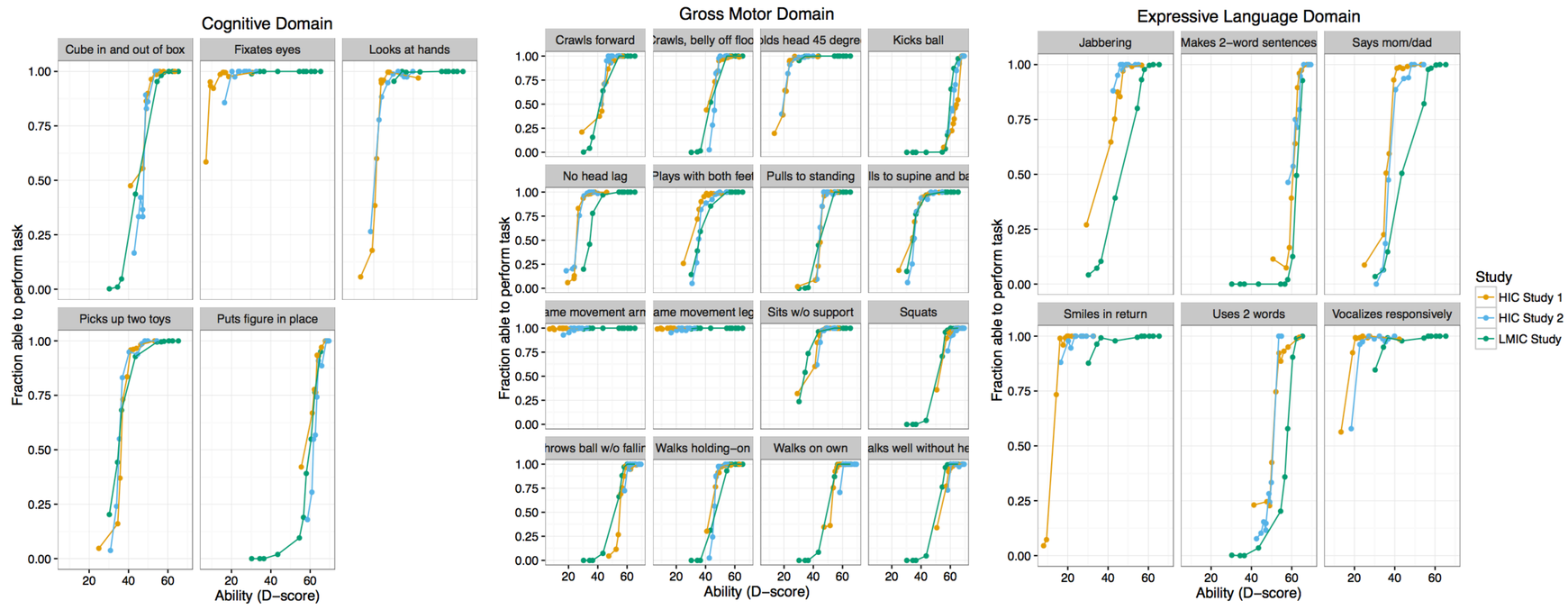


Figure 2: Rasch Model

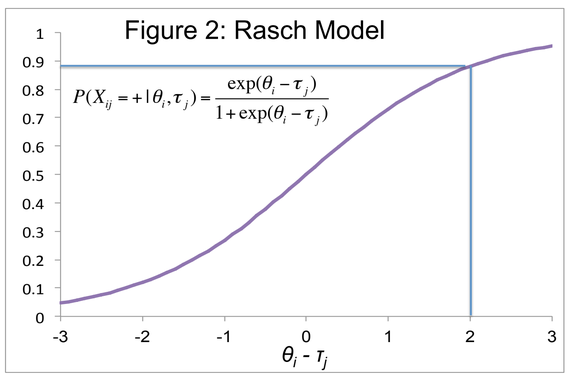


Figure 3: Scatterplots of D-scores based on matched- and full-sets of items in HIC Study 1

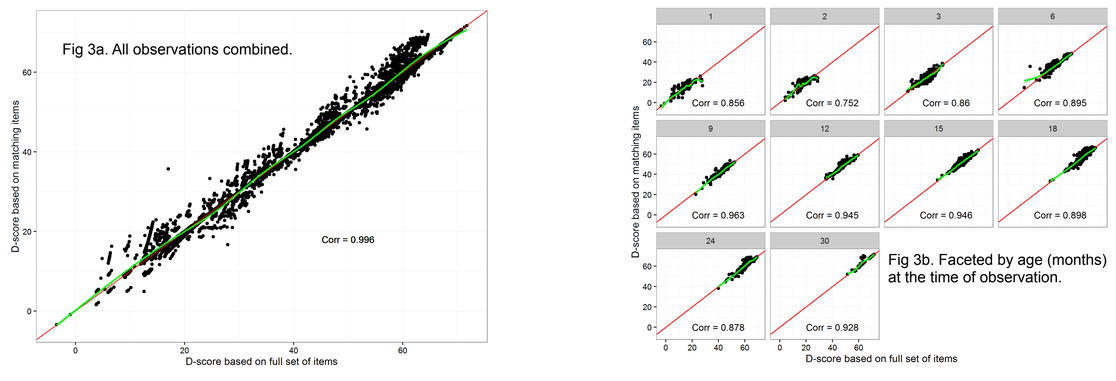


Fig 3b. Faceted by age (months) at the time of observation.

Figure 4: Discrimination plot all items in LMIC study

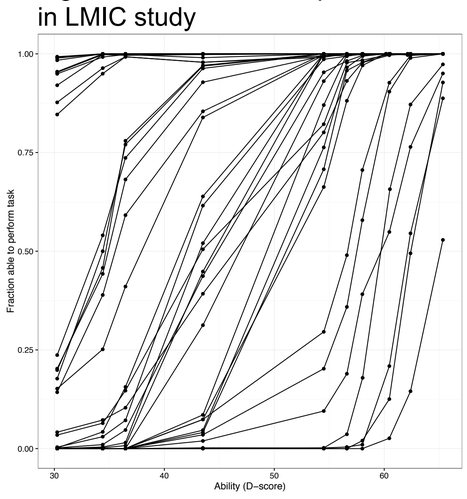


Figure 5. Mean D-score vs. Age

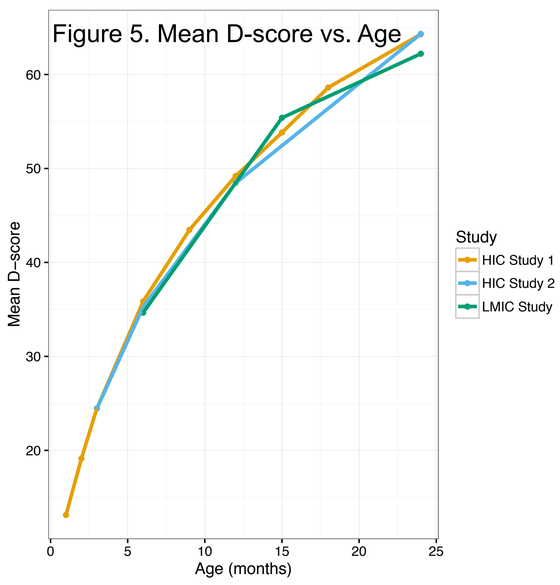
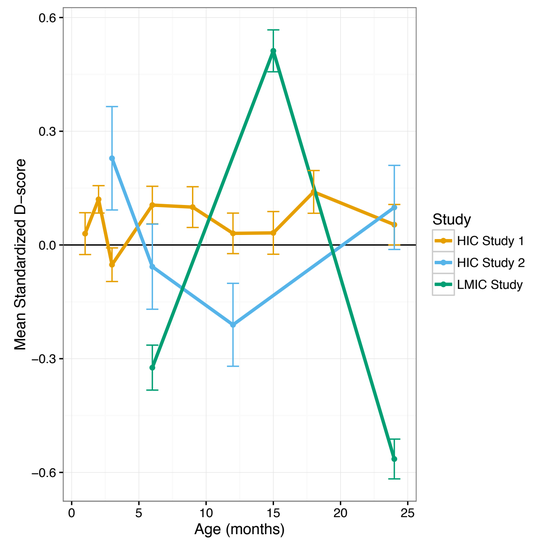


Figure 6. Mean standardized D-score vs. Age.



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 284 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[1] (an item-response theory model) (Figure 2).
- Specifically, the probability of a positive response to each item (X_{ij}) was modeled as a function of the difference between a child’s ability (θ_i) and an item-level difficulty (τ_j).
- The item-level difficulty values (τ_j) were previously estimated using data from one of the HIC studies.[1,2]
- A child’s D-score at each age was the derived as the expected *a posteriori* estimate of θ_i, conditional on the item responses, the τ_j 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 based on 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).

- D-scores in all 3 studies increased consistently as children matured. (Figure 5)
- Average standardized scores were lower at 6 and 24 mo and higher at 15 mo in LMIC than HIC. (Figure 6)

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.

Acknowledgments

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

References

1. Jacobusse G, van Buuren S, Verkerk PH. An interval scale for development of children aged 0-2 years. *Stat Med.* 2006;35(13):2272-2283.
2. van Buuren S. Growth charts of human development. *Stat Methods Med Res.* 2014;23(4):346-368.

Sponsored by the Healthy birth, growth and development (HBGD) initiative representing Discovery & Translational Sciences, Integrated Development, Integrated Delivery, Nutrition, Agriculture, Water, Sanitation & Hygiene, Maternal, Newborn & Child Health, Enteric & Diarrheal Diseases, and Pneumonia Project Strategy Teams. HBGD*ki* was conceived of and is directed by shasha.jumbe@gatesfoundation.org