In-depth analyses of longitudinal patterns from the Bayley Scale of Infant Development (Bayley-III)

M. ten Hooven – Radstaake¹, A. Sadhwani², A.C. Wheeler³, M. DeRamus⁴, A. Rietman⁵, S.E. Mous⁶, L. Hiruma⁷; C. Okoniewski⁸.

¹, ⁵, ⁶. ENCORE Erasmus MC, Sophia Children’s Hospital, Dr. Molewaterplein 40, 3015 GD Rotterdam, the Netherlands.
². Boston Children’s Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.
³, ⁴, ⁷, ⁸. UNC, 101 Renee Lynne Ct. Carrboro, Chapel Hill, NC 27510, USA.

The Bayley-III is often used to assess the development of individuals with AS. However limited information is known about the longitudinal patterns over time. Using data from three sites which closely follow individuals with AS (BCH, UNC and Erasmus), we assessed patterns of responding in 53 children with AS. We analyzed differences in raw score and age equivalents between assessments. We also conducted stability analyses by analyzing the number of pass-to-fail items (items done correct in the first and wrong in the second assessment) and fail-to-pass items (items done wrong in first and correct in second assessment) in children who increased, unchanged or declined their raw score.

Results: Preliminary analysis shows that less than half of the children improved their age-equivalent with more than three months, within a median time between assessments of 12–22 months (depending on subscale). Differences between scales and several individual deviations were found. A clear association between “time between assessment” and “difference in raw score” was not found. Only a few children progressed twice in a row.

Overall children who increase their scores were younger. Analysis on the scoring patterns revealed many instabilities, for instance on the Cognitive scale 17/27 children who increased their raw score had pass-to-fail items and 7/16 children who declined their raw score had fail-to-pass items. Pass-to-fail items were found in both relative difficult and easy items. There were differences between the scales in the degree of stability between assessments. We will discuss how the differences and instabilities in scoring over time can be explained and also consider differences between subtypes. We will debate what these results mean for the reliability and validity of the Bayley-III as an outcome measure. Based on our results we will propose recommendations that should be taken to increase both its validity and reliability.