

Data Quality Monitoring of Bayley Assessments: Error Detection in Rare Neurodevelopmental Disorder Trials

Laura Miller, PhD¹, Whitney Wood, PhD¹, and Pam Ventola, PhD^{1,2}

¹Cogstate; ²Yale Child Study Center

Methodological Issue Addressed

Ensuring interpretable, reliable developmental outcome data in rare-disease trials, where small sample sizes magnify the impact of even minor administration or scoring errors can be challenging. This is particularly problematic for the Bayley-III, given its procedural complexity and multiple scoring decision points that increase vulnerability to avoidable variability.

Background

Neurodevelopmental outcome measures such as the Bayley-III require high inter-rater reliability and strict adherence to standardized administration procedures. In trials of rare disorders, decentralized data collection across sites with variable assessor experience introduces risk for administration and scoring errors. While centralized data quality monitoring approaches are increasingly adopted, relatively few studies describe how identified errors are evaluated, corrected, and quality-controlled at the reviewer level.

This analysis examines error frequency, magnitude of score change, and type of error that led to a score change across Bayley-III administrations.

Method

Data were aggregated from two rare neurodevelopmental disorder clinical trials. Each administration underwent structured data quality review by trained clinical reviewers.

Reviewers completed standardized training and annual recalibration against gold-standard reviews, with all reviewers meeting the predefined performance threshold of >90% agreement.

Method

Reviewers reviewed record forms and video administrations of the Bayley to evaluate the following:

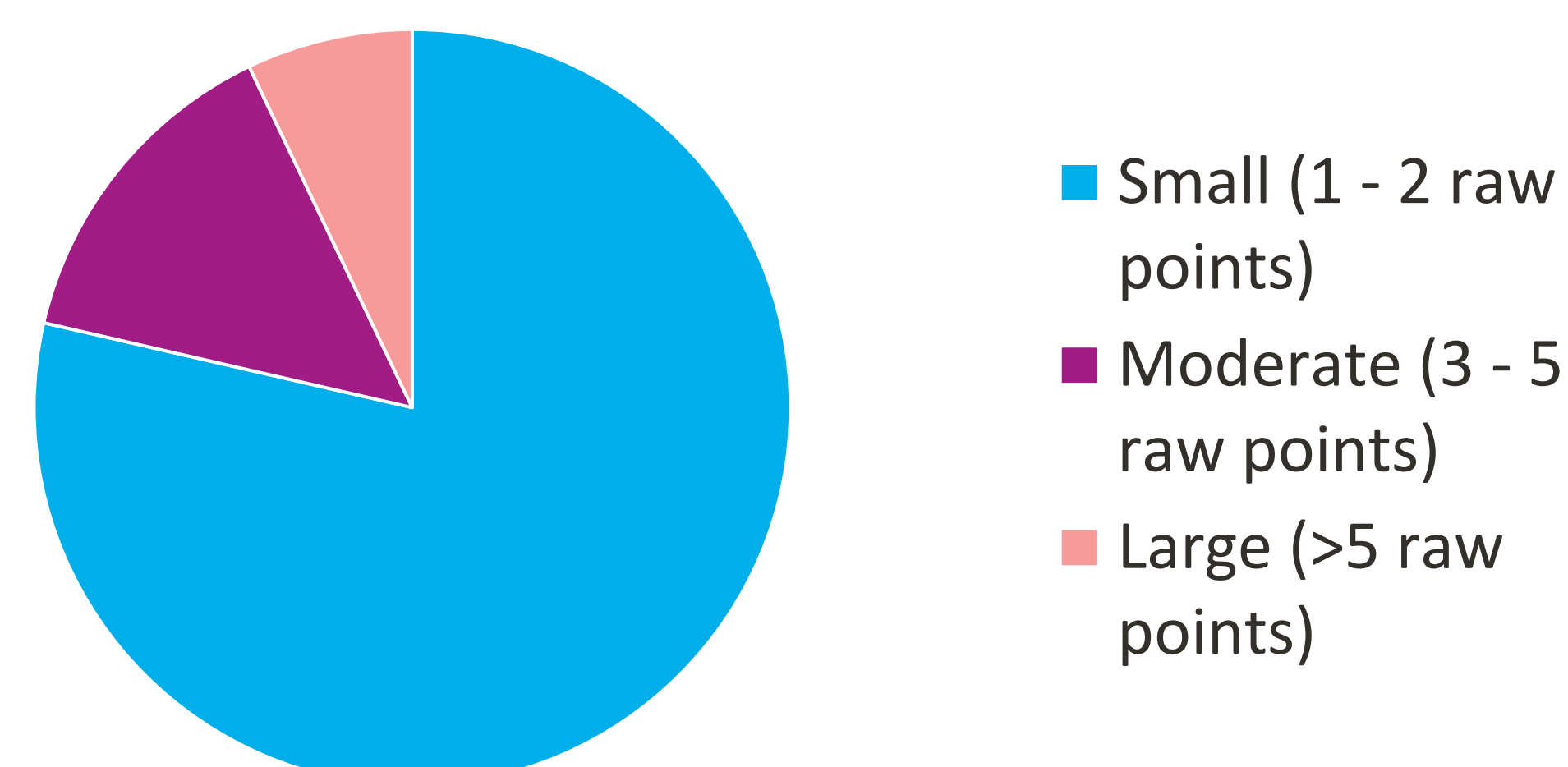
- Correct start point used
- Basal Established
- Ceiling Established
- All items scored between Basal and Ceiling
- Raw scores calculated correctly
- Correct administration procedures
- Item delivery accuracy
- Verbatim prompts
- Scoring fidelity
- Appropriate engagement (e.g., rapport/bedside manner)

Descriptive analyses were performed across administrations. The primary endpoint was the proportion of administrations requiring score correction following data quality review.

Results

Across 133 total administrations (Trial A: $n = 54$; Trial B: $n = 79$), 51.13% ($n = 68$) contained no identified errors, while 48.87% ($n = 65$) had at least one error. Data quality monitoring identified 29 administrations requiring score correction, representing 21.80% of all administrations and 44.62% of administrations with at least one identified error.

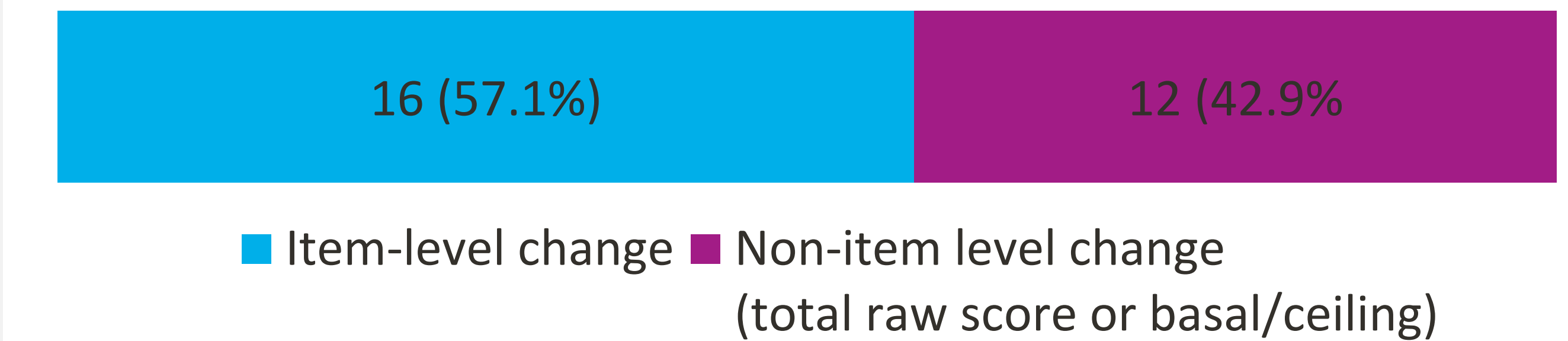
Score correction magnitude was evaluated to assess potential impact. One corrected administration reflected an identified error in which a missing score was assigned a value of zero and was excluded from magnitude categorization due to the absence of a numerical score change. Among the remaining 28 numerical corrections, 22 (78.57%) were categorized as small (1–2-point change in raw score), 4 (14.28%) as moderate (3–5-point change), and 2 (7.14%) as large (>5-point change).



Results

Score change corrections were then further broken down to determine how many administrations with a score change correction included an item level score change. Of 28 administrations with a score change, 16 administrations had at least 1 item level score change. The remaining 12 had score changes due to incorrectly adding the total raw score or incorrect calculations based on basal and ceiling rules.

Score Change Administrations ($n = 28$)



Conclusion

1. A structured data quality monitoring framework systematically identified and resolved Bayley administration and scoring errors.
2. Quantifying correction magnitude and applying formal reviewer calibration improved methodological rigor, reviewer reliability, and data interpretability.
3. Findings support integrating structured data quality monitoring with reviewer reliability.

References

- Bayley N. The Bayley scales of infant and toddler development. 3rd edn. San Antonio, TX: Harcourt Assessment, Inc, 2006.
- Berg, A. T., Ludwig, N. N., Wojnaroski, M., Chapman, C. A., Hommer, R., Conecker, G., ... & Downs, J. (2024). FDA patient-focused drug development guidances: considerations for trial readiness in rare developmental and epileptic encephalopathies. *Neurology*, 102(1), e207958.
- Phillips, D., Wilson, C., & Fernandez, V. (2026). Defining Bayley Scales of Infant and Toddler Development Third Edition (BSITD-III) meaningful change and item relevance in children with neuronopathic MPS II: a caregiver interview-based study. *Orphanet Journal of Rare Diseases*.

