

Comparative Performance of Video and Record-Form Review Methods for Bayley Administration Fidelity

Submitter Whitney Wood

Affiliation Cogstate

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

I have used the poster abstract template to develop my abstract. Yes

Methodological Issue Being Addressed Ensuring the fidelity of Bayley administrations is essential to data integrity in rare neurodevelopmental disorder (NDD) trials. Central reviewers typically assess administration quality using either video review, which captures full examiner-participant interactions, or record-form review, which focuses solely on scoring and documentation. It remains unclear which review method is more effective at identifying errors or capturing the full scope of administration quality. This study directly compares the two approaches to determine which identifies more errors in clinical trial settings.

Introduction The Bayley Scales of Infant and Toddler Development are commonly used as developmental endpoints in rare NDD trials where sample sizes are small and measurement error has a large impact on interpretability. Administration errors such as inaccurate start or stop points, incorrect item delivery, and scoring inaccuracies can threaten endpoint validity. While record-form review offers insight into scoring and documentation accuracy, it cannot evaluate examiner behavior. In contrast, video review allows full observation of administration procedures but requires greater resources. Understanding the relative sensitivity of each method is essential for optimizing central-review strategies and ensuring high-quality developmental data.

Methods Bayley-III and Bayley-IV administrations (N = 520) from several ongoing and completed NDD trials were independently evaluated using two approaches: Record-Form Review Only (n = 408 administrations) and Video Review (n = 112 administrations). Administration review type was determined by trial-specific availability and monitoring plans. Each administration was reviewed using either record form review or video review only. No administrations received both review types.

In both review approaches, trained central reviewers evaluated the following: correct start point used, basal established, ceiling established correctly, all items scored between basal and ceiling, and raw scores correctly calculated. For the Video Review administrations only, central reviewers also evaluated components related to correct administration procedures, item delivery accuracy, verbatim prompts, scoring fidelity, and appropriate engagement (e.g., rapport).

For our analyses, we classified each administration dichotomously as having errors or no errors. The two methods were compared using an unadjusted chi-square test of independence. Analyses did

not adjust for protocol, age group, Bayley version (III vs. IV), site, or reviewer level variability. Secondary descriptive analyses examined proportion of record form errors vs. performance-based errors identified in Video Review administrations. Confidence intervals were calculated using the Wilson score intervals as appropriate for binomial proportions.

Results We found a significant association between review method and error detection, $\chi^2(1) = 18.90$, $p = .000014$. Video review identified a substantially higher proportion of administration errors (55 of 112 administrations; 49%; 95% CI: 40% - 58%) compared with record-form review (112 of 408 administrations; 27%; 95% CI: 23.3% - 32%). Of the 55 video review administrations with errors, 27 of 55 administrations (47%; 95% CI: 36.4% - 61.9%) contained errors identifiable through record-form review only (e.g., basal/ceiling errors, scoring calculation errors), while 28 of 55 administrations (53%; 95% CI: 38.1% - 63.6%) contained performance based errors that were detected only through video review (e.g., incorrect item delivery, scoring fidelity, verbatim prompts). These findings indicate that video administration review was more sensitive in this dataset to identifying administration and procedural errors than record-form review alone.

Conclusion Video review identified a higher proportion of Bayley administration errors than record-form review in detecting Bayley administration errors, particularly errors related to procedural deviations that cannot be captured through documentation alone. Record-form review remains valuable for identifying scoring inaccuracies, but reliance on record forms alone may underestimate administration error rates. These findings should be interpreted cautiously due to limitations such as non-randomized design, sampling imbalance, and unadjusted analyses, but generally these results suggest that video review may offer greater sensitivity for monitoring administration quality in NDD trials.

Co-Authors

Whitney Wood¹, Laura Miller¹, Jess Segura¹, Pam Ventola²

¹ Cogstate

² Cogstate, Yale Child Study Center

Keywords

Keywords
Bayley administration fidelity
Error detection
Neurodevelopmental disorder trials
Central Monitoring

Guidelines I have read and understand the Poster Guidelines

Disclosures None