

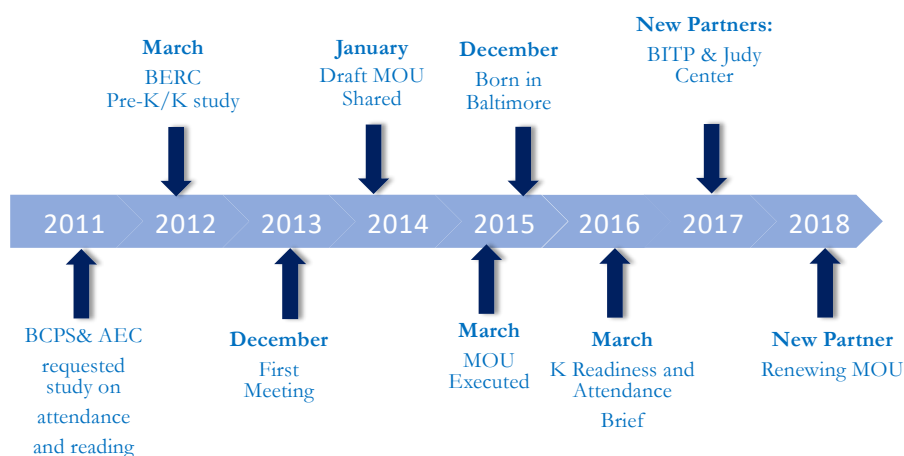
The History of The Baltimore Early Education Data Collaborative and Methodology Used

Faith Connolly, Jeffrey Grigg & Stephanie DeSouza

June 2018

The Baltimore Early Education Data Collaborative began after the release of BERC's report Early Elementary Performance and Attendance in Baltimore City Schools' Pre-Kindergarten and Kindergarten in March 2012. That report introduced numerous partners who worked in early education in Baltimore and the through conversations about how much we could learn if we shared our data, led to the creation of the Early Education Data Collaborative. This brief provides the technical background on the data and methods used in our work.

Figure 1. Timeline of the Early Education Data Collaborative.



Data Sources

In December 2013, BERC invited Baltimore City agencies serving young children and their families to form the Early Education Data Collaborative. By integrating information on a series of cohorts of children born to Baltimore City residents, we hope to understand the experiences of young children in Baltimore and how those experiences relate to later educational outcomes. Participating agencies included: Baltimore City Health Department, Baltimore City Public Schools, The Ark, Baltimore City Head Start, Family League, and Maryland Family Network.

Linking Records Through A Matching Procedure

Records were linked using child's first and last name and birth date, first, using exact matching methods. Although participating agencies generally use the official first and last names from the child's birth certificate, typos and data entry discrepancies prevented some records from being matched exactly. Initially unmatched records were then matched using a probabilistic (or "fuzzy") matching method to assign a score to each potential match from 0 to 1, with 1 indicating an exact match (Harron, Goldstein, & Dibben, 2016; Wasi & Flaaen, 2015). All candidate matches were required to match exactly on date of birth, the "fuzzy" protocol was applied only to names. All matches with a score above 0.6 were reviewed manually for accuracy.

of the match; matches below a score of 0.67 nearly always discarded. Analysis of a subsequent cohort resulted in comparable match rates. 64% of children born in Baltimore in the 2007-2008 cohort were identified in at least one other agency's files, and birth records were matched to 76% of the kindergarten students, 77% pre-kindergarten students, and 77% of all children who attended Head Start.

After the matching was performed, all personal information was removed from the data; only de-identified data were used for analysis.

Limitations

While it was required that children who were part of home visiting were born to mothers living in the city, our match rate for the home visiting was 88%, not 100%. Therefore, we believe that approximately 10-12% of failures to match records are due to matching issues (false negatives).

We are able to report only on the data from our partners. Children from organizations that did not participate are excluded and may bias any findings or determinations.

Agencies do not have a standard protocol for handling children with multiple surnames, which are more common among children with Spanish-language names (e.g., Gabriel Jose Garcia Marquez). Differences in handling these names could lead to higher rates of match failure.

Mothers were never identified in these data, so siblings (including twins) cannot be identified.

Maps

Geographic analyses were used to show where mothers lived at the time of their child's birth by the census tract—Baltimore City has over 200—in which they lived at the time of the child's birth.

References

- Harron, K., Goldstein, H., & Dibben, C. (Eds.). (2016). *Methodological Developments in Data Linkage*. West Sussex, UK: John Wiley & Sons, Ltd.
- Wasi, N., & Flaaen, A. (2015). Record Linkage Using Stata: Preprocessing, Linking, and Reviewing Utilities. *Stata Journal*, 15(3), 672-697.

*The authors wish to thank to the support provided by
The Harry and Jeanette Weinberg Foundation to do this work.*