Baby's First Years: Statistical Analysis Plan for Phase Two

November 17, 2023

NOTE: Age 6 and Age 8 measures and hypotheses will be added prior to the start of each data collection wave, after the protocol is fully developed and piloting completed.

Project Summary

In the Baby's First Years (BFY) study, one thousand infants born to mothers with incomes falling below the federal poverty threshold in four metropolitan areas in the United States were assigned at random within each of the metropolitan areas to one of two cash gift conditions. The sites are: New York City, the greater New Orleans metropolitan area, the greater Omaha metropolitan area, and the Twin Cities. IRB and recruiting issues led to a distribution of the 1,000 mothers across sites of 121 in one site (the Twin Cities), 295 in two of the other sites (New Orleans and Omaha) and 289 in New York. (We have also randomly sampled 80 of the participating families in the Twin Cities and New Orleans to participate in an in-depth qualitative study, but do not elaborate on those plans in this document.)

Mothers were recruited in postpartum wards of the 12 participating hospitals shortly after giving birth and, after consenting, were administered a 30-minute baseline interview. They then were asked to consent to the cash gifts. The "high-cash gift" treatment group mothers (40% of all mothers) are receiving unconditioned cash payments of $333 per month ($4,000 per year) via debit card for 76 months. Mothers in the "low-cash gift" comparator group (60% of all mothers) are receiving a nominal payment – $20 per month, delivered in the same way and also for 76 months. The 40/60 randomization assignment is stratified by site, but not by hospitals, within each of the four sites.

BFY was originally formulated to study the effects of monthly unconditional cash transfers on child development for the first three years of life, with the cash gifts set to be distributed for 40 months (3 years, 4 months). In response to the COVID-19 pandemic and the need to postpone in-person research activities, the cash transfers were extended for an additional year, through 52 months (4 years, 4 months), enabling us to postpone in-person direct child assessments to age 4. Subsequently, payments were extended a second time, such that they will now be provided to participants for a total of 6 years, 4 months. Interviews conducted at child ages 1, 2 and 3 are providing information about family functioning as well as several maternal reports of developmentally-appropriate measures of children's cognitive and behavioral development. A lab visit at age 4 provided high-quality measures of child well-being as well as maternal response to a questionnaire that resembled the one administered at ages 1, 2 and 3.

The current analysis plan focuses on impact analyses of lab-based assessments at child ages 6 and 8, which were proposed for Phase 2 of the BFY project.

Conditional on participants' consent and our success in securing agreements with state and county agencies, we are also collecting state and local administrative data regarding parental employment, utilization of public benefits such as Medicaid and Supplemental Nutrition Assistance Programs (SNAP), and any involvement in child protective services. (We have worked with state and local officials to ensure to the extent feasible that the study's cash gifts are not considered countable income for the purposes of determining benefit levels from social assistance programs.)

The compensation difference between families in the high and low cash gift groups has the potential to boost family incomes by $3,760 per year, an amount shown in the economics and developmental psychology literatures to be associated with socially significant and policy-relevant improvements in children's school achievement. After accounting for likely attrition, a total sample size of 800 at ages 6 and 8 years, divided 40/60 between high and low cash-gift groups, provides sufficient statistical power to detect meaningful differences in cognitive, emotional and brain functioning, and key dimensions of family context (see below).

At the ages 6 and 8 lab in-person visits at universities we will administer validated, reliable and developmentally sensitive measures of language, executive functioning and socioemotional skills. We will also collect the same measure of resting EEG-that we did at age-4, to measure young children's brain activity at rest. We will also include a new ERP measure, to assess task-based brain activity, though that measure is still under development. We are currently in the process of finalizing the full set of measures that will be gathered from the child and mother.

The family process measures that we will gather are based on two theories of change surrounding the income supplements: that increased investment and reduced stress will facilitate children's healthy development. We are obtaining measures of both of these pathways annually. *Investment pathway*: Additional resources enable parents to buy goods and services for their families and children that support cognitive development. These include higher quality housing, nutrition and non-parental child care; more cognitively stimulating home environments and learning opportunities outside of the home; and, by reducing or restructuring work hours, more parental time spent with children. *Stress pathway:* A second pathway is that additional economic resources may reduce parents' own stress and improve their mental health. This may allow parents to devote more positive attention to their children, thus providing a more predictable family life, less conflicted relationships, and warmer and more responsive interactions.

Analysis Plan

Pre-registered Hypotheses. We originally preregistered hypotheses with clinicaltrials.gov within a month after recruitment began (May, 2018) and in September, 2018, preregistered hypotheses with the Registry of Effectiveness Studies and the AEA RCT Registry. We intend to pre-register hypotheses for ages 6 and 8 shortly after our measures are finalized and well before the anticipated start of Age 6 field work in July, 2024.

Hypothesis Testing and Power Analysis. Our Phase 2 key aims are to evaluate the impacts of income supplementation on validated, reliable, and developmentally-sensitive measures of reading and math achievement, self-regulation, and socioemotional functioning, as well as lower rates of special education and grade retention, compared with children in the low-cash gift group at child ages 6 and 8 – this is Aim 1 in our Phase 2 NICHD application. Aim 2 calls for measurement of developmentally-sensitive electroencephalographic-based measures of brain functioning at child ages 6 and 8. Aim 3 is focused on data related to family expenditures, neighborhood quality, parent stress and parenting practices, and children's time spent in nonparental care gathered at child ages 6 and 8.

All of our pre-registered hypotheses focus on full-sample impacts, although we will also estimate in exploratory analyses moderation of impacts by gender, race/ethnicity (African American, Latino, White), family structure at birth and depth of poverty at birth (income to needs $\leq .5$ or not). Before conducting these main analyses, all measures will be examined for psychometric equivalence across race/ethnicity and whether Spanish or English is a primary language spoken at home and we will compare high- and low-cash gift groups within site on all baseline characteristics to confirm balance among the age 6 and age 8 sample with completed data.

Our basic empirical approach will use the survey and neuroscience data to compare the pooled cross-city $333/month and $20/month groups on a wide range of family process and child outcome measures. Because of random assignment, the low-cash gift group average outcomes enable us to identify the average outcomes corresponding to the counterfactual state that would have occurred for individuals in the high-cash gift group if they had not been offered the additional $313/month income supplement. Therefore, differences in outcomes for the high- compared with the low-cash gift group (after random assignment) can be interpreted as estimates of causal treatment effects of the $313/month higher income (regardless of whether treatment-group participants actually expend all of the funds.) These are commonly known as intent-to-treat effects.

Estimation strategy. We illustrate our approach to estimation in a simple regression framework. The "Intent-To-Treat effect" (ITT) is captured by the estimate of the coefficient π_1 in a regression of some child or family process outcome (Y) on a dichotomous indicator for assignment (Z) to the high payment group as in (1).

$$(1)\ Y = Z\pi_1 + X\beta_1 + \varepsilon_1$$

We have experienced extremely low rates of "non-compliance" with the offer of cash gifts paid via the debit cards, with only 1of the 1,000 participants not yet using their debit card by child age 3. We will adjust standard errors using robust variance estimation techniques (Cameron et al. 2008). We will estimate (1) without and then with baseline demographic child and family characteristics (X) to improve the precision of our estimates by accounting for residual variation. These baseline measures, all gathered prior to random assignment, have been checked for adequate variation and sufficient independence from other baseline measures. They include: dummy variables for three of the four sites; mother's age, completed schooling, household income, net worth, general health, mental health, race and Hispanic ethnicity, marital status, number of adults in the mother's household, number of other children born to the mother, whether the mother smoked or drank alcohol during pregnancy and whether the father is currently living with the mother; and child's sex, birth weight, gestational age at birth and birth order.

We will apply our regression estimation strategy to the assessment-based measures of cognitive, language, self-regulation, and socio-emotional functioning and EEG measures of brain activity gathered at ages 6 and 8. To investigate family process impacts, we will apply our estimation strategy to maternal and family measures gathered at child ages 6 and 8

Attrition. The greatest threat to internal validity is potential bias from sample attrition overall, within site, and differential attrition rates by treatment status overall and within site. We will carefully track response rates by site, by treatment status across sites, and then treatment status within site. Response rates have been very high at ages 1-4. Of the original 1,000 recruited participants, we secured interviews with 931 at age 1, 922 at age 2, and 922 again at age 3. At age 4 we gathered data from 890 of the 990 families (5 mothers and 5 children are deceased) – 846 in-person at the university labs or other locations; and 44 by telephone.

We will also conduct sensitivity checks to evaluate whether missing data might be biasing estimates. Most sample attrition that is systematically related to our outcomes of interest (Y) would presumably also be related to the distribution of baseline characteristics (X), and so bias due to sample attrition would be evident if our estimates are sensitive to conditioning on baseline characteristics. Some attrition may be due to time-varying (or unobserved) characteristics, and we can approach this problem in two ways. First, we will examine the sensitivity of our results to worst-case bounds, which enable us to bracket the true effects of our treatment without imposing any assumptions about the unobserved outcomes of participants (Manski, 1989; Manski, 1990; Manski, 1995). A second approach to addressing the problem of missing data will be to use multiple imputation strategies with all available data, (including all survey and administrative data on outcomes and predictor variables). Multiple imputation is an appropriate method if, conditional on all observed information, data are missing at random. Finally, because we have permission to collect administrative data from over 75% of mothers, we will be able to compare survey respondents and survey non-respondents on formal earnings and receipt of income from social programs.

Interpretation of parameters. The coefficients obtained in our regression models will be used to quantify the causal effects of the $313/month difference in income supplementation on ages 6 and 8 child brain circuitry, cognitive development and socioemotional functioning. We will use the same methods to generate causal impact estimates for the family processes in each of the conceptual pathways. Examining the possible explanatory mechanisms in this way uses a series of separate regression equations to estimate program effects on possible treatment mediators, rather than estimating a structural-equation mediation model, and has been effectively used to infer possible mediation in comparable studies. This approach is preferred because it preserves the exogenous (experimental) variation in income generated by random assignment. The underlying insight is that randomization occurred with respect to receipt of the cash gifts and not on the basis of the proposed pathway mediators. With the potential for multiple mediators, a causal interpretation cannot be given to mediational models without very strong, often implausible, assumptions that there are no unobserved confounds of the association between the mediator and outcome. Still, the pattern of impacts can yield important insight as to which processes are likely to be present and absent and set the stage for future analyses.

Statistical power. The compensation difference between families in the high- and low-cash gift groups amounts to $313 per month and $23,788 over the course of the 76 months. The annual equivalent of this amount is in the range of income increases associated with child impacts of around .20 sd in studies of welfare experiments and the EITC (Duncan, Morris & Rodrigues, 2011; Morris, Duncan, Clark-Kauffman, 2005; Dahl & Lochner, 2012). After accounting for likely 20% attrition in the age-6 lab visit, and in the absence of adjustments for sample clustering within hospitals or increased precision owing to the inclusion of baseline covariates in our impact estimates, the sample size of 800 at age 6, divided 40%/60% between high and low payment groups, provides 80% statistical power to detect a .219 sd impact at p <.05 in a two-tailed test on cognitive functioning and family processes. The use of baseline covariates in estimation models will improve this power, while the use of bootstrap standard errors will decrease it.

Multiple comparisons. One strength of this study is the collection of survey, neuroscience lab and administrative data on a wide range of outcomes and explanatory pathways. However, the probability of rejecting a true null hypothesis for at least one outcome is greater than the significance level used for each test. We will address the possibility of false positives while minimizing the reduction in statistical power to detect meaningful effects. Best-practice methods differ across disciplines so we will draw from multiple approaches with the goal of ensuring that results from one approach are consistent with results from others (Romano & Wolfe, 2005; Porter, 2018; Benjamini, 2010; Holm, 1979, Westfall & Young, 1993; Schochet, 2008). Where possible we have aggregated measures used to test our pre-registered hypotheses into indexes. In the case of related measures that cannot be aggregated into a single index, we will estimate the statistical significance of the entire family ("familywise error rate") using stepdown resampling methods in Westfall and Young (1993; Westfall, Tobias, Wolfinger, 2011). Pre-registered clusters of measures are identified with grey bars in appendix tables.

Data release. We are releasing data and documentation from the study to the research community approximately 18 months following the end of each data collection wave to enable independent researchers to pursue replication, mediation, moderation as well as other related analytic questions.

Age 6 and 8 Resting EEG Hypotheses and Analysis Plan.

Following publication of Age-1 resting EEG treatment impacts (Troller-Renfree et al., 2022), we amended the Age-4 resting EEG analysis plan to include primary and secondary hypotheses, and will continue to employ this analysis plan at age-6. The original preregistration of Age-1 EEG data included hypotheses across multiple frequency bands. However, due to the inability to reach many participants in-person due to the onset of the pandemic, as well as the rejection of artifact-laden EEG files, and the high correlation between EEG bands as well as the expected effect size and consistency of the hypothesized effects, we were left with inadequate statistical power for multiple hypothesis testing across bands. As we had uniform, directional hypotheses for all three mid- to high-frequency bands, we updated this analysis plan to instead include an index of mid- to high-frequency power (described below; primary hypothesis) as well as more traditional neuroscientific investigation of power within bands (described below; secondary hypothesis). Please see the history of preregistrations, including analysis plans, for more detail.

For our primary hypothesis, we will test whether the high-cash gift group has more mid- to high-frequency power than the low-cash gift group, we will create a single composite measure that aggregates across the portion of the spectrum defined by the three mid-to-high-frequency bands (alpha, beta, and gamma power). Because this approach is focused on estimating intent-to-treat differences in a single index score, there is no need for multiple-testing adjustments. Covariates will include all preregistered covariates as well as the number of artifact-free epochs contributed by each participant. Models will be examined with and without preregistered baseline covariates as above, and we will conduct sensitivity checks to evaluate whether missing data might be biasing estimates, as described above.

As to secondary hypotheses, consistent with the methods used by another prominent RCT examining an early-life intervention on EEG activity (Debnath, Tang, Zeanah, Nelson, & Fox, 2020; Marshall, Fox, & BEIP Core Group, 2004; Vanderwert, Marshall, Nelson, Zeanah, & Fox, 2010; Vanderwert, Zeanah, Fox, Nelson, & III, 2016), we will explore band-specific and regional effects using mixed-design analyses of variance (mixed-ANOVA). Our secondary hypothesis is that there will be an intervention effect on frontal gamma spectral power between the low-cash gift group and high-cash gift group. Covariates will include all preregistered covariates as well as the number of artifact-free epochs contributed by each participant.

In addition, to explore all regional-frequency effects, we will perform separate mixed-ANOVAs for each frequency band of absolute and relative power with region (frontal, central, parietal, occipital) as a within-subject factor, and group (low-cash, high-cash) as the between-subjects factor. Greenhouse–Geisser correction will be applied for violations of sphericity. Post hoc comparisons will be performed for significant main effects of group. Any main and interaction effects not involving group will not be followed up. Multiple-adjustment corrections will be applied for all post hoc comparisons. Covariates will include all preregistered covariates as well as the number of artifact-free epochs contributed by each participant.

References

Benjamini, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. Biometrical Journal, 52(6), 708–721. https://doi.org/10.1002/bimj.200900299

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90(3), 414-427. https://doi.org/10.1162/rest.90.3.414

Dahl, G. B., & Lochner, L. (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. American Economic Review, 102(5), 1927–1956. https://doi.org/10.1257/aer.102.5.1927

Debnath, R., Tang, A., Zeanah, C. H., Nelson, C. A., & Fox, N. A. (2020). The long-term effects of institutional rearing, foster care intervention and disruptions in care on brain electrical activity in adolescence. *Developmental science*, *23*(1), e12872. https://doi.org/10.1111/desc.12872

Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random assignment experiments. Developmental Psychology, 47(5), 1263-1279. http://dx.doi.org/10.1037/a0023875

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics, 6(2), 65–70. https://www.jstor.org/stable/4615733

Manski, C. F. (1989). Anatomy of the Selection Problem. The Journal of Human Resources, 24(3), 343–360. https://doi.org/10.2307/145818

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. The American Economic Review, 80(2), 319–323. https://www.jstor.org/stable/2006592

Manski, C. F. (1995). Learning about social programs from experiments with random assignment of treatments. Institute for Research on Poverty Discussion Papers 1061-95, University of Wisconsin Institute for Research on Poverty.

Marshall, P. J., Fox, N. A., & BEIP Core Group. (2004). A Comparison of the Electroencephalogram between Institutionalized and Community Children in Romania. *Journal of Cognitive Neuroscience, 16*(8), 1327–1338. https://doi.org/10.1162/0898929042304723

Morris, P. A., Duncan, G. J., & Clark-Kauffman, E. (2005). Child well-being in an era of welfare reform: the sensitivity of transitions in development to policy change. Developmental Psychology, 41(6), 919–932. https://doi.org/10.1037/0012-1649.41.6.919

Porter, K. E. (2018). Statistical Power in Evaluations That Investigate Effects on Multiple Outcomes: A Guide for Researchers. Journal of Research on Educational Effectiveness, 11(2), 267–295. https://doi.org/10.1080/19345747.2017.1342887

Romano, J. P. & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. Econometrica, 73(4), 1237-1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x

Schochet, P. Z. (2008). Guidelines for multiple testing in impact evaluations of educational interventions. Final report. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED502199

Troller-Renfree, S. V., Costanzo, M. A., Duncan, G. J., Magnuson, K., Gennetian, L. A., Yoshikawa, H., ... & Noble, K. G. (2022). The impact of a poverty reduction intervention on infant brain activity. *Proceedings of the National Academy of Sciences*, *119*(5), e2115649119.

Vanderwert, R. E., Marshall, P. J., Nelson, C. A. III, Zeanah, C. H., & Fox, N. A. (2010). Timing of intervention affects brain electrical activity in children exposed to severe psychosocial neglect. *PLoS ONE, 5*(7), Article e11415. https://doi.org/10.1371/journal.pone.0011415

Vanderwert, R. E., Zeanah, C. H., Fox, N. A., & Nelson, C. A. (2016). Normalization of EEG activity among previously institutionalized children placed into foster care: A 12-year follow-up of the Bucharest Early Intervention Project. *Developmental Cognitive Neuroscience, 17,* 68–75. https://doi.org/10.1016/j.dcn.2015.12.004

Westfall, P. H., Tobias, R. D., & Wolfinger, R. D. (2011). Multiple comparisons and multiple tests using SAS, second edition. Cary, NC: The SAS Institute.

Westfall, P. H. & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment. Hoboken, New Jersey: John Wiley & Sons.