

# **Baby's First Years: Summary, Pre-registered Hypotheses, Analysis Strategies and Paper Plans**

June 17, 2021

## **Project Summary**

One thousand infants born to mothers with incomes falling below the federal poverty threshold in four metropolitan areas in the United States are being assigned at random within metropolitan area 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. The high cash gift treatment group mothers (40% of all mothers) will receive unconditioned cash payments of \$333 per month (\$4,000 per year) via debit card for 52 months. Mothers in the low cash gift comparator group (60% of all mothers) receive a nominal payment – \$20 per month, delivered in the same way and also for 52 months. The 40/60 randomization assignment is stratified by site, but not by hospitals, within each of the four sites.

Mothers are being recruited in maternity wards of the 12 participating hospitals shortly after giving birth and, after consenting, are administered a 30-minute baseline interview. They then are asked to consent to the cash gifts. The three follow-up waves of data collection conducted at child ages 1, 2 and 3 will provide information about family functioning as well as developmentally appropriate measures of children's cognitive and behavioral development. An additional feature of our ages 1-3 data collection plans is that we will attempt to randomly assign a designated interview date within a one-month interval centered on the child's birthday. This provides variation in the timing of outcome data with respect to participants' receipt of the cash gift that will enable us to learn more about the incremental value of a stable predictable monthly infusion of cash.

We will collect information about the mother and child in the home when the child is 12 and 24 months of age (with information collected via phone during the COVID-19 restrictions on in-person data collection). We had planned to assess and interview mothers and children in research laboratories at each site at age 3. Due to COVID-19, the age 3 data collection is being conducted via phone and in-person interviews and laboratory-based assessments will be conducted at child ages 45-48 months. Conditional on participants' consent and our success in securing agreements with state and county agencies, we will also collect 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 also have plans to randomly sample 80 of the participating families in two of the sites (the Twin Cities and New Orleans) to participate in an in-depth qualitative study, but do not elaborate on those plans in this document.

The compensation difference between families in the high and low cash gift groups will

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. (We have worked with state and local officials to ensure to the extent feasible that our cash gifts are not considered countable income for the purposes of determining benefit levels from social assistance programs.) After accounting for likely attrition, our total sample size of 800 at age 3 years, divided 40/60 between high and low payment groups, provides sufficient statistical power to detect meaningful differences in cognitive, emotional and brain functioning, and key dimensions of family context (see below).

Cognitive and emotional development measures will be gathered at 12, 24, 36, and 45-48 months of age. At the age 45-48 months lab visit we will administer validated, reliable and developmentally sensitive measures of language, memory, executive functioning and socioemotional skills. We will also collect direct measures of young children's brain development at ages 1 and 45-48 months. Measures and preregistered hypotheses about them as well as family based measures are shown in the two tables at the end of this document.

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 will obtain data measuring 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 preregistered hypotheses with [clinicaltrials.gov](https://clinicaltrials.gov) within a month after recruitment began (May, 2018) and in September, 2018, preregistered hypotheses with the [Registry of Effectiveness Studies](https://www.effectivepsychology.org/) and the [AEA RCT Registry](https://www.aea-rct.org/). Appendix Tables 1 and 2 detail our original hypothesized impacts and which groups of measures will be subject to multiple testing adjustments. Appendix Tables 3 and 4 incorporate minor changes to the tables that were originally posted in our pre-registrations. These changes are mostly made to data collection at age 2, with a few changes to age 3 data collection. There were no changes to age 1. Appendix Tables 5 and 6 incorporate minor changes to reflect the COVID-19 disruptions that impacted data collection at age 2, and altered data collection plans at age 3 and ages 45-48 months.

*Hypothesis Testing and Power Analysis.* Our key aims are to evaluate the impacts of income supplementation on: validated, reliable, and developmentally-sensitive measures of cognitive, language, memory, self-regulation, and socio-emotional functioning at child ages 1 (a

small subset), 2 and 3 (a larger subset), and ages 45-48 months (almost all) – this is Aim 1 in our NICHD application; developmentally-sensitive electroencephalographic-based measures of brain functioning at child ages 1 and 45-48 months (Aim 2); and family expenditures, food insecurity, housing and neighborhood quality, parent stress and parenting practices, and child care arrangements gathered at child ages 1, 2, 3, and 45-48 months (Aim 3).

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 successful implementation of random assignment.

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 enables 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 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  $\pi_1$  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).

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

Consistent with experiences from a 30-family pilot study we conducted in 2014, we anticipate extremely low rates of “non-compliance” with the offer of cash gifts paid via the debit cards.

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, will first be 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 adult 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, memory, self-regulation, and socio-emotional functioning at child ages 2 and 3, and the EEG measures of brain activity at ages 1 and 45-48 months and ERP measures of brain activity at ages 45-48 months (see Appendix Tables 5 and 6). To investigate family process impacts, we will apply our estimation strategy to measures of stress physiology, family expenditures, food insecurity, housing and neighborhood quality, mothers' executive function, parent stress and parenting practices, and child care arrangements gathered at child ages 1, 2, 3, and 45-48 months as shown in Appendix Tables 5 and 6 and described in the section on paper plans.

*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. Any early signs of differential attrition will be expediently addressed through small, strategic adjustments in survey follow-up efforts, including use of financial incentives, or more tailored strategies such as using on-the-ground reconnaissance techniques to locate individuals. Based on the successes in our pilot study, our investigators' prior experience with the Survey Research Center, and because of the continued contact with all participants the debit card ensures, we anticipate high response rates in later data collection (80+% at 36 months) with little to no differential attrition.

If necessary, we will consider a two-stage sampling procedure at the final stages of our data collection efforts during each wave in order to minimize attrition-related biases. The procedure calls for randomly subsampling from the remaining difficult-to-reach nonrespondents and concentrating resources and efforts to locate them. Analysis weights will be developed to adjust for the possible two-stage survey response sampling. This weighting approach has been successfully implemented in comparable studies. In addition to case-based nonresponse we also anticipate the usual (i.e., infrequent but not nonexistent) item-based nonresponse owing to refusals, interview breakoffs, etc.

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 expect relatively high rates (~80%) of baseline consent to collect administrative data, 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 age-1 and 45-48 month 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 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 \$16,276 over the course of the 52 months. 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, 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 3, 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. Based on exploratory analyses of age-3 cognitive outcomes in the Fragile Families study, we expect that these two offsetting factors will have little net impact on the size of our estimated standard errors.

*Multiple comparisons.* One strength of our proposal 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 will release data and documentation for our study to the research community at the end of each data collection wave once data are cleaned and coded, to enable independent researchers to pursue replication, mediation, moderation as well as other related analytic questions.

## 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>
- 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.
- 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>
- 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.

For Appendix Tables 1, 2, 3, and 4 see [Statistical Analysis Plan uploaded in July 2020.](#)