

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

June 27, 2019

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 will likely lead to a distribution of the 1,000 mothers across sites of roughly 115 in one site (the Twin Cities) and 295 in each of the three other sites. 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 40 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 40 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 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. At age 3, mothers and children will be assessed and interviewed in research laboratories at each site. 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, and 36 months of age. At the age-three 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 3. 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 within a month after recruitment began (May, 2018) and have since (September, 2018) preregistered hypotheses with the [Registry of Effectiveness Studies](#) as well. We intend to submit a revised set of hypothesis before we begin collecting age-1 impact data in July, 2019 to clinicaltrials.gov, the Registry of Effectiveness Studies and in the [AEA RCT Registry](#). Appendix Tables 1 and 2 detail our hypothesized impacts, the data collection waves in which we will be gathering data to test them, and which groups of measures will subject to multiple testing adjustments. Differences from our original preregistration are relatively minor – for example, when technology enabled us to conduct an EEG during our age 1 home visit, we added a hypothesis about that.

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 (a larger subset) and 3 (almost all) – this is Aim 1 in our NICHD application; developmentally-sensitive electroencephalographic-based measures of brain functioning at child ages 1 and 3 (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 and 3 (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 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).

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

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. Each of 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 3 and ERP measures of brain activity at age 3 (see Appendix Table 1). 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 and 3, as shown in Appendix Table 2 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 3 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 \$12,520 over the course of the 40 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 step-down resampling methods in Westfall and Young (1993; Westfall, Tobias, Wolfinger, 2011). Pre-registered clusters of measures are identified with grey bars in Appendix Tables 1 and 2.

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.

Appendix Table 1

Measure description	Measure source	Psychometrics	Wave preregistered <u>Primary</u> Outcome	Wave preregistered <u>Secondary</u> Outcome	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Language Development					
Language Milestones	Squires et al., 2009	sensitivity .86 specificity .85		1,2,3	Measured using ASQ- Communication Subscale
Child Vocalization	Xu et al., 2009	internal consistency .69-.86 test-retest reliability .76		2	Measured using LENA processing software
Verbal Comprehension	Wechsler, 2012	internal consistency .95 test-retest reliability .86-.92		3	Measured by WPPSI-IV- Vocabulary subtest
Language Processing	Golinkoff et al., 2017		3		Measured by Quick Interactive Language Screener (QUILS)- Language Processing Subscale
Communicative Development	Fenson, 2002	internal consistency .85		2	Measured by MacArthur Communicative Development Inventories
Executive Function & Self-Regulation					
Self-Regulation	Smith-Donald et al., 2007	internal consistency of assessor report (not full assessment) .82-.93	3		Measured by the Preschool Self-Regulation Assessment - PSRA (parent report and child assessment)
Executive Function	Carlson, 2017 OR Wechsler, 2012 OR Willoughby et al., 2011	MEFS: validity .92 test-retest .93 WPPSI-IV: internal consistency .95 test-retest reliability .86-.92	3		Minnesota Executive Function Scale OR WPPSI-IV Working Memory OR EF Touch Executive Functioning
Socio-Emotional Processing					
Social-Emotional Problems	Briggs-Gowan et al., 2004	internal consistency .65-.79 test-retest reliability .87	3	1,2	Measured by the Brief Infant-Toddler Social and Emotional Assessment (BITSEA)- Problem Scale
Behavior / Emotional Problems	Achenbach et al., 2000	parent report reliability .80	3	2	Measured by the Child Behavior Checklist
Social-Emotional Behavior	Roggman et al., 2013; Griffen & Friedman, 2007			1,2	Measured using NICHD SECCYD parent-child-interaction task coding scheme, with child codes Positive Mood, Negative Mood, Activity Level, Sustained Attention, Positive Engagement
IQ					
Intelligence	Wechsler, 2012	internal consistency .95 test-retest reliability .86-.92	3		Measured by WPPSI-IV
Brain Function					
Resting Brain Function	Tomalski et al., 2013; Otero et al., 2013; Marshall et al., 2004	n/a	3	1	Measured by EEG
Language-Related Brain Function	Tomalski et al., 2013; Otero et al., 2013; Marshall et al., 2005	n/a		3	Measured by electroencephalogram
Health: BMI					
Body Mass Index (BMI)	Kuczmarski, 2000	n/a	3		Measured by CDC scales
Health: Physiological Stress					
Physiological Stress	Ursache et al., 2017; Meyer et al., 2014; Davenport et al., 2006	n/a		2	Measured by hair cortisol
Health: Sleep					
Sleep problems	Yu et al., 2012	reliability .9	3	1,2	Measured by PROMIS Sleep Disturbance- Short Form adapted from ECHO; Additive index of the following items: 1. trouble falling asleep (0: never; 1: 1-2 nights; 2: 3-6 nights; 3: every night) 2. sleeping through night (reverse coded) 3. problem with sleep

Appendix Table 1

Measure description	Measure source	Psychometrics	Wave preregistered <u>Primary</u> Outcome	Wave preregistered <u>Secondary</u> Outcome	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Health: Other Indicators					
Overall Health, Medical Care, Diagnosis of Condition or Disability	Halim et al., 2013	n/a	3	1,2	<p>Additive index of the following items*:</p> <p>1. Child's overall health? (4: excellent, 3: very good, 2: good, 1: fair, or 0: poor)</p> <p>2. About how many times in the last year did you take child to a doctor because [he/she] was sick? 0-1 times, 2-5 times, 6+</p> <p>3. About how many times in the last year did you take child to a doctor because [he/she] was hurt or injured?</p> <p>4. Did you ever have to take child to the Emergency Room because [he/she] was sick, hurt or injured? (Y/N)</p> <p>5. How many times ER?</p> <p>6. Has child been diagnosed with any health condition or disability since birth? (Y/N)</p> <p>*factor analysis of items will be conducted to scale the index</p>
School Achievement & Behavior					
School test scores for target children and siblings	Administrative data	n/a	School age (target child)	School age (siblings)	
Student behavioral data for target children and siblings	Administrative data	n/a		School age (target child and siblings)	

Appendix Table 2

Measure description	Measure/Item source	Psychometrics	Wave preregistered <u>Primary Outcome</u>	Wave preregistered <u>Secondary Outcome</u>	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Household Economic Hardship					
Index of economic stress	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	<p>Additive index of dichotomous variables (higher score=more stress):</p> <ol style="list-style-type: none"> 1. worried about expenses? (0: occasionally or never; 1: frequently or more) 2. whether spent more than income? (0: no; 1: yes) 3. missed rent or mortgage (0 if homeless; 1 if missed rent or mortgage) 4. Set aside rainy day funds for 1 mo (0: Yes; 1: No) 5. Ability to cover expenses for 1 mo with loss of income (0: Yes; 1: No) 6. in past 12 mos, missed payments for water, gas, oil, electricity? (0: no; 1: yes) 7. in past 12 mos, gas, water, electricity ever shut off? (0: no; 1: yes) 8. Since child's birth, have you ever been evicted or forced to leave? (0: No; 1: Yes).* 9. needed medical or dental care and did not get it? (0=none; 1=yes) <p>*changes to "in the past 12 months" for surveys at ages 24 and 36 months</p>
Household Poverty rate					
Index of food insufficiency	Economic Research Service, USDA, 2012			1, 2, 3	<p>Additive index of 5 dichotomized items (higher score=more food insecurity):</p> <ol style="list-style-type: none"> 1. Food didn't last, no \$ for more (0: Never true, 1: sometimes or often true) 2. Can't afford balanced meals (0: Never true, 1: sometimes or often true) 3. Cut size or skip means (0: No; 1: Yes) 4. Eat less than should (0:No; 1: Yes) 5. Hungry (0:No; 1: Yes)
Social Services Receipt					
Number of Benefits received by mother	Study PIs			1, 2, 3	<p>Additive index of dichotomized items (higher score=more benefits received):</p> <ol style="list-style-type: none"> 1. Food stamps SNAP (0: not currently receiving; 1: currently receiving) 2. Free or reduced childcare 3. Early Head Start or HS 4. Women, Infants and Children (WIC) 5. State Unemployment 6. Cash assistance/TANF 7. Medicaid coverage for self 8. Housing assistance 10. LIHEAP / heat/AC assistance
Mother's Labor Market and Education Participation					
Time to labor market reentry from birth	Current Population Survey			1, 2, 3	<p>Continuous outcome: # of months until mom's reentry into labor market from birth of child derived from the following items:</p> <ol style="list-style-type: none"> 1. did you ever work for pay since child's birth?* 2. in what months did you work for pay? <p>*changes to "in the past 12 months" for surveys at ages 24 and 36 months</p>
Time to full-time labor market reentry from birth	Current Population Survey			1, 2, 3	<p>Continuous outcome: # of months until mom's full-time reentry into labor market from birth of child derived from the following items:</p> <ol style="list-style-type: none"> 1. did you ever work full time since child's birth?* 2. in what months did you work full time? <p>*changes to "in the past 12 months" for surveys at ages 24 and 36 months</p>
Mother's education and training attainment	Current Population Survey			1, 2, 3	<p>Dichotomous variable indicating that mother participated in education and/or job training activities since birth*</p> <p>*changes to "the past 12 months" for surveys at ages 24 and 36 months</p> <p>**will be adding new items for surveys at age 24 and 36 months</p>
Child-Focused Expenditures					
Index of child-focused expenditures (since birth)	Lugo-Gil, Yoshikawa, 2006			1	<p>Additive index of the following dichotomous items (higher score=more purchased):</p> <p>Since child's birth, purchased...</p> <ol style="list-style-type: none"> 1. Crib? 2. Car seat? 3. High chair? 4. Safety covers for outlets? 5. Latches for cabinets? 6. Gate? 7. Smoke detector? 8. books (yes/no)?
Index of child-focused expenditures (in past 30 days)	Lugo-Gil, Yoshikawa, 2006			1, 2, 3	<p>Continuous dollar amount:</p> <p>Past 30 days, total \$ amount spent on... 9. books 10. toys 11. clothes 12. diapers 13. videos*</p> <p>*products will be adjusted for child age at the age 24 and 36 month data collections</p>
Cost of paid child care	National Study of Early Care and Education			1, 2, 3	<p>Out of pocket spending on child care last week.</p> <p>1. altogether, about how much money did you spend out-of-pocket on all of [CHILDNAMEF]'s child care arrangements last week?</p>
Use of center-based care	National Study of Early Care and Education			1, 2, 3	<p>1. Has child spent any time in childcare or day care? (Y/N)</p>
Housing and Neighborhoods					
Index of perceptions of neighborhood safety	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	<p>Additive index of two items (higher score=feels more safe).</p> <ol style="list-style-type: none"> 1. how safe during day? (3: very safe, 2: safe, 1: unsafe, 0: very unsafe) 2. how safe during night? (3: very safe, 2: safe, 1: unsafe, 0: very unsafe)
Index of housing quality	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	<p>Additive index of 7 items (higher score=higher quality):</p> <ol style="list-style-type: none"> 1. Bad walls (0: big problem; 1: small problem; 2: not problem) 2. bad plumbing 3. rodents 4. cockroaches 5. bad windows 6. bad heat 7. overall condition (3: excellent, 2: good 1: fair, 0: poor)
Homelessness	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	<p>Additive index of two dichotomized items (higher score=more homelessness):</p> <ol style="list-style-type: none"> 1. Since child's birth, have you been homeless?* (0: Yes; 1: No) 2. Since birth, have you been in a group shelter?* (0: Yes; 1: No) <p>*changes to "in the past 12 months" for surveys at ages 24 and 36 months</p>
Excessive Residential mobility	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	<p>Moved three or more times since birth of baby* (Y/N)</p> <p>*changes to "in the last 12 months" for surveys as ages 24 and 36 months</p>

Appendix Table 2

Measure description	Measure/Item source	Psychometrics	Wave preregistered Primary Outcome	Wave preregistered Secondary Outcome	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Neighborhood poverty	Census			1, 2, 3	# of residents below poverty line in census tract divided by total number of residents in census tract
Family and Maternal Perceived Stress Perceived stress: Perceived Stress Scale (PSS)	Cohen et al., 1994, 1983	alpha: .86		1, 2, 3	Additive index of 9 items (0: never; 1: almost never; 2: sometimes; 3: fairly often; 4: very often): 1. upset because of something unexpected 2. felt unable to control important life things 3. felt nervous and stressed 4. confident in ability to handle personal probs (reverse coded - rc) 5. couldn't cope with all things to do 6. control of irritations in life (rc) 7. "on top of things" (rc) 8. angered bc of things outside control 9. could not overcome difficulties
Parenting stress: Aggravation in Parenting Scale	PSID-Child Development Supplement	alpha: .71		1, 2, 3	Additive index of 7 items (0: Strongly agree-5: Strongly disagree): 1. confidence in parenting abilities 2. feels good about parenting abilities 3. thinks good parent 4. kids will say she was wonderful 5. giving up more for kids than ever expected 6. feels trapped (rc) 7. unable to do different things bc of kids (rc)
Maternal Happiness and Optimism					
Global happiness	The General Social Survey from NORC			1, 2, 3	One-item with 3-point response scale "Taken altogether, how happy are you these adys?" (0: not happy; 1: pretty happy; 2: very happy)
Maternal Agency: HOPE scale	Snyder et al., 1991	alpha: .86 test-retest: .81		1, 2, 3	Additive index of 9 items with 5-point response scale (0: definitely false; 5: definitely true) 1.think of ways to get out of a jam 2.. energetic pursuit of goals 3. lot of ways around any problem 5. ways to get what's important 6. solves problems 7. past has prepared me for future 8. pretty successful in life 9. meets goals set for oneself
Maternal Physiological Stress					
Maternal hair cortisol: from sample of hair that is at least 15mg and ~3cm long	Ursache et al., 2017		2	1	Measured using a sample of hair that is $\geq 15\text{mg}$ I weight and $\sim 3\text{cm}$ long; analyzed with sensitive and specific enzyme-linked immunosorbent assay; assay readout converted to pg cortisol per mg dry hair weight
Maternal Mental Resources					
Maternal cognitive resources: Flanker Inhibitory Control and Attention Test	Zelazo et al., 2013	test-retest: .92		2	Additive score of two outcome vectors (accuracy and response time)
Maternal Mental Health					
Index of maternal depression: PHQ-8	Kroenke & Spitzer, 2002			1, 2, 3	Additive index of 8 items (0: not at all; 1: several days; 2: more than half of days; 3: every day) 1. little interest or pleasure doing things 2. feeling down, depressed, hopeless 3. trouble sleeping or sleep too much 4. feel tired and no energy 5. poor appetite or overeating 6. feel like a failure 7. trouble concentrating 8. moving slowly or fidgety
Index of maternal anxiety: Beck Anxiety Inventory	Steer & Beck, 1997	alpha: .92 test-retest: .75		1, 2, 3	Additive index of 21 common anxiety symptom items (0: not at all; 1: mildly; 2: moderately; 3: severely bothersome)
Maternal Substance abuse					
Alcohol and cigarette use	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	Additive index of the following items (0: never in last year; 1: less than 1x per month; 2: several times per month; 3: several times per week; 4: everyday): 1. How often do you smoke cigarettes? 2. How often drink alcohol?
Opioid use	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	Number of times per week used opioids?
Chaos in Home					

Appendix Table 2

Measure description	Measure/Item source	Psychometrics	Wave preregistered <u>Primary Outcome</u>	Wave preregistered <u>Secondary Outcome</u>	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Index of chaos in the home: Home Environment Chaos Scale	Evans et al., 2005	alpha: .77 test-retest: .93		1, 2, 3	Additive index of 20 items (higher score=more chaos): (0: not true; 1: true) 1. can find things (reverse coded - rc) 2. little commotion in home (rc) 3. always rushed 4. can "stay on top of things" (rc) 5. always late 6. "zoo" in home 7. can talk w/o interruption (rc) 8. always a fuss 9. family plans don't work out 10. can't hear oneself think at home 11. drawn into others' arguments 12. can relax at home (rc) 13. phone takes up a lot of time 14. atmosphere is calm at home (rc) 15. regular morning routine (rc) 16. eat together during daily (rc) 17. evening routine with child (rc) 18. regular late afternoon routine with child (rc) 19. child goes to bed at regular time (rc) 20. set aside for talking with child daily (rc)
Maternal Relationships					
Physical Abuse	Fragile Families and Child Wellbeing Study			1,2,3	1. Ever abused? (1: yes; 0: no)
Frequency of Arguing	Fragile Families and Child Wellbeing Study			1,2,3	1. How often argue about things that are important to you? (1: never; 2: almost never; 2: sometimes; 3: fairly often; 4: very often)
Relationship quality	Fragile Families and Child Wellbeing Study			1, 2, 3	Additive index of the following items (higher score=higher qual rel) 1. Partner fair and willing to compromise? (3: Often; 2: sometimes; 1: never) 2. partner expressed affection or love? (3: Often; 2: sometimes; 1: never) 3. partner insulted or criticized you or your ideas (0: Often; 1: sometimes; 2: never) 4. partner made you feel down or bad about yourself during an argument? (0: Often; 1: sometimes; 2: never) 5. partner encouraged or helped you to do things that were important to you? (2: Often; 1: sometimes; 0: never) 6. partner isolated you? (0: Often; 1: sometimes; 2: never) 7. partner hurt you physically (0: Often; 1: sometimes; 2: never) 8. partner sexually abused you? (0: Often; 1: sometimes; 2: never) 9. partner listened to you? (3: Often; 2: sometimes; 1: never) 10. partner made you feel afraid? (0: Often; 1: sometimes; 2: never) 11. partner threatened or hurt your child/children? (0: Often; 1: sometimes; 2: never)
Maternal Physical Health					
Global health	Idler & Benyaminini			1, 2, 3	One item with 5-point response scale "overall, how would you describe your health..." (0: excellent-5 poor)
Sleep	MTO; Kling, Liebman, Katz, 2007			1, 2, 3	Additive index of the following items (higher score=higher qual sleep): 1. Quality of sleep (0: very poor-5: very good) 2. Difficulty falling asleep (0: not at all; 5: very much) (rc) 3. Felt tired (0: not at all-5: very much) (rc)
Mother's BMI	CDC scales			3	Calculated by dividing weight by stature
Parent-Child Interaction Quality					
Adult word count: LENA	Xu et al (2009), LENA			1, 2	
Conversational turns: LENA	Xu et al (2009), LENA			1, 2	
Index of mother's positive parenting behaviors	Roggman, et al., 2013; Griften & Friedman, 2007; Belsky, et al., 2007	inter-rater reliability varies by domain: .69-.80; alpha: .78	2	1	Measured using PICCOLO coding of parenting behaviors from three sub-scales (affection, responsiveness, encouragement and teaching) with responses ranging from 0: absent, 1: barely, 2: clearly
Maternal Epigenetic Age					
Epigenetic age	Fiorito et al., 2017			2	Measured by the Horvath Method
Maternal DNA Methylation					
DNA methylation	Hughes, et al., 2018; Cao-Lei et al., 2014			2	
Frequency of Parent Child Activity					
Self-Report of Parent-child activities	Rodriguez & Tamis-LeMonda, 2011			1, 2	Additive index of 4 items with response scale (lower score=higher frequency of activities): 1. read books (0: everyday; 1: a few times/week; 2: a few times/month; 4: rarely or never) 2. tell stories 3. play together 4. play groups
Maternal Discipline					

Appendix Table 2

Measure description	Measure/Item source	Psychometrics	Wave preregistered <u>Primary Outcome</u>	Wave preregistered <u>Secondary Outcome</u>	Relevant Items (All measures between grey lines will be subject to multiple testing adjustments)
Spanking discipline strategy: one-item, yes or no question	Reichman et al., 2001			1, 2	1. In past month, have you spanked child due to misbehavior (1: yes; 2: no)