Statistical Analysis Plan
Household Income and Child Development in the First Three Years of Life
(Baby's First Years)
May 25, 2018
NIH R01 HD087384

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 2 and 3 (Aim 1); developmentally-sensitive electroencephalographic-based measures of brain circuitry at child age 3 (Aim 2); and family expenditures, food insecurity, housing and neighborhood quality, routines and time use, parent stress and parenting practices, and child care arrangements gathered at child ages 1 and 2 (Aim 3). Most of the efforts proposed here focus on full-sample impacts, although we will also estimate in exploratory analyses moderation of impacts by gender, race/ethnicity (African American, Latino, White), site, family structure at birth and depth of poverty at birth (income to needs ≤ .5 or not).

Before conducting these main analyses, all measures will be examined for psychometric equivalence across race/ethnicity and home language and we will compare treatment and control samples on all baseline characteristics to confirm successful implementation of random assignment. We will release data and documentation for our study to the research community to enable independent researchers to pursue replication, mediation, moderation as well as other related analytic questions.

Our basic empirical approach will use the survey and laboratory data to compare the pooled cross-city $333/month treatment and $20/month control groups on a wide range of family process and outcome measures. Because of random assignment, the control group average enables us to identify the average outcomes corresponding to the counterfactual state that would have occurred for individuals in the treatment group if they had not been offered the additional $313 (=$333-$20)/month income supplement. Therefore differences in outcomes for the treatment group compared with the control 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; 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 treatment group as in (1).

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

We will adjust standard errors by site using bootstrapping methods described in Cameron et al. (2007). We condition on baseline demographic child and family characteristics (X) to improve the precision of our estimates by accounting for residual variation. Consistent with our pilot study experience, because all of the women in the study will have consented to the opportunity to receive either $20/month or $333/month, and because we will regularly receive data about card use and thus be able to carefully monitor card use and ensure replacement of lost or malfunctioning cards, we will be able to adjust for take-up of the income supplement payments if necessary. We do not anticipate high rates of "non-compliance."

To accomplish Aims 1 and 2, we will apply our regression estimation strategy to the assessment-based measures of cognitive, language, memory, self-regulation, and socioemotional functioning at child ages 2 and 3, and the EEG and ERP measures of brain activity at age 3. To accomplish Aim 3, we will apply our strategy to measures of stress physiology, family expenditures, food insecurity, housing and neighborhood quality, routines and time use, mothers' executive function, parent stress and parenting practices, and child care arrangements gathered at child ages 1 and 2. Given our data collection plan, data for Aim 3 will be available before data for Aims 1 and 2. The timing of our analyses will proceed accordingly.

Our Aim 3 analyses are based on an "investment pathway" model in which families with larger cash gifts invest some or all of the additional income in ways that enrich the environments in which their children are raised. These differing timed "investments" may be independent of each other, but that can be can be combined into summary measures. We lay out our plan for grouping conceptually-related family processes in Table 1. Key elements of the investment pathway include financial hardship, neighborhood and housing quality, child-related enrichment expenditures and parent-child interactions supporting language and cognitive development. Key elements of the stress pathway include mother and child stress physiology, mothers' physical and mental health, family chaos, and quality of parenting. Some family characteristics, such as maternal cognitive resources, do not fit neatly into one pathway, although they may be related to both pathways. In other cases, measures might align with conceptual pathways in a complex fashion. For example, increased employment coupled with higher quality child care may constitute an "investment," but so might reduced employment with more parental care. As a result, we approach these "other" family processes as exploratory hypotheses (Table 1). An important first task will be using psychometric approaches such as confirmatory factor analysis to evaluate the extent to which the elements of each proposed conceptual pathway can be combined into a summary measure, given measurement across differing time periods and dimensions of behavior. This work will be an important foundation for subsequent analyses evaluating the conceptual pathways.

A final proposed analytic strategy worth highlighting concerns the estimation of short-term effects of income supplementation on measures of maternal cognitive resources. These short-term income effects are estimated by predicting maternal executive function ("bandwidth") as a function of the time since our last monthly payment, expecting that cognitive load and related distractions will increase with the length of time between the interview and the most recent payment. We will randomly assign the target interview dates from between 1 and 30 days following the child's second and third birthdays, thus generating random variation in the interval between our most recent payment and the Flanker assessments of maternal executive function. We anticipate using an instrumental variables estimation strategy in which the predicted interval between the monthly payment and the interview is estimated as a function of target minus payment date (plus controls) and, in the second stage, flanker scores are regressed on predicted intervals, plus controls.

Attrition. The greatest threat to internal validity is potential bias from sample attrition overall, and differential attrition rates by treatment status. Following our survey contractor's (the Survey Research Center at the University of Michigan) highly successful implementation of protocols in the Moving to Opportunity housing mobility experiment that achieved a 90% response rate at the 10-12 year follow-up, we will carefully track response rates by site, by treatment status, and then treatment status within site (Gebler et al, 2012). Any early signs of differential attrition will be expediently addressed through small, strategic adjustments in survey follow-up efforts, including use of financial and comparable incentives, or more tailored strategies such as using on-the-ground reconnaissance techniques to locate individuals. Based on the successes in our pilot study, and because of the continued contact with participants the debit card ensures, we anticipate high response rates in later data collection (80+% at 36 months).

If necessary, we will consider a two-stage sampling design for the follow-up such that once a good balanced response rate has been achieved, we will subsample from the remaining difficult-to-reach nonrespondents and increase resources and efforts to locate them. Analysis weights will be developed to adjust for the differential birth parity at enrollment and possible two-stage survey response sampling. This weighting approach has been successfully implemented in comparable studies. 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 the all available data, (including all survey and administrative data). Multiple imputation is an appropriate method if, conditional on all observed information, data are missing at random (Graham, 2009). Finally, because we expect very high rates 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-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 (Gennetian & Miller, 2002; Huston et al., 2005). 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 income supplements 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 (Green, Ha & Bullock, 2010; Valerie & VanderWeele, 2013). 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 allocation of our sample is: 40% (400 mothers and infants) to the $333/month treatment condition and 60% (600 mothers and infants) to the $20/month control condition. The compensation difference between families in the experimental and control 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; 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% into the experimental group and 60% into the control groups, provides 80% statistical power to detect a .209 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, which are expected to yield roughly offsetting effects.

Multiple comparisons. One strength of our proposal is the collection of laboratory, survey 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 (Romano & Wolf, 2005). 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 (Kling, Liebman & Katz, 2007; Benjamini, 2010; Westfall & Young, 1993). Following standard practice, we will first consider the statistical significance of individual treatment effects in isolation ("per-comparison significance"). Second, we will estimate the statistical significance of the entire family of related measures in our cluster groups ("familywise error rate") using step-down resampling methods (Westfall & Young, 1993) rather than the highly conservative Bonferroni correction (see Table 1 for a listing of cluster groups). Third, we will make summary statements about the effects of income supplementation on our family and child outcome clusters by estimating effects on constructed summary measures (informed by psychometric techniques such as confirmatory factor analysis). The summary measure approach incurs the least loss of statistical power (Kling, Liebman & Katz, 2007); the extent to which statistical power is reduced and minimum detectable effect sizes increased with the step-down resampling familywise approaches depends on the number and covariance of outcomes, and thus is difficult to project a priori.

References

Benjamini, Y. Simultaneous and selective inference: current successes and future challenges. *Biometrical Journal* **52**, 708-721, doi:10.1002/bimj.200900299.PMID 21154895 (2010).

Cameron, A. C., Gelbach, J. B. & Miller, D. L. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics* **90**, 414-427 (2008).

Dahl, G. B. & Lochner, L. The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review* **102**, 1927-1956, doi:10.1257/aer.102.5.1927 (2012)

Duncan, G. J., Morris, P. A. & Rodrigues, C. Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology* **47**, 1263-1279, doi:10.1037/a0023875 (2011). PMCID: PMC3208322.

Gebler, N., Hudson, M. L., Sciandra, M., Gennetian, L. A. & Ward, B. Achieving MTO's high effective response rates: Strategies and tradeoffs. *Cityscape* **14**, 57-86 (2012).

Gennetian, L. A. & Miller, C. Children and welfare reform: A view from an experimental welfare program in Minnesota. *Child Development* **73**, 601-620, doi:10.1111/1467-8624.00426 (2002).

Graham, J. W. Missing data analysis: Making it work in the real world. *Annual Review of Psychology* **60**, 549-576, doi:10.1146/annurev.psych.58.110405.085530 (2009).

Green, D. P., Ha, S. E. & Bullock, J. G. Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *The Annals of the American Academy of Political and Social Science* **628**, 200-208 (2010).

Huston, A. C. *et al.* Impacts on children of a policy to promote employment and reduce poverty for low-income parents: New hope after 5 years. *Developmental Psychology* **41**, 902-918 (2005).

Kling, J. R., Liebman, J. B. & Katz, L. F. Experimental analysis of neighborhood effects. *Econometrica* **75**, 83-119, doi:10.1111/j.1468-0262.2007.00733.x (2007).

Manski, C. F. Anatomy of the selection problem. *Journal of Human Resources*, 343-360 (1989).

Manski, C. F. Nonparametric bounds on treatment effects. *The American Economic Review*, 319-323 (1990).

Manski, C. F. *Learning about social programs from experiments with random assignment of treatments*. Vol. 9505 (Institute for Research on Poverty, University of Wisconsin--Madison, 1995).

Romano, J. P. & Wolf, M. Stepwise multiple testing as formalized data snooping. *Econometrica* **73**, 1237-1282 (2005).

Valeri, L. & VanderWeele, T. J. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods* **18**, 137, doi:10.1037/a0031034 (2013).

Westfall, P. H. & Young, S. S. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279 (John Wiley & Sons, 1993).

Table 1: Conceptually-Related Family Processes

FAMILY INVESTMENT PATHWAY MEASURES

Economic wellbeing
Household income
Indicators of economic hardship
Food insufficiency
Assets and debt
Total household expenditures
Neighborhood quality
Neighborhood poverty
Perceptions of neighborhood safety (safety, victimization)
Housing quality
Crowding/number of rooms
Type of housing
Housing problems
Child-related enrichment expenditures
Children's books and toys
Out-of-pocket nonparental care
Parent-child interaction and environment
Infant/Toddler HOME Inventory (stimulating toys and activities, home organization)
Parent-child language
Mean length of utterance
Types and tokens from videotaped interaction (free play and clean-up task)

FAMILY STRESS PATHWAY MEASURES

Family stress
Chaos in the home
Parental stress index
Family stress index
Maternal hair cortisol
Maternal salivary cortisol
Mother's health and resources
Maternal physical health
Maternal mental health
Maternal cognitive resources ("bandwidth")
Sensitivity of parenting
Parental warmth as indexed on the Infant/Toddler HOME Inventory
Child protective services
Synchronicity of maternal and child cortisol reactivity
Child stress measures
Child hair cortisol
Child salivary cortisol

RELATED FAMILY PROCESSES

Parental work histories and schedules: total hours (full or part time), number of jobs, days worked, regularity of work schedule

Nonparental care experiences: number and type of providers, hours in care, regularity of care, qualities of care
Maternal romantic relationships: marital status, relationship quality with biological father and/or other romantic partner, presence of domestic violence