I. General

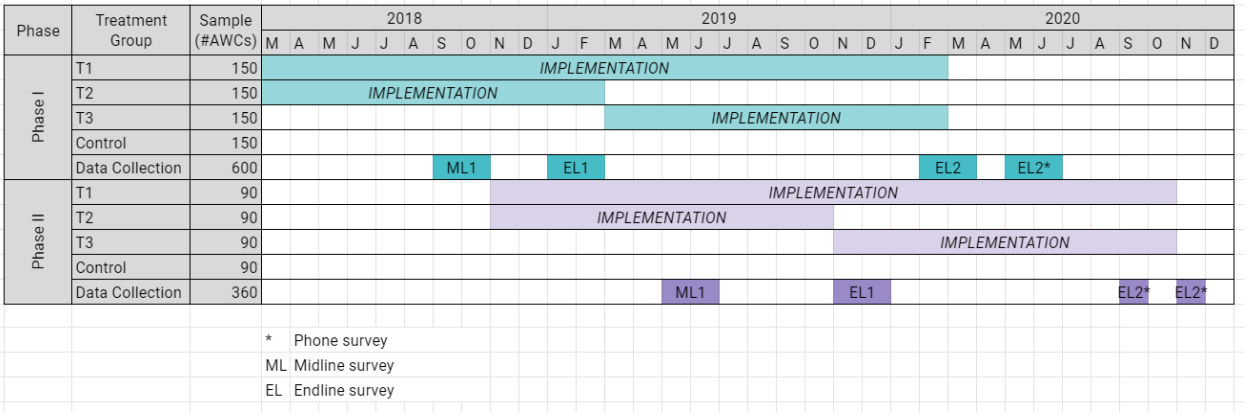
Project Title: Targeted Unconditional Framed Cash Transfers for Early Childhood Health and Nutrition Outcomes

Principal Investigators: Karthik Muralidharan (University of California, San Diego); Paul Niehaus (University of California, San Diego); Sandip Sukhtankar (University of Virginia); Jeffrey Weaver (University of Southern California)

Project Overview: The study is conducted across 960 anganwadi centres in 8 districts of Jharkhand, India. Districts 1-5 fall under Phase 1 of the study, and districts 6-8 fall under Phase 2. The target population is children under two years of age and their mothers. The main outcome variables are child anthropometrics (weight-for-age, height-for-age, and weight-for-height-for-age) and maternal nutrition (dietary diversity and caloric intake). Data on nutrition is collected as part of a longer household survey, while weight and height are recorded in a separate anthropometric survey conducted by highly trained enumerators. All instruments have been shared.

Data Description: All datasets were prepared using R and Stata 15. The variable “uid” uniquely identifies the mother in all datasets. However, the observations in the data are at the child level, which means there are a few duplicate uids, where the respondent has given birth to twins. Through all rounds of surveys, we have used -87 to record “other” responses that the respondent may give which are not prespecified options, -88 for “did not know,” and -89 as “did not answer.” Detailed information about each dataset is below, followed by information pertaining to the consumption module and indices in the household survey.

Data Collection Timelines:



II. Midline Data Publication

- P1Y1 midline anthropometrics: The survey was conducted in the 5 Phase 1 districts from 24th September 2018 to 11th December 2018. 3306 households were part of the study sample in this phase, of which we were able to successfully survey 2914. The main reasons for unsuccessful attempts were miscarriages/ child deaths (176), or respondents not being available (182). 1 household refused to participate. Data was collected on the mother's height, child's height, mother's weight, child's weight, and child's morbidity. The child's height was not recorded in this survey. As a standard protocol, the mother's weight alone and with her child were recorded twice each. If the discrepancy between two readings was larger than 0.5 kilograms, it was recorded a third time.
- P2Y1 midline anthropometrics: The survey was conducted in the 3 Phase 2 districts from 10th June 2019 to 9th August 2019. 2429 households were part of the study sample in this phase, of which we were able to successfully survey 1903. The main reasons for unsuccessful attempts were miscarriages/ child deaths (192), respondents not being available (189), or children being over the target age group (124). 9 households refused to participate. Data was collected on the mother's height, child's height, mother's weight, and child's weight. As a standard protocol, all anthropometrics were recorded twice each. If the discrepancy between two readings was large (greater than 0.5 kilograms in case of weight and greater than 0.5 centimetres in case of height), it was recorded a third time.
- P2Y1 midline household: The survey was conducted in the 3 Phase 2 districts from 10th June 2019 to 9th August 2019. 2429 households were part of the study sample in this phase, of which we were able to successfully survey 1894. The main reasons for unsuccessful attempts were miscarriages/ child deaths (192), respondents not being available (195), or children being over the target age group (124). 10 households refused to participate. Data was collected on the mother's morbidity, child's morbidity, mother's dietary diversity, child's dietary diversity, mother's receipt of government services and schemes, mother's interactions with financial institutions, and household's expenditure on food and other items.

III. Endline I Data Publication

- P1Y1 endline anthropometrics: The survey was conducted in the 5 Phase 1 districts from 23rd January 2019 to 2nd March 2019. 3133 households were part of the study sample in this phase, of which we were able to successfully survey 2811. The main reasons for unsuccessful attempts were respondents not being available (233), children being over the target age group (43), or miscarriages/ child deaths (31). 1 household refused to participate. Data was collected on the mother's weight, child's weight, child's height, and child's morbidity. The mother's height was not recorded. As a standard protocol, all anthropometry measures were recorded twice each. If the discrepancy between two

readings was large (greater than 0.5 kilograms in case of weight and greater than 0.5 centimetres in case of height), it was recorded a third time.

- P1Y1 endline household: The survey was conducted in the 5 Phase 1 districts from 23rd January 2019 to 2nd March 2019. 3133 households were part of the study sample in this phase, of which we were able to successfully survey 2798. The main reasons for unsuccessful attempts were respondents not being available (246), children being over the target age group (43), or miscarriages/ child deaths (31). 2 households refused to participate. Data was collected on the mother's morbidity, child's morbidity, mother's dietary diversity, child's dietary diversity, mother's receipt of government services and schemes, mother's interactions with financial institutions, and household's expenditure on food and other items.
- P2Y1 endline anthropometrics: The survey was conducted in the 3 Phase 2 districts from 17th November 2019 to 14th December 2019. 2112 households were part of the study sample in this phase, of which we were able to successfully survey 1881. The main reasons for unsuccessful attempts were respondents not being available (189), children being over the target age group (12), or miscarriages/ child deaths (10). 9 households refused to participate. Data was collected on the mother's height, child's height, mother's weight, and child's weight. As a standard protocol, all anthropometrics were recorded twice each. If the discrepancy between two readings was large (greater than 0.5 kilograms in case of weight and greater than 0.5 centimetres in case of height), it was recorded a third time.
- P2Y1 endline household: The survey was conducted in the 3 Phase 2 districts from 17th November 2019 to 14th December 2019. 2112 households were part of the study sample in this phase, of which we were able to successfully survey 1876. The main reasons for unsuccessful attempts were respondents not being available (190), children being over the target age group (12), or miscarriages/ child deaths (10). 12 households refused to participate. Data was collected on the mother's morbidity, child's morbidity, mother's dietary diversity, child's dietary diversity, mother's receipt of government services and schemes, mother's interactions with financial institutions, and household's expenditure on food and other items.

Notes on Consumption Module

- Large outliers for nutrient z-scores: There are a handful of cases where the mother's nutrient z-score is particularly high. These are a direct result of the respondent's recorded consumption of given ingredients, rather than a peculiarity with the way the z-scores are calculated. The variables as of now are not winsorized, but it would be reasonable for the user to do so for any analysis.
- Large outliers for calorie variables: There are several cases where the respondents' estimated number of calories consumed far exceeds what is reasonably expected of an individual. We winsorize all calorie variables at the 99th percentile. As such, all subsequent construction of variables (e.g. nutrients, nutrient indexes, etc.) are based on these winsorized variables.

Index Construction

- Empowerment index (0 - 5) [section D in instrument]
    - The empowerment index contains five components: whether the respondent needs permission to visit relatives; visit the market; purchase medicine for their child; whether the respondent considers themselves financially dependent; and who in the household decides what to spend money on. Note that for Phase 2, we collect responses to four additional questions, but do not include them in the construction of the index in order to maintain congruence with the Phase 1 index. The variables used to construct the empowerment index are emp_visit_permission emp_mkt_permission emp_financially_dep emp_meds_decision and emp_fin_decision.
- Depression index (0 - 5) [section C in instrument]
    - The depression index is constructed using five components that ask: In the past two weeks, did the respondent feel nervous or stressed; unable to complete their daily tasks; get poor sleep; have little energy; struggle to concentrate during the day. Note that Phase 1 and Phase 2 components are identical. The variables used to construct the empowerment index are e4_1_1 e4_1_2 e4_1_3 e4_1_4 and e4_1_5.
- Dietary diversity module [derived from section B in instrument]
    - Nutrient Index: The nutrient index is calculated in three steps, and each of the intermediate variables are included in the published dataset. First, we calculate the percentage of recommended nutrients consumed by the mother based on the total grams of each ingredient consumed in the 24-hour recall module. Using data from digitized Indian Food Composition Tables (IFCT), we are able to estimate nutrient values for each ingredient. We sum nutrient values across all ingredients to create a variable that is the total amount of each nutrient consumed by the mother. We then calculate the percentage of recommended nutrient uptake based on whether the mother is a) pregnant, b) lactating (0-6 months), c) lactating (6-12 months), or d) engaged in moderate work. These recommended intakes are based on values given in the National Institute of Nutrition guidebook. Next, we turn these recommended percentages into nutrient z-scores. The z-score calculation is based on the method outlined in Kling et al. (2007). For each nutrient, we calculate the control mean and control standard deviation, then for nutrient *n* consumed by individual *i*, we calculate the z-score to be $(n_i - n_{control\_mean}) / n_{std\_dev}$.
    - DQI-I: The Dietary Quality Index-International (DQI-I) values are based on this document and are calculated based on the 24-hour recall module in the survey. The complete DQI-I score is out of 100 and is the sum of the variety, adequacy, moderation, and balance subcomponent scores, enumerated below.
        - Variety (0 - 15)
            - 5 food groups: meat/poultry/fish/egg, dairy/beans, grains, fruits, and vegetables. Each food group is awarded either 0 or 3 pts. 3 points are awarded if at least 1 item from that group was consumed. The variety score is calculated using variables

var_dairybeans_mom var_nonveg_mom var_fruits_mom var_vegetables_mom and var_grains_mom.

- Variety: protein (0 - 5)
    - 6 sources: meat, poultry, fish, dairy, beans, and eggs. If 3 or more sources consumed: 5 pts; if 2 sources consumed: 3 pts; 1 source consumed: 1 pts; 0 sources consumed: 0 pts. The variety protein score is calculated using variables var_prot_meat_mom var_prot_poultry_mom var_prot_fish_mom var_prot_dairy_mom var_prot_beans_mom and var_prot_eggs_mom.
- Adequacy (0 - 40)
    - 8 groups: vegetables, fruit, grain, fiber, protein, iron, calcium, and vitamin C. Between 0 and 5 points awarded for each of the 8 adequacy groups based on whether the mother consumed the Recommended Daily Allowance (RDA). These RDA values are obtained from the National Institute of Nutrition guidebook and are based on the number of servings consumed and the overall energy (calorie) intake of each individual. For example, >= 2 servings of fruit for an individual with an intake of <= 1700 calories corresponds to a score of 5. In that example, < 2 servings of fruit correspond to a score of 0. These scoring criteria depend on each food group and on the caloric intake of the respondent. The adequacy score is calculated using variables fruits_adeq_mom veg_adeq_mom grains_adeq_mom fiber_adeq_mom prot_adeq_mom iron_adeq_mom calc_adeq_mom and aacid_adeq_mom.
- Moderation (0 - 30)
    - 5 groups: total fat, saturated fat, cholesterol, sodium, and empty calorie foods. Between 0 and 6 points awarded for each of the 5 moderation groups, depending on percentage of RDA met. For example, an individual whose visible fat calorie intake is <= 20 percent of total calories consumed would receive a moderation score of 6, > 20 percent and <= 30 percent would receive a score of 3, and > 30 percent would receive a score of 0. These scoring criteria depend on each food group. The moderation score is calculated using variables moderation_vfat moderation_satfat moderation_chol moderation_sodi and moderation_empty_cal.
- Balance (0 - 10)
    - 2 groups: macronutrient ratio, fatty acid ratio. Between 0 and 6 points awarded, depending on the ratio of macronutrients and between 0 and 4 points awarded depending on the ratio of fatty acids consumed. For example, an individual whose percentage calories from carbohydrates are between 55 and 65% of total calories consumed, and whose percentage calories from protein are between 10 and 15% of total calories consumed, and whose

percentage calories from visible fat are between 15 and 25% of total calories consumed, would receive a score of 6. These scoring criteria depend on each food group. The balance score is calculated using variables balance_macronut_ratio and balance_fatty_acid_ratio.

- Minimum meal frequency is a dichotomous indicator based on WHO and DHS calculations and is defined as follows: an appropriate meal frequency is defined as breastfeeding two to three times a day between 6 and 8 months, increasing to three to four times a day between 9 and 23 months, with nutritious snacks offered once or twice a day as desired.
- Mother's dietary diversity (MDD-M) is a measure of whether or not women 15 to 49 years of age have consumed at least five out of seven defined food groups the previous day or night. The seven food groups are: grains, roots and tubers; legumes and nuts; dairy products (milk, yogurt, cheese); flesh foods (meat, fish, poultry and liver/organ meats); eggs; vitamin A rich fruits and vegetables; other fruits and vegetables. We create two separate measures using this information, a continuous variable indicating the number of food groups consumed (0 - 7) and a dichotomous indicator of whether dietary diversity was met. The MDD-M is calculated using variables mddmgrains mddmlegumesnuts mddmdairy mddmfleshfood mddmeggs mddmvitamina and mddmotherfruitveg.
- Child dietary diversity (MDD-C) is defined as the proportion of children 6 to 23.9 months of age who receive foods from 4 or more food groups: grains, roots and tubers; legumes and nuts; dairy products (milk, yogurt, cheese); flesh foods (meat, fish, poultry and liver/organ meats); eggs; vitamin A rich fruits and vegetables; other fruits and vegetables. We create two separate measures using this information, a continuous variable indicating the number of food groups consumed (0 - 7) and a dichotomous indicator of whether dietary diversity was met. The MDD-C is calculated using variables mddcgrains mddclegumesnuts mddcdairy mddcfleshfood mddceggs mddcvitamina and mddcotherfruitveg.