STATISTICAL ANALYSIS PLAN

Intention-to-Treat analysis for stunting in 1-year old children

Version 1.2
August 18, 2022

Household air pollution and health: A multi-country LPG stove intervention trial (HAPIN)

Trial Registration: NCT02944682

Modification History:

Version	Date	Changes
1.0	3/24/22	
1.1	5/20/22	Minor (secondary analyses): Fixed an error in the categories of gestational age at intervention used in secondary analyses to match birthweight analysis: from {9–12, 13–19 weeks} to {9–17, 18–29 weeks}.
1.2	8/18/22	Minor (secondary analysis): Fixed an error in the subscript numbering of coefficients in the interaction model. Replaced the coefficient numbers so that there is no overlap in numbering: $\ldots + \beta_3 X_i \times t_{ij} + \sum_{1}^{9} \beta_{k+3} I_{ik}$

1. INTRODUCTION

This document contains the statistical analysis plan (SAP) for stunting outcomes at one year of age for the HAPIN trial. Stunting (i.e., a child's height for a given age and sex that is two standard deviations below the standard) is one of the four primary outcomes. The goal of the SAP is to avoid data-driven analyses during and at the end of the study to the extent possible.

1.1. Background and Rationale

Globally, nearly 3.8 billion people rely on solid fuels for cooking and heating, the vast majority in low- and middle-income countries (LMICs). The resulting household air pollution (HAP) is one of the most important environmental risk factors in the 2019 Global Burden of Disease, accounting for an estimated 2.3 million premature deaths annually, largely among women and young children. Previous interventions have provided cleaner biomass-based cookstoves but have failed to reduce exposure to levels that produce meaningful health improvements. There have been no large-scale field trials with liquefied petroleum gas (LPG) cookstoves, likely the cleanest scalable intervention.

This study will provide evidence — including costs and implementation strategies — to inform national and global policies on scaling up LPG stoves among vulnerable populations. Ultimately, this will facilitate deeper policy-level discussions as well as identify requirements for initiating and sustaining HAP interventions globally.

1.2. HAPIN Study Overview

The aim of the HAPIN study is to conduct a randomized controlled trial of LPG stove and fuel distribution in 3200 households in resource-poor settings in four LMICs (India, Guatemala, Peru, and Rwanda) to deliver rigorous evidence regarding potential health benefits across the lifespan. Each intervention site will recruit 800 pregnant women (aged 18-<35 years, 9 to <20 weeks gestation), and will randomly assign half their households to receive LPG stoves and an 18-month supply of LPG. Controls will not receive the intervention at the commencement of the trial and are anticipated to continue cooking with solid biomass fuels; they will be

compensated for their participation in the study. The mother and offspring will be followed until the child is 12 months old. In households with a second, non-pregnant older adult woman (aged 40 to <80 years) we will also enrol and follow her during the 18-month follow-up period to assess cardiopulmonary, metabolic, and cancer outcomes. To optimize intervention use, we will implement behavior change strategies. We will assess cookstove use, conduct repeated personal exposure assessments to HAP ($PM_{2.5}$, black carbon, carbon monoxide), and collect dried blood spots (DBS) and urinary samples for biomarker analysis and biospecimen storage on all participants at multiple time points. The primary outcomes are birth weight, severe pneumonia, and stunting at age 12 months in the child, and blood pressure in the older adult woman.

1.3. Study Objectives

The HAPIN study will address the following specific aims: (1) using an intent-to-treat analysis, determine the effect of a randomized LPG stove and fuel intervention on health in four diverse LMIC populations using a common protocol; (2) determine the exposure-response relationships for HAP and health outcomes; and (3) determine relationships between LPG intervention and both targeted and exploratory biomarkers of exposure/health effects.

2. STUDY METHODS

2.1. Trial Design

HAPIN is a randomized, two-arm intervention trial with parallel assignment. Study sites in the four countries (Guatemala, India, Peru, Rwanda) were selected and evaluated based on activities conducted in the formative research. HAPIN uses a rolling recruitment process whereby each International Research Center (IRC) will enroll 800 pregnant women (one per household) and an additional approximately 120 older adult women (this will vary by IRC) from the same households who meet inclusion/exclusion criteria (Section 4.1). Key characteristics of each study site is given in Table 2 of the HAPIN design publication (Clasen et al. 2020).

Recruitment and enrollment will occur over approximately 15 months at ~53 pregnant women/8 older adult women per month per IRC. All participants will be followed longitudinally for ~18 months (until the child is age 1).

2.2. Randomization

To ensue balance between arms, households have been randomly allocated to intervention or control arms as and when they consent to participate. To maintain balance of treatment assignments within each study site at the IRCs, a total of 10 randomization strata are implemented as follows.

- The India IRC randomization list is stratified by the two study sites
- The Peru IRC randomization list is stratified by the six study sites
- Guatemala and Rwanda have one site each.

Separate randomization lists have been generated for each field team conducting randomization at each IRC. Two randomization lists are produced for each of those field teams: one for households that include an older adult woman (OAW), and one for households that do not. Additional details on randomization of households can be found in the HAPIN protocol.

2.3. Sample Size Considerations

For the primary outcome, stunting, the power for the hypothesis test for a difference in the proportion of stunting was approximated by

$$\Phi\left(-Z_{1-\alpha/2}+\frac{|logRR|}{\sqrt{\frac{1-p_1}{np_1}+\frac{1-p_2}{np_2}}}\right)$$

where $\boldsymbol{p_1}$ is the risk of stunting in the control group, $\boldsymbol{p_2}$ is the risk of stunting in the intervention group, $\boldsymbol{RR}$ is the true relative risk (i.e. $\boldsymbol{p_2/p_1}$), and $\boldsymbol{n}$ is the common sample size for each of the intervention and control groups. The following assumptions, based on previous studies, were made for power calculations: for stunting, we assumed a risk among the control group of $\boldsymbol{p_1 = 0.30}$. We also took into consideration a 10% attrition rate over one-year follow-up, resulting in an effective sample size of n=1440 per treatment arm for stunting.

	Parameter	Sample Size Per Arm	-20%	-10%	Original	+10%	+20%	Previous Studies Estimate (95% CI)
Relative risk for stunting	p_1	1600	0.79	0.80	0.81	0.83	0.84	0.79 (0.70-0.89)[c]

[c] This comprehensive review article of the health effects of HAP exposure by Bruce et al. (2013) is based on 2 two observational studies for moderate stunting (Z-score < 2 SD).

2.4. Trial Framework
HAPIN is a superiority trial. The primary intention-to-treat analysis is a test of statistical significance to evaluate whether the outcome data are consistent with the assumption of there being no difference between the intervention and control arms. Exposure-response analysis between stunting and exposure during the one-year follow-up period will be conducted as a separate analysis per the original aims of the study.

2.5. Statistical Interim Analyses and Stopping Guidance
No interim analysis will be conducted.

2.6. Timing of Analysis
Analyses will be conducted once data collection are complete and the SAP has been approved and registered.

2.7. Timing of Outcome and Covariate Assessments
Each participating household was followed from enrollment until the index child reached (or would have reached, assuming a live birth and continued vitality) his/her first birthday. For the purposes of this analysis plan for stunting the follow up is through the first year of life. In addition to baseline assessments conducted at recruitment, personal exposures to household air pollution were conducted during pregnancy at 24-28 and 32-36 weeks of gestation, and three times during the first year of life: at <3 months, 6 months and 12 months of age

3. STATISTICAL PRINCIPLES
3.1. Confidence Intervals and P-Values
We will use 95% confidence intervention for the purposes of estimation of health effects.

Intention-to-treat analysis of the primary outcome (stunting as a dichotomous variable) will utilize a two-sided test at an α-level of 0.0125. The Bonferroni correction for multiple testing, while conservative, is used to control for familywise type I error rate to be 0.05 under any dependence structure among the four HAPIN primary outcomes.

Analysis of all secondary outcomes will use an α-level 0.05 to identify statistical significance. If subgroup analyses have more than two categories, simultaneous hypothesis tests will be used.

3.2. Adherence and Protocol Deviations
All homes in the intervention arm will be equipped with Stove Use Monitoring Systems (SUMS) from Geocene on their traditional stoves, as well as a subset of approximately 80 homes in the control arm. SUMS data will be collected longitudinally. Compliance will be checked every two weeks when SUMS data is downloaded.

Behavioral reinforcements (messages and materials) were delivered when intervention households showed any use of their traditional stoves. We flagged households that are using their traditional stove one or more times over the previous two-week monitoring period. After flagging these households, we probed members of the participating household to ascertain reasons for non-compliance and intervene as necessary. At all behavioral reinforcement visits, a brief questionnaire was conducted to identify the barriers to LPG stove use in the household and document the messages and materials used to address those barriers. Once specific reasons/factors were determined, personalized behavior change reinforcements were delivered.

The primary analysis will be an intention-to-treat (ITT) effect of the intervention on stunting and not a per protocol analysis that considers adherence.

3.3. Analysis Populations

All analyses in this Statistical Analysis Plan will be ITT. Primary and secondary outcomes are listed below in section 5.2. For each outcome, the analysis will include all children who have a valid height measurement (*complete-case analysis*). We define loss to follow-up as any reason that contributes to a missing outcome value, including death or withdrawal of the child prior to the first year of life and no valid height. The same population will be used for exposure-response analyses.

Secondary analyses will use various subsets of the study to examine effect modification.

4. TRIAL POPULATION

4.1. Eligibility

Children will be eligible to participate in the study if they fulfill the following inclusion and exclusion criteria at screening:

Inclusion criteria:

- Offspring of a confirmed pregnancy (hCG positive blood or urine test)
- Offspring of pregnant women aged 18 to <35 years (via self-report)
- Household uses biomass stove predominantly
- Lives in study area
- Offspring of pregnant women 9 – <20 weeks gestation at recruitment confirmed by ultrasound
- Offspring of a Singleton pregnancy (one fetus)
- Was viable fetus with normal fetal heart rate (120-180 beats per minute) at time of ultrasound
- Mother Continued pregnancy at the time of randomization confirmed by self-report
- Mother agrees to participate with informed consent

Exclusion criteria:

- Mother currently smokes cigarettes or other tobacco products
- Mother plans to move permanently outside study area in the next 12 months
- Mother uses LPG stove predominantly, or is likely to use LPG predominantly, in the near future

We enrolled only 1 child per household.

4.2. Recruitment

The following information will be included in the CONSORT flow diagram. All counts will be reported as total and by IRC.

- Reasons for exclusion when assessed for eligibility
    - Not pregnant/no viable fetus
    - Mother outside of age range
    - Does not/will not primarily cook with biomass
    - Planned to move/moved away
    - Unwilling to participate
    - Gestational age out of range
    - Not a singleton
    - Smoker
    - Not in study area

    - Withdrawn by study team/not pursued further
- Participants determined to be ineligible after randomization
- Reasons for exits after randomization
    - Voluntary withdrawal
    - Withdrawn by study team
    - Moved away
    - Pregnancy loss (termination/miscarriage/stillbirth)
- Reasons for exclusion due to missing data
    - Heights outside of window
    - Missing height measurements

4.3. Withdrawal/follow-up

The study will record reasons for exit classified into several categories:

- Not eligible
- Participant voluntary withdrawal
- Withdrawn by study team
- Moved away from study area
- Deceased
- Lost to follow up
- Mother abortion/miscarriage/stillbirth/child death
- Other

For exits due to non-eligibility, voluntary withdrawal and withdrawal by study team, several pre-specified reasons will be used, as well as the option to fill in other reasons. The last completed visit will also be recorded. Reasons for withdrawal and loss to follow-up will be ascertained as soon as possible.

4.4. Baseline Participant Characteristics

For the ITT analysis, baseline characteristics will be summarized by trial arm, overall and separately by each IRC as defined by Table 1. Means and standard deviations will be calculated for continuous variables and percentages will be calculated for categorical variables. Missing data will be reported as a separate category.

Table 1. Baseline characteristics to be reported

Variables	Type	Definition/Assessment Methods
Mother's age (years)	Categorical	Calculated as the date at baseline minus the date of birth. Date at baseline is assigned by the date of visit if not missing. Categorized as <20, 20-24, 25-29, 30-35
Nulliparous (Never having given birth before)	Categorical	If A1 = 1 or (A1 = 0 and A5 = 0 and A6 = 0) then nulliparity = 1; else if A1 ne . then nulliparity = 0; else if A1 eq . then nulliparity = .; A1 = Is this your first pregnancy? A5 = How many of your children were born alive? A6 = How many of your children were stillborn? Yes / No / Missing
Mother's highest level of education completed	Categorical	• No formal education or some primary school • Primary school or some secondary school incomplete • Secondary school or vocational or university/college • Missing
Mother height	Continuous	Average height calculated from two closest heights measurements
Mother's body mass index (BMI)	Continuous	BMI calculated as the average weight (kg) divided by the average height squared (m^2)
Mother's hemoglobin level	Continuous	

Household food insecurity score	Categorical	Categories (corresponding score): • Food secure (0) • Mild (1,2,3) • Moderate (4,5,6) / Severe (7,8) • Missing See http://www.fao.org/3/as583e/as583e.pdf
Mother's minimum diet diversity	Categorical	Categories (corresponding diet diversity score): • Low (< 4) • Medium (4-5) • High (>5) • Missing
Gestational age (weeks)	Continuous	Calculated as the date at baseline minus the date of screening ultrasound plus gestational age at screening, and then divided by 7
Number of people who sleep in this house	Continuous	
Second-hand smoking	Categorical	Whether someone other than the pregnant woman in household smokes (smoking of the pregnant mother was an exclusion criteria) (yes/no/missing)
Assets	Categorical	Responses for each of the following 5 items: TV, radio, mobile phone, bicycle, and bank account. (Yes / No / Missing

5. DATA ANALYSIS

In this section we provide the analysis approach for the intention-to-treat aim. The primary outcome for both approaches is length at 12 months. We present the primary analysis for each aim, along with effect modification and secondary analyses (alternative model specifications, secondary outcomes).

5.2. Outcome Definitions

This section describes each primary and secondary outcome, including data collection approaches and calculations for derived outcomes. Recumbent length was measured to the nearest 0.1 cm using a measuring board. If the first and second length measurements differed by >0.7 cm, a third measurement was taken. The final length will be calculated as the average of the two closest values. Infant length was assessed quarterly (3, 6, 9 and 12 months of age) as described above using the same procedures.

The primary outcome is stunting at 12 months of age represented in Z-scores. Stunting will be assessed using length-for-age z-scores (LAZ) based on the 2006 WHO Child Growth Standards. Stunting is defined as LAZ that is 2 SD below the median (i.e., LAZ < - 2). We will also examine linear growth through age 12 months as a secondary outcome. Implausible values and outliers are identified by a LAZ Z-score falling outside the range of (-6, 6).

Research Question:

Determine the effect of the LPG stove/fuel/educational messaging intervention on:

1. Stunting at 12 months (primary outcome, dichotomous: stunting defined LAZ at 12 months that is 2 SDs below the median WHO MGRS)
2. Length-for-age Z-scores at 12 months (secondary outcome, continuous)
3. Severe stunting at 12 months (dichotomous outcomes defined LAZ at 12 months that is 3 SD below median WHO MGRS)
4. Stunting at 6 months (secondary outcome, dichotomous: stunting defined as LAZ at 6 months that is 2 SDs below the median WHO MGRS)

Sensitivity analysis:

1. Stunting between 3 – 12 months (dichotomous)

Endpoints/outcomes of interest

1. Stunting at 12 months (primary outcome, dichotomous variable)
2. Length-for-age Z (LAZ) score at 12 months (secondary, continuous variable);

3. Severe stunting (secondary outcome, dichotomous variable)
4. Stunting at 6 months (secondary outcome, dichotomous variable).

5.3. Intention-to-Treat Analysis
Primary analysis:
1. For stunting (LAZ < -2) at 12 months of age, we will build a log-binomial regression model to measure the relative risk of stunting between intervention and control groups (intervention/control) with the following test statistic under the null hypothesis that the coefficient (in this case, it is the log RR) is zero (i.e., $\beta_1 = 0$) and adjusted for strata using indicator variables:

$$log(E[LAZ_i(t_i = 12, X_i) < -2]) = \beta_0 + \beta_1 X_i + \sum_{1}^{9} \beta_{k+1} I_{ik}$$

where LAZ < -2 is stunting at 12 months for child $i = \{1, \cdots, n\}$, $X_i = 0$ if control and $= 1$ if intervention for child i, $\hat{\beta}_0$ is the log risk of stunting at 12 months for controls, $\hat{\beta}_1$ is the log relative risk of stunting at 12 months between intervention and control participants, and $\hat{\beta}_2$ through $\hat{\beta}_{10}$ are the coefficients for the indicator variables representing nine strata (with one stratum as reference, for a total of 10 strata). In the event of non-convergence, a log-Poisson model will be fit, with sandwich variance estimator at the individual level.

The test for $\beta_1 = 0$ is a Z-test with test statistic as follows:

$$Z = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

To examine if the effect is heterogenous across subpopulations <u>in secondary analyses</u> (i.e., the intervention works better or worse in all subgroups), we plan to include interaction analyses by:

- **Pre-randomization**
  - International Research Center
  - Maternal height
    - We will evaluate interaction with maternal height in three ways:
      - As a continuous variable to test for the interaction between intervention and maternal height (most powerful test)
      - Interaction based on <50th percentile, ≥50th percentile threshold (≥152 cm)
      - Interaction based on using previously accepted/published cut-offs of <151 cm, 151-154 cm, 155 cm
  - Child sex
  - Socioeconomics
    - We will calculate a socioeconomic index based on assets, indicators of wealth and water/sanitation variables using principal components analysis. The first principal component will be used as a socioeconomic index (SESPCA). We will evaluate interaction with socioeconomic status in two ways:
      - As a continuous variable to test for the interaction between intervention and SESPCA (most powerful test)
      - Interaction based on <50th percentile, ≥50th percentile SESPCA score
  - Baseline food insecurity score
    - We will evaluate interaction with baseline food insecurity in two ways:
      - As a continuous variable to test for the interaction between intervention and food insecurity score (most powerful test)
      - Interaction based on <50th percentile, ≥50th percentile food insecurity score
- **Post-randomization**
  - Gestational age (GA)

        - We will use GA at time of receiving the intervention
            - We will evaluate interaction with GA in two ways:
            - As a continuous variable to test for the interaction between intervention and gestational age (most powerful test)
            - Interaction based on two categories: weeks 9-17 and weeks 18-29.
    - Birthweight
        - We will evaluate for interaction with birthweight in three ways:
            - As a continuous variable to test for the interaction between intervention and birthweight or birthweight-for-age Z-scores (most powerful tests)
            - Interaction based on <2500 g and ≥ 2500 g (definition of low birthweight)
            - Interaction based on birthweight-for age Z-score < -1.28 and ≥ 1.28 (10th percentile, i.e., small for gestational age)
    - Exclusive breastfeeding ≥ 6 months (vs. < 6 months)
    - Adherence to the intervention using the fraction of cooking days when the biomass stove was not used based on the SUMS Geocene data *in lieu* of the intervention variable.
        - We will evaluate for interaction with adherence in two ways:
            - As a continuous variable using the fraction of cooking days without biomass stove use.
            - Categorizing the fraction of cooking days without biomass stove use into three groups: <0.9, 0.9 – 0.98, ≥0.99 – 1.

Analyses of secondary outcomes:

2. For length for age at 12 months, we will build a linear regression model to measure the difference in Z-scores between intervention and control (intervention – control) with the following test statistic under the null hypothesis that the coefficient is zero (i.e., $\beta_1 = 0$).

$$LAZ_i\ (t_i = 12, X_i) = \beta_0 + \beta_1 X_i + \sum_1^9 \beta_{k+1} I_{ik} + \varepsilon_i$$

where LAZ is length for age at 12 months for child $i = \{1, \cdots, n\}$, $X_i = 0$ if control and $= 1$ if intervention for child i, $\hat{\beta}_0$ is the average LAZ at 12 months for controls, $\hat{\beta}_1$ is the difference in LAZ at 12 months between intervention and control participants, $\hat{\beta}_2$ through $\hat{\beta}_{10}$ are the coefficients for the indicator variables representing nine strata (with one stratum as reference, for a total of 10 strata), and $\varepsilon_i \sim N(0, \sigma^2)$ is independent and identically distributed.

The test for $\beta_1 = 0$ is a t-test with test statistic as follows:

$$T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

where $se(\hat{\beta}_1)$ is obtained from the square root of the second element of the trace of the matrix $\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ where $\boldsymbol{X}$ is matrix of covariates.

Since the intervention may have had an important effect on length *in utero*, we will include LAZ at birth to separate pre-natal from post-natal effects. In this model:

$$LAZ_i\ (t_i = 12, X_i) = \gamma_0 + \gamma_1 X_i + \gamma_2 LAZ_i(t_i = 0) + \sum_1^9 \gamma_{k+2} I_{ik} + \varepsilon_i$$

LAZ is length for age at 12 months for child $i = \{1, \cdots, n\}$, $LAZ(t = 0)$ is length-for-age at birth, $X_i = 0$ if control and $= 1$ if intervention for child i, $\hat{\gamma}_0$ is the average LAZ at 12 months for controls above and beyond the effect

of $LAZ(t=0)$, $\hat{\gamma}_1$ is the difference in LAZ at 12 months between intervention and control participants above and beyond the effect of $LAZ(t=0)$, and $\varepsilon_i \sim N(0, \sigma^2 I_{n\times n})$ is independent and identically distributed, and $\hat{\gamma}_3$ through $\hat{\gamma}_{11}$ are the coefficients for the indicator variables representing nine strata (with one stratum as reference, for a total of 10 strata).

If $LAZ(t=0)$ has a large percentage of data missing (>10%), then we will replace this variable with birthweight-for-age Z-score as a proxy.

To separate post-natal from prenatal effects of the intervention on length at 12 months, we will conduct a mediation analysis whereby the ratio of $\frac{\hat{\gamma}_1}{\hat{\beta}_1}$ represents the fraction of the intervention effect that is due to the post-natal period (i.e., direct effect) and $1 - \frac{\hat{\gamma}_1}{\hat{\beta}_1}$ represents the fraction of the intervention effect that is due to the prenatal period (i.e., indirect effect). We will utilize the bootstrap (using 2000 samples with replacement) to estimate 95% percentile bootstrap CIs for these ratios. These analyses will be conducted using the mediate function in the mediate package.

To demonstrate that the overall effect is consistently seen in all subpopulations (or if the intervention works better or worse in all subgroups), we plan to include interaction analyses by (see above for definitions):

- **Pre-randomization**
  - International Research Center
  - Maternal height
  - Child sex
  - Socioeconomics
  - Baseline food insecurity score
- **Post-randomization**
  - Gestational age at the time of intervention
  - Birthweight
  - Exclusive breastfeeding ≥ 6 months (vs. < 6 months)
  - Adherence to the intervention using % of days without biomass stove use based on the SUMS Geocene data *in lieu* of the intervention variable.

3. For severe stunting (LAZ < -3) at 12 months of age, we will build a log-binomial regression model to measure the relative risk of stunting between intervention and control groups (intervention/control) with the following test statistic under the null hypothesis that the coefficient (in this case, it is the log RR) is zero (i.e., $\beta_1 = 0$) and adjusted for strata using indicator variables:

$$log(E[LAZ_i(t_i = 12, X_i) \ < -3]) = \beta_0 + \beta_1 X_i + \sum_{1}^{9} \beta_{k+1} I_{ik}$$

where LAZ < -3 is severe stunting at 12 months for child $i = \{1, \cdots, n\}$, $X_i = 0$ if control and $= 1$ if intervention for child i, $\hat{\beta}_0$ is the log risk of stunting at 12 months for controls, $\hat{\beta}_1$ is the log relative risk of stunting at 12 months between intervention and control participants, and $\hat{\beta}_2$ through $\hat{\beta}_{10}$ are the coefficients for the indicator variables representing nine strata (with one stratum as reference, for a total of 10 strata). In the event of non-convergence, a log-Poisson model will be fit, with sandwich variance estimator at the individual level. Given that severe stunting is an uncommon event, estimation of parameters for then 10 strata may yield small values that diverge to infinity. Alternatively, we will adjust for IRC (3 parameters with one IRC as reference) rather than adjust for the strata.

The test for $\beta_1 = 0$ is a Z-test that looks as follows:

$$Z = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

To demonstrate that the overall effect is consistently seen in all subpopulations (or if the intervention works better or worse in all subgroups), we plan to include interaction analyses by (as defined above):

- **Pre-randomization**
  - International Research Center
  - Maternal height
  - Child sex
  - Socioeconomics
  - Baseline food insecurity score
- **Post-randomization**
  - Gestational age at the time of intervention
  - Birthweight
  - Exclusive breastfeeding ≥ 6 months (vs. < 6 months)
  - Adherence to the intervention using % of days without biomass stove use based on the SUMS Geocene data *in lieu* of the intervention variable.

4. For stunting at six months of age, we will build a log-binomial regression model to measure the relative risk of stunting between intervention and control groups (intervention/control) with the following test statistic under the null hypothesis that the coefficient (in this case, it is the log RR) is zero (i.e., $\beta_1 = 0$) and adjusted for strata using indicator variables:

$$log(E[LAZ_i(t_i = 6, X_i) < -2]) = \beta_0 + \beta_1 X_i + \sum_{1}^{9} \beta_{k+1} I_{ik}$$

where LAZ < -2 is stunting at 6 months for child $i = \{1, \cdots, n\}$, $X_i = 0$ if control and $= 1$ if intervention for child i, $\hat{\beta}_0$ is the log risk of stunting at 6 months for controls, $\hat{\beta}_1$ is the log relative risk of stunting at 6 months between intervention and control participants, and $\hat{\beta}_2$ through $\hat{\beta}_{10}$ are the coefficients for the indicator variables representing nine strata (with one stratum as reference, for a total of 10 strata). In the event of non-convergence, a log-Poisson model will be fit, with sandwich variance estimator at the individual level.

The test for $\beta_1 = 0$ is a Z-test that looks as follows:

$$Z = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

To demonstrate that the overall effect is consistently seen in all subpopulations (or if the intervention works better or worse in all subgroups), we plan to include interaction analyses by (as defined above):

- **Pre-randomization**
  - International Research Center
  - Maternal height (<50th percentile, ≥50th percentile)
  - Child sex
  - Socioeconomics
  - Baseline food insecurity score
- **Post-randomization**
  - Gestational age at the time of intervention
  - Birthweight
  - Exclusive breastfeeding ≥ 6 months (vs. < 6 months)
  - Adherence to the intervention using % of days without biomass stove use based on the SUMS Geocene data *in lieu* of the intervention variable.

Secondary analysis of longitudinal data:
Length-for-age Z-score at 12 months may be missing in ~ 15-20% of children due to the COVID-19 pandemic (percentage missing to be confirmed by DMC). We propose conducting a secondary analysis that includes

information about stunting in earlier time periods to utilize information for all children. For stunting between 3 – 12 months age, we will build a multivariate generalized linear mixed model with a log link function (binomial distribution) and random effects to measure the average difference in the risk of stunting between control and intervention (intervention – control):

$$log\big(E\big[LAZ_{ij}\left(t_{ij}, X_i\right) < -2\big]\big) = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \beta_2 X_i + \sum_1^9 \beta_{k+2} I_{ik}$$

where LAZ_{ij} is = 1 if child $i = \{1, \cdots, n\}$ is stunted (i.e., LAZ < -2) and = 0 if not at time t where $j = \{3, 6, 9, 12\}$; $X_i = 0$ if control and $= 1$ if intervention for child i; $\hat{\beta}_0$ is the log risk of stunting at birth; $\hat{\beta}_1$ is the difference in log risk of stunting between time j and birth; $\hat{\beta}_2$ is the average difference in log risk of stunting between the intervention and control arms (intervention – control); $\hat{\beta}_3$ through $\hat{\beta}_{11}$ are the coefficients for the indicator variables representing nine strata (with one stratum ais reference, for a total of 10 strata); b_{0i} represents the random intercept where $b_0 \sim N(0, {\sigma_0}^2)$; b_{0i} represents the random slope where $b_1 \sim N(0, {\sigma_1}^2)$. In the event of non-convergence:

- A log-Poisson model will be fit, with sandwich variance estimator at the level of each observation of each child.
- We will only fit a random intercept and ignore the slope.
- We will consider the use of generalized estimating equations.

We will also evaluate an interaction between intervention and age using the following model:

$$log\big(E\big[LAZ_{ij}\left(t_{ij}, X_i\right) < -2\big]\big) = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \beta_2 X_i + \beta_3 X_i \times t_{ij} + \sum_1^9 \beta_{k+3} I_{ik}$$

To display the results of model, we will plot the relative risk (and corresponding 95% CI) as a function of exact age to visualize age-specific intervention effects.

To demonstrate that the overall effect is consistently seen in all subpopulations (or if the intervention works better or worse in all subgroups), we plan to include interaction analyses by (as defined above):

- **Pre-randomization**
  - International Research Center
  - Maternal height (<50th percentile, ≥50th percentile)
  - Child sex
  - Socioeconomics
  - Baseline food insecurity score
- **Post-randomization**
  - Gestational age at the time of intervention
  - Birthweight
  - Exclusive breastfeeding ≥ 6 months (vs. < 6 months)
  - Adherence to the intervention using % of days without biomass stove use based on SUMS Geocene data *in lieu* of the intervention variable.

Missing Data. We will conduct complete-case analysis for the outcome data regardless of the amount of missing data. We will provide proportions of missing data for key variables discussed above.

5.4. Analysis Replication Plan

The intention-to-treat and exposure-response analyses will be replicated by two independent analysts (Larry Moulton/Shakir Hossen/William Checkley and Howard Chang/Dong-Yun Kim). Secondary analyses of any outcome related to sensitivity analyses (i.e., alternative health model specifications, alternative covariate specification) will not be replicated.

The replication team will receive the following from the Data Management Core (DMC).

1. A cleaned analytic dataset where exclusions have been applied following the CONSORT diagram. The dataset will also include maternal characteristics at baseline, covariates for subgroup analysis and covariates to include in the exposure-response analyses. Two databases will be provided:
    a. A blinded dataset where the analysis is done by one of the independent analysts (Shakir Hossen) before the unblinded dataset is replicated.
    b. An unblinded database for replication and final analysis purposes.
2. A table summarizing maternal characteristics at baseline (overall and by IRC).
3. The set of outcomes (primary and secondary) and subgroup analysis to be replicated.
4. For the exposure-response analysis only, the list of pre-specified covariates to be included in the regression models and forms of the exposure-response function.

Specific replication tasks include:

1. Replicate summary statistics (e.g., mean, standard deviation, percentages, proportion missing) in the baseline characteristic table.
2. Replicate intention-to-treat analyses for primary and secondary outcomes according to models specified in Section 5.3.
3. Replicate results from interaction analyses (intention-to-treat only).