STATISTICAL ANALYSIS PLAN

Intention-to-Treat and Exposure-Response Analysis for Birth Weight

Version 1.6
September 13, 2021

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

Trial Registration: NCT02944682
Protocol Version: HAPIN - Main Study Protocol v11.0 28 May 2021

Modification History: NA

1. INTRODUCTION

This document contains the statistical analysis plan (SAP) for birth weight of the HAPIN Study. Birth weight 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 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 the most important environmental risk factor 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 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 will be followed along with her child until the child is 1 year 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 in order 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 1 year 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, 2-arm intervention trial with parallel assignment. Study sites in the four countries (Guatemala, India, Peru, Rwanda) have been 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, birth weight, the power for the hypothesis test for difference in mean birth weight is approximated by

$$\Phi\left(-Z_{1-\alpha/2}+\frac{|\Delta|}{\sqrt{2\sigma^2/n}}\right)$$

where Φ is the cumulative distribution function of a standard normal distribution, Z_q is the qth quantile of the standard normal distribution, Δ is the true difference in mean between the control and intervention group, σ^2 is the common variance in birth weight among treatment and control groups, and n is the common sample size for each of the treatment group and the control group.

We assume birth weight has a standard deviation σ = 437 and evaluated this assumption by $\pm$10% and $\pm$20%. We assume a 10% attrition over the 1-year follow-up period, resulting in an effective sample size of $n = 1440$ per arm for birth weight. Hypothesis tests are two-sided at an α-level of 0.0125. The minimal detectable difference in mean birth weight with > 80% power ranges from 43 to 65 grams. The power evaluation does not consider randomization strata, which may increase power by reducing residual variation in birth weight. A previous randomized trial of improved cookstoves in Guatemala (RESPIRE) estimated the difference in birth weight of 89 grams (95% CI: -27, 204).

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 birth weight and exposure during pregnancy 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
All analysis 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 are to be followed from enrollment until the index child reaches (or would have reached, assuming a live birth and continued vitality) his/her first birthday. For the purposes of this analysis plan for birth weight the follow up is through birth. In addition to baseline measurements at recruitment, women have two further assessments at 24-28 and 32-36 weeks of gestation.

3. STATISTICAL PRINCIPLES

3.1. Confidence Intervals and P-Values
All confidence intervals will be presented at 95% confidence.

Intention-to-treat analysis of the primary outcome (birth weight as a continuous 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 family-wise type I error rate to be 0.05 under any dependence structure among the four HAPIN primary outcomes.

Subgroup analysis will use an α-level 0.05 to identify statistically significant effect modifications. If the effect modifiers have more than two categories, simultaneous hypothesis tests will be used.

Analysis of air pollution exposure-response associations and analysis of secondary outcomes will use an α-level 0.05 to identify statistical significance.

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

Behavioral reinforcements (messages and materials) will be delivered when intervention households show any use of their traditional stoves. We will flag households that are using their traditional stove one or more times over the previous two-week monitoring period. After flagging these households, we will probe members of the participating household to ascertain reasons for non-compliance and intervene as necessary. At all behavioral reinforcement visits, a brief questionnaire will be 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 are determined, personalized behavior change reinforcements will be delivered.

The intention-to-treat analysis of birth weight will not consider adherence.

3.3. Analysis Populations
The primary analysis of primary outcome and secondary outcomes will be intention-to-treat (ITT). For each outcome, the analysis will include all recruited pregnancies that have a valid outcome measurement (*complete-*

case). We define loss to follow-up as any reason that contributes to a missing outcome value, including death of the mother prior to birth, miscarriage, stillbirth, no valid birth weight measured and withdrawal from study prior to birth. The same population will be used for exposure-response analyses.

Secondary analysis may use various subsets of the study to examine effect modification.

4. TRIAL POPULATION

4.1. Eligibility

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

Inclusion criteria:

- Confirmed pregnancy (hCG positive blood or urine test)
- Aged 18 to <35 years (via self-report)
- Uses biomass stove predominantly
- Lives in study area
- 9 – <20 weeks gestation confirmed by ultrasound
- Singleton pregnancy (one fetus)
- Viable fetus with normal fetal heart rate (120-180 beats per minute) at time of ultrasound
- Continued pregnancy at the time of randomization confirmed by self-report
- Agrees to participate with informed consent

Exclusion criteria:

- Currently smokes cigarettes or other tobacco products
- Plans to move permanently outside study area in the next 12 months
- Uses LPG stove predominantly, or is likely to use LPG predominantly, in the near future

If two pregnant women live in the same household and are interested in participating, the one with the earliest gestational age will be chosen to participate.

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
  - Birth weights excluded, outside of 24hr window
  - Birth weights missing, no measured birthweight

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 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 intervention versus control arms, 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 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 intentional to treat and exposure-response aims. The primary outcome for both approaches is birth weight. 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 outcomes, including data collection approaches and calculations for derived outcomes.

The primary outcome is birth weight in grams.

Birth weight is measured in duplicate to the nearest gram within 24 hours of birth by a trained fieldworker or nurse. Newborns are weighed naked or in a pre-weighed blanket, typically at the health facility where infants are delivered. If the 2 weight measurements differ by more than 10 g, then a third weight measurement is taken. If the HAPIN birth weights are missing, invalid or measured over 24 hours after birth, non-HAPIN medical professional measurements at birth from medical records or birth certificate are used. Implausible values and outliers are identified by:

- the birth weight-for-gestational age z-score falling outside of (-6, 5);
- the length-for-gestational age z-score falling outside of (-6, 6);
- the weight-for-length z-score falling outside of (-5, 5);
- the head-circumference-for gestational age z-score falling outside of (-5,5).

Children are excluded if their gestational age at birth is greater than 300 days because a Z-score cannot be calculated.

Secondary Outcomes are given in Table 2.

Table 2. Definition and assessment methods for secondary outcomes		
Parameter	**Timing**	**Assessment Method**
Birth weight among full term infants	At birth	Birth weight among births with gestational age >= 37 weeks.
Z-scores for birth weight	At birth	Z-score for weight adjusted by gestational age defined using INTERGROWTH tables (intergrowth21.tghn.org).

Low birth weight	At birth	Dichotomized with low birth weight defined as < 2500 grams

5.3. Intention-to-Treat Analysis

Table 3 summarizes the intention-to-treat (ITT) analysis methods of each outcome. All analyses will adjust for 10 randomization strata using dummy variables. For the intention-to-treat analyses of any outcome, no baseline covariate-adjusted effects will be estimated.

Table 3. Statistical models for intentional-to-treat analysis of primary and secondary outcomes

Outcome and Study Aims	Statistical Models
Birth weight ITT analysis	Linear regression model with indicator for intervention and indicators for randomization stratum indicators.
Birth weight INTERGROWTH Z-score ITT analysis	Linear regression model with indicator for intervention and indicators for randomization stratum indicators.
Low birth weight ITT analysis	Log-binomial model with indicator for intervention and indicators for randomization stratum indicators.

Subgroup Analysis. Effect modification analyses will be conducted using interaction terms between the indicator variable for the intervention (study arm, control or intervention) and the effect modifiers. The list of pre-specified subgroup analyses for the ITT analysis is given in Table 4. A joint statistical test will be conducted to detect effect modification at a type I error rate of 0.05.

Table 4. Definition for variables for subgroup analysis for intention-to-treat analysis

Parameter	Subgroup Definitions
Infant sex	
International Research Center	Guatemala, India, Peru, Rwanda
Gestational age at enrollment	Early versus late intervention defined by (1) median gestational age when the intervention was installed, and (2) by first trimester versus later (12 weeks and under versus > 12 weeks)

Additional Analysis. If imbalance between control and intervention groups for a baseline covariate (Section 4.4) suggests problems with randomization, and the covariate is a potential confounder, covariate-adjusted effects will be evaluated as a sensitivity analysis.

Missing Data. Our primary approach to missing outcome data will be a complete-case analysis by excluding participants without a birth weight record. It is anticipated that missing birth weight will be less than 5% and balanced between intervention arms.

5.4. Exposure-Response Analysis

For each pollutant ($PM_{2.5}$, black carbon and CO), time-weighted exposures will be estimated using 24-hr personal measurements at baseline, first and second follow-up visits. For participants in the control group, we an average will be calculated from all available measurements. For participants in the intervention group, gestational days prior to LPG installation will be assigned the baseline measurements, and gestational days following LPG will be assigned to the average of all post-randomization measurements (up to 2). For the intervention group, if the baseline measurement is missing the mother will be excluded from the analysis.

In the exposure-response analyses, all models will be adjusted for the potential confounders given in Table 5. Confounder selection are based on conceptual directed acyclic graphs, the associated minimal set to eliminate

confounding, and previous studies. Additional covariates are added to explain variance in the outcome (birth weight).

Table 5. A priori covariate adjustments in exposure-response analyses		
Parameter	**Type**	**Subgroup Definitions**
International Research Center and randomization stratification	Categorical	Randomization strata within Guatemala, India, Peru, Rwanda
Infant sex	Binary	
Mother's age at baseline (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'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
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)

Table 6 summarizes the exposure-response (ER) analysis methods of each outcome.

Table 6. Statistical models for exposure-response analysis of primary and secondary outcomes	
Outcome	**Statistical Models**
Birth weight in grams	Linear regression model with covariate adjustment given in Table x.

	Non-linear associations between birth weight and exposure will be evaluated via (1) log transformation of the exposure, (2) categories based on quartiles, (3) parametric splines (knots selection based on quartiles), and (4) penalized smoothing splines. Model selection of the ER function will be based on prediction criteria (i.e., AIC) and ease of interpretability.
Low birth weight and very low birth weight	Log-binomial regression will be applied similar to the analysis of continuous birth weight.

Subgroup Analysis. Subgroup (effect modification) analyses will be conducted using interaction terms between the indicator variable for the intervention (study arm, control or intervention) and the effect modifiers. The list of pre-specified subgroup analyses for the exposure-response analysis is given in Table 7.

Table 7: Definition for variables for subgroup analysis for exposure-response analysis

Parameter	**Subgroup Definitions**
Infant sex	Male, Female
International Research Center	Guatemala, India, Peru, Rwanda

Additional Analyses. The following sensitivity analyses will be conducted:

- Include the following additional covariates in the model (added one at a time to the base model above): marital status, number of people in household, assets. See Table 1 for variable definitions.
- For participants in the intervention group with missing baseline exposure measurement, assign the group mean as the baseline to these visits.
- Examine trimester-specific exposures.
- In the exposure-response analysis, remove participants form the control group without baseline pollutant measurements.

Missing Data. For missing outcome, a complete-case analysis will be carried out by excluding participants without a birth weight record. Missing confounder information will be addressed with the use of a missing categorical variable for each covariate (i.e., the missing by indication approach). In the exposure-response analysis, participants without time-weighted pollutant exposures as defined in Section 5.4 will be excluded in a complete-case analysis.

5.5. Analysis Replication Plan

Selected components of the intention-to-treat and exposure-response analyses will be replicated by an independent analyst. 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.
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 exposure-response analyses for primary and secondary outcomes according to models specified in Section 5.4.
4. Replicate results from effect modification analyses (intention-to-treat only).