Study Protocol

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

Protocol Date: November 17, 2023

Protocol Version: 17.0

NCT02944682

MAIN STUDY PROTOCOL

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

Version 17.0, November 17, 2023

Contents

Abbreviations 9
1. Study Site Selection, Recruitment and Eligibility 10
1.1. Recruitment 10
1.2. Inclusion and Exclusion Criteria 11
2. Baseline Survey and Assessments 12
2.1. Baseline Survey: pregnant women 12
2.2. Baseline Survey: older adult women 12
2.3. Baseline Health Assessment: pregnant women 12
2.3.1. Prenatal exams 12
2.3.2. Ultrasound measurement to assess gestational age and fetal growth 13
2.4. Baseline Health Assessment: older adult women 13
2.4.1. Blood pressure 13
2.4.2. Carotid intima media thickness (CIMT)/brachial artery reaction time (BART) 14
2.4.3. St. George Respiratory Questionnaire (SGRQ) 14
2.4.4. Short Form 36 (SF-36) Questionnaire 14
2.5. Baseline Biomarker Assessment (urine and blood samples): pregnant women and older adult women 15
2.6. Baseline Exposure Assessment: pregnant women and older adult women 15
3. Randomization/Allocation of Interventions 16
3.1. Intervention (Experimental) Arm 18
3.2. Control Arm 18
3.3. Duration of Treatment Period 20
4. Follow-up Assessments 21
4.1. Trial Period 21
4.2. Extended Follow Up 22
5. Health Outcomes 23
5.1. Data collection during pregnancy period 23
5.1.1. Pregnant women 23
5.1.2. Older adult women 23
5.2. Data collection during child's first year of life 24
5.2.1. Child 24
5.2.2. Mother 32
5.2.3. Older adult women 32
5.3 Data collection in cases of death 33

5.4 Data collection during extended follow-up period 33
5.4.1. Recruitment and informed consent....33
5.4.2. Child 34
5.4.3. Mother 36
5.4.4. Older adult woman....37
6. Details of Radiologic Procedures 39
6.1. Training to Perform Ultrasound 39
6.2. Training to perform periodic assessment and maintenance of the ultrasound system...39
6.3. Equipment 39
6.4. Storage of images, quality control, and transfer of quantitative interpreted ultrasound data to the central Data Management Center....39
7. Details of Exposure Assessment Procedures....41
7.1. Instrumentation....42
7.2. Ambient Sampling of Outdoor Air and Long-term Kitchen Sampling 43
7.3. Exposure Data Quality Assurance and Control....44
7.4. Exposure Assessment Training....44
7.5 Additional Measurements in a Sub-sample 44
8. Details of Stove Use Monitoring 46
9. Details of Specimen Collection, Processing, and Storage....47
9.1. Pregnant Women and Older Adult Women Urine Sample Collection....48
9.2. Pregnant Women and Older Adult Women DBS Sample Collection 49
9.3. Pregnant Women, Older Adult Women, and Child Capillary Blood Sample Collection....49
9.4. Child Urine Sample Collection 49
9.5. Child DBS Sample Collection....49
9.6. Short-term Storage of Biospecimens at Intervention Research Centers....50
9.7. Pregnant Women and Older Adult Women Urine Sample Processing....50
9.8. Child Urine Sample Processing 50
9.9. DBS Sample Processing....50
9.10. Long-term storage in Biorepository 50
9.11 Sub-study of Additional Biospecimen (buccal cells, oral rinse, nasal turbinate and venous blood) Sample Collection from a Subset of Older Adult Women in Peru and Guatemala 51
9.11.1. Buccal cell scrape and oral rinse 51
9.11.2. Nasal turbinate brush samples....51
9.11.3. Peripheral blood....51
10. Details of Specimen Analysis 53

10.1. Dried blood samples 53
10.1.1. Endothelial and Cardiovascular Markers and Inflammation Markers 53
10.1.2. Markers of Oxidative Stress 53
10.1.3. Hemoglobin (Hb) and Hemoglobin A1c (HbA1c) 53
10.1.4. Tumor-Associated Antigen Antibodies 53
10.1.5. Cytochrome P450 Induction 54
10.1.6. Lipids 54
10.1.7. Metabolomics 54
10.1.8. MiRNA Arrays 54
10.2. Capillary Blood 54
10.2.1. Hemoglobin 54
10.2.2. Hemoglobin A1c 55
10.3. Urine 55
10.3.1. Analysis of Polycyclic Aromatic Hydrocarbon (PAH) Metabolites 55
10.3.2. Analysis of 3-OH Cotinine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), levoglucosan, 8OH-deoxyguanosine (8OHdG), and Volatile Organic Chemicals (VOC) metabolites 55
10.3.3. Heavy Metals Analysis 55
10.4. Sub-study of Additional Biospecimen Samples Among Older Adult Women 55
10.4.1. Genomic markers 56
10.4.2. Microbiome 56
10.4.3. Epigenetics (DNA methylation) 56
10.4.4. Metabolomics 57
10.4.5. MicroRNA 57
10.4.6. Clinical immune/inflammatory biomarkers 57
11. Human Subjects, Research Ethics and Protection of Human Subjects 58
11.1. Adverse Event Reporting 58
11.2. Potential Risks 58
11.3. Procedures for Protection of Human Subjects 59
11.4. Procedures for Ensuring Confidentiality 60
11.5. Potential Benefits of the Proposed Research to Human Subjects and Others 60
12. Centralized Data Management 61
12.1. Data Management Standard Operating Procedures 61
12.2. Data Quality and Security 61
12.3. Analytic Pipelines 62

12.4. Web Portal 62
13. Data Analysis Plan—Trial Period 63
13.1. Intention-to-treat analysis for primary outcomes (Aim 1) 63
13.2. Study power for primary outcomes 64
13.3. Exposure-Response Analyses (Aim 2) for Primary Outcomes 67
13.4. Secondary and Sensitivity Analyses for the Primary Outcomes 68
13.5. Analysis of secondary outcomes 69
13.6. Missing Data 69
13.6.1. General analytic strategy 69
13.6.2 Assessing the impact of COVID-19 on data completeness and revised strategies for analysis in the presence of missing data. 70
13.7. Economic Evaluation 71
13.8. Implementation Fidelity 71
13.9. Collection of additional contact information at the end of the trial 72
14.Data Analysis Plan—Extended Follow Up Period 73
14.1 Statistical analysis plan for Aim 1 (Intention-to-Treat Analysis) 73
14.2 Aim 2 Statistical Approach 73
14.3 Exposure-Response Analysis (Aim 3) 73
14.4. Power Considerations 74
15. Timeline 75
15.1 Trial period 75
15.2 Extended follow up period 75
References 76
Appendix 1 80
Summary of Formative Research 80
1. Scoping Phase (Phase I) 80
1.1 Goals 80
1.2 Activities 80
1.3 Results on Site Selection 81
1.4 Results on Behavior Change Strategies 83
2. Pilot Intervention Phase (Phase II) 84
2.1 Goals 84
2.2 Progress and Results 85
3. Pilot Assessment Procedures Phase (Phase III) 88
3.1 Goals 88

3.2. Results 88
4. Respiratory Rate/Pulse Oximetry Assessment Phase (Phase IV) 90
4.1 Goals 90
4.2 Accomplishments 90
4.3 Planned timeline for Phase IV 90
Appendix 2: Pilot Study CONSORT Diagrams 92
Appendix 3: SAPs for Main Outcomes 95

Executive Summary

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 2016 global burden of disease, accounting for an estimated 2.6 million 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.

The aim of this 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) and to conduct extended follow up of the cohort in order to deliver rigorous evidence regarding potential health benefits across the lifespan.

In the trial period, 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 enroll 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 and urinary samples for biomarker analysis and biospecimen storage on all participants at multiple time points. The primary outcomes for the trial are birth weight, incidence of severe pneumonia in the first 12 months of life, and stunting at 1 year in the infant, and systolic blood pressure in the older adult woman. Our secondary outcomes are maternal blood pressure, preterm birth, fetal growth, gestational age at birth, infant growth (as a continuous outcome), infant development, and diastolic blood pressure, pulse pressure, mean arterial pressure, CIMT, BART, St. George Respiratory Questionnaire, and SF-36 (quality of life) in the older adult woman. Additional secondary pneumonia outcomes in children include the following: (1) WHO Integrated Management of Childhood Illnesses (IMCI) non-severe pneumonia, (2) WHO IMCI severe pneumonia, (3) Pocketbook WHO non-severe pneumonia, (4) Pocketbook WHO severe pneumonia, (5) Hospitalization for respiratory illness, (6) Lung ultrasound or chest radiograph pneumonia, (7) hypoxemic pneumonia).

We 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.

After participating children reach one year of age, we will continue to follow consenting HAPIN participants annually in India, Guatemala and Rwanda to assess the longer-term health effects of the intervention and their association with exposure to household air pollution. Previous evidence suggests that the benefits of reduced exposure during the first, critical year of development will continue even if the intervention ends. We will continue to follow children and adult participants using methods employed during the HAPIN trial period. The primary outcomes in the extended follow up period are linear growth and neurodevelopment among HAPIN children. The specific aims of the extended follow up period are the following: (1) determine the effects of a 500-day LPG stove and fuel intervention (gestation and first year of life) on child development and growth through age 5; (2) characterize personal exposures to fine particulate matter (PM2.5), carbon monoxide (CO) and black carbon through age 5 and determine the longer-term effects of the intervention on exposure; and (3) evaluate exposure-response between exposure during several critical periods of development and growth. In

addition, Guatemala has an additional aim of determining the effects of the intervention on child respiratory health.

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.

Abbreviations

BART	brachial artery reactivity testing
BC	black carbon
BEC	HAPIN behavior and economics core
BMI	body mass index
CO	carbon monoxide
CCA	common carotid artery
CHW	community health worker
CIC	HAPIN clinical imaging core
CIMT	carotid intima-media thickness
CITI	Collaborative Institutional Training Initiative
CRP	C-reactive protein
DBS	dried blood spot
ECM	enhanced children's microPEM
EFW	estimated fetal weight
FHM	fetal heart motion
GLPGP	global liquefied petroleum gas partnership
HAP	household air pollution
HAZ	height-for-age z-score
HB	hemoglobin
IRC	intervention research center
ICAM-1	intercellular adhesion molecule 1
IL-6	interleukin 6
IMCI	integrated management of childhood illness
IYCF	infant and young child feeding
LAZ	length-for-age z-score
LMIC	low- and middle-income country
LOD	limit of detection
LPG	liquefied petroleum gas
LUS	lung ultrasound
MOH	ministry of health
MPO	myeloperoxidase
PBS	phosphate buffer solution
PID	personal identification number
PM2.5	fine particulate matter
QA	quality assurance
QALY	quality-adjusted life year
QC	quality control
RSD	relative standard deviation
SD	standard deviation
SF-36	short form 26
SGRQ	St. George Respiratory Questionnaire
SOP	standard operating procedure
SRU	Sri Ramachandra University
SUM	stove use monitor
TIP	trial of improved practice
UPAS	ultrasonic personal aerosol sampler
VCAM-1	vascular cellular adhesion molecule 1
VOC	volatile organic chemical
WHO	World Health Organization

1. Study Site Selection, Recruitment and Eligibility

Study sites in the four countries (Guatemala, India, Peru, Rwanda) were selected and evaluated based on activities conducted in the formative research (Phase I Scoping and Phase II Pilot Intervention) – see Appendix for results.

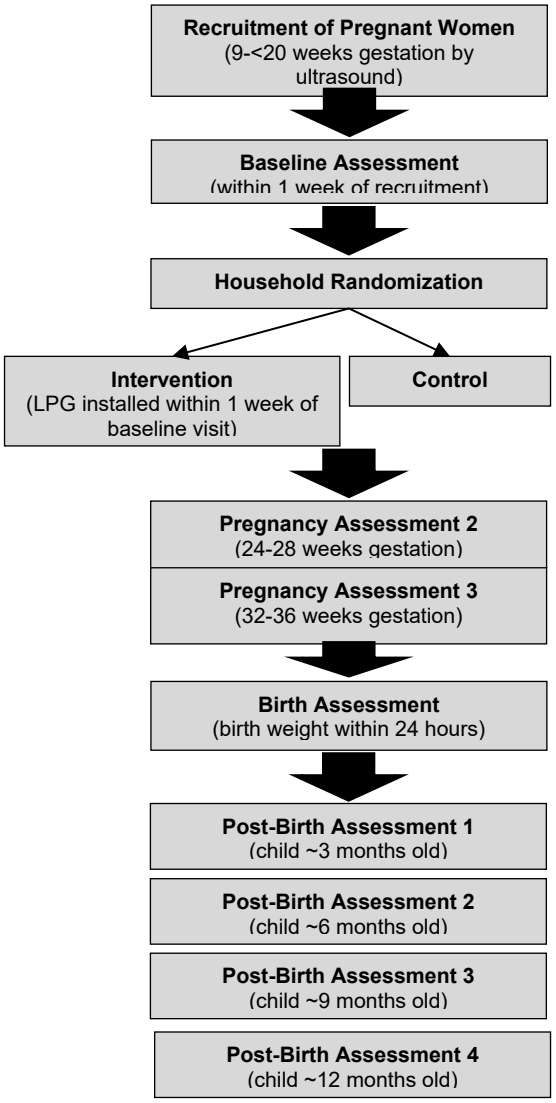
1.1. Recruitment

Participant flow is summarized in **Figure 1**. We will use 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 (see **Section 1.2.**). Recruitment/enrollment will occur over approximately fifteen 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), and the total duration of field activities is planned to span 3.5 years. At the end of the first year of enrollment we will assess the need to enroll additional pregnant women to increase the statistical power for childhood pneumonia. Recruitment has been piloted in Phase II of the formative research (see Appendix).

We will work in cooperation with ministries of health (MoH) to identify eligible pregnant women by working with antenatal clinics and/or community health workers (CHWs). Since the coverage of antenatal care varies across the study sites, CHWs will also help identify women to include those who do not visit clinics thus improving generalizability.

A screening questionnaire will be administered by study staff to confirm eligibility. Informed consent will be obtained from all study participants using standard procedures. Participants will be able to withdraw from the study at any time. We expect to have less than 10% loss to follow-up during pregnancy and early childhood, which is feasible and has been accomplished successfully at our IRCs and in previous studies by study investigators (e.g., Smith et al. 2011, Thompson et al. 2011). Post-birth, we anticipate that loss to follow-up

Figure 1. Participant flow diagram for the pregnant woman and her child. An older adult woman will be enrolled in an estimated 15% of households in addition to the pregnant woman and her child. Exposure and health assessment time points for that older adult woman will overlap with those for the pregnant woman and child (e.g., Pregnancy Assessments 1-3 and Post-Birth Assessments 1-4).



by the 12-month follow-up visit will be about 10% as well. Communication strategies and activities that enhance community engagement will be developed and adapted for each IRC to improve retention.

1.2. Inclusion and Exclusion Criteria

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 youngest gestational age will be chosen to participate.

An **older adult woman** in the same household will be eligible to participate in the study if they fulfill the following inclusion and exclusion criteria at screening:

Inclusion criteria:

- Aged 40 to <80 years (via self-report)

Exclusion criteria:

- Currently smokes cigarettes or other tobacco products
- Pregnant (via self-report)
- Plans to move out of her current household in the next 12 months

If two or more older adult women live in the same household and are interested in participating, one woman will be randomly selected to participate (the one with the next birthday[month and day]).

2. Baseline Survey and Assessments

Our primary outcomes for the trial are birth weight, incidence of severe pneumonia in the first 12 months of life, and stunting at 1 year in the infant, and systolic blood pressure in the older adult woman. Our secondary outcomes are maternal blood pressure, preterm birth, fetal growth, gestational age at birth, infant growth (as a continuous outcome), infant development, and diastolic blood pressure, pulse pressure, mean arterial pressure, CIMT, BART, St. George Respiratory Questionnaire, and SF-36 (quality of life) in the older adult woman. Additional secondary pneumonia outcomes in children include the following: (1) WHO Integrated Management of Childhood Illnesses (IMCI) non-severe pneumonia, (2) WHO IMCI severe pneumonia, (3) Pocketbook WHO non-severe pneumonia, (4) Pocketbook WHO severe pneumonia, (5) Hospitalization for respiratory illness, (6) Lung ultrasound or chest radiograph pneumonia, (7) hypoxemic pneumonia).

Baseline and health visits will be conducted at the participant's home in the morning at the end of the 24-hour exposure sampling period. For brachial artery reactivity testing (BART)/ carotid intima-media thickness (CIMT), assessed at baseline and the sixth visit (child ~12 months old), we will assist the older adult woman to travel to the location where the measurements will be conducted (varies across IRCs). Assessment procedures have been piloted in Phase II and Phase III of the formative research (See Appendix).

2.1. Baseline Survey: pregnant women

Following recruitment and the obtention of informed consent, a baseline survey will be administered by a trained field worker or nurse. This baseline survey covers a wide range of topics including household composition, socioeconomic and demographics data, and pregnancy-related data. Pregnant women will also be asked about their clinical history including medication use. A dietary questionnaire to assess household food insecurity will also be administered.

2.2. Baseline Survey: older adult women

The baseline survey for the older adult women will collect data on dietary intake, physical activity, and clinical history including medication use.

2.3. Baseline Health Assessment: pregnant women

2.3.1. Prenatal exams

Women will have a prenatal exam at <20 weeks gestation (baseline), and again at 24-28 weeks and 32-36 weeks. A nurse or trained field worker will measure resting blood pressure, and maternal weight and height. Any woman with pregnancy-related problems will be referred to the nearest health center or hospital.

Specifically, blood pressure will be measured in triplicate using an automatic monitor. In analyses, the average of all 3 readings will be used. If a participant is found to have a systolic blood pressure $\geq$140 mmHg and/or a diastolic blood pressure $\geq$90 mmHg, their blood pressure will be checked again during the same visit. If a participant has a systolic blood pressure $\geq$140 mmHg and/or a diastolic blood pressure $\geq$90 mmHg on these two measurements they will be referred to the nearest health center or hospital to receive age appropriate treatment. If a participant has systolic blood pressure < 80 mmHg or a diastolic blood pressure < 40 mmHg then they will be referred to the nearest health center or hospital.

2.3.2. Ultrasound measurement to assess gestational age and fetal growth

Ultrasound measurement will be conducted by trained personnel in a clinic or home setting at <20 weeks gestation to determine eligibility and assess fetal growth. An abdominal transducer with frequency of 5 MHz will be used. This provides sufficient resolution to view early stages of fetal development. Measurements will include femur length, abdominal circumference, head circumference, biparietal diameter and crown-rump length. An estimated date of delivery is derived from these individual measures. If a yolk sac is identified at the baseline examination, but an embryo is not visualized (~7 weeks gestation), the fetal ultrasound will be reassessed 2 weeks after the initial measurement. No sex determination will be conducted during the ultrasound examinations. Images and all fetal measurements will be stored digitally and will be compared to international standards for fetal growth established by the INTERGROWTH-21st Project (Papageorghiou et al. 2014a). The fetal crown-rump length will be used if CRL is <8.5cm (85 mm or 9+0 to 13+6 weeks); If CRL is ≥8.5 cm (85mm or 14+0 to 19+6 weeks) then Head Circumference (HC) will be used to establish the gestational dating for the remainder of the pregnancy using a representative international standard (Papageorghiou, et al. 2014b).

2.4. Baseline Health Assessment: older adult women

The following health outcomes will be assessed in all the older adult women in the subsample of 120 households per IRC (this will vary by IRC). The older adult women will be asked to fast overnight prior to the scheduled health assessment when having the BART test (occurring in Peru only, at the baseline and final visit).

BART and CIMT will be undertaken in special ad hoc satellite locations in reasonable proximity to the study subjects, not at HAPIN field offices. The locations are identified with the help of local government or health officials. Transportation is provided solely for participants in in the BART and CIMT sub-studies. This is partly due to the need for fasting (in the case of BART), and is an effort to minimize the inconvenience of attending our designated testing centers for these outcomes. The details by IRC are provided below.

Rwanda: CIMT measurements will be conducted at the health center nearest to the participant's home and we will provide roundtrip transportation.

Guatemala: CIMT measurements will occur in health centers and health posts near participants' homes.

Peru: BART/CIMT measurements will be conducted at the community center nearest to the participant's home and we will provide roundtrip transportation.

India: CIMT measurements will be conducted at one of the three HAPIN ultrasound centers and we will provide roundtrip transportation.

2.4.1. Blood pressure

Following the same protocol as with the pregnant women (see **Section 2.3.1.**), a nurse or trained field worker will measure resting blood pressure. Specifically, blood pressure will be measured in triplicate using an automatic monitor. In analyses, the average of all 3 readings will be used. If the older adult woman is found to have a systolic blood pressure ≥140 mmHg and/or a diastolic blood pressure ≥90 mmHg, their blood pressure will be checked again during the same home visit. If the older adult woman has a systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg on these two measurements they will be referred to the

nearest health center or hospital to receive age appropriate treatment. If the older adult woman has a mean arterial pressure < 65 mmHg or a diastolic blood pressure < 40 mmHg and has either lightheadedness, dizziness, has experienced an episode of fainting, is having chest pain, shortness of breath or an irregular heartbeat, then they will be referred to the nearest health center or hospital.

2.4.2. Carotid intima media thickness (CIMT)/brachial artery reaction time (BART)

To assess cardiovascular health via assessment of plaques, carotid ultrasound will be performed with a portable ultrasound by trained sonographers. Bilateral carotid artery ultrasound of the posterior wall of the distal 1 cm of the common carotid artery (CCA) will be performed as previously described (Stein et al. 2008). CIMT will be conducted following a standardized protocol: two segments will be obtained for each individual, one posterior CCA segment for each carotid artery. The two mean CIMT values will be averaged (Stein et al. 2008). This score has been previously shown to be a reliable and reproducible index of extracranial atherosclerosis (Espeland et al. 1995, Espeland et al. 1996, Crouse et al. 1995).

Plaques will be determined for the common, bulb, and internal carotid segments and defined as focal areas projecting into the vessel lumen with a ≥50% thickness than the bordering areas. Victor Davila-Roman (Washington University-St. Louis) will lead the Clinical and Imaging Core and will review CIMT studies. He will be assisted by Lisa de las Fuentes (Washington University-St. Louis). Both Davila-Roman and de las Fuentes will be blinded to clinical information at the time of image processing and analysis. The intra- and inter-observer ICC for CIMT measurements in Davila-Roman's Laboratory are 0.90 and 0.99, respectively (de las Fuentes et al. 2008, de la Fuentes et al. 2009, Gu et al. 2008). If an atherosclerotic plaque is discovered (e.g., suspicion of severe carotid disease) that participant will be referred to their local health care system for additional evaluation.

In Peru only, we will utilize high-frequency ultrasound visualized BART (Corretti et al. 2002) to monitor endothelial function. BART measures endothelial function via flow-mediated dilatation to reactive hyperemia following the release of arm blood-flow occlusion. In this test, baseline artery diameter is measured, then a blood pressure cuff is inflated in induce distal arm ischemia for 5 minutes and after released, post-occlusion brachial artery diameter is measured. The ratio of post- to pre-occlusion artery diameter represents endothelial function where lower values indicate worse endothelial function.

2.4.3. St. George Respiratory Questionnaire (SGRQ)

We will assess respiratory symptoms using the Saint George's Respiratory Questionnaire (SGRQ)(Jones et al. 1992), which measures impaired health and perceived well-being among individuals with chronic airway disease and offers many advantages for our study, namely: (1) easy to use; (2) can be used to quantify changes in health following treatment, (3) it is not limited to individuals with chronic obstructive pulmonary disease, and (4) it provides a standard metric that can be used for easy comparison across our four diverse settings. Our team of researchers has successfully used the SGRQ in multiple resource-poor settings in previous studies (Miranda et al. 2012, Sherpa et al. 2015).

2.4.4. Short Form 36 (SF-36) Questionnaire

We will administer the SF-36 questionnaire (http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html) to evaluate quality of life. The SF-36 is a public use survey that has been used widely to assess functional status and calculate health-related quality

of life measures (Brazier et al. 2002). It assesses health across eight dimensions and will be used to estimate the quality-adjusted life years (QALYs) saved from the intervention.

2.5. Baseline Biomarker Assessment (urine and blood samples): pregnant women and older adult women

Urine and dried blood spots (DBS) (via finger prick) will be obtained from all participating pregnant mothers and older adult women to assess biomarkers. The research methods are detailed in **Section 9**.

The urine samples will be spot samples. Clean-catch urine will be collected in standard sterilized urine collection cups. Up to 21 mL of urine will be transferred to aliquots in cyrovials, labeled, and refrigerated within 4 hours of collection. As soon as possible but within 12 hours, urine samples will be frozen and kept at -20°C until shipment to the Biomarker Core at Emory, except that samples collected in India will be analyzed by SRU in order to comply with local regulations restricting export of samples.

DBS will be collected on 5 spot Guthrie cards. Fingers will be sterilized with an alcohol wipe then pricked with a sterile lancet and the first drop of blood will be wiped away. All 5 spots will be collected. Cards will be labeled and dried at room temperature (20-25°C) on top of a paper towel or a DBS rack for a minimum of 24 hours without the use of external heat or a fan. Dried cards will be placed in ziptop bags into which desiccant packets have been placed. Urine and DBS will be shipped, express shipment to the Biomarker Core at Emory.

2.6. Baseline Exposure Assessment: pregnant women and older adult women

Both the pregnant women and the older adult women will be monitored for 24-hour personal exposure to HAP ($PM_{2.5}$, carbon monoxide [CO], black carbon [BC]). We will also collect data on temperature and humidity. The methods are detailed in **Section 7**.

3. Randomization/Allocation of Interventions

As we will follow a rolling recruitment process over a 15-month period, and to ensue balance between arms, households will be randomly allocated to intervention or control arms as and when they consent to participate. The randomization procedure is household-based and may be stratified on other variables, such as geographical region within a site.

I. Generation and Distribution of Randomization Lists

The HAPIN Data Management Center (DMC) will generate a list of random treatment/control assignments following procedures described in HAPIN SOP 'DMC-07: Generation of Randomization Lists and Envelopes'.

To maintain balance of treatment assignments within each study site at the IRCs,

a. The India IRC randomization list will be stratified by the two study sites
b. The Peru IRC randomization list will be stratified by the six study sites

Guatemala and Rwanda have one site each.

Separate randomization lists will be generated for each field team conducting randomization at each IRC. Two randomization lists will be produced for each of those field teams: one for households that include an older adult woman (OAW), and one for households that do not.

For each randomization list, sealed envelopes will be prepared by the HAPIN DMC at Emory University containing the treatment/control assignments, following the procedures outlined in HAPIN SOP 'DMC-10 Production of Randomization Envelopes'. The outside of each envelope will be labeled with the following information: IRC, site identifier (where appropriate), field team number, the indicator 'A' (or another indicator, as specified by each individual IRC) for households that include an OAW and 'M' (or another indicator, as specified by each individual IRC) for households that do not, and a sequence number between 1 and 800 corresponding to placement on the randomization list. Inside each envelope there will be a card labeled either 'Intervention: Gas Stove Now' or 'Control: Gas Stove Later' or with other appropriate wording supplied by IRCs in the language and or script of the country. The envelope will also contain another card, identical to the first, which will be left in the household after randomization. Inside the envelope, the printed sides of the cards will face each other to reduce the chance of the assignment being visible from the outside of the envelope.

Each IRC will receive the sealed assignment envelopes from the DMC. The envelopes will be separated by site, OAW/non-OAW status and team number. For example, an IRC may have two different sites and two field teams within each site that may be conducting randomization on any given day. Thus, the IRC would receive four separate sets of envelopes for each site (sets for Team 1 OAW, Team 1 non-OAW, Team 2 OAW and Team 2 non-OAW). Each set of envelopes will be ordered by sequence number and that order must be maintained during storage and handling of the envelopes in order to preserve the integrity of the randomization list on which the set is based.

The DMC will produce and ship enveloped for each required set to the IRCs in batches. Envelope shipments will occur at least every four months until sample size targets have been met. Sufficient numbers of envelopes will be supplied in each batch to allow for the proper conduct of the study.

II. Randomization of Households

Randomization will occur after all baseline questionnaires and procedures have been completed. A verbal confirmation that the pregnancy is continuing will be obtained prior to randomization.

Each day that households will be randomized by a field team, the procedures below will be followed:

1. The HAPIN data manager will direct the field team to one set of OAW envelopes, and one set of non-OAW envelopes, from which that field team should draw for the day. The field team will not use any other sets of envelopes during that day; however, on subsequent days, the field team will not need to draw from the same set of envelopes – a different set of envelopes for that same site may be chosen by that field team on subsequent days.

   The field team will log out (i) a set of envelopes from the OAW household envelopes corresponding to the number of households to be visited plus 8 additional envelopes, and (ii) a set of envelopes from the non-OAW household envelopes corresponding to the number of households to be visited plus 8 additional envelopes.

2. The field team will keep the logged-out envelopes in sequence at all times, except during the presentation of envelopes to the household for selection (step 5 below). Keeping the envelopes in sequence is important in order to maintain the integrity of the randomization list on which the set of envelopes is based.
3. In the household, the field team member will gather all family members present in the household who will participate in the selection of an envelope (to include the pregnant woman at a minimum). The field team member will remind the household of the study goals and show the contents of two (unsealed) sample envelopes, one containing a control assignment, the other an LPG assignment.
4. A field team member will select the first 6 sealed envelopes from the set of sealed, sequenced envelopes. The envelopes must be chosen from either the OAW envelopes, or the non-OAW envelopes, as appropriate for the household in question. The field team member will lightly tap the envelopes on the bottom edge to ensure that the cards within them fall to the bottom of the envelope and are in a similar position in all envelopes.
5. The field team member will lay the envelopes on a table or clean surface, with the non-labeled side face up, and ask the pregnant woman (and/or her family, as appropriate) to point to one envelope. It is important that all family members present in this step agree upon and/or approve of which envelope is selected. The envelope may not be handled (e.g., held up to the light, shaken or felt), as this may compromise the randomization process and might also damage or otherwise spoil the envelopes.
6. The pregnant woman (or someone else in the household if the pregnant woman asks someone else to pick up the envelope for her) will pick up the selected and agreed-upon envelope from the 6 presented envelopes. The envelope will not be opened until step 9 below.
7. The envelopes that were not selected will be removed from the table and re-stored, in sequence, in the set from which they were taken.
8. The HAPIN household ID will be recorded by a field team member on the label on the selected envelope. The site, OAW or non-OAW status, team number and envelope number of the selected envelope will be recorded on the Household Randomization CRF S6.
9. The selected envelope will be opened by the participant. The envelope must be opened by hand by pulling the flap and separating it from the rest of the envelope in such a manner that the “Tamper Resistant” seal is also torn. Letter openers or other implements should not be used. The cards will be removed and examined to reveal treatment status. The treatment status will be recorded on the Household Randomization CRF S6.
10. One of the cards will be replaced in the envelope. If, in step 9, the tamper-resistant seal was not torn during the opening of the envelope, the field team should tear the seal off the envelope as much as possible. The field team will keep the envelope separate from sealed envelopes, and will return it to the HAPIN office at the end of the day’s field activities.
11. The family will be given the other card indicating the randomization assignment. A field-team member will write the household ID on the card, as well as any other information deemed necessary by the IRC (e.g., signatures, date of randomization, etc.)
12. The process continues for each eligible household on the agenda for the day.
13. When the day’s agenda is complete, the remaining unopened envelopes must be returned and logged back into the IRC HAPIN Office. The unused envelopes from the day will be replaced in sequence into the set of remaining assignment envelopes in the office so they will be the first to be selected for the next recruitment visit.

The list of selected envelope IDs will be recorded at the IRC HAPIN Office. The selected envelopes will be securely stored at the office. The HAPIN DMC will perform periodic audits of the used envelopes.

III. **Spoiled or Wasted Envelopes**

If an envelope is worn, soiled or otherwise damaged so that its appearance is visibly altered prior to being selected by any households, it should be removed and not used in household randomizations. In this event, the date of removal must be noted on the label, 'XXXXX' marked on the study ID portion of the label, and the envelope returned to the HAPIN IRC office for secure storage and future DMC audits. The envelope must not be opened, and its contents must not be examined or removed by anyone other than HAPIN DMC staff. The 'Spoiled Randomization Envelope Report' must be completed in LabKey by the IRC data manager.

3.1. Intervention (Experimental) Arm

The intervention consists of a high-quality, locally available LPG stove having at least two burners, a continuous supply of LPG fuel for ~18 months, and the promotion of stove use on an exclusive basis for cooking. The intervention will be provided free of charge to all intervention households upon enrollment; control households will be provided compensation as described in **Section 3.2** below. We have investigated the costs and availability of stoves and LPG fuel in all of the four settings in which the trial will be conducted, and are confident that we can procure and provide the same to all participating intervention households. We will work with government agencies, LPG suppliers and our research partner, the non-profit Global LPG Partnership (GLPGP), to reduce the costs of procuring LPG. On each visit to provide additional gas cylinders in Rwanda, Guatemala and Peru, we will examine stove condition, perform any repairs necessary, and measure and record weight of LPG cylinders in order to anticipate the need for refills. (This cannot be done in India as the public sector oil marketing company is responsible for stove installation, cylinder refills and repairs; study staff can facilitate those aspects for intervention participants.) We will monitor the rate of LPG usage by calculating average kilograms of LPG used per day (using data on gas cylinder weights and number of days between installation and exchange of each cylinder). We will follow up with households who show rates of LPG use that are significantly above or below the average to identify potential issues with sharing of gas cylinders (over-use of LPG) or potential use of traditional stoves (under-use of LPG). If traditional stove use is detected at this visit, an LPG cooking promoter will be sent to the household within one week to conduct a tailored training and behavioral intervention (behavioral reinforcement as described below).

Based on stove usage data (see **Section 8**), behavioral reinforcements (messages and materials) will be delivered when households show any use of their traditional stoves. Participants with low compliance will be probed to determine their reasons for stacking. Once specific reasons/factors are determined, personalized behavior change reinforcements will be delivered. A SOP is being developed to outline the behavior change reinforcement strategy. 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 participants to figure out the reasons for non-compliance and intervene as necessary. This process will be described in a SOP that can be easily used by the field team. 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.

3.2. Control Arm

Control households will not receive the intervention at the commencement of the trial, and are anticipated to continue to cook primarily with solid biomass fuels. However, control households will receive compensation designed to meet three aims: (i) comply with applicable ethics requirements for treatment of controls, (ii)

compensate control participants for the burden associated with this study, with the view of minimizing losses to follow-up, and (iii) offset the economic advantage to intervention households accorded by the provision of free stoves and fuel. Compensation has been designed based on a uniform set of trial-wide principles, then adapted to the particular needs of each IRC based on extensive local formative research and requirements stipulated by applicable IRBs. Study-wide principles include:

1. Retention strategies should not have the potential for impacting the outcomes of interest. For example, directly providing foodstuffs would not be helpful for the design of the study since it could impact child growth and other HAPIN-relevant outcomes.
2. Any strategy to be incorporated into the trial cannot intentionally incentivize practices or behaviors that might put the control group at risk of adverse health outcomes. For example, incentives cannot be conditioned on continued use of biomass stoves. On the other hand, they can be conditioned on continued participation in the trial.
3. The strategy should be created in consideration of the potential for other adverse effects, such as disputes about who receives and can spend the incentive. Cash rewards, and more specifically cash transfer programs have accrued years of experience and common issues have been discussed.
4. The strategies must be approved by applicable IRB/ethics committees and the HAPIN DSMB

The major challenge of this strategy has been to address the economic gradient caused by the supply of fuel. This has been a major issue from the beginning, for several reasons. First, there is uncertainty about the extent to which the provision of LPG actually creates and genuine economic disparity. Most of our study communities do not purchase fuel; rather, they collect it themselves, in many cases from abundant and readily accessible supplies of wood near their homes. Accordingly, there is not a clear monetary difference even if there are time-savings. Second, there is uncertainty about the actual effect of such time-savings on our study outcomes, since it is not clear how this additional time would be spent or how this might translate into health effects.

IRC-specific incentives are as follows:

<u>Guatemala</u>
Guatemala plans to provide a catalogue of items of varying value from which control households can choose. The principle to be applied is that the amount of time the household would have to wait to receive an item would increase in proportion to its value. For relatively high priced items such as a refrigerator, dining table and chairs, or clothes wardrobe, the household would have to wait until the end of the trial. For smaller items (e.g., clothes iron, set of dishes, blender) the household would have to wait less time (exactly how much time is still to be defined). Households would receive the catalogue upon assignment to the control arm of the study and would earn a certain number of points every month for full participation in the study. The household could decide whether to spend its points on smaller items throughout the trial or on a larger item at the end.

<u>India</u>
The SRU Ethics Committee specifically directed that at the end of the trial, the India IRC research team must provide control households with the exact equivalent of what it is providing to intervention households. Hence the India IRC will provide control households with an LPG connection and stove identical to that received by intervention households as well as an amount of LPG equal to that received by intervention households. The SRU ethics committee stipulated that the India team is not allowed to provide monetary compensation to either intervention or control households except to compensate for the equivalent of lost wages due to participation in trial activities. The India IRC research team will pay the LPG distributor the full amount allocated for the LPG connection, the LPG stove, and 18 months of refills for each control participant at the end of their respective 18 month follow-up period. The amount can only be used for LPG refills, but it is fully accessible to the participant. The distributors have agreed that they will stamp an LPG passbook every month for the controls as an innovative way to reaffirm their support to provide the LPG refills at the end of the participants' respective study period.

Peru

The Peru IRC plans to provide control households with the following items at the indicated times:

Item	Timing
New hat for pregnant woman	24-28 weeks gestation
New hat for older woman	
Diaper bag with baby clothes, baby blanket	At birth visit
New sweater for mother	6 month post-birth visit
New sweater for adult woman	
Child hat, toy	
LPG stove + 1 tank of LPG	End of study

Rwanda

The Rwanda IRC plans to provide control households with the following items at the indicated times:

Item	Timing
Kitenge cloth for pregnant woman	24-28 weeks gestation
Kitenge cloth for older woman	
Outfit of clothing for newborn	Birth visit
Large towel for carrying the child	Birth visit
Baby blanket	Birth visit
Outfit for baby	3 month post-birth visit
Outfit for baby	6 month post-birth visit
Baby sheet to cover baby when carried on back	6 month post-birth visit
Utensils	6 month post-birth visit
Baby shoes	9 month post-birth visit
Additional utensils	9 month post-birth visit
Outfit for baby	9 month post-birth visit
Toothbrush	9 month post-birth visit
Toothpaste	9 month post-birth visit
Choice of 1 of the following items: • LPG stove, one cylinder, and 3 LPG refills (total: 60 kg LPG) • A solar home system • Cash deposit	End of study

3.3. Duration of Treatment Period

The intervention and follow-up period for included households will be approximately 18 months (remainder of pregnancy, plus 1 year for the child).

4. Follow-up Assessments

4.1. Trial Period

For the trial, each participating household will be followed from enrollment until the index child reaches (or would have reached, assuming a live birth and continued vitality) his/her first birthday, a total of about 18 months. During this period, we will be collecting data (at different frequencies[1]) for the various primary and secondary outcomes, including health, exposure to HAP, biomarkers, and intervention compliance and acceptability. These are summarized in **Table 1**.

Table 1. Schedule of exposure and outcome assessments during trial period.
PW=pregnant woman; C=child; O=older adult woman in household; NM=new mother

Child age (study time point)	**<20 weeks gestation** (baseline)			**24-28 weeks gestation** (1-3 months post-randomization)			**32-36 weeks gestation/birth** (3-5 months post-randomization)			**<3 months old** (~9 months post-randomization)			**~6 months old** (~12 months post-randomization)			**~9 months old[2]** (~15 months post-randomization)			**~12 months old** (~18 months post-randomization)		
	PW	**C**	**O**	**PW**	**C**	**O**	**PW**	**C**	**O**	**NM**	**C**	**O**	**NM**	**C**	**O**	**NM**	**C**	**O**	**NM**	**C**	**O**
Personal exposure																					
24-h $PM_{2.5}$, BC, CO	X		X	X		X	X		X	X	X	X	X	X	X				X	X	X
Urinary biomarkers	X		X	X		X	X		X		X	X		X	X					X	X
Primary outcomes																					
Birth weight								X birth													
Severe Pneumonia*																					
Stunting																				X	
Systolic blood pressure	X		X	X		X	X		X	X		X	X		X	X			X		X
Secondary outcomes																					
Maternal hypertension	X			X			X														
Diastolic blood pressure			X			X			X			X			X			X			X
Mean arterial pressure			X			X			X			X			X			X			X
Pulse pressure			X			X			X			X			X			X			X
Fetal growth		X			X			X													
Gestational age at birth								X birth													
Child linear growth (continuous)								X birth			X			X			X			X	
Preterm birth								X birth													
Child development											X			X			X			X	
WHO non-severe pneumonia*											X			X			X			X	
WHO severe pneumonia*											X			X			X			X	
Hospitalization for respiratory illness*											X			X			X			X	
WHO Pocketbook non-severe pneumonia*											X			X			X			X	
WHO Pocketbook severe pneumonia*											X			X			X			X	
Hypoxemic pneumonia*											X			X			X			X	
Lung ultrasound or chest radiograph pneumonia*											X			X			X			X	
Hypoxemic pneumonia											X			X			X			X	
BART**/CIMT			X																		X
SGRQ			X																		X
SF-36			X																		X

[1] Visit windows are defined as +/-14 days from the target visit date for anthropometry measurements, and +/-28 days from the target visit date for exposure, biomarker and blood pressure measurements. The visit window for all procedures for the last visit (18 months post-randomization) is +/-28 days. Visits for biomarker and blood pressure must nevertheless coincide with exposure assessment.
[2] Due to COVID-19, some additional measurements/samples may be taken at the 9 month visit, since previously scheduled visits may have been skipped due to lockdown.

Chronic disease biomarkers			X			X			X			X			X						X
Metabolomics/ microRNA***														X	X						
Covariates																					
Socio-demographics	X		X																		X
Clinical history	X		X	X		X	X		X	X	X	X		X	X		X	X		X	X
Weight/height/BMI	X		X	X		X	X		X			X			X						X
Diet/food security	X		X																X		X
IYCF											X			X			X			X	
Hemoglobin (anemia)	X			X			X				X									X	
Biospecimens collected																					
Urine	X		X	X		X	X		X		X	X		X	X					X	X
Dried blood spots	X		X	X		X	X	X birth	X		X	X		X	X					X	X

Abbreviations: PW=pregnant woman; C=child; O=older adult woman in household; NM=new mother; BART, brachial artery reactivity testing; BC, black carbon; BMI, body mass index; CIMT, carotid intima-media thickness; CO, carbon monoxide; IYCF, infant and young child feeding practices; LAZ, length-for-age z-score; $PM_{2.5}$, fine particulate matter; SF-36, Short Form 36 survey; SGRQ, St. George Respiratory Questionnaire. *Recorded whenever children present to HAPIN health facilities with respiratory symptoms.**BART in Peru only. ***Metabolomics/microRNA biomarker discovery in 100/site for the older adult woman and child.

4.2. Extended Follow Up

After completion of the initial trial described above, consenting participants in Guatemala, India and Rwanda will be enrolled in an extended follow up cohort study in order to assess longer-term effects of the intervention and of associations between household air pollution and health. The primary outcomes for the extended follow up period are child linear growth and neurodevelopment. The secondary outcomes are specified in Table 2 together with the schedule for visits. The detailed procedures for the annual follow up visits are described in Section 5.4 below.

Table 2. Schedule of exposure and outcome assessments during extended follow up.

Study Time Point	**24-MONTH VISIT**									**36-MONTH VISIT**									**48- and 60-MONTH VISITS**					
	Guatemala			**Rwanda**			**India**			**Guatemala**			**Rwanda**			**India**			**Guatemala****		**Rwanda**		**India**	
	M	**C**	**O**	**M**	**C**	**O**	**M**	**C**	**O**	**M**	**C**	**O**	**M**	**C**	**O**	**M**	**C**	**O**	**M**	**C**	**M**	**C**	**M**	**C**
Personal exposure																								
24-h $PM_{2.5}$, CO and BC	X	X	X		X		X	X	X		X			X			X			X		X		X
Urinary biomarkers		X		X	X			X																
DBS biomarkers		X		X	X			X																
Health Outcomes																								
Neurodevelopment*		X			X			X			X			X			X			X		X		X
Growth (height and weight)	X	X	X	X	X	X		X	X	X	X			X			X			X		X		X
BMI	X	X	X	X	X	X		X	X	X	X			X			X	X		X		X		X
Blood pressure	X		X		X	X			X	X									X	X***		X***		X***
CIMT			X																					
Lung function											X									X		X***		X***
Covariates																								
Recent acute respiratory infection																				X		X		X
Diarrheal infection																				X		X		X
Care-seeking																				X monthly		X		X
Vaccine status											X			X			X			X		X		X
IYCF											X			X			X			X		X		X
Diet/food security	X			X			X			X			X			X			X		X		X	
MPSS										X			X			X			X		X		X	
FCI																			X		X		X	
Anxiety and depression																			X		X		X	
Mental health																			X		X		X	
Sociodemographics										X			X			X			X		X		X	

Abbreviations: M=mother; C=child; O=older adult woman in household; BC, black carbon; BMI, body mass index; CIMT, carotid intima-media thickness; CO, carbon monoxide; FCI, Family Care Indicators; IYCF, infant and young child feeding practices; MPSS, Maternal Perceived Stress Scale; $PM_{2.5}$, fine particulate matter. *CREDI at 24 months; MDAT at 36, 48 and 60 months. **All measures repeated at an additional 54-month visit in Guatemala. ***60-month visit only.

5. Health Outcomes

5.1. Data collection during pregnancy period

5.1.1. Pregnant women

Following the same procedures as baseline (see **Section 2.3.**), women will have two further prenatal exams at 24-28 and 32-36 weeks gestation. We will measure her weight at each of these visits to track gestational weight gain. Any pregnancy-related problems will be referred to the nearest health center or hospital.

Fetal growth via ultrasound
Women will have two additional ultrasound measurements to assess fetal growth at 24-28 and 32-36 weeks of gestation, using standard fetal imaging parameters. These will include confirmation of fetal viability via fetal heart motion (FHM) documented by M-mode imaging, and ascertainment of placental position and fetal presentation (vertex/breech/transverse/variable).

At the time of the second exam at 24-28 weeks, we will estimate fetal weight (EFW).

At the third trimester exam at 32-36 weeks, standard fetal biometry and EFW will again be obtained. Abnormal fetal growth will be defined as EFW less than the 10th percentile for gestational age based on the recently completed international standard for fetal growth (INTERGROWTH-21st) produced by the WHO (Papageorghiou,et al 2104a).

In addition, we will document fetal amniotic fluid volume in the second and third exams both as the amniotic fluid index (sum of the deepest vertical pockets of amniotic fluid in each quadrant of the abdomen) and as the single deepest vertical pocket (DVP). Per current obstetric recommendations, the DVP will be used as the primary measure of amniotic fluid volume (Reddy et al. 2014). Since clinical practices regarding antenatal surveillance of fetal wellbeing may differ by country, we will develop indications for assessment of fetal wellbeing specific to each study site. When indicated, we will obtain an ultrasound based Biophysical Profile, a score of 0 to 8 representing 2 points each for the presence of standardized fetal movements, tone, breathing and amniotic fluid volume (ACOG 2014).

Urine and blood samples
Two additional samples of urine and DBS (via finger prick) will be obtained from the pregnant mother at 24-28 and 32-36 weeks.

5.1.2. Older adult women

The older adult women will be asked to fast overnight prior to the scheduled health assessment when the BART test is being administered.

Blood pressure
Following the same procedures as baseline (see **Section 2.4.**), blood pressure measurements will be taken at the same two visits during the pregnancy period on the older adult woman.

Urine and blood samples
Additional samples of urine and DBS (via finger prick) will be obtained during these visits as well.

5.2. Data collection during child's first year of life

5.2.1. Child

Birth visits

We will ask family members of the pregnant woman to contact a designated HAPIN staff member when the woman goes into delivery. Given that our Rapid Assessment Survey conducted as part of Phase I (See Appendix) indicated a majority of women at the IRCs give birth in health centers, and that we will already be in close contact with these facilities as part of recruitment and enrollment, we will also ask health center staff to notify HAPIN staff when a HAPIN participant goes into a health center for delivery. At this visit, staff will assess the neonate and mother. Medical records may be used to determine delivery details (data and time of birth, sex, etc). We will refer all participants with health-related problems to the health center or hospital using established protocols. We will also examine gestational age at birth and preterm birth as secondary outcomes.

Birth weight/Anthropometry

Birth weight will be assessed by a trained nurse or health worker at the <24-hour visit during the changing of newborn into pre-weighed clothing to keep the infant at least partially covered to prevent cold stress. At the same time, anthropometric measures will be taken. Recumbent length will be measured to the nearest 0.1 cm using a measuring board. If the first and second length measurements differ by >0.7 cm, a third measurement should be taken. Birth weight will be measured to the nearest 10 g using a digital electronic scale. Older infants will be measured in their mother's arms using the TARE function, to the nearest 100 g. If the first and second measurements differ by >100 g (for birth weight) or >0.2 kg (for older infants), a third measurement should be taken. Head circumference will be measured to the nearest 0.1 cm. If the first and second measurements differ by >0.5 cm, a third measurement should be taken. Calibration of scales and infantometers, using standard weight and height measures, will be done weekly in the office. The mother will receive written documentation of the newborn's weight on a prepared health feedback form. In cases where birth weight cannot be measured by a trained nurse or health worker within 24 hours of birth, the measurement from the hospital/clinic medical record will be used.

Infant weight and length will be assessed quarterly (3, 6, 9 and 12 months of age) as described above using the same procedures. The primary outcome measured is linear growth failure, or stunting, evaluated at 1 year. Stunting will be assessed using height-for-age z-scores (HAZ) based on the 2006 WHO Child Growth Standards (WHO 2006a). Stunting is defined as HAZ <-2 SD below the median z-score. We will also examine linear growth through age 1 as a secondary outcome.

Child Development Milestones

We will assess child development at the same intervals as the anthropometry visits (3, 6, 9, and 12 months of age). We will use the Caregiver Reported Early Childhood Development Instrument (CREDI) to assess population-level early child development outcomes. The CREDI is a population-level measure of early childhood development (ECD) for children from 0-3 years of age. It measures gross and fine motor, expressive and receptive language, cognition and social-emotional development. We will use the long form which provides domain specific questions and is recommended for research studies. The advantages that the CREDI has over the Developmental Milestones Checklist, which was previously suggested, is that the CREDI includes measures of cognition, does not require adaptation, has been designed to be "culturally and linguistically neutral", has been implemented in over 17 countries (including Guatemala and India) and administered to over 17,000 children. The validation of the short form, which is recommended for use in population surveys, was recently published in 2017. The CREDI is open-source and the website has materials as well as code for scoring items. We have been in contact with Dr. Dana Charles McCoy, who developed the CREDI, who has provided additional materials. She shared the draft manuscript of the long form validation which is currently under review. (https://sites.sph.harvard.edu/credi/)

The CREDI assesses five domains of child development: 1) motor development (fine and gross motor), 2) language development (expressive and receptive language), 3) cognitive development (executive function, problem solving and reasoning, and pre-academic knowledge), 4) socio-emotional development (emotional and behavioral self-regulation, emotional knowledge, and social competence), and 5) mental health (internalizing and externalizing behaviors). The CREDI can be used as a population-level measure of early child development for children from birth to three years of age. The content of the CREDI is also culturally neutral and allows for cross-context comparisons while still being psychometrically valid and reliable. The CREDI has been testing on 8022 children living in 17 low and middle-income countries and found to have high reliability and construct validity (McCoy, Sudfeld, Bellinger 2017). We will be using the "long-form" version of the CREDI.

The assessment will be conducted through a parent interview where age-appropriate and developmentally-appropriate questions about the child's milestones and behaviors will be asked of their caregivers.

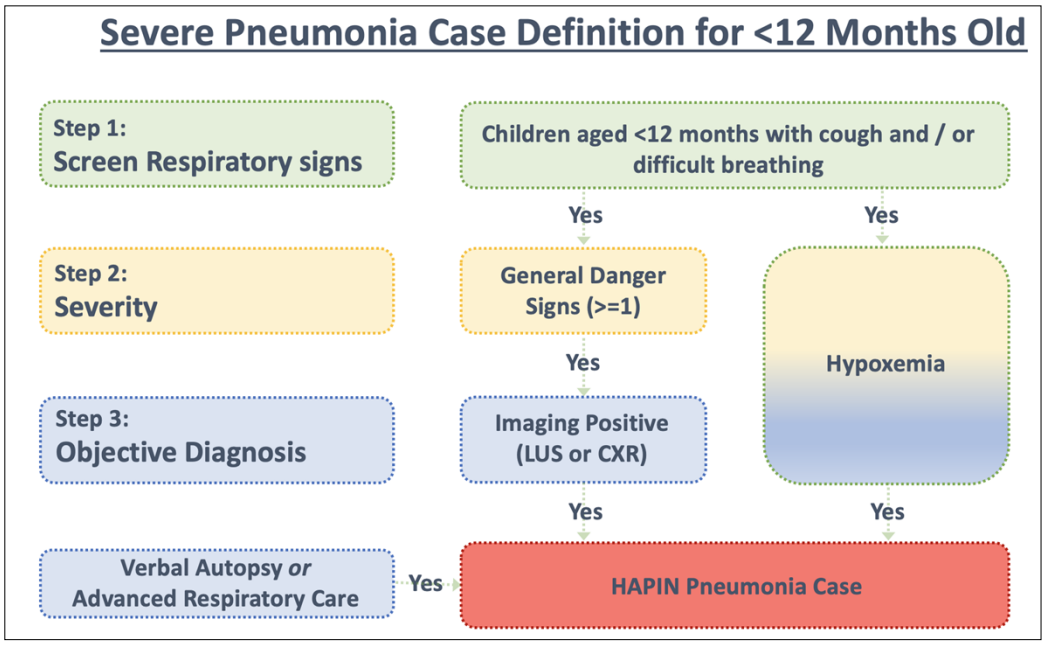
Because child stimulation may affect child development, we will also administer the Family Care Indicators Interview (UNICEF), a 9-item measure that ascertains child stimulation. The Interview asks if an adult engaged with the child in the past 3 days: read or looked at books, told stories, sang songs, took child outside of home, played with child, and spent time counting, drawing or naming things. All assessments will be done through a 30 minute parent interview.

Infant pneumonia

HAPIN pneumonia Case Definition (see Simkovich SM, et al., 2020)
We will define HAPIN pneumonia as a case of severe pneumonia adapted from the revised WHO classification of childhood pneumonia (World Health Organization, 2014), which includes three algorithms for the case definition: (1) the presence of cough and/or difficult breathing and at least one general danger sign and primary endpoint pneumonia on a lung ultrasound or chest x-ray, or (2) the presence of cough and/or difficult breathing and hypoxemia (measured via pulse oximetry, SpO_2), or (3) children who died but their death is attributed to pneumonia by verbal autopsy. General danger signs for all participants include any of the 5 signs of: (1) unable to drink or breastfeed, (2) vomiting everything, (3) convulsions, (4) lethargy or unconscious, (5) stridor at rest. There are 4 additional general danger signs for participants <2 months of age (neonatal danger signs): (1) unable to feed, (2) grunting, (3) not moving at all or moves with stimulation only, (4) severe chest indrawing. Severe acute malnutrition is also considered a danger sign for all participants, and is defined as a calculated weight-for-length z score < -3 or weight-for-age z score < -3 (if length unavailable or age <60 days).

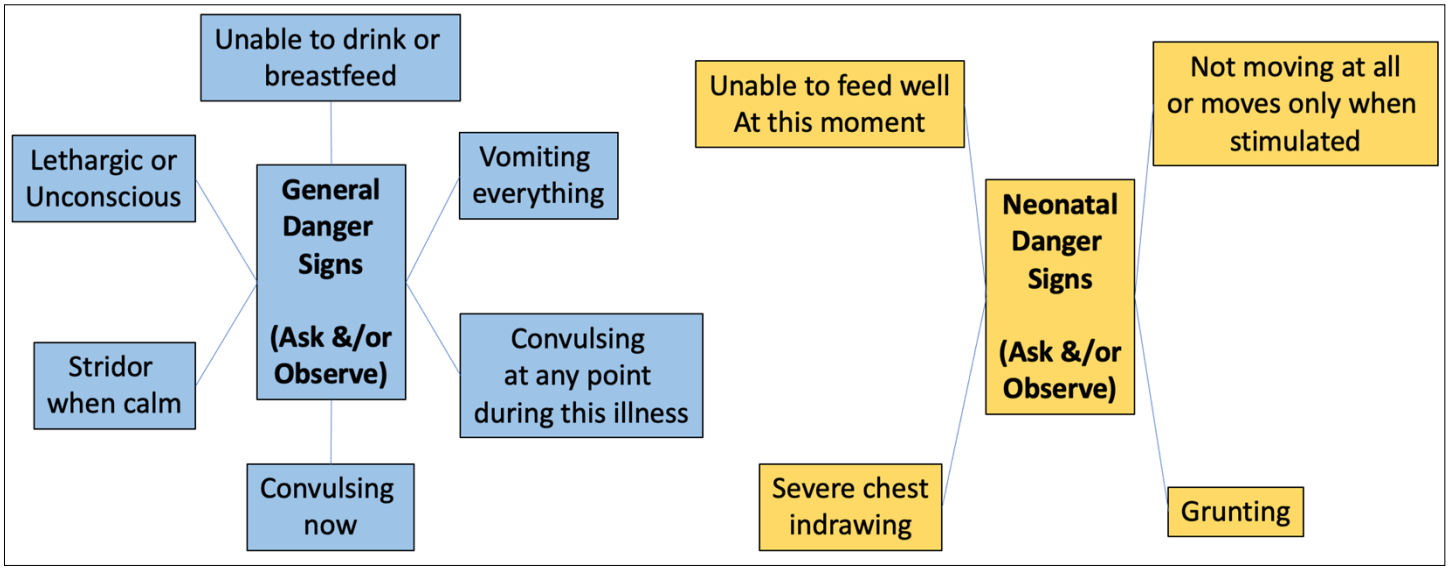
Lung ultrasound will be the imaging modality obtained unless there are logistical or regulatory reasons this cannot happen and then we will obtain a chest x-ray. We will define tachypnea based on current World Health Organization thresholds (≥60 breaths/minute for 0-2 month olds and ≥50 breaths/minute for 3-11 month olds). We will determine hypoxemia based on the physiologic threshold of ≤92% for altitudes <2,500 meters above sea level and ≤86% for altitudes ≥2,500 meters above sea level. Children are also considered hypoxemic, regardless of oxygen saturation, and assumed to have cough and/or difficult breathing if they are receiving advanced respiratory support at the time of evaluation, which includes any of the following: intubation and mechanical ventilation, non-invasive ventilation with continuous positive airway pressure support (CPAP), non-invasive ventilation with bi-level positive airway pressure support (BIPAP), or high-flow nasal cannula oxygen. Children who die but are determined to have had pneumonia or symptoms and/or signs of a respiratory illness consistent with pneumonia by verbal autopsy will also be considered a case of severe pneumonia. Children are permitted to meet the above case definition for severe pneumonia multiple times if the repeat event occurs either >14 days after the hospital discharge date or, if the hospital discharge data is not available or the child was not hospitalized, >30 days after the date the child met criteria for severe pneumonia. Implausible values and outliers identified from the variables that comprise the case definition will be considered missing.

Figure 2. Severe pneumonia case definition



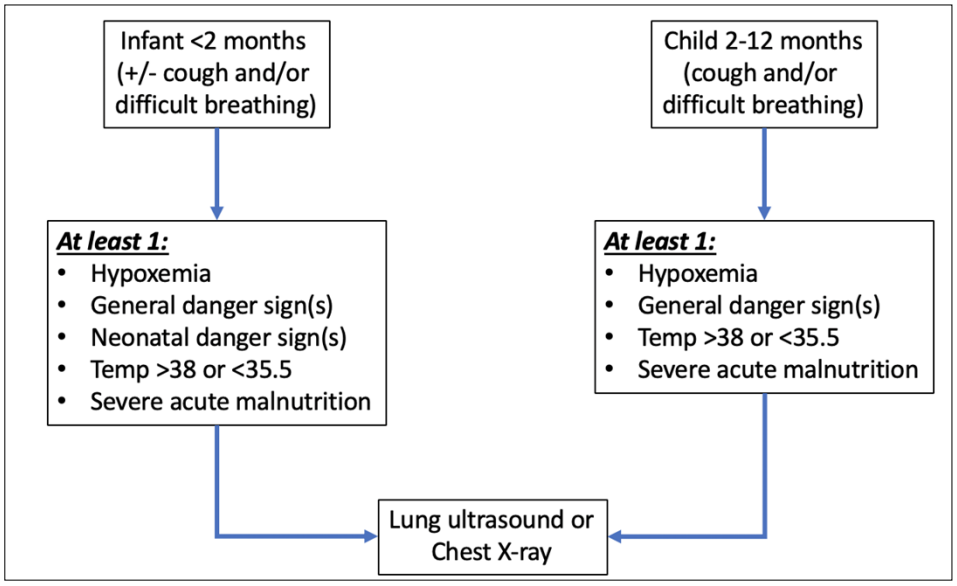
If a HAPIN child presents to a healthcare facility, across all sites, a HAPIN team member (either a HAPIN study physician or field team member, or a HAPIN-trained physician or healthcare worker located at a participating health facility) will be notified by the healthcare facility staff. Each IRC has designed an identification system for mothers to bring to the facility to identify these children. If a child is 2-12 months, field staff can observe or mother can report a cough and/or history of difficulty breathing (developed <14 days ago). If the child has observed and/or reported cough and/or difficulty breathing and/or the child is a neonate, then the child will be assessed for WHO general danger signs (and if neonate then neonatal danger signs). A field staff member will ask and observe for each danger sign (Figure 3).

Figure 3. Pneumonia danger signs



All children will be assessed for vitals (temperature, respiratory rate, pulse oximetry) and lower chest wall indrawing. Respiratory rate will be measured by study staff directly observing the respiratory pattern and manually counting the number of breaths over two 60 second periods. The child will be considered to have fast breathing for age (tachypnea) if the average respiratory rate per minute is ≥60 breaths/minute for infants <2 months of age, ≥50 breaths/minute for infants 2 to <12 months of age. If only one 60 second respiratory rate count is available we will include that measurement. Anthropometry will be performed per HAPIN protocol on all children. If a child has any general danger signs (or for neonates any general danger sign or neonatal danger sign) and/or temperature >38º Celsius or < 35.5º Celsius and/or acute severe malnutrition, and/or is hypoxemic (as defined above), then a lung ultrasound or chest x-ray will be obtained (Figure 4).

Figure 4. Imaging algorithm



Lung Ultrasound

We plan to use Sonosite EDGE ultrasounds (Sonosite, Bothell, WA). For the purposes of image acquisition, we will divide the right and left chest into three zones; anterior, lateral, and posterior. Each zone can be subdivided into superior and inferior. The examiner will record 6-second video clips while fanning across and down each zone in both sagittal and transverse probe orientations, respectively. Sagittal and transverse recordings of each of the 12 zones will result in a total of 24 videos recorded. In previous work, we have established that this exam takes less than 10 minutes to perform (Chavez et al 2014, Ellington et al 2016). The examiner will have the ability to make on-site diagnosis and record the full evaluation for further over read and quality assurance. At the end of each day, all LUS videos will be uploaded to TRICE (TRICE Imaging, Del Mar, CA), a cloud-based system that securely manages and archives medical images for remote over-read and quality assurance. We will randomly select two readers out of a panel of four LUS readers, consisting of expert sonographers, for determination of consolidation. In the case of disagreements between the primary and secondary reader, all four readers will be asked to provide an independent evaluation first followed by an unblinded consensus assessment. We will define Primary End-point Pneumonia (PEP) as:

a. Consolidation that is either >1 intercostal space in size or >1 cm in width or height with or without the presence of air bronchograms (hyperechoic areas) or shred sign.
b. Pleural effusion of any size that has any of the following spatially located findings:

- Consolidation of any size
- Pleural abnormalities
- Interstitial infiltrate (≥ 3 B-Lines)

In addition, an expert sonographer will evaluate 10% of all LUS videos, and if any reader shows <80% agreement with the expert radiologist then additional training and standardization will be necessary.

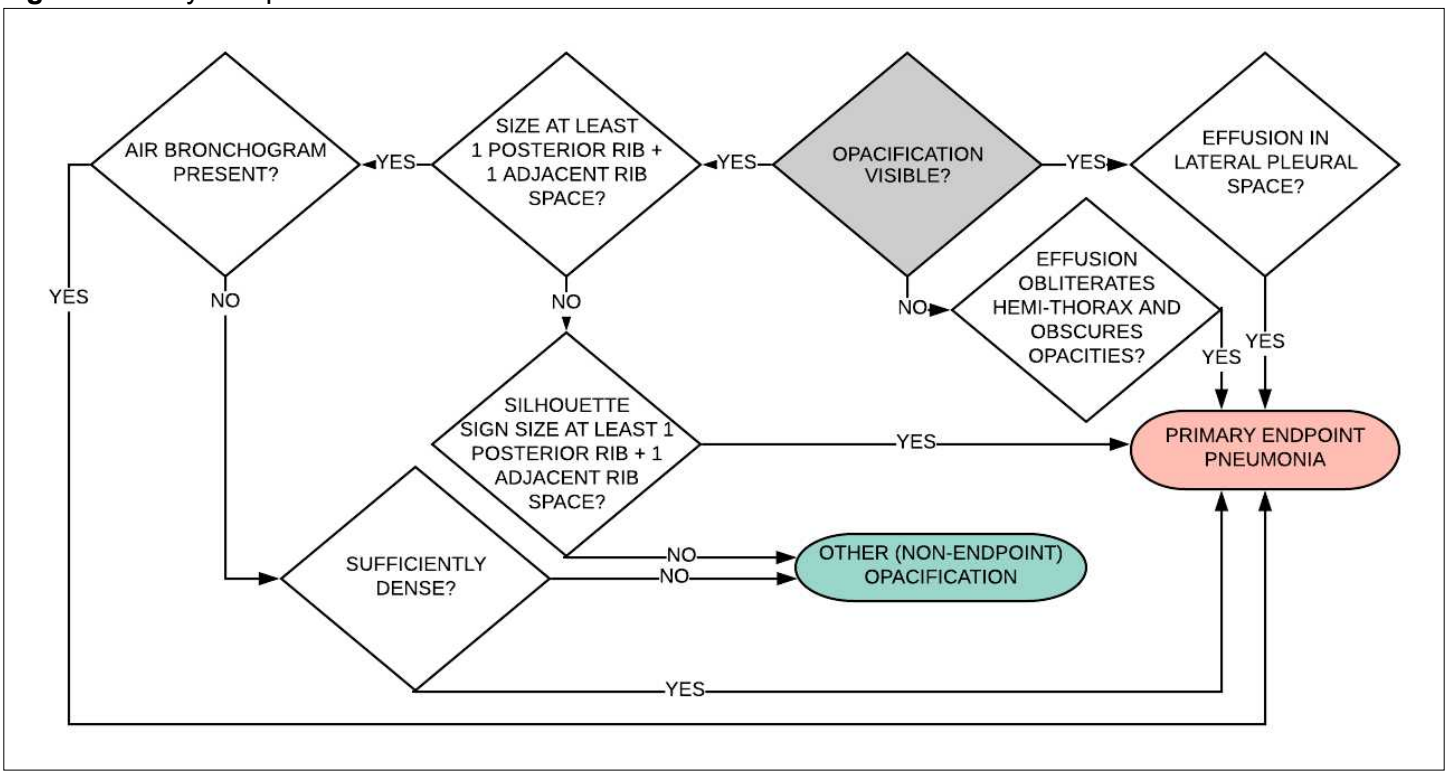
Chest X-ray

In situations where a lung ultrasound cannot be performed (i.e. ultrasound not available, sonographer not available) a copy of the CXR taken as a diagnostic tool during the clinical encounter will be obtained. Interpretation follows the HAPIN X-ray interpretation protocol. The protocol defines pneumonia as primary endpoint pneumonia.

If an image is obtained digitally, it will either be uploaded in TRICE (TRICE Imaging, Del Mar, CA), REDCap, or printed onto film. If an image is not able to be obtained digitally and the film is obtained, then it will need to be scanned in by study staff in a 4,800 dpi scanner or digitized by a camera with at least 20 megapixels and uploaded to TRICE or REDCap. Each IRC will randomize each image to 2 readers who have completed and become certified in HAPIN x-ray interpretation training. Each reader will be notified via email that the image is available for interpretation and will directly enter the interpretation on the HAPIN X-ray Assessment Form in TRICE or REDCap. If there is disagreement between the interpretation of an x-ray, then two additional readers (adjudicators) will be assigned to interpret the image. If discordancy in interpretation persists then a 5th reader will finalize the image interpretation.

Each image will be read in a standardized approach based on guidelines by the WHO X-ray interpretation (WHO 2005) and updated based on the WHO CRES (Mahomed 2017). Each radiologist who will be interpreting images will be asked to complete a certification course. This course was developed by HAPIN clinicians with the guidance of experts in WHO chest x-ray interpretation. During the course, a reader is asked to complete a pre-test module of 60 x-rays, then attend a series of lectures to educate on quality control of imaging, definitions of primary end-point pneumonia and other non-endpoint opacifications, (Figure 5). To become certified each reader must correctly interpret 80% of 60 images from the CRES (WHO CRES group) image library as determined by consensus readers. Every 4-6 months, readers will be provided via email a summary of how to interpret x-rays and be asked to interpret 20 x-rays. To stay certified, each reader must obtain a score of 80% correct. Remediation will be conducted as necessary.

Figure 5. X-ray interpretation



Pneumonia Case Ascertainment for Each IRC

India

In our study sites, all severe pneumonia cases end up in secondary and tertiary care hospitals (both public and private). Therefore, they are our case ascertainment sources for HAPIN. However, our dialogues with various stakeholders, especially the pediatricians, showed us that the diagnosis of severe pneumonia is not standardized across the hospitals. We will, therefore, emphasize the importance of diagnosing severe pneumonia using uniform guidelines (WHO and HAPIN criteria) and train the medical doctors on this by continuing medical education programs. Clinician health managers will be liaising with the medical doctors for this aspect and they will also co-ordinate with public and private hospital administrators for the seamless data flow. Trained data abstractors will be sent to abstract data on severe pneumonia from medical records of both secondary and tertiary care public and private hospitals on a periodic basis.

Guatemala

Per MoH protocol, severe pneumonia cases seen in health post/centers should always be referred by clinic staff to the main public hospital in Jalapa. A small fraction of these cases may choose to go to a private clinic in Jalapa but given the level of poverty, we do not anticipate many will choose to pay for services when services are free at the public hospital. We will ask families to contact us when a child is seen in a health post/center for a respiratory illness and follow-up with their personnel if the child is referred for hospital admission. Our clinical field staff (doctors or nurses) will examine the child within 4-6 hours of being seen at the health post/center for a respiratory illness. The child will be examined at the health post/center, the public/private hospital or in the home (if the family refuses to take the child to the hospital). We will make home visits to ask about any respiratory illnesses or clinic visits in the past month at one month of age, then quarterly.

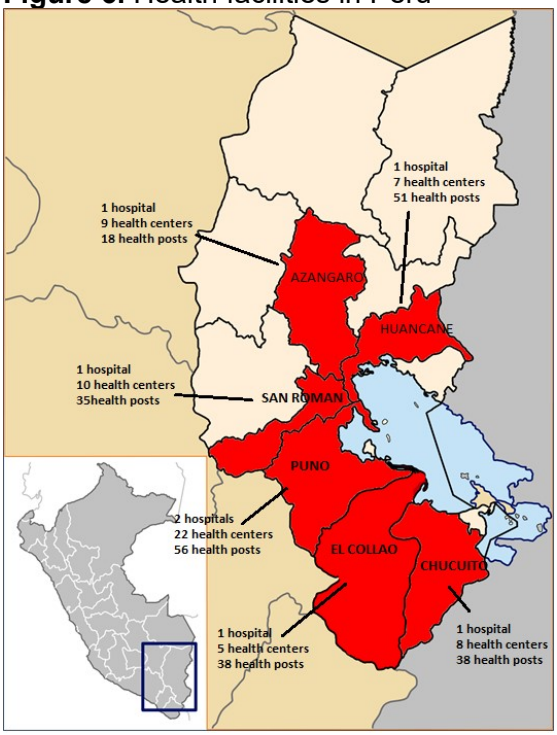
The Guatemalan treatment manual (treatment norms for first and second levels, published by MoH) identifies severe pneumonia as: subcostal retractions and one or more of the following: tachypnea (1-2 months, >

60/minute; 2 mo-1 yr, > 50/minute), nasal flaring, axillary temperature > 38 C, positive auscultation signs (wheezes and crackles) and “quejido” or fussiness. Pneumonia is tachypnea and one or more of the following: cough, fever, and positive auscultation signs (wheezes and crackles). These definitions are obviously different than our HAPIN definition. The diagnosis of pneumonia is not standardized, and nurses with minimal training who staff the health centers/posts will not be able to differentiate between the two. We will provide in-services to MoH providers about our study and our case ascertainment, but we will use our field staff and the HAPIN definition to classify severe pneumonia for children with respiratory illnesses in our study. If we need to abstract data from medical records, clinical field staff will do so, but our experience is that medical records are poorly documented, leading to incorrect information or missing data.

Peru

In Peru, we have identified all potential health facilities where children could be treated for pneumonia, and have learned during our formative research that there are 7 hospitals where children with severe pneumonia will be referred for treatment and management (see Figure 6 for a map of the catchment study area). In light of the change of strategy for the ascertainment of pneumonia, we propose to develop and train a specialized team of six health care workers and 1 supervising physician to be responsible for the screening of pneumonia at these seven health care facilities. In coordination with MINSA and ESSALUD, through our general agreements with the directors at each of the hospitals, our pneumonia team will be permanently stationed at the hospitals during the study period and will be available to assess study children immediately according to our HAPIN protocol for the presence of WHO severe pneumonia, and conduct pulse oximetry and lung ultrasound assessment. Study team members will be in close communication with attending physicians at the hospitals, who will help to screen for children who are study participants in HAPIN. We will ask attending physicians to call the on-call study member if there are any after-hours cases. We will work on a shift schedule in pairs at the hospitals (2 of the hospitals can be covered by 1 pair of team members from our pneumonia team given the distance), and rotating schedule for after-hours cases (every seven day call schedule for our pneumonia team). A smaller team of 7 people allows us to adequately undergo standardization and re-standardization, to ensure the highest possible quality of data. If we need to abstract data from medical records, trained field staff will do so.

Figure 6. Health facilities in Peru



Rwanda

Per MoH protocol, severe pneumonia cases seen by community health workers (CHW) at the village-level or in health posts/centers should always be referred by clinic staff to the public district hospital, Rwinkwavu Hospital. All severe pneumonia cases in our study area should be referred to this hospital. Currently, we are assessing the percentage of cases that might present to health centers and/or public hospital in neighboring districts, but we anticipate this to be an extremely small percentage. There are three health centers and six health posts in our study area.

Pneumonia and severe pneumonia are diagnosed using the WHO IMCI algorithm. Pneumonia is treated with Amoxicillin at the village and/or health centre level. Nurses at health centres can provide ill children with severe pneumonia with pre-transfer medicines such as Gentamicin and Ampicillin. While all ministry of health nurses are trained in WHO IMCI algorithm, the training might not be standardised across health centres. We will conduct presentations prior to the start of the study with all health centres and health post centres as well as the head of CHW in the area, to introduce the study, our case definition and provide a refresher training on the WHO IMCI pneumonia algorithm.

We will have HAPIN nurses based at the hospital and all three health centers during working hours and they will be able to assess children arriving to the health facility with signs of respiratory illness. Our staff will be able to access the diagnosis made by the health facility and will conduct their own assessment after the facility staff have made their assessment, to determine if the child has WHO severe pneumonia and/or HAPIN severe pneumonia. We will be calling health facilities on a daily basis to determine if any children were assessed with signs of respiratory illness during after-hours and we will attempt to observe these cases early in the morning. If we need to abstract data from medical records, study staff will do so, but our experience is that medical records are often poorly documented, leading to incorrect information or missing data. We will have initial meetings with heads of health centers to reinforce the importance of record keeping and will be conducting visits at ages 1 month, 3 months, 6 months, 9 months and 12 months to reinforce this message.

Pneumonia Surveillance

Surveillance for WHO severe pneumonia cases will occur at all health facilities in the HAPIN study area where children with severe pneumonia are commonly treated or hospitalized**.** These facilities will be identified in the catchment areas of the HAPIN study by each IRC team through interviews with local physicians at government and private health facilities, and they will be georeferenced for mapping purposes. One or more HAPIN team members (either a HAPIN study physician or field team member, or a HAPIN-trained physician or healthcare worker located at a participating health facility) will then be identified at each of these facilities. These HAPIN team members will be invited to participate in the study, and if they agree then they will be provided with standardized training on how to perform a history of respiratory symptoms (i.e., how to ask about self-reported history of cough and difficulty breathing in the last 7 days), how to evaluate respiratory signs (how to assess respiratory rate and how to evaluate for chest indrawing) and danger signs (how to observe or ask about danger signs), and how to measure SpO_2 using the Masimo Rad-G device. In addition, HAPIN team members will perform LUS whenever possible or request a CXR whenever possible. HAPIN team members will need to pass the standardization exercise before they are able to perform evaluations for HAPIN pneumonia cases, followed by re-standardization tests that will take place every 6 months. Children who present with pneumonia will be evaluated by HAPIN team members as soon as possible after initial presentation.

Child Health Status Visits

At five time-points (post-birth 1 month, 3 months, 6 months, 9 months and 12 months), child health status visits by our field staff will also evaluate for self-reported health care visits or hospitalizations since the last visit due to respiratory disease, followed by a rapid assessment of respiratory rate and SpO_2. The primary objective of these visits is to identify children who may have had a case of pneumonia since the previous visit that we did not capture during our surveillance in health facilities. This will serve as a stop-gap in our surveillance system. In the rare circumstance that our field staff identify a child who meets criteria for WHO pneumonia during these

visits, this child will be referred to the nearest health facility for evaluation, and potentially included as a HAPIN case after evaluation. The other objective is to capture the respiratory rate and SpO_2 of healthy children to determine the reference range for each IRC.

Pneumonia Education Materials
An educational poster has been designed to illustrate six danger signs for WHO/HAPIN pneumonia. There are four versions of this poster as the illustrator has made each illustration culturally specific to each IRC (i.e. Indian, African and Latin illustrations). These posters are designed to be hung in homes to not only serve to educate women on the danger signs of pneumonia, but to serve as a constant reminder to be aware of these danger signs in their children.

Pneumonia Deaths
Pneumonia deaths (and all deaths in the participants) will be recorded; if information is available to confirm that the case would fit the HAPIN case definition it will be included in our primary case definition. For infants who die at home, we will also conduct verbal autopsies to help understand the cause of death.

Urine and blood
DBS will be collected from the child at birth, and approximately 3, 6, and 12 months of age, and urine will be collected at approximately 3, 6, and 12 months. For urine, a sterile cup or a pediatric urine collection bag will be left with the mother. Pediatric urine collection bag instructions inform the mother to affix the bag over the child's urethral area on the morning of a scheduled visit. The urine collection bag should not remain on the child for more than 3 hours. The mother will be instructed how to remove and store the urine bag until the site visit. The urine sample will be collected upon each site visit and stored in a cooler until return to the local HAPIN satellite laboratory where it will be divided into multiple aliquots for storage. DBS will be collected as outlined below except the prick will be made in the heel for children <6 months in accordance with WHO guidelines.

5.2.2. Mother

Twenty-four hour personal sampling data for PM2.5 and CO will be collected from the mother at 3, 6, and 12 months after the child birth. In addition, blood pressure data will be obtained at the same time.

Starting in June 2020, we will administer a questionnaire (CRF) on impacts of COVID-19 in all HAPIN households with the enrolled pregnant woman/mother. The questionnaire focuses on behaviors related to prevention of COVID-19 transmission; economic impacts of the pandemic; and changes in access to food, fuel, health services, and cooking practices. The CRF will be administered either in person or via mobile phone, based on local recommendations in each IRC. CRF implementation is expected to take approximately 30 minutes. The CRF will be administered at least once in all IRCs; additional rounds of data collection using the same CRF may be deemed necessary depending on the progression of the COVID-19 pandemic in each IRC.

5.2.3. Older adult women

The older adult women will be asked to fast overnight prior to the scheduled health assessment when scheduled for a BART test.

Blood pressure
Following the same procedures as baseline (see **Section 2.4.**), blood pressure measurements will be taken at <6, and 12 months after the index child is born.

Urine and blood samples

Following the same procedures as baseline (see **Section 2.5.**), samples of urine and DBS (via finger prick) will be obtained at multiple intervals (see Table 1 and 2).

CIMT/BART (BART in Peru only) via ultrasound
Following the same procedures as baseline (see **Section 2.4.**), CIMT/BART will be obtained at ~12 months after the index child is born.

SGRQ
Following the same procedures as baseline (see **Section 2.4.**), the SGRQ will be obtained at ~12 months after the index child is born.

SF-36 Questionnaire
Following the same procedures as baseline (see **Section 2.4.**), the SF-36 questionnaire will be obtained at ~12 months after the index child is born.

The same dietary intake, physical activity, and clinical history questionnaires administered during baseline will be re-administered at ~12 months after the index child is born. The medication use, including hypertensives, component of clinical history will be collected at multiple intervals (see Table 1).

5.3 Data collection in cases of death

Verbal autopsies will be conducted on all stillbirths and participant deaths. HAPIN verbal autopsy forms – for stillbirth, neonatal, child, maternal, and older adult woman deaths - have been modified from the St. John's Research Institute Million Deaths Study (Jha et al, 2005). These forms have all been approved by the Emory IRB. All study staff conducting verbal autopsy interviews and collecting data have attended an online training by Dr. Prem K. Mony, Vice Dean, St. John's Research Institute, Bangalore, India. Ideally, verbal autopsies are conducted within 2-3 months after death, but some verbal autopsies may be conducted up to 24 months after death, which may be after study exit. This is usually done when a family moved away temporarily from the study area or we lost contact with the family during the period of the study. Due to COVID-19 lockdowns, starting in March 2020, some verbal autopsies are being conducted by phone.

5.4 Data collection during extended follow-up period

Following the trial, we will continue to follow the cohort of consenting HAPIN participants in India, Guatemala and Rwanda to assess the longer-term effects of the intervention and associations between health outcomes and exposure to household air pollution. The outcomes during the extended follow up period are child linear growth and neurodevelopment along with the additional outcomes in Table 2. In addition, Guatemala has an additional aim of determining the effects of the intervention on child respiratory health. As described below, we will continue to follow children and adult participants at least annually using methods employed during the HAPIN trial period. The visit window for the 24-month visit is +/- 90 days, with a preferred window of +/- 28 days. The visit window for 36-month, 48-month, 54-month (Guatemala only) and 60-month visits is -28 days/+90 days. Procedures for the extended follow up period are described below and summarized in Table 3.

5.4.1. Recruitment and informed consent

All households who complete the HAPIN trial will be approached to determine the mother's interest in participating in, and having her child participate in, this cohort study. At study exit in the HAPIN trial, participants (the mothers) are asked if they are willing to be contacted in the future for follow-up studies. For those who agree, contact information is collected for the participants themselves as well as for three close

relatives or friends expected to know their whereabouts. As of December 2020, 99% of participants agreed to provide their contact information.

When contact is made, study staff will give a brief description of the study to gauge interest. If the mother/father/guardian is interested in participating, study staff will arrange a home visit. At this visit, details of the study procedures and potential risks and benefits will be explained to the parent/guardian as well as the older adult woman (if one resides in the household). Participation in the study is voluntary. A participant's decision of whether or not to enroll her/himself, and their child, in this study does not affect their or their child's medical care at the community health clinics and/or hospitals. Written informed consent, in the native language, will be obtained at recruitment. Due to the child's age, this will be in the form of parental/guardian consent for the participation of the child. Consent forms will follow the format used in the original HAPIN trial, and will include a description of the study, the respondent burden, the potential risks and benefits, and names and contact information for the local PD/PI and other local responsible individuals who can be contacted for additional study details. For illiterate participants, the consent form will be read and the consent process done in the presence of a witness. All participants will receive a copy of the consent form for their personal records.

5.4.2. Child

24-month Visit
All sites: Anthropometry and child development will be measured according to the procedures outlined in **Section 5.2.1**.

Urine and DBS samples will be collected according to procedures outlined in **Section 9.4** (field teams will have the option to leave the urine cups with the mother on the day prior to collection or to collect a spot sample during the visit) and **9.5**. The child will also either be monitored for 24-hour personal exposure to HAP ($PM_{2.5,}$ CO and BC) indirectly, as detailed in **Section 7**, or, if the child is able to wear the air monitor themselves, they will be monitored directly. Neurodevelopment will be measured using the CREDI tool, according to procedures outlined in **Section 5.2.1**.

The child will either be monitored for 24-hour personal exposure to HAP ($PM_{2.5,}$ CO and BC) indirectly, as detailed in **Section 7**, or, if the child is able to wear the air monitor themselves, they will be monitored directly. In a random subsample of 10-30% of children, personal exposure measurements will be conducted every 6

36-month Visit
All sites: Child weight will be measured with a Seca 876 scale. It can be measured either by having the child stand (alone) on a scale, or by weighing the mother and child together (using the tare function). If the first and second weights differ by more than 0.1 kg, then a third measurement will be taken. Height will be measured with a Seca 213 stadiometer. If the 2 height measurements differ by more than 0.7 cm, then a third height measurement will be taken.

The child will either be monitored for 24-hour personal exposure to HAP ($PM_{2.5,}$ CO and BC) indirectly, as detailed in **Section 7**, or, if the child is able to wear the air monitor themselves, they will be monitored directly. In a random subsample of 10-30% of children, personal exposure measurements will be conducted every 6 months.

Child neurodevelopment will be measured using the **Malawi Developmental Assessment Tool (MDAT).** In the HAPIN trial, the CREDI tool was used, but CREDI only measures children < 36 months. The MDAT measures gross motor (39 items), fine motor (42 items), language/cognition (40 items) and social skills (36 items). Originally developed and validated in rural Malawi, it has now been used in over 25 countries with more than 8,000 children as both a clinical and research tool. The MDAT is a continuous test with start and stop rules. Most items are administered directly to the child; however, those that are not easily observed (e.g., child

speaks in full sentences; child understands sharing with others; child can dress self) are administered by parent report. Children receive either a pass or fail for each item, and summed pass scores can produce a composite score as well as domain-specific scores. The MDAT is an open access measure, and the authors provide international norms for scoring and comparison as well as a standardized training package and requirements for procuring the kit of objects used in administration. The majority of kit objects are common items in LMICs (e.g., cups, balls, plastic bottles, bottle-caps, soap, locally made toys) that can be accessed locally or made. The MDAT has demonstrated sensitivity in detecting change in developmental status in studies with HIV-infected children on ART, malnourished children, premature infants and children with cerebral malaria [PK1]. The MDAT has been validated and is currently in use in Guatemala (in consultation with Dr. Patricia Kariger) and Rwanda. The MDAT has also been used in South Asia in Nepal and Bangladesh in the Every Newborn – Simplified Measurement Integrating Longitudinal Neurodevelopment & Growth study (Day, 2019) and by the Childhood Acute Illness and Nutrition Network in Pakistan (CHAIN Network, 2019).

We will translate and culturally adapt the MDAT for each IRC following official recommendations for translation and adaptation (Wild, 2005) so that it will be ready to implement at the 36-month assessment. Before piloting and start of enrollment, an extensive MDAT training for field research staff will be conducted virtually for each IRC by an expert. During training, video recordings (at least one for each MDAT administrator) will be shared with the expert, who will evaluate administration of the tests and provide feedback to ensure administrators are fully trained to conduct the test.

Information may also be collected from medical charts. If the child is sick and is taken to a health facility, an evaluation of the child's respiratory health may also occur. *Guatemala only:* If there are signs of respiratory disease during the evaluation, an ultrasound may be taken to view the child's lungs.

Guatemala only: Lung function measurements will be made using the forced oscillation technique **(FOT)** with the Tremoflo C-100 device with disposable mouthpieces. FOT is a technique that can identify early changes in the airways. This technique is non-invasive and uses sound waves of different frequencies superimposed on the patient's normal breathing to measure respiratory resistance and reactance, two components of respiratory impedance. The FOT device measures the relationship between externally applied pressure waves and the resulting air flow to measure respiratory impedance. Values produced at high frequencies correspond to the proximal and large airways, and values produced at low frequencies correspond to distal and small airways (Ram et al, 2019).

<u>48-month Visits (and 54-month Visits in Guatemala only)</u>
All sites: Anthropometry and child neurodevelopment will be conducted according to procedures described above (36-month visit).

Personal exposure measurements will be collected using the **Ultrasonic Personal Air Sampler (UPAS),** which is a small PM2.5 filter sampler. The UPAS has the ability to quietly run for 30-35 hours without external battery packs. The device uses a cyclone for PM2.5 size separation, and a small piezoelectric pump to provide airflow. Airflow is measured internally using high-sensitivity pressure sensors. The device is configured for operation over Bluetooth using a specially designed Android/iPhone app, depending on the version. There are many versions of the UPAS currently in use: UPAS version 2 (UPAS V2), UPAS version 2 Plus (UPAS V2+), and UPAS version 2.1 (UPAS V2.1). For the purpose of this project, UPASes deployed will be the UPAS version 2.1**.** The inlet system, also known as the cyclone, used, will be the Generation 2 Inlet System. In a random subsample of 10-30% of children, personal exposure measurements will be conducted every 6 months.

Guatemala only: Lung function measurements will be made using the FOT technique described above. Additionally, lung function tests will be performed both before and after receiving a 400 µg inhalation dose of salbutamol, a bronchodilator that relaxes the respiratory muscles and opens the airways. After the first test, the child will be given a puff of salbutomol and will rest ~20 minutes before a second test is conducted. This

method assesses the reversibility of lung function resistance, for example, due to respiratory obstruction or inflammation. Before implementing this technique, Dr. Eric McCollum will visit the study site to train the clinical personnel on performance of the reversibility testing and use of salbutamol, as well as awareness of possible secondary effects, such as dizziness, tachypnea, or headache.

As during the HAPIN trial, monthly communication (via telephone or household visit) may occur, to collect information on respiratory illnesses. If the child is sick and is taken to a health facility, an evaluation of the child's respiratory health and collection of information from the child's medical records will occur.

Household (kitchen) monitors, as described in **Section 8**, may be installed as well.

India only: Quarterly health and stove-use checks will occur. SUMS may be installed on stoves and data periodically downloaded.

60-month Visits
All sites: Anthropometry, child neurodevelopment (MDAT), personal exposure measurements (UPAS) and lung function measurements (FOT technique) will be conducted according to procedures previously described. In addition, blood pressure will be measured following procedures described in **Section 2.4**. Child-sized blood pressure cuffs will be used.

Child breath sounds will be recorded and interpreted to identify signs of lower respiratory disease, such as wheeze, crackles, and stridor. Breath sound recordings will be made using the **Feelix** digital stethoscope (Sonavi Labs, Baltimore MD), a practical, non-invasive tool for detecting, amplifying, and recording body sounds, such as breath sounds. Digital stethoscopes like the Feelix device provide a potentially valuable innovation for research because they allow independent and objective analysis of breath sounds, thus improving the accuracy of diagnosis. Previous studies have shown that trained health workers can effectively use the Feelix to record lung sounds (Ahmed 2022; McCollum 2020). The Feelix device will be used to record sounds during 10-second intervals at four different locations: upper front right, upper front left, upper back right, and upper back left of the Thorax. There is no risk for the participant. A standardized diagnosis of breath sounds will be based on both physician interpretation and computerized automated analysis. If any signs of lower respiratory disease are detected, the participant will be notified and referred to seek follow-up healthcare.

Information may also be collected from medical charts.

Monthly or quarterly communication (via telephone or household visit) may occur, to collect information on respiratory illnesses. If the child is sick and is taken to a health facility, an evaluation of the child's respiratory health may also occur. *Guatemala only:* If there are signs of respiratory disease during the evaluation, an ultrasound may be taken to view the child's lungs.

5.4.3. Mother

24-month Visit
All sites: Some of the questionnaires administered at the main HAPIN study follow-up visits (some slightly modified) will be administered annually.

Rwanda: Urine and DBS samples will be collected according to procedures outlined in **Section 2.5**. Periodic health and stove-use checks may occur. Height and weight will be measured, and Body Mass Index (BMI) calculated. SUMS may be installed on stoves and data periodically downloaded.

India: The mother will be monitored for 24-hour personal exposure to HAP ($PM_{2.5}$, CO and BC) as detailed in **Section 7**. Household (kitchen) monitors, as described in **Section 8**, may be installed as well.

Some stove monitoring units (SUMS) were left on stoves at study exit (see **Section 8**). Data was not downloaded from the SUMS after study exit. If households re-enroll for the extended follow-up, this data will be downloaded.

Guatemala: Height and weight will be measured, and Body Mass Index (BMI) calculated. Blood pressure will be measured according to procedures detailed in **Section 2.4**. The mother will be monitored for 24-hour personal exposure to HAP ($PM_{2.5}$, CO and BC) as detailed in **Section 7**. Household (kitchen) monitors, as described in **Section 8**, may be installed as well.

36-month Visit:
All sites: Some of the questionnaires administered at the main HAPIN study follow-up visits (some slightly modified) will be administered annually. A smaller number of questionnaires, about child health and stove use, will be administered quarterly.

Guatemala: Height and weight will be measured, and BMI calculated. Blood pressure will be measured according to procedures detailed in **Section 2.4**.

India: Periodic health and stove-use checks may occur.

48- and 60-month Visits
All sites: Some of the questionnaires administered at the 36-month follow-up visit (some slightly modified) will be administered annually, along with some new questionnaires.

5.4.4. Older adult woman

24-month Visit
All sites: Following the same procedures as baseline (see **Section 2.4**) and follow-up, blood pressure measurements will be taken. Some of the questionnaires administered at follow-up visits (some slightly modified) will also be administered. Height and weight will be measured, and BMI calculated.

India: The OAW will also be monitored for 24-hour personal exposure to HAP ($PM_{2.5}$, CO and BC) as detailed in **Section 7.**

Guatemala: The OAW will also be monitored for 24-hour personal exposure to HAP ($PM_{2.5}$, CO and BC) as detailed in **Section 7.** CIMT measurements will be conducted as described in **Section 2.4.2**.

Table 3. Study procedures at 24-, 36-, 48- and 60-month visits

	24-month visit			36-month visit			48-, 54- (Guatemala only) and 60-month visits		
	Guat-emala	Rwanda	India	Guat-emala	Rwanda	India	Guat-emala	Rwanda	India
				Child					
Anthropometry	X	X	X	X	X	X	X	X	X
Neuro-development	X	X	X	X	X	X	X	X	X
DBS	X	X	X						
Urine	X	X	X						
Blood pressure							X[1]	X[1]	X[1]
Personal HAP exposure	X	X	X	X[2]	X[2]	X[2]	X[2]	X[2]	X[2]
Medical charts				X[3]	X[3]	X[3]	X[3]	X[3]	X[3]
Lung function				X			X	X	X
Lung ultrasound				X[3]			X[3]		
Monthly or quarterly communication or visit							X	X	X
				Mother					
Questionnaires (re: mother & child)	X	X	X	X	X	X	X	X	X
Anthropometry	X	X		X					
DBS		X							
Urine		X							
Blood pressure	X			X					
SUMs		X	X						
Personal HAP Exposure	X		X						
Kitchen monitors/ exposure	X		X						
Stove use checks		X				X			X
				OAW					
Questionnaires	X	X	X						
Anthropometry	X	X	X						
DBS									
Urine									
Blood pressure	X	X	X						
Personal HAP exposure	X		X						
CIMT	X								

[1] At 60-month visit only.
[2] In a random subsample of 10-30% of children, personal exposure measurements will also be conducted every 6 months.
[3] Only if needed.

6. Details of Radiologic Procedures

6.1. Training to Perform Ultrasound

Ultrasound training will occur at each site and will consist of several phases.

Phase 1 is the identification and training of appropriate ultrasound technical personnel (i.e., sonographers).

Phase 2 involves on-site ultrasound didactic and hands-on supervised training (~1-week introductory course in basic ultrasound and different modalities) of personnel for acquisition of all imaging biomarkers. People with medical background (nurses, midwives, physicians) will be targeted for training, depending on the local circumstances at each site. Topics to be covered include: ultrasound system and transducers, basic anatomy and physiology of organs being imaged, and specific topics related to ultrasound imaging of the heart, blood vessels, lungs and fetus.

Phase 3 will consist of submission of test certifying cases for each imaging modality with assessment of quality and written feedback; every sonographer is supposed to pass the certifying cases with 90% or better quality, as assessed by the Clinical and Imaging Core physicians. If necessary, additional on-site training will be scheduled as needed on a case-by-case basis.

Phase 4 will consist of an ongoing quality control (QC) program in which all images are reviewed by Clinical and Imaging Core physicians and written feedback is given for every study.

6.2. Training to perform periodic assessment and maintenance of the ultrasound system

Sonographers will be trained to conduct periodic system maintenance and performance assessment of the ultrasound system, particularly the transducer and other quality assurance (QA) activities to assure uniformity of imaging at all IRCs. It is assumed that each IRC has a maintenance agreement for the ultrasound system.

6.3. Equipment

Each site will have three or more portable ultrasound systems, which will be Sonosite Edge (Sonosite, Bothell, WA). Each ultrasound system will have the following two transducers: Curvilinear 5-2 MHz for abdomen/ob-gyn applications, and Linear 13-6 MHz for BART, CIMT and lung ultrasound.

6.4. Storage of images, quality control, and transfer of quantitative interpreted ultrasound data to the central Data Management Center

Development of infrastructure that will facilitate digital transfer of the ultrasound studies from the IRC to the Clinical and Imaging Core at Johns Hopkins University. Approximately 4,000 ultrasound studies per site (16,000 total) will be performed over a period of about 3.5 years. After each ultrasound study is performed at each IRC, the following steps will occur prior to analysis by the Clinical and Imaging Core:

1. The ultrasound study is downloaded to the IRC workstation by the sonographer. Each site will also store the ultrasound studies on hard drive locally at the individual sites.

2. The IRC sonographer logs onto the TRICE website through a secure channel. Any updated information, notifications, or personalized messages are displayed.

3. The authenticated user enters the appropriate ultrasound study data (i.e., study number, study site, sonographer's name) on a web-based form. This information is associated with the ultrasound study (i.e., DICOM file) and placed on the queue for night transfer to both the Johns Hopkins University server and the Clinical and Imaging Core.

4. All ultrasound studies performed on any given day are transferred automatically each day.

5. After completion of the data transfer, the IRC and the Clinical and Imaging Core are notified of the data transfer upon log-in to the web-based system; by then, full back-up of the files to a remote location has been completed. The storage server will be used to track data transfer and store a database of Clinical and Imaging Core ultrasound reports. It will contain a web-based interface for management and reporting. Problems associated with data transfer are identified by the Clinical and Imaging Core and the appropriate actions are taken to rectify.

The ultrasound readings and quantitative measurements will be performed at the Clinical and Imaging Core. The Clinical and Imaging Core will assume the responsibility for the supervision and monitoring of protocol adherence, including the periodic written performance reviews to the individual sonographers at the IRCs. The Technical Supervisor and the investigators at the Clinical and Imaging Core will assume this responsibility, and will be directly supervised by the Clinical and Imaging Core Co-Directors. A written critique of the IRC sonographers' performance will be provided during the pilot phase. The critique will address the following areas: technical quality of the studies (i.e., number of complete segments), optimal image acquisition to allow visualization of all anatomic structures, the quality of the 2D and Doppler signals, the success in downloading the images to the local workstation and the server. For those sonographers who may experience problems in meeting the requirements, additional evaluations may include: 1) assessment of reproducibility of imaging: same subject imaged by two different sonographers during same sitting, same subject imaged by the same sonographer on two separate sittings; 2) the reproducibility of off-line measurements of their own studies, and of the studies performed by other sonographers. The overall quality of images stored on digital clips will be assessed by two readers to determine adherence to protocol. After the quantitative measurements have been derived at the Clinical and Imaging Core, they will be sent to the Data Management Center.

7. Details of Exposure Assessment Procedures

Two approaches will be used to measure personal exposure. For all adults, exposure will be measured with instrumentation on the participant. For infants for whom it is impractical to measure directly due to their size, a micro-environmental approach will be used. For these cases, the infant will be outfitted with a small location beacon (described below), and the personal exposures will be estimated by integrating corresponding area concentrations in the house with time spent in the respective locations. In both cases, equipment will be deployed for 24 hours (compliance for personal sampling will be checked from the PM monitor's accelerometry data). Pollutants measured will include $PM_{2.5}$, CO, and BC.

Pairs of surveyors will do installations, with one focusing on equipment installation and the other on interviews and participant interactions. The sampling team at each site will have a field supervisor, who will be responsible for coordinating the surveyors, checking their work, and interfacing with the research team. The sampling team will work continuously, rotating through the study group over the course of the year, thereby covering potential seasonal changes in exposure and stove usage.

Exposure for pregnant women will be measured at baseline, 24-28 weeks gestation, and 32-36 weeks gestation (**Figure 7**). The index child will be measured at 1-3 months of age, approximately 6 months of age, and approximately 12 months of age. The older adult women will be measured at the same time points as the mother during pregnancy, and as the child after birth.

Procedures for exposure assessment have been piloted in the Phase II formative research (see Appendix).

Figure 7. Exposure assessment timeline including frequency of assessment for intervention and control households. (Area measurements will also be conducted when using indirect exposure assessment for children)

Study Group	Measurement	Pre-intervention (<20 weeks gestation)	18-26 Weeks gestation	30-36 weeks gestation	<3 months[#]	6 months	12 months	Total Repeats
Intervention Group (n=400)	Pregnant woman (PM2.5, CO, BC)	■	■	■				3
	Child (PM2.5, CO, BC)				■	■	■	3
	Other woman, 25% of households (PM2.5, CO, BC)	■	■	■	■	■	■	6
	Traditional stove usage	SUMS →						Continuous
	LPG Usage	Cylinder tracking* →						Continuous
Control Group (n=400)	Pregnant woman (PM2.5, CO, BC)	■	■	■				3
	Child (PM2.5, CO, BC)				■	■	■	3
	Other woman, 25% of households (PM2.5, CO, BC)	■	■	■	■	■	■	6
	Traditional stove usage	SUMS+ →						Continuous

[#] Will attempt to capture as close to neonatal period as logistics allow

*Will also monitor via questionnaire and SUMS in a subsample

+Will conduct SUMS monitoring in subsample

7.1. Instrumentation

Accurate $PM_{2.5}$ exposure estimates are fundamental to this proposal, and sampling mothers and children in the remote areas is often challenging. Our primary instrument for collecting $PM_{2.5}$ is well suited for these measurements due to its combination of small size, and quiet operation. The Enhanced Children's MicroPEM (ECM) (**Figure 8**), developed by Research Triangle Institute, is a gravimetric sampler weighing approximately 150 g and capable of operating continuously at 0.3 L/min for up to 48 hours. The ECM is virtually silent during operation and participants can wear the sampler on a shoulder band or in a pocket on a customized garment within their breathing zone. The ECM measures $PM_{2.5}$ gravimetrically with a filter and contains a calibrated mass-flow element, a 6-axis accelerometer (to log activity rate and to verify the user complies with wearing the sampler). The ECM also measures and real-time $PM_{2.5}$ with a nephelometer (light scattering sensor), in addition to temperature and pressure. Secondary instrumentation for measuring $PM_{2.5}$ may include gravimetric pump and filter systems, and other small nephelometers. Use of these other $PM_{2.5}$ systems, while suitable for personal monitoring, will be prioritized for use in area sampling (e.g., the kitchen and sleeping area

measurements described below), such that ECMs remain the primary instrument for $PM_{2.5}$ personal sampling of the mother and the older adult woman.

For infants for whom the ECM is impractical to use due to their size or participant preference, a micro-environmental approach will be used, whereby area sampling via 2 ECMs will occur in the kitchen and sleeping areas.

Figure 8. Enhanced Children's MicroPEM (ECM). About the same weight as an iPhone 6.



The child will be outfitted with a coin-sized location beacon (Roximity, USA) linked to receivers in these same areas. Personal exposures for the child will be estimated by integrating corresponding area concentrations with time spent in the respective locations. Real-time area $PM_{2.5}$ concentrations will be estimated with the ECMs nephelometer. The real-time $PM_{2.5}$ concentrations will be corrected by normalizing their response with other ECM measurements where gravimetric and real-time measurements are made simultaneously. Real-time carbon monoxide concentrations will be measured with EL-USB-CO300 monitors (Lascar Electronics, UK). The optical attenuation method is a non-destructive technique that has previously been validated in the field and will allow us to preserve sample integrity for subsequent analyses.

In addition, the new mother will wear a personal sampler at the same time that the children are monitored, for $PM_{2.5}$ and CO. The purpose is to ascertain the child's exposure when the child is next to the mother but outside of the kitchen or sleeping area. The mother will also wear a beacon receiver so we can ascertain when the child is next to the mother. The mother will be instructed to leave the sampling vest near her child if she leaves home without the child. If the mother does not consent to wearing the sampling vest during these monitoring periods, she will be instructed to place a bag with the air and proximity monitors near where the child is expected to spend the majority of the time.

Children will be able to wear their own sampling vest by the time they are 24- to 36-months of age, so at that point their exposure will be monitored directly. The personal sampler used will be either the ECM or another approved monitoring device.

Real-time carbon monoxide concentrations will be measured for all women's personal and area-monitoring locations with EL-USB-CO300 monitors (Lascar Electronics, UK). They will also be used for infant monitoring when feasible. The monitor is about the size and shape as a large pen and makes no noise.

7.2. Ambient Sampling of Outdoor Air and Long-term Kitchen Sampling

Primary ambient sampling will be made using the E-Sampler (Met One Instruments, USA) or comparable systems. The ambient monitoring systems will measure 24-hour $PM_{2.5}$ with integrated filter measurements every 6th day. The E-Sampler also collects continuous real-time data with a light scattering sensor, not just during the 24-hour gravimetric sampling period. The systems will be installed in at least three locations representative of the field sites to track sectoral changes in ambient $PM_{2.5}$ concentrations. An additional network of lower-cost $PM_{2.5}$ monitors – utilizing the PurpleAir II-SD continuous, real-time monitor (PurpleAir, LLC, USA) – will be deployed to augment E-Samplers and provide a more spatially and temporally resolved picture of $PM_{2.5}$ pollution in HAPIN communities.

Long-term term kitchen concentrations of $PM_{2.5}$ (on the order of months) will be made with the SweetSense Air sensor in Rwanda (n=100). SweetSense has been using their home monitoring systems with stove projects in

Rwanda for several years, and it will send stove use and air quality data via mobile network for cloud-based processing and automated reporting. Their platform has integrated the PATS+ sensor for continuous $PM_{2.5}$ monitoring.

7.3. Exposure Data Quality Assurance and Control

Gravimetric filters will be stored in freezers, weighed on ≤1 µg resolution microbalances, and adjusted for changes in field blank weights (3% of total samples, ~ 200 blanks per site). ECM sample flows will be logged continuously, and those differing by more than 15% from the set flow rate will be discarded. Compliance of wearing monitors will be checked via accelerometer data. CO gas monitors will be calibrated monthly.

7.4. Exposure Assessment Training

Berkley Air Monitoring Group will lead the exposure and stove use trainings, and provide technical support throughout the project. Berkley Air has conducted dozens of capacity building workshops and trainings in developing countries, including projects with the USEPA, CDC, and Global Alliance for Clean Cookstoves. Workshops will be conducted at the beginning of all study periods at all study locations. These workshops will include theoretical background and hands on practical training on instrumentation, data management, data analysis, and field protocols. For HAP exposure measurements, local teams will be trained to operate instrumentation and equipment needed for sample collection. During field campaigns, monitoring support will be provided through follow-up trainings, telephone conferences, and email communication.

7.5 Additional Measurements in a Sub-sample

In a 10% random subsample of study participants (pregnant women, older women, children, new mothers), we will double the number of exposure measurements. The purpose of these additional measurements is three-fold, 1) to validate our current/usual number of measurements as a good estimate of long-term average exposure, 2) to provide additional data to help model personal $PM_{2.5}$ for use in other studies where personal measurements are not possible, and 3) to validate our indirect measurements on children.

Specifically, we will double the number of samples (of $PM_{2.5}$, black carbon, CO) in a random subsample of 80 pregnant women, 80 children, and 80 new mothers per site (10% sample of total population), and in 20 older women per site (also a 10% sample), of the same type described above. This would be, then, over the entire follow-up, 6 samples in pregnant women, 6 samples in children, 6 samples in new mothers, and 12 samples in older women (Table 4). The additional samples will be collected at approximate mid-points between the times the regular samples are to be taken. Kitchen area samples will also be collected during these visits for the additional measurements. In the case of children, kitchen samples would be collected in any case as part of the indirect method of measurement using Beacon monitors. The sampling will be conducted with the same instrumentation described in **Sections 7.1 and 7.2** above. Additional SUMs monitoring of stove use will also be conducted in all these households on both biomass and LPG stoves, following the procedures outlined in **Section 8**.

In addition, as a way to validate the indirect method of estimating infant exposure described in section 7.1 above, we will add indirect sampling of $PM_{2.5}$ and CO for new mothers in the 10% sample. Indirect measurements will be conducted as described in **Section 7.1**. Briefly, for women measured via a personal measurement, we will add Beacon transmitter to be worn by the new mother over 24 hours, which will transmit a signal to a Beacon receiver. The transmissions will be used to determine the location of the new mother over

24 hours. At the same time area samplers will measure $PM_{2.5}$ and CO in the kitchen and the bedroom. Personal exposures in the kitchen and bedroom for the new mother will then be estimated by integrating corresponding area concentrations with time spent in the respective locations. These indirect estimates will then be compared to the direct personal measurement from the same mother for her time in the kitchen and bedroom.

Randomization will be done monthly at each IRC by study personnel, by randomly assigning numbers to all households recruited monthly, ranking them by the random number, and choosing the first 10%. Randomization will be stratified by intervention and control arm. This would then be approximately 6 household per monthly recruitment cycle who would be recruited for extra sampling. However, if needed for logistic reasons, the rates of recruitment could be increased in the early part of follow-up. Specifically, personal sampling equipment demands are expected to less during the first half of the study as recruitment builds up towards the full sample size, and therefore it may be more practical to conduct the extra sampling towards the beginning of the study. Participants selected for additional sampling will then be recruited via a special consent form explaining the additional measurements. Participants may be offered additional compensation for their extra time, with the exact form of the compensation being different depending on country, and with the compensation being consistent with compensation already offered for participation in the main study. In the case where participants choose not to participate in the additional sampling, investigators will go back to the randomized list to seek the next participant on the list.

Table 4. Timing for measurements for the sub-sample with additional (highlighted) measurements. PW=pregnant woman, C=child, O=older adult woman in household, NM=new mother

Child age	**<20 weeks Gest.**			**23 weeks Gest.**			**26 weeks Gest.**			**28 weeks Gest.**			**30 weeks Gest.**			**35 weeks Gest.**			**2 months old**			**4 months old**			**6 months old**			**8 months old**			**10 months old**			**12 months old**		
	PW	C	O	PW	C	O	PW	C	O	PW	C	O	PW	C	O	PW	C	O	NM	C	O	NM	C	O	NM	C	O	NM	C	O	NM	C	O	NM	C	O
Personal exposure, **24-h $PM_{2.5}$, BC, CO**	X		X	X		X	X		X	X		X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

8. Details of Stove Use Monitoring

All homes in the intervention arm will be equipped with Stove Use Monitoring Systems (SUMS) on their traditional stoves and subset of approximately 80 homes in the control arm. Homes in the intervention arm will be equipped with SUMS on their traditional stoves and a subset of approximately 80 homes will also have their LPG stoves monitored. The 80 homes per IRC will be the same homes where the additional exposure measurement will take place (see **Section 7.5**). Use of the LPG stoves in intervention arm households will also be tracked through project records of the distribution of and use of cylinders. Use monitoring will begin immediately after recruitment, when households in the intervention arm are still using their traditional stoves. It will continue throughout the 18-month intervention period. Compliance will be checked every two weeks when SUMS data is downloaded. The stove use monitors to be deployed will be Geocene Dots, small data loggers with high temperature k-type thermocouple leads, which attach to the various stoves used in a home. The high temperature thermocouples allow for placement near or within the combustion zone, which provides distinct, clear temperature traces for converting into stove use estimates. Stove use will be defined as times when the temperature is increasing or stable, and significantly above ambient temperature. The SUMS data will allow us to determine the level of compliance achieved by the behavior change strategies.

SUMS may be left on stoves after study exit in case follow-up studies are conducted. If left, the SUMS data from the post-exit time period will not be used/accessed until such time that a new consent that includes permission from the participant to access that data is administered. The data may be downloaded to a confidential server in the interim since there is limited storage space on the SUMS, but the data will not be accessible.

Originally, there was a plan to conduct qualitative, in-depth interviews with a subset of primary cooks from households in the intervention arm in order to understand factors influencing use of LPG stoves. These were scheduled to occur near the end of households' participation in the trial, but prior to study exit. This timing coincided with the onset of COVID-19 and related lockdowns; therefore, study staff were not able to go to homes to conduct these interviews. These interviews will not be conducted.

9. Details of Specimen Collection, Processing, and Storage

In conjunction with the Intervention Trial, biospecimens (urine and dried blood spots [DBS]) will be collected from the pregnant woman, index child, and, in an estimated subsample of 15% of households, an older adult woman. These samples will be collected at the developmentally-relevant intervals outlined in Table 5.

Table 5. Samples to be collected from each participant at relevant time points. PW=pregnant woman, C=child, O=older adult woman in household, NM=new mother

Child age (study time point)	**<20 weeks gestation** (baseline)			**24-28 weeks gestation** (1-3 months post-randomization)			**32-36 weeks gestation/birth** (3-5 months post-randomization)			**<3 months old** (~9 months post-randomization)			**~6 months old** (~12 months post-randomization)			**~9 months old** (~15 months post-randomization)			**~12 months old** (~18 months post-randomization)		
	PW	C	O	PW	C	O	PW	C	O	NM	C	O	NM	C	O	NM	C	O	NM	C	O
DBS	X		X	X		X	X	X	X		X	X		X	X					X	X
Urine	X		X	X		X	X		X		X	X		X	X					X	X

Prior to biospecimen collection, field staff at all four sites will be trained and tested in standardized sample collection, processing, and storing procedures. This training will also include Bloodborne Pathogen for Researchers and Biosafety training modules. Our standardized protocols will be strictly followed for biospecimen collection. Universal precautions for preventing transmission of infectious disease will be strictly followed during biospecimen collection and processing. This includes the use of gloves, sterilization and disposal of expendables that come in contact or may come in contact with biological material, proper disposal of biohazardous material, and the use of other personal protective equipment (e.g., safety glasses, face shields, lab coats), where warranted. The timetable for field processing, handling and storage of specimens is provided in the table above. The samples should be kept at -20°C for no more than 6 months. For samples from Guatemala, Peru, and Rwanda (and the relevant samples being shipped from India), shipments will be made at a minimum every 6 months to Emory where samples can be stored at -80°C indefinitely (Table 6).

Table 6. Processing and Storage Conditions for Biosamples.

	Time at room temperature before processing	Time in refrigeration before processing	Short term storage before shipping to Emory
Plasma *derived from venous whole blood (VWB)	**Immediately** pipette off layer after centrifugation of VWB	Can store under refrigeration **for up to 8 hrs before** processing	Storage at minimum temp of -20°C for up to 6 months
Buffy coat (BC)/Red blood cells (RBC) *derived from VWB	**Immediately** pipette off layer after centrifugation of VWB	Can store under refrigeration **for up to 8 hrs before** processing	Storage at minimum temp of -20°C for up to 6 months
Urine	**4 hr maximum** (including time at the participant's house prior to team's arrival). May store an additional 4 hr under refrigeration but no more than 4 hr at RT.	Can store under refrigeration for up to 4 hrs (room temperature + refrigeration can be up to 8 hrs prior to processing); IRC may choose to have cold storage in vehicles	Storage at minimum -20°C for up to 6 months
Dried Blood Spots	Until dry (~24 hrs depending upon climate. More humid climates require longer drying, arid climates require less drying)	Do not refrigerate	Storage at minimum -20°C for up to 6 months
Buccal cells	**4 hr maximum.** May store an additional 4 hrs under refrigeration.	Storage for up to 4 hrs under refrigeration before processing	Storage at minimum -20°C for up to 6 months
Oral Rinse	**4 hr maximum.** May store an additional 4 hrs under refrigeration.	Storage for up to 4 hrs under refrigeration before processing	Storage at minimum -20°C for up to 6 months
Nasal turbinate	**Must be placed in RNAprotect® Cell Reagent immediately after collection**	May be refrigerated for up to 4 hrs before freezing	Storage at minimum -20°C for up to 6 months

9.1. Pregnant Women and Older Adult Women Urine Sample Collection

Women will be provided with a sterile urine collection cup (e.g. 100 mL or 500 mL) and instructed to urinate briefly to waste before collecting the remainder of the urine void in the cup. The time of urine collection and time of previous urine void (if known) will be recorded on the collection log and the total volume collected will be estimated and recorded. The urine will be transferred in its entirety to 4 cryovials (2*4-mL tubes with 3 mL urine; and 2*10-mL tubes with 7.5 mL urine) and will be labeled. The labelled bottles will be secured in freezer boxes and will be stored in a cooler with ice packs until they can be transported to short-term freezer storage at the IRC's laboratory. Samples will be stored at -20°C until shipment on dry ice to the Biomarker Core (see Table 6).

9.2. Pregnant Women and Older Adult Women DBS Sample Collection

A finger from the non-dominant hand will be swabbed with a sterile alcohol wipe. The sterile lancet will be used to puncture the skin and the initial drop of blood will be wiped away with an alcohol swab. Blood from the finger will be allowed to drip onto 5 standard spots on a Guthrie DBS card. The hand and arm will be massaged and warmed until a large drop of blood appears on the finger. The first drop will be wiped away and the next drops will be quickly applied to an unfilled spot on the card. The finger should not be squeezed nor should it touch the card to deposit the spot. Bleeding can be encouraged by massaging the arm or warming the hands. The spots should be filled by gravity drop of the blood droplets. A second finger stick will be collected if fewer than three spots are achieved with the first finger stick. No more than two finger sticks will be performed per participant per visit. Each spot will contain 75 µL to <100 µL blood so the total blood collection will be less than 1 mL. Cards will be labeled and dried at room temperature (20-25°C) on top of a paper towel or DBS rack for a minimum of 24 hours without the use of external heat or a fan. Dried cards will be placed in labeled ziptop bags (1 card per bag) into which a desiccant pouch and a humidity indicator card will be placed (see Table 6).

9.3. Pregnant Women, Older Adult Women, and Child Capillary Blood Sample Collection

Blood will be collected using the same finger stick used for the DBS (protocol described above in 9.2. Pregnant Women and Older Adult Women DBS Sample Collection). Before collecting the DBS, the field staff will collect 5-50 µL of blood from the finger into the Hemocue (for hemoglobin measurement in pregnant women and children) or A1CNow+® (for HbA1c measurement in other women) blood collection capillary cuvette. The cuvettes will be analysed in the field following collection as described in **Section 10.2**. Capillary Blood below. A second puncture for adults AND children is allowed if the hemoglobin result is <7.0 g/dL, which suggests severe anemia. A recheck with a second puncture will allow this result to be confirmed.

9.4. Child Urine Sample Collection

For urine, a sterile urine collection cup (e.g. 100 mL or 500 mL) or pediatric urine collection bag will be left with the mother with instructions. Pediatric urine collection bag instructions include to affix the bag over the child's urethra area on the morning of a scheduled visit. The urine collection bag should not remain on the child for more than 3 hours. The mother will be instructed how to remove and store the collected urine sample (cup or bag) until the site visit. The urine sample will be collected upon each site visit and stored in a cooler until return to the local HAPIN satellite laboratory where it will be divided into multiple aliquots (same as described above for the adult participants) for storage. The length of time the urine bag was on the child will be recorded. The urine will be chilled within 4 hours and frozen within 12 hours for short-term storage at -20°C until shipped to the Biomarker Core. For long term storage, the urine will be kept at -80°C (see Table 5).

9.5. Child DBS Sample Collection

World Health Organization (WHO) and American Association of Clinical Chemistry (AACC) recommends heel stick for children up to and including 6 months, and finger stick for children older than 6 months. A finger or heel for children <6 months in accordance with WHO guidelines will be swabbed with a sterile alcohol wipe. The sterile lancet will be used to puncture the skin and the initial drop of blood will be wiped away with an alcohol swab. Blood will be allowed to drip onto 5 standard spots on a Guthrie DBS card. The hand or foot can be warmed and massaged until a large drop of blood appears on the finger or heel and the drop will be quickly applied to an unfilled spot on the card. The fingers or heel should not be squeezed or "milked." The drops should fall by gravity; the fingers or heel should not touch the card. For heel stick DBS collection, the blood drop itself may touch the card, but not the heel. Only one finger or heel stick will be performed per visit for dried blood spots for children (note: a second puncture is allowed only if the hemoglobin result is <7.0 g/dL, section 9.3 on hemoglobin). Each spot will contain a minimum of 40 µL, and up to 75-100 µL blood so the total blood

collection will be less than 1 mL. Cards will be labeled and dried at room temperature (20-25°C) on top of a paper towel or DBS rack for a minimum of 24 hours without the use of external heat or a fan. Dried cards will be placed in labeled ziptop bags (1 card per bag) into which a desiccant pouch and a humidity indicator card will be placed (see Table 6).

9.6. Short-term Storage of Biospecimens at Intervention Research Centers

The DBS and urine samples will be stored in -20°C freezers at the IRCs prior to shipment to the Biomarker Core. The DBS and urine samples will be shipped express shipment to Emory University, where they will be kept at -80°C long term or sent to BioLINCC repository.

9.7. Pregnant Women and Older Adult Women Urine Sample Processing

After collection, urine will be divided into 4 aliquots (we anticipate storing 4 cryovials (2*4-mL tubes with 3 mL urine; 2*10-mL tubes with 7.5 mL urine). Each vial will be labeled with a unique barcode/ID label, stored in durable freezer boxes in a -80°C freezer and will be logged into the freezer map inventory system for continued monitoring of location/storage parameters. When a vial is requested for analysis, the vial will be logged out to the proper person using the freezer mapping inventory system and a full chain-of-custody will be retained for each sample. Any residual sample will be returned to the freezer with a notation that it has undergone an additional thaw-refreeze cycle.

9.8. Child Urine Sample Processing

Same as **Section 9.7** Pregnant Women and Older Adult Women Urine Sample Processing

9.9. DBS Sample Processing

DBS samples from all participants will be transferred to a new ziptop bag, and re-labeled with fresh humidity recorder/desiccant pouches. Current humidity at receipt will be recorded in the freezer mapping inventory system as will location and number of spots retained. When ready for use, a DBS card will be removed (still frozen) and a spot will be cut off and placed in a similar ziptop bag with desiccant and humidity recorder before giving to the appropriate person for analysis. All handling and chain-of-custody of DBS will be recorded.

9.10. Long-term storage in Biorepository

The Biomarker Core will develop and maintain a long-term biorepository to store biological specimens (DBS and urine) in a manner that retains sample integrity and participant confidentiality by using barcodes/ID numbers assigned by the Trial Coordinating Center at the time of participant enrollment to identify samples.

Urine samples will be aliquoted into 4 cryovials (2*4-mL tubes with 3 mL urine; 2*10-mL tubes with 7.5 mL urine) to enable access to the samples without repeated thaw-refreeze cycles to maintain sample integrity. DBS cards will be stored in ziptop bags with a desiccant pouch to prevent deterioration of samples.

The biorepository consists of four secure -80˚C freezers linked to a monitoring and alarm system that generates automatically transmitted alarms (via email and phone) when the temperature deviates within 5˚C of the target temperature. In the case of freezer failure, a fifth “overflow” freezer will be used to temporarily house the samples until the freezer is repaired. Regular freezer maintenance (i.e., defrosting and calibrating temperature gauges and alarms) will be performed every six months and documented. The freezers will be fully mapped with each sample location (shelf, box, position) and volume. The map will be retained and updated, as needed, using the Biomarker Core’s Laboratory Information Management System linked to a web portal to which study collaborators have ready access.

Each freezer will remain locked with access granted only by the Biomarker Core Co-Directors in accordance with the study protocol or at the request of the study PIs or Steering Committee. Access to the study samples outside of the current protocol will require a vetting process similar to that of NHLBI BioLINCC repository. All procedures for collection, labeling, and storage are compatible with the NHLBI BioLINCC repository. A limited dataset will be provided to NHLBI BioLINCC as well.

9.11 Sub-study of Additional Biospecimen (buccal cells, oral rinse, nasal turbinate and venous blood) Sample Collection from a Subset of Older Adult Women in Peru and Guatemala

A sub-study of additional biomarkers, designed in collaboration with colleagues at the National Cancer Institute, will be conducted among approximately 200 older adult women in the intervention and control arms from the Peru and Guatemala IRCs (400 women total). It should be noted that Peru and Guatemala, based on data through 2018, will have higher rates of recruitment of older women than Rwanda and India, with approximately 25% of households in Peru and Guatemala having OAW (200 per IRC). Samples for this sub-study will be collected for all 200 subjects per IRC at baseline before randomization, and at follow-up, approximately 18 months following the intervention (therefore, at the first and seventh visits for the older adult woman). Additional samples include buccal cells, oral rinse, nasal turbinate, and peripheral (venous) blood. The collection of these additional samples is anticipated to add approximately 15 minutes to the standard biomarker collection protocol.

9.11.1. Buccal cell scrape and oral rinse

The oral rinse and buccal scrapes are both relatively high cell yield techniques. For the oral rinse, participants will vigorously rinse their mouth and the rinsates are collected in a centrifuge tube. The tube is centrifuged and the pellet and supernatant are removed to separate cryovials, labeled and frozen. Two buccal cell scrapes will be collected by gently scraping the buccal mucosa on both sides of the mouth with a small plastic collection spoon. One is placed into phosphate buffered saline (PBS) and one into RNAprotect® Cell Reagent immediately after collection, kept cold on ice, and then frozen at the IRC laboratory. Collection is gentle and causes no discomfort to study participants. These samples yield substantial amounts of human DNA, microbial DNA, and RNA. All samples will be frozen at -20°C until shipment to the biorepository at Emory where they will be stored at -80°C until analysis (see Table 6).

9.11.2. Nasal turbinate brush samples

Nasal turbinate brush samples can be collected using a soft cytobrush on each turbinate and placed into phosphate buffered saline (PBS) in a tube and labeled. The nasal turbinate sample will be placed into RNAprotect® Cell Reagent immediately after collection, kept cold on ice, and then frozen at -20°C until shipment to the biorepository at Emory where they will be stored at -80°C until analysis (see Table 6).

9.11.3. Peripheral blood

A 5-mL peripheral (venous) blood sample will be collected in an EDTA vacutainer tube by standard clinical venipuncture of a cubital vein. The sample will be transported to the IRC laboratory within 8 hours of collection. The sample will be centrifuged for 15 minutes at 1500-1800 g and plasma can be pipetted into three 0.5 mL cryovials, and the remainder pipetted into a fourth cryovial. The buffy coat/red blood cell can be pipetted into two additional and separate 2 mL cryovials. All samples will be labeled and frozen at -20°C until shipment to the biorepository at Emory where they will be stored at -80°C until analysis (see Table 6).

Table 7. Samples to be collected in *Additional Biospecimens Sub-study from a Subset of Older Adult Women in Peru and Guatemala*. PW=pregnant woman, C=child, O=older adult woman in household, NM=new mother.

Child age (study time point)	**<20 weeks gestation** (baseline)			**24-28 weeks gestation** (1-3 months post-randomization)			**32-36 weeks gestation/birth** (3-5 months post-randomization)			**<3 months old** (~9 months post-randomization)			**~6 months old** (~12 months post-randomization)			**~9 months old** (~15 months post-randomization)			**~12 months old** (~18 months post-randomization)		
	PW	C	O	PW	C	O	PW	C	O	NM	C	O	NM	C	O	NM	C	O	NM	C	O
Buccal cell scrape & oral rinse			X																		X
Nasal turbinate			X																		X
Venous blood			X																		X

10. Details of Specimen Analysis

10.1. Dried blood samples

Analyses of DBS samples will be prioritized based upon number of spots available for analysis as follows with insurance that at minimum one spot is retained in the biorepository. For most biomarkers, we will evaluate a random subset (n=25) of older adult woman, 6 repeated measures, to inform sample size required for main analysis.

Pregnant Women: Hb/HbA1C

Older Adult Women: CRP, endothelial markers, other inflammation/oxidative stress markers, P450, p53 TAA. Note: metabolomics and miRNA selection will be dependent upon spot availability.

Child: metabolomics/miRNA

10.1.1. Endothelial and Cardiovascular Markers and Inflammation Markers

(Table 8) DBS are soaked in 250 µL phosphate buffered saline (PBS) to release blood from the spot. The diluted samples (maximum aliquot) are placed in 96-well plates, in duplicate, and prepared according to the standard assay protocol. The resulting reaction products are analyzed on a multiplex plate reader with a full set of calibrants and quality control samples. Values are averaged before reporting. Calculated values of replicates that have differences greater than 20% will be repeated. Similarly, samples are repeated if the quality control samples indicate a failure of proper measurement.

Table 8. Target biomarkers of endothelial and cardiovascular function, and inflammation.

Biomarker	Purpose
Intercellular adhesion molecule 1 (ICAM-1)	Endothelial marker of cardiovascular function
Vascular cellular adhesion molecule 1 (VCAM-1)	Endothelial marker of cardiovascular function
Endothelin-1	Endothelial marker of cardiovascular function
E-selectin	Endothelial marker of cardiovascular function
C-reactive protein (CRP)	Inflammation marker
Interleukin 6 (IL-6)	Inflammation marker
F2-Isoprostanes	Inflammation markers

10.1.2. Markers of Oxidative Stress

DBS are diluted with 200 µL PBS to release blood from the spot. Myeloperoxidase (MPO) are measured using commercially available immunoassays.

10.1.3. Hemoglobin (Hb) and Hemoglobin A1c (HbA1c)

DBS are diluted with PBS to release blood from the spot. Hemoglobin and hemoglobin A1c are measured using a high performance liquid chromatography-tandem mass spectrometry. The percentage HbA1c is the ratio of HbA1c to Hb.

10.1.4. Tumor-Associated Antigen Antibodies

Antibodies associated with the p53 signaling pathway have been associated with respiratory cancers as well as other cancers. A commercially available p53 antibody array will be used that evaluates 196 highly specific

antibodies related to the p53 signaling pathway. This array contains site-specific and phospho-specific antibodies, enabling evaluation of tyrosine and serine/threonine phosphorylation at specific sites.

10.1.5. Cytochrome P450 Induction

Cyp450 will be measured using standard spectrophotometry and the ferrous CO difference spectrum. The difference between absorbance at 450 nm wavelength of totally reduced P450 (artificially bound to CO using sodium dithionite) and the sample. Using the Beer-Lambert law and an average extinction coefficient of 91, 000 $M^{-1}cm^{-1}$, the concentration of Cytochrome P450 in the blood sample can be calculated.

10.1.6. Lipids

HDL cholesterol, LDL cholesterol, total cholesterol, and triglycerides will be measured using standard clinical methods (e.g., ultra-centrifugation and fractionation of cholesterol).

10.1.7. Metabolomics

One blood spot will be dissolved in PBS and subsequently extracted with acetonitrile containing internal standards. Samples are analyzed using liquid chromatography-high resolution mass spectrometry enabling accurate mass calculation. Data are subjected to linear regression and principle component analyses to identify features that are significantly related to exposure status with an initial false discovery rate of 20%. Data are matched to standard metabolomics databases to tentatively identify over 10,000 features (unique mass to charge ratios coupled with retention times).

10.1.8. MiRNA Arrays

One spot will be dissolved in 250 µL PBS to release blood from the spot. The sample will be processed according to the manufacturer's guidelines of TagMan Array kits. Database searches will be performed to identify the miRNA present.

10.2. Capillary Blood

Pregnant Women: Hb

Older Adult Women: HbA1c

Child: Hb

10.2.1. Hemoglobin

The HemoCue® Hb 201 system, a point of care (POC) instrument from HemoCue AB, will be used to ascertain hemoglobin (Hb) among the pregnant women and children. Once the small, portable device is turned on and before collecting the DBS, the field staff will collect blood from the finger into the Hemocue blood collection device. This device is then inserted into the instrument analyzer and results are displayed after 15-60 seconds. The participant will receive results in writing on a prepared health feedback form. The field technician will explain the results using a prepared script. This will be collected at three visits for the pregnant woman (baseline,1 - 3 months and 3 – 5 months post-randomization), and two visits for the child (6 months and 12 months of age).

10.2.2. Hemoglobin A1c

The A1CNow+® system, a POC instrument from Chek Diagnostics, will be used to ascertain HbA1c among the older adult women. Before collecting the DBS, 5 ul of blood will be inserted into the small, portable instrument blood collector apparatus. Within 5 minutes, the HbA1c results are available. The participant will receive results in writing on a prepared health feedback form. The field technician will explain the results to the participant using a prepared script. This will be collected at four visits (baseline, 3-5 months, 12 months, and 18 months post-randomization).

10.3. Urine

Exposure biomarkers (especially for children whose urine may be limited) will be prioritized as follows: polycyclic aromatic hydrocarbon (PAH) biomarkers, levoglucosan, volatile organic chemicals (VOC) biomarkers, heavy metals, and tobacco-related biomarkers.

10.3.1. Analysis of Polycyclic Aromatic Hydrocarbon (PAH) Metabolites

The target PAH metabolites are urinary 1-OH pyrene and 1- and 2-naphthols. 0.5 mL urine will be aliquoted into a test tube and spiked with isotopically labeled internal standards for automatic recovery correction and normalization of mass spectral data. 1 mL urine is aliquoted and spiked with isotopically labeled isotopes. 1 mL 1M NaAc buffer is added along with 100 µL b-gluruconidase/sulfatase. The solution is incubated at 37C for 17 hours. The hydrolysate is extracted using solid phase extraction (mixed mode cartridges) and the extracts are concentrated. The extract will be subjected to a derivatization process with diazomethane to protect the polar hydroxyl groups then the concentrated extract will be analyzed using GC-MS/MS.[45] PAH metabolites do not respond well to LC-MS/MS analysis thus require a derivatization and GC-MS/MS analysis. Each analyte has either separation by time or separation by mass (or both). Two precursor → product ions (quantification and confirmation ion transitions for MS/MS) and its isotope analogue. The limits of detection (LODs) for all analytes are 0.1 ng/mL with relative standard deviations (RSDs) of less than 12%.

10.3.2. Analysis of 3-OH Cotinine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), levoglucosan, 8OH-deoxyguanosine (8OHdG), and Volatile Organic Chemicals (VOC) metabolites

1 mL urine is aliquoted and spiked with isotopically labeled isotopes. 1 mL 1M NaAc buffer is added along with 100 µL b-gluruconidase/sulfatase. The solution is incubated at 37C for 17 hours. The hydrolysate is extracted using solid phase extraction (mixed mode cartridges) and the extracts are concentrated for analysis by high performance liquid chromatography-tandem mass spectrometry with electrospray ionization. Similar to our PAH analysis, two precursor → product ions (quantification and confirmation ion transitions are monitored for each analyte and its isotope analogue. The LODs are in the low pg/mL range with RSDs less than 10%.

10.3.3. Heavy Metals Analysis

A spot is soaked 200 uL in PBS to liberate the dried blood. The samples are digested using nitric acid on a heater block at 90C for 1 hour to break down molecular components of the blood. The digested matrix is diluted with buffer and injected onto the ICP-MS. Multiple isotopes of each metal are monitored to ensure reliable results. Yttrium is typically used as an internal standard. The method LODs range from 0.1-0.5 ng/mL with RSDs of <10%.

10.4. Sub-study of Additional Biospecimen Samples Among Older Adult Women

The additional collection of buccal cells, oral rinse, nasal turbinate, and peripheral (venous) blood samples is described in **Section 9.11**. Specimen analysis is described here:

10.4.1. Genomic markers

mRNA expression and microRNA will be analyzed in buccal and nasal turbinate cell samples using the Gene Chip Human Exon 1.0 ST array from Affymetrix and through sequencing.

10.4.2. Microbiome

Frozen aliquots of oral rinses will be shipped on dry ice to the sequencing vendor (MR DNA Lab, Shallowater, TX) for sample preparation, generation of sequencing libraries, DNA sequencing, and initial sequence analysis. DNA will be purified using the industry standard MoBio Power Soil kit (Qiagen, Carlsbad, CA). The 16S rDNA sequence provides a bacteria-specific target suitable for the identification of bacteria, either individually, in mixed bacterial populations, or in complex samples containing multiple taxonomic kingdoms. For this project, the V4 region of the purified DNA will be amplified with 515F/806R primers containing library adapters and barcodes. These amplification products will be subjected to Illumina sequencing to generate a minimum of 20,000 reads. Sequencing reads will be processed by MR DNA Lab using a proprietary, QIIME-like, pipeline. Briefly, primer and barcode sequences will be trimmed from raw reads and QC analysis removes short reads, and reads with ambiguous bases or nucleotide homopolymer stretches longer than six bases. Sequences that pass QC will be binned based on possessing at least 97% sequence identity and any chimeric sequences are removed. Using these binned sequences, the operational taxonomic units (OTUs) will be identified based on BLASTn results against a database of curated 16S sequences at various taxonomic levels. To calculate alpha diversity, OTU data will be analyzed with a custom Excel spreadsheet to calculate richness, Inverse Simpson index, and Shannon index. To facilitate direct comparisons between samples, the alpha diversity values will be converted to number of effective species. Significance will be assessed through single factor ANOVA testing. Open source R-scripts will be used for more advanced statistical analyses, e.g. linear correlations between various factors and PCOA analysis. Total bacterial burden will be determined by MR DNA Lab using qPCR quantitation of total 16S counts. From this value, the absolute levels of each OTU in the sample can be extrapolated.

10.4.3. Epigenetics (DNA methylation)

DNA samples (~500 ng) randomized across 96-well plates will be provided to the Integrated Genomics Facility at Emory, which provides fee-for-service DNA modification, array processing and data acquisition using the Illumina Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA). Resulting array data will be downloaded through secure FTP to by Emory researcher and HAPIN collaborator Dr. Carmen Marsit's laboratory. We will implement the most up-to-date R/Bioconductor package minfi (currently v 1.22.1, which can accommodate the EPIC platform) in the R statistical environment. Emory's laboratory has developed an extensive QA/QC pipeline to assess data quality and to perform the requisite data normalization procedures to address potential technical issues. The reproducibility of this assay has been demonstrated with correlation coefficients of >0.99 across replicate samples (manufacturer and in Dr. Marsit's lab). The new Illumina Infinium MethylationEPIC array is designed to provide a truly genome-wide examination of DNA methylation by simultaneously profiling the methylation status for >850,000 CpG sites at single-nucleotide resolution. This new technology in addition to including over 90% of the CpG sites covered in the prior Illumina Infinium 450K array also covers CpG sites outside of CpG islands, non-CpG sites found to be methylated in human stem cells, enhancer regions, DNase hypersensitive sites, and miRNA promoter regions. Utilizing a commercially available array allows for open-source sharing and comparison of data across studies, which will allow for external validation and inclusion in meta-analyses with consortium groups.

DNA methylation in peripheral leukocyte DNA samples will also be processed and arrayed as described here for the buccal cell samples through the Integrated Genomics Core at Emory. As above, strict QA/QC procedures will be implemented, and in all analyses, we will implement the methods to estimate and control for blood cell composition developed Dr. Marsit and colleagues and available in the R minfi package (Houseman et al, BMC Bioinformatics, 2012).

10.4.4. Metabolomics

A broad metabolomic analysis of plasma samples is a key technology that will be applied to these samples, and the technology is rapidly evolving to being able to identify thousands to tens of thousands of compounds. Analyses will be conducted in HAPIN collaborator, Dean Jones' laboratory.

10.4.5. MicroRNA

Plasma will also be assessed for soluble microRNA levels using small RNA sequencing (RNA-seq). The technology to detect and quantify a spectrum of soluble RNA species is continuing to improve. Dr. Marsit's laboratory has developed significant expertise in the isolation of nucleic acids from peripheral plasma exosomes utilizing the Norgen Plasma/Serum Circulating and Exosomal RNA purification kit, which is a two-step affinity purification followed by vesicle lysis and RNA purification (including miRNA) using a column-based method. They have optimized these procedures to utilize small volumes down to 25 µL. Although the gold-standard for exosomal purification would be ultracentrifugation followed by electron-microscopy to confirm exosomal components, the limited volume availability of samples hinders feasibility for application to human population studies. RNA quantity is assessed using the Nanodrop Spectrophotometer, and integrity assessed using the Agilent Bioanalyzer. Isolated RNA will be separated into single use aliquots and kept in RNAse-free water at -80oC. RNA for all experiments will be utilized within 2 weeks of extraction as NCI has noted loss in miRNAs stability with extended storage once extracted from tissues. Total RNA samples will be provided to the Integrated Genomics Facility at Emory, which provides fee-for-service small RNA sequencing using the Illumina miSEQ platform through cooperation with the Hudson-Alpha Genomics Institute. Sequencing quality and mapping will be assessed through defined pipelines established in Dr. Marsit's laboratory. Small RNAseq will provide comprehensive profiling of the small RNA transcriptome in peripheral exosomal samples and can be used to identify those small RNA that may be acting as peripheral signals of exposure and contributing to defined outcomes.

10.4.6. Clinical immune/inflammatory biomarkers

Panels of plasma immune/inflammatory markers (e.g. CRP, MIP-1D, CXCL11, IL-16) will be analyzed using the Luminex platform.

11. Human Subjects, Research Ethics and Protection of Human Subjects

11.1. Adverse Event Reporting

To assess for unexpected risks that may be associated with the study protocol, participants will be monitored for adverse events. Investigators at the IRCs will determine if any clinical adverse experiences occur during the duration of the study because of procedures or the intervention itself. The investigator will evaluate and determine if changes in status among pregnancy, growth and development among children, or blood pressure among the older adult woman are important and different from that which is expected in the course of the study. If important and unexpected adverse experiences occur, they will be recorded as an adverse event. Serious adverse events (SAEs) will include any death and any complication associated with the intervention or procedure, including LPG tank ruptures and burns related to LPG stoves or tank ruptures, or any other event that fits the standard definition of an SAE. An SAE report will be reported to the Data Coordinating Center, the IRC PI, and the IRC medical monitor (MM) within 48 hours. A designated timeline of review (for relatedness and expectedness) and feedback will be followed, as described in the Data Safety Monitoring Plan (DSMP). SAEs will be reported to the Emory IRB and to local IRBs and Ethics Committees per their reporting requirements.

A Data Safety Monitoring Board (DSMB) will be created to monitor safety and efficacy in accordance with NIH Guidelines. We will not pre-define stopping rules given the low level of risk and the need to complete the follow up period to demonstrate the outcomes of interest. Nevertheless, we will adhere to decisions by the DSMB to suspend the trial based on clear findings of efficacy or futility.

11.2. Potential Risks

The potential risks from data collection are minimal, and include mild discomfort during the finger prick (adults and children) or heel stick (children) for dried blood spots, minimal inconvenience from wearing exposure monitoring instruments and from the ultrasound assessments, and the identification of new health problems that may be upsetting to participants or their families. The exposure experienced by the subjects in this study will be the everyday level experienced in the natural setting of their homes. Most data are obtained by questionnaire, observation, or non-invasive procedures and pose no risk of physical harm to the pregnant woman, offspring, or non-pregnant woman. Furthermore, many of the data collection tools and protocols have been used in previous studies by several team members in the proposed countries without evidence of adverse effects. No adverse effect in humans of using ultrasounds has been reported and the study will be conducted using strict guidelines in accordance with the ALARA (as low as reasonably achievable) principle of the American Institute for Ultrasound in Medicine. Rarely people develop arm discomfort while the blood pressure cuff is inflated, and redness or skin irritation from the electrodes used for the ultrasound. The only minimally invasive components are the finger prick (adults and children) or heel stick (children) for dried blood spots, and the urine sample collection. These biological specimens will be obtained following SOPs. Appropriate biohazard and infection control procedures, including use of single-use lancets will be enforced to avoid risk to the subjects. In the event that new health problems are identified, findings will be explained to the participant or participant's caregiver (in the case of the children) and appropriate follow-up with a local health care provider will be advised.

The study includes the provision of LPG stoves and gas to participants. LPG stoves are generally considered safe. The possibility of adverse events will be minimized by:

- Using well-reputed stove and gas providers
- Providing new cylinders and valves where possible

- Providing new and high quality connecting hoses and regulators (the latter are used to ensure constant pressure of the cooking appliance).
- Wherever possible, the chosen stoves will have an automatic ignition (piezoelectric igniter) to avoid use of matches or lighters to start the flame.
- Placing barcodes on cylinders for identification, and providing refills in these same cylinders when possible
- Including testing for leaks and connecting/disconnecting cylinders in field assessment team training
- Installing the stove and fuel cylinder in a secure location and providing instructions to householders about its operation and safety precautions (see details in paragraph that follows)
- Examining stove condition regularly, performing any repairs necessary, and measuring and recording weight of LPG cylinders in order to anticipate need for refills.

At the time of delivering the LPG stove and fuel cylinder, householders will receive complete instructions about their use and procedures necessary to prevent accidents. The installer will ensure that it is placed in a stable, secure location, that the stove and fuel cylinder are properly connected to prevent leaks, and that the household members are familiar with use and safety procedures. These procedures will be reviewed with household members within 2 weeks of installation and at least annually thereafter during the course of the study.

Subjects will be free to withdraw from the study at any time without penalty, prejudice, or coercion. We will explain to the women that the study visits should not replace their routine prenatal and postnatal exams, and that their access to care will not be affected by their participation in the study. Detailed procedures for the notification of any abnormal findings of clinical relevance as well as non-emergency and emergency referrals are described in SOPs disseminated by the Trial Coordinating Center.

11.3. Procedures for Protection of Human Subjects

The Institutional Review Board (IRB) at the Trial Coordinating Center (Emory University) will work with each IRC to ensure that Human Subjects/Ethics approvals are obtained prior to enrollment of participants in order that all participants are safe from harm and their rights are protected. The original main trial study protocol was reviewed and approved by IRBs or Ethics Committees at Emory, SRU and the Indian Government Ethics Committee, UVG and Guatemalan Government Ethics Committee, LSHTM and the Rwandan Government Ethics Committee, and JHU and PRISMA (Peru IRC). All necessary amendments to this original main trial study protocol are submitted on an ongoing basis. Reliance agreements with all US-based institutions are in place with the Emory IRB. All approved study documents and approval letters are available to study investigators via the web portal (Box), and maintained by the Trial Coordinating Center.

Details of the study procedures and potential risks and benefits will be explained at recruitment. Participation in the study is voluntary. A participant's decision of whether or not to participate in this study does not affect their medical care at the community health clinics and/or hospitals. Written and verbal informed consent, in the subject's native language, will be obtained from all subjects at recruitment and, for the pregnant woman, will include parental consent for the participation of her child. Consent forms will follow the format we have used in our previous studies in the population of these regions/countries, and will include a description of the study, the respondent burden, the potential risks and benefits, and names and contact information for the local PD/PI and other local responsible individuals who can be contacted for additional study details. For illiterate participants, the consent form will be read and the consent process done in the presence of a witness. All participants will receive a copy of the consent form for their personal records. Subjects will not receive compensation for their participation in the study.

Under the direct supervision of IRC PD/PIs, local Field Supervisors and Project Coordinators will monitor the study for safety and ensure that protection is assured throughout data collection. In particular, they will pay close attention to any adverse outcomes in women and children. The Trial Coordinating Center will develop a common protocol for monitoring and reporting outcomes, which will be implemented by the ICs. All local study teams will meet at least monthly and summary reports will be reported to the Trial Coordinating Center at regularly scheduled Steering Committee calls.

This study will employ a DSMB, which can request access to unblinded data, call for meetings with the study investigators (either by telephone conference or in person), and suspend the study if required at any time.

11.4. Procedures for Ensuring Confidentiality

The study will employ standard methods for protecting the confidentiality of participants through the use of personal identification numbers (PIDs) on all study materials, password-protected computer data files, and locked file cabinets for storing hard copies of interview and other study materials (e.g., signed consent forms). All study investigators will undergo training in Human Subjects Research Ethics. All key personnel will be CITI (Collaborative Institutional Training Initiative) certified per Emory IRB requirements. This training includes procedures to minimize the potential for breaches of confidentiality, including but not limited to ensuring that all files are closed, that interviews are conducted in private settings, and that no conversations about individual study participants occur in public settings.

Study subjects will be assured that information gathered during interviews and health and exposure assessments is kept confidential. Names and other easily recognizable identifiers will be collected electronically by local study staff for the purpose of following the participant during the study. . No personal information (e.g., name, address, telephone number) will leave the local IRCs or be shared with the Trial Coordinating Center. Numeric study identifiers are included so that data from the several instruments can be linked; however, these are not meaningful to casual observers without access to the original study logs. The same procedures are followed for all biological samples. The study will also obtain a Certificate of Confidentiality from the National Institutes of Health, to cover all new data collected during the study. The purpose of a Certificate of Confidentiality is to prevent anyone other than study investigators from gaining legal access to data collected for research purposes. Study subjects can thus be assured that information revealed by them during an interview or collected as part of the examination will be kept confidential and cannot be subpoenaed or obtained by legal means except with their permission.

11.5. Potential Benefits of the Proposed Research to Human Subjects and Others

All participants will benefit from receiving an LPG stove and an 18-month supply of LPG. This has been shown to reduce HAP exposure in previous studies, thus has the potential to improve health. Although expected to be infrequent, additional benefits may come from the detection of previously unknown health problems that, with treatment, will result in better future health for participants.

The proposed study will provide robust evidence on the potential health benefits of a LPG stove and gas intervention. In addition, it will provide the first comprehensive data on exposure-response relationships between HAP and key maternal, child, and adult health outcomes, including birth weight, pneumonia, growth, preterm birth, child development, maternal blood pressure during pregnancy, adult blood pressure, and several other cardiopulmonary and chronic disease measures and biomarkers. Results will inform the degree of reduction in HAP exposure needed in order to meaningfully improve neonatal and child health outcomes, which are key global health targets. Consequently, results should be useful to the global effort to reduce the health risks associated with HAP exposures due to residential combustion of biomass fuels.

12. Centralized Data Management

Data management requires focused effort and oversight by a central data center for the overall project. All sites have a strong history of involvement in detailed epidemiologic studies of home-based air pollutants, providing a strong basis for local data quality. The Trial Coordinating Center Data Management Core (DMC), jointly with site-specific data managers, will collect survey and other field data on smartphone/tablet-based platforms that incorporate GIS positioning and QR readers, allowing data collection teams to scan barcodes placed on LPG stoves, fuel tanks, exposure monitoring equipment, and biospecimens.

12.1. Data Management Standard Operating Procedures

The DMC will develop and maintain SOPs for all data management activities. These will be reviewed, approved, and documented centrally, and disseminated to all study sites with associated training as necessary. The full set of up-to-date SOPs will be available via the web portal developed and maintained by the Trial Coordinating Center. An initial standard operating procedure (SOP) will formalize the process for drafting, reviewing, approving, and revising subsequent SOPs. Once posted, the DMC will facilitate any necessary online or on-site training relating to any associated data collection, storage, transfer, and analysis protocols. Data management SOPs will be in place for the Biomarker Core, the Exposure Monitoring Core, the Clinical and Imaging Core, and the Social Sciences and Behavioral Economics Core. We note that data management SOPs are in addition to laboratory, clinical, or other procedural SOPs and will detail specific elements of data storage, entry, quality, transfer, processing, and security.

After evaluation by the Biomarker Core Quality Assurance officer, data will all be uploaded into the Biomarker Core LIMS system for permanent storage along will all relevant metadata. This database is password-protected, backed up daily and administered by a full time data base manager. Any alterations in data after uploading into the database will be documented as to why the correction was needed with full metadata. Only the database manager and Biomarker Core Co-Directors will have write access to the database other than general uploading of analytic or metadata.

Data reports with all relevant data and metadata will be made available for each of the analytic aims of the study to the relevant study investigators after consultation of exact data needs. Each data set will be maintained with the specific data analysis to ensure consistency in analytics.

12.2. Data Quality and Security

The heterogeneous nature of the study sites presents a challenge for data management and coordination. To our advantage, all sites have considerable experience in initiating, conducting, and completing large-scale epidemiologic studies and trials, and have experience in establishing and maintaining secure data collection and storage protocols and practices. The Trial Coordinating Center and Data Management Core will work closely with each site to establish an overall data management plan customized to local needs and systems but maintaining consistent standards of data quality and security. Specifically, the Trial Coordinating Center will work with on-site data managers to develop a standard database framework, shared data dictionary, and common implementation. While the main study database will be maintained in REDCap (for forms) and the study LIMS (for laboratory data), due to limited internet connectivity at some study sites, we may implement a local instance of the study database (e.g., REDCap, MS Access), as necessary, on a secure server at each site, with data transfer and QA/QC protocols and SOPs in place. Household data will be collected via approved case report forms, implemented on password-protected tablets where possible, and on equivalent paper forms where not.

Data QA/QC protocols will include structured double-data entry and regular entry error review. The Trial Coordinating Center and Data Management Core will work with local sites to establish, document, and implement the relevant data collection SOPs, data QA/QC SOPs, data storage and access SOPs, and data transfer SOPs. All ultrasound image data will be transferred (via secure protocol) to Johns Hopkins University and managed, analyzed, and disseminated to co-investigators according to a specified ultrasound image data SOP via Trice. All other data will be encrypted and transferred from the site servers to Emory University via secure protocol (e.g., direct data entry via REDCap, or secure FTP, via remote, password-protected login to local secure servers, or cloud-based options) where data will be compiled into an overall SQL relational database. Cloud solutions hold great promise, (e.g., Amazon Web Services) and they are under rapid development at Emory University and elsewhere to provide viable, secure data storage but they will require all sites to have reliable access to the systems and will likely only provide part of the overall solution, at least for the first several years. Data transfer protocols will be automated to provide regular (e.g., daily, weekly) transfers of study data to the Trial Coordinating Center thereby minimizing the impact of any local data loss at each site.

12.3. Analytic Pipelines

For each report, manuscript, or presentation associated with the study, the data management team will extract and curate an associated analytic data set (and associated code) with a unique identifier. Linking a single analytic data set to each research product ensures reliability and reproducibility of results and allows for clear and clean updates to the data set. The archived analytic “pipeline” will include the analytic dataset, the database extraction code, any preprocessing code (SQL/SAS/R/etc.), and the final analysis code associated with each report, manuscript, or presentation.

12.4. Web Portal

To enable study communication and disseminate findings and results, the Trial Coordinating Center will establish a web portal allowing secure online communication between study investigators. The portal will also include public information on the overall project and serve as a repository for study output in the form of reports, manuscripts, and presentations. In addition, the portal will have a password-protected component providing a discussion forum, collaborative workspace, adverse event reporting, an analytic dataset repository, and administrative documentation (e.g., SOPs, recruitment reports, etc.).

13. Data Analysis Plan—Trial Period

This section summarizes the proposed data analysis plans for the trial period. More detailed statistical analysis plans (SAPs), both intention-to-treat and exposure response, will be pre-specified prior to unblinding the corresponding health endpoints by study arm assignment. For the primary outcomes only, these SAPs will be published in advance and any amendments or changes documented. These final SAPs for the primary outcomes are included in Appendix 3 (3.A: Birthweight, intention-to-treat (ITT) and exposure-response; 3.B.: Stunting, ITT; 3.C.: Pneumonia, ITT; and 3.D: Blood pressure, Older adult women, exposure-response; and 3.E.: Blood pressure, older adult women, ITT).

For the extended follow up period, SAPs shall also be pre-specified.

13.1. Intention-to-treat analysis for primary outcomes (Aim 1)

For Aim 1, we will perform two-sided tests for each primary outcome separately at an α-level of 0.0125. The Bonferroni correction for multiple testing, while conservative, controls for family-wise type I error rate to be 0.05 under any dependence structure among the four primary outcomes. The randomization should preclude the need to adjust for any confounders. However, we will compare potential confounders between the intervention and control arms and adjust for potential confounders as appropriate, if needed.

Birth weight will be treated as a continuous outcome and we will carry out a two-sample t-test with the following test statistic under the null hypothesis of no difference in mean birth weight:

$$T_1 = \frac{\bar{y}_2 - \bar{y}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\bar{y}_j =$ sample mean for the control group (j = 1) and the intervention group (j = 2)
$s_j^2 =$ sample variance for the control group (j = 1) and the intervention group (j = 2)
$n_j =$ sample size for the control group (j = 1) and the intervention group (j = 2).

After excluding any women in blood pressure medication at baseline or at any time during follow up, which we anticipate at approximately 5%, repeated measurements of blood pressure of the older adult woman will be analysed using a model to examine differences in mean systolic blood pressure post-randomization, while adjusting for baseline systolic BP. Specifically, let y_i denote the blood pressure for woman i. The regression model is given by

$$\bar{y}_i^{Post} = \beta_0 + \beta_1 X_i + \beta_2\left(y_i^{Base} - \bar{y}^{Base}\right) + \varepsilon_i$$

where for individual i, $\bar{y}_i^{Post}$ is the average post-randomization blood pressure measurements, X_i is an indicator variable (0 for control and 1 for intervention), $\left(y_i^{Base} - \bar{y}^{Base}\right)$ is the centered baseline measurement of systolic blood pressure, and $\varepsilon_i \sim N(0, \tau_i^2)$ represents independent Normal error. The parameter of interest β_1 captures differences in average blood pressure due to the intervention, accounting for baseline systolic blood pressure.

The variance τ_i^2 is a function of the number of post-randomization measurements for individual i and the between-measurement correlations. Specifically, τ_i^2 is given by

$$\tau_i^2 = \frac{\sigma^2}{m_i}\left[1 + (m_i - 1)p - m_i\theta^2\right]$$

where m_i is the number of post-randomization measurements, σ^2 is the cross-sectional between-individual variance of systolic blood pressure at any time point, p is the intra-class correlation of systolic blood pressure post-randomization and θ is the correlation between baseline and post-randomization measurements. We note that p and θ may be different between the control and intervention groups. Given estimates of p and θ, we can estimate β_1 , for example, using weighted least-square to account for different variances due to the differing number of repeated measurements across individuals and correlations. A Wald's test will be conducted under the null hypothesis of $\beta_1 = 0$.

For the binary primary outcome of stunting at 12 months of age, we will carry out a test for relative risk between control and interventional group with the following test statistic under the null hypothesis that the relative risk is zero:

$$T_2 = \frac{\log \widehat{RR}}{\sqrt{\frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{n_1} - \frac{1}{n_2}}}$$

where $x_j =$ number of cases for the control group (j = 1) and the intervention group (j = 2)
$n_j =$ sample size for the control group (j = 1) and the intervention group (j = 2)
$\widehat{RR} = (x_2/n_2)\,/(x_1/n_1)$.

For the primary outcome of the incidence of HAPIN pneumonia episodes (see Section 5.2.1) over a one-year follow-up period, we will use a Poisson regression model based on child-days at risk of follow up, to derive an incidence rate ratio for the intervention vs. control groups. The use of Poisson regression and rate ratios, rather than a relative risk of first occurrence of pneumonia, is motivated by observing that the occurrence of multiple severe pneumonias episodes in the study children, which occurs often enough to warrant taking the multiple events into account. We will account for multiple severe pneumonias episodes by estimating a dispersion parameter using quasi-likelihood estimation.

For a trial of this size, even though covariate imbalance observed post hoc will be most likely due to a random phenomenon, the primary analyses will include indicator variables to account for the stratification scheme, i.e., the stratification of random sampling with different geographical areas in India and Peru, as well as individual IRC sites.

Adjustment for baseline covariates known to be strongly associated with outcomes has the potential to increase power. However, the research team has concluded that the challenge to pre-specifying a limited set of covariates and the functional forms of their associations with outcomes may introduce additional uncertainty in this trial. Hence no covariate-adjusted effects will be estimated for the primary analyses, except for blood pressure where baseline measurements will be used to account for within-individual correlations.

We also plan to conduct analyses by subpopulation by specifying stratified/interaction analyses by variables like age (endpoints in older adult women), BMI (endpoints in older adult women), infant sex (birth weight, stunting).

13.2. Study power for primary outcomes

Note that all power calculations for the primary outcomes include the Bonferroni correction. The power calculations for the primary intent-to-treat analysis do not account for any potential confounders if they are deemed to be necessary.

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, 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.

For power calculation for blood pressure, we assume $m_i = m = 5$ and $p = \theta = 0.76$ for all individuals. The power for the hypothesis test for the regression slope β_1 for n individuals per arm can be approximated by (Frison and Pocock 1992):

$$\Phi\left(-Z_{1-\alpha/2}+\frac{|\beta_1|}{\sqrt{\frac{2\sigma^2}{nm}[1+(m-1)p-mp^2]}}\right).$$

The above approximation assumes that β_0 and β_2 are known (for large trials, these parameters can be estimated with good precision), and the coefficient β_2 applies to both the intervention and control group (i.e. θ is identical between groups).

The power for the hypothesis test for relative risk of pneumonia and stunting is 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 p_1 is the risk of pneumonia/stunting among the control group, p_2 is the risk of pneumonia/stunting among the intervention group, RR is the true relative risk (i.e. p_2/p_1), and n is the common sample size for each of the treatment group and the control group. For pneumonia, power for Poisson regression is also approximated by the equation for relative risk above assuming that follow-up times are the same for all participants.

The following assumptions, based on previous studies, are made for power calculations.

- For birth weight, we assume a standard deviation $\sigma = 437$ [Thompson et al. 2011]
- For blood pressure, we assume a standard deviation $\sigma = 12$ (unadjusted for baseline) [McCracken et al. 2007] and an intraclass correlation of $\rho = 0.30$.
- For pneumonia, we assume a 1-year cumulative incidence rate among the control group as $p_1 = 0.06$ [a weighted average from Mackenzie et al. 2014, Gupta et al. 2010, Mortimer et al. 2007, Broor et al. 2017, Farooqi et al. 2015]. The actual incidence rate accounting for multiple events over the entire follow-up period is expected to be higher.
- For stunting, we assume a risk among the control group of $p_1 = 0.30$ [Danaei et al. 2016].

We assume a 10% attribution over the 1-year follow-up period, resulting in an effective sample size of $n = 1440$ per treatment arm for birth weight, stunting, and pneumonia. For blood pressure we assume an effective

sample size of $n = 340$ per treatment arm due to an additional 5% exclusion of women who will be on medication at recruitment or during the study period. Each woman will provide $m = 5$ measurements. Given the above sample sizes, estimated power as a function of detectable effects are given in Figure 9.

Figure 9. Power as a function of effect size for the four primary outcomes given planned sample sizes (α = 0.0125)

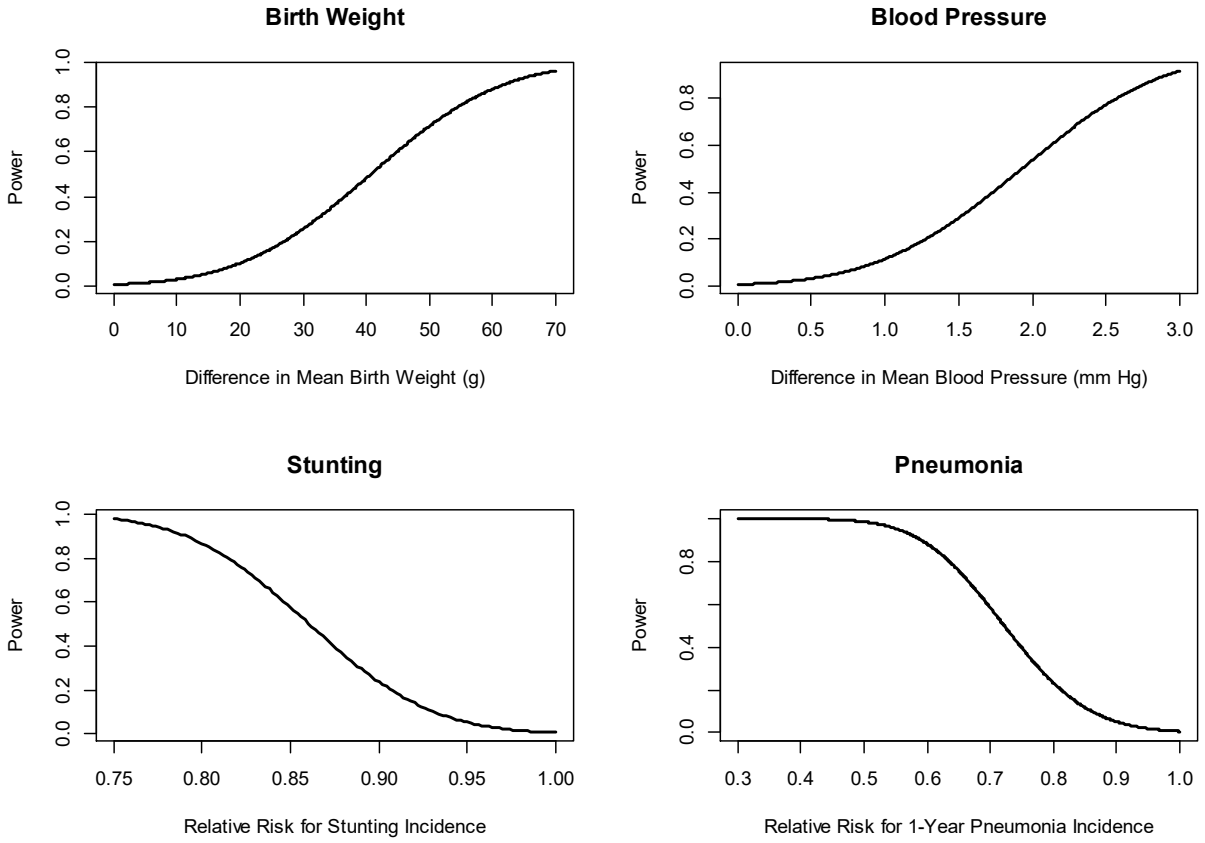


Table 9 summarizes the minimal detectable difference in mean (for birth weight and blood pressure) and minimal detectable relative risk (for stunting and pneumonia) associated with an 80% power and a type I error rate α of 0.0125. We also examine the impact of power when key population parameters are set to $\pm 10\%$ and $\pm 20\%$ of their values assumed above. Note that for blood pressure we show power calculations for only 200 OAW per arm, although our estimated OAWs will number 240 per arm (120 per IRC, 60 in each arm). Initial recruitment data suggests that the number of OAWs may be less than what we originally assumed (15% vs 25%). Nonetheless, with that sample size, Table 8 shows we have good power to detect a smaller difference than has been previously reported in the literature.

Table 9. Minimum Detectable Effects with 80% Power and α = 0.0125.*

	Parameter	Sample Size Per Arm	-20%	-10%	Original	+10%	+20%	Previous Studies Estimate (95% CI)
Change in mean birth weight (g)	σ^2	1600	43	49	54	60	65	89 (-27, 204)[a]
Change in mean systolic blood pressure (mm Hg)	σ^2	200	2.30	2.44	2.58	2.70	2.82	3.7 (-0.6, 8.1)[b]
Relative risk for stunting	p_1	1600	0.79	0.80	0.81	0.83	0.84	0.79 (0.70-0.89)[c]
Relative risk for HAPIN severe pneumonia	p_1	1600	0.60	0.62	0.64	0.66	0.67	0.67 (0.45, 0.98)[d]

*We examined the impact of power when key population parameters (variance of continuous measures σ^2, incidence rates among controls p_1 for relative risks) are set to ±10% and ±20% of their values. Calculations assume a 10% attrition during follow-up.
[a]Thompson et al. (2011) is a sub-study of a randomized trial of improved cookstoves in Guatemala (RESPIRE), in which those using stoves were compared to those not using stoves after adjusting for confounders.
[b]McCracken et al. (2007) is a study of 120 older women (>38) from the same randomized trial (RESPIRE), which adjusted for confounders.
[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 s.d.).
[d]Smith et al. (2011), from the RESPIRE randomized trial, provides the estimated RR of 0.67, the RR for clinician diagnosed severe pneumonia for children under 18 months of age. Our estimated background rate of 0.09 for controls is based on our observed severe pneumonia rate in both arms (treatment and control) in the HAPIN study (using the HAPIN severe pneumonia definition), with 20% of the child person time observed, and assuming the RESPIRE RR of 0.67 for intervention vs. controls.
For a fuller description of the literature and the choice of three of these effect estimates (birthweight, blood pressure, infant pneumonia), see Steenland (2018).

Previously reported effects of cookstove intervention are also provided. We note that the current trial, with the use of LPG, is expected to result in a larger decrease in exposure and in lower post-intervention PM concentrations, than these previous studies; hence, we expect a stronger true effect size than those reported previously.

13.3. Exposure-Response Analyses (Aim 2) for Primary Outcomes

For Aim 2, exposure-response associations (using measured personal $PM_{2.5}$ and CO concentrations for pregnant mother / older adult women / child, as appropriate for each primary and secondary outcome) will be estimated within the intervention and the control groups separately, as well as in a combined analysis. Potential confounders will include, but not limited to, age, sex, body mass index, socioeconomic status, education, tobacco use, secondhand smoke, physical activity, nutrition, and season. For each outcome we determine potential confounders using previous literature and directed acyclic graph (DAG) methods. We will determine final models by evaluating change in effect estimates for meaningful changes. We will present crude and adjusted estimates and 95% confidence intervals.

The general regression model is given by

$$g(E[Y_i]) = \beta_0 + \beta_1 X_i + \boldsymbol{\gamma Z}_i$$

where Y_i is the primary outcome of interest, $g(\,)$ is the appropriate link function (identity for birth weight and blood pressure, logistic for stunting and pneumonia), X_i is the continuous exposure of interest, and $\boldsymbol{Z}_i$ is the vector of confounders. The set of confounders will be chosen differently for each outcome and sensitivity analyses will be conducted to examine the robustness of the parameter of interest β_1 against different confounder controls. For repeated measurements of blood pressure an individual-level random intercept will be included.

The exposure of interest X_i will also depend on the outcome. For birth weight, a gestational exposure will be obtained by averaging the 2 or 3 measurements during pregnancy. For stunting and pneumonia, we will consider average gestational and average first-year of life exposures. For systolic blood pressure, we will consider time-varying exposures, available at the time of blood pressure measurement.

13.4. Secondary and Sensitivity Analyses for the Primary Outcomes

We will examine several secondary analyses for the primary outcomes. The following sensitivity analyses will be conducted:

1. We will evaluate the first episode of pneumonia over a one-year follow-up period.
2. For pneumonia we will also plan to evaluate 7 alternative pre-specified case definitions as secondary outcomes including: (1) WHO Integrated Management of Childhood Illnesses (IMCI) non-severe pneumonia, (2) WHO IMCI severe pneumonia, (3) Pocketbook WHO non-severe pneumonia, (4) Pocketbook WHO severe pneumonia, (5) Hospitalization for respiratory illness, (6) Lung ultrasound or chest radiograph pneumonia, (7) hypoxemic pneumonia.
3. We will apply recent advances in causal inference to estimate the intervention's effect due to the reduction in air pollution level via a principal stratification approach. This approach overcomes the challenge of the pollutant concentration being a post-randomization variable.
4. For blood pressure, we will additionally account for potential use of blood pressure medication at each measurement via the method described by Tobin et al. 2005 (method j, censored normal). These authors have shown via simulation that exclusion of those on medication leads to a bias towards the null. Their method (j) assumes that those on medication have a true underlying blood pressure greater than some minimum, generally the level above which medication is typically given.
5. We will examine difference in BP change as a function of time (i.e. linear divergent intervention effect) using the same approach that controls for baseline measurements.
6. Non-linear associations between air pollution concentration and health outcomes will be examined using generalized additive models and other spline-based approaches (Aim 2).
7. We will remove participants whose cotinine measurements indicate exposure to secondhand smoke and/or active smoking.
8. We will conduct analyses removing identified extreme values (top and bottom 1% and 5%) of exposure and health measurements (for continuous measurements).
9. We will remove participants with known substantial exposure to sources of air pollution exposure other than the cooking stove in the home. We will consider marginal structural models as a way for adjusting for drop-out/non-compliance.

11. For birth weight, additional analyses will include (1) restricting to full-term births only, (2) adjusting by gestational weeks, and (3) using z-scores for birth weight derived from INTERGROWTH tables (https://intergrowth21.tghn.org).

Sub-group / interaction analyses will be conducted by IRC, age (endpoints in older adult women), BMI (endpoints in older adult women), infant sex (birth weight, stunting).

13.5. Analysis of secondary outcomes

Analyses will be conducted for all binary and continuous secondary outcomes using similar methods as the primary outcomes without the multiple-testing adjustment of type I error rate.

Random effect models will be used to analyze for most repeated outcome measurements (baseline and post-intervention) to account for within-subject correlation Intervention effect on average difference will be examined, child development, and most biomarkers. Intervention effect on the rate of change (as a function of age or time post-randomization) will be examined for fetal growth and childhood linear growth. Analyses of maternal blood pressure, blood pressure among new mothers, and secondary blood pressure outcomes in the older adult women will be conducted analogously to the analysis of systolic blood pressure in older women described above.

Sub-group analysis, exposure-response function, and their associated sensitivity analyses will follow similar modeling approaches as described for the primary outcomes.

13.6. Missing Data

13.6.1. General analytic strategy

We will record the number, timing, pattern, and reason for missing data. Missing data may occur due to subject-initiated drop out, investigator-initiated withdrawal, missing visits, mechanical failure and input data error. We will evaluate whether the missing data patterns are differential between intervention and control groups, as well as evaluate the temporal pattern using Kaplan-Meier curves. For all analyses, we will assume missing data to be missing at random (MAR) a priori. The MAR mechanism specifies that the complete data distribution can be modeled using the observed data.

For the intention-to-treat analyses, where the outcome is only measured at a single time point inverse probability weighting will be employed to account for missing outcome data. Specifically, a logistic regression will be fit to associate an indicator variable of the missing outcome variable with a set of predictors. Inverse of the predicted probabilities from this model will be used as weights in estimating intervention effect. Time-varying weights will be used for the pneumonia incidence rate to account for varying follow-up time until drop-out/withdrawal. We will also consider last-value-carry-forward using blood pressure measured at 3-month or 9-month post randomization if strong within-subject correlation is detected among participants.

For the analysis of repeated measurements, the use of random-effect models that incorporate subject-specific random effects will account for both outcomes and covariates (e.g. air pollution exposures and maternal BMI) under the MAR mechanism. We will also consider sensitivity analyses for handling missing data using multiple imputation. Multiple imputation aims to directly impute missing outcomes/covariates and, importantly, allows incorporation of additional uncertainty of the imputation model in our effect estimates.

In some instances, data may not be missing at random. For example, still birth and early neonatal death can be considered a mechanism for missing outcome data that's not missing at random, since there may be more stillbirths among the control group. Formal causal inference is challenging due the fact that the primary outcome, birth weight, is not defined if early death occurs. While stillbirths are rather rare (25/1000 births in LMIC countries), and would not be expected to greatly influence our analysis of birthweight, we will address the issue of not missing at random using in secondary analyses using the 'survival average causal effect' (SACE) (Tchetgen et al. 2015). This method adjusts for the propensity of still birth/early deaths as a function of intervention status and baseline covariates.

13.6.2 Assessing the impact of COVID-19 on data completeness and revised strategies for analysis in the presence of missing data.

Overall, the HAPIN trial has experienced excellent adherence and very little loss-to-follow up (as summarized in regular reports to the Data Safety Monitoring Board (DSMB)). However, the COVID-19 pandemic had clear impacts on access to study participants across all four HAPIN IRCs. Where circumstances and nature of measurement allowed, data elements will be collected via alternative mechanisms (e.g., phone interviews to complete case report forms (CRFs), the use of hospital records of birthweight (taken by HAPIN-trained hospital staff on HAPIN scales)). Where in-person visits are required (e.g., administration of the Caregiver Recorded Early Development Instrument, exposure measurements, blood pressure, anthropometry), data are documented as missing due to COVID-19. /All alternative measurement approaches are documented and reviewed by the HAPIN Steering Committee as part of its detailed COVID-19 Risk Assessment of activities. Protocols for such alternative measurements are documented and stored.

The analytic strategies detailed in 13.6.1 provide a template for managing missing data in general for the HAPIN study. However, the specific pandemic patterns differed between IRCs, the ubiquitous nature of the pandemic resulted in unanticipated amounts and types of missing data necessitating the following updated details. Recall that the primary outcomes of interest include blood pressure for the mother and older adult women, birthweight, pneumonia, and 12-month stunting for the children. Our primary goal is to assess and mitigate any impact of missing data on the two main summaries of primary effects of treatment, namely: the intent-to-treat analyses, and the exposure-response analyses. Our proposed strategies aim to maintain study validity (Will affect estimates remain unbiased in the presence of missing data?) and study power (Will the remaining sample size be sufficient to reliably detect anticipated effect levels?).

With respect to study validity, our original design relies on randomized treatment assignment to minimize potential confounding factors. While specific missingness patterns may differ between IRCs, data missing due to the pandemic are unlikely to be differentially distributed between treatment arms, resulting in nondifferential missingness across arms. This should reduce the number of observations but should not do so in a manner such as to induce bias in either direction. This said, while we do not anticipate bias, in order to fully assess this situation, we will document indicators of missing elements to allow calculation and comparison of rates and types of missingness impacted by COVID-19 by study arm within each IRC.

To assess validity of the ITT analyses, we will conduct sensitivity analyses comparing effect estimates prior to COVID-related lockdowns (beginning approximately mid-March 2020 in each of the four IRCs) to results obtained including pre- and post-COVID observations. While such comparisons will likely not be powered to assess statistical significance in differences, the summaries will provide an assessment of the stability of estimates compared to the uncertainty associated with the estimates. Summaries of missing data rates by IRC, and the results of these sensitivity analyses will be included in primary manuscripts to address potential concerns regarding COVID-related impacts on study results.

To assess validity of the exposure-response studies, we will assess data completeness for estimated exposure trends across study visits, documenting changes in estimate uncertainties associated with anticipated reduced precision during COVID lockdowns due to increased missingness. Indicators of pre-COVID and COVID period

observations will allow estimation of variation of effect estimates by lockdown periods. For exposure-response functions, uncertainty bounds will account for reduced sample sizes associated with any concentrations of missingness for certain ranges of exposures. Should questions of bias arise, these descriptive data allow calculation of potential adjustments via expansion of the inverse probability weighting techniques defined above. With respect to missing due to COVID-19, the probability of missingness due to COVID-19 will, in most cases, involve logistic regressions with time-varying effects of covariates interacting with COVID-19 lockdown periods.

With respect to study power for primary outcomes, for any interim power estimates prior to the end of data collection, HAPIN investigators will monitor whether the increases in data missingness due to COVID-19 lockdowns and access adjustments result in reduced power to detect the anticipated effect sizes from the original study design. These comparisons mirror the original ~~simulation-based~~ power analyses defining the overall HAPIN study design and sample size targets. These results maintain the same anticipated effect sizes (based on the literature at the time of the HAPIN design), but update the primary study outcomes to reflect the new reduced anticipated sample sizes due to missing data during COVID, as well as the pooled exposure and outcome distributions observed in the HAPIN IRC populations to date. Initial results from these analyses (given aggregate data accrued at that time) were presented to the DSMB in April 2020 and an updated set in October 2020. These “conditional power analyses” provide a tool to document any power loss due to further reduced sample sizes (e.g., due to COVID-19 resurgence or related access limitations) for each primary outcome and its externally-defined effect sizes.

13.7. Economic Evaluation

Economic evaluation will involve the assessment of program costs (administration, training, and implementation), health system costs (due to medical attention), and costs assumed by the households (health expenditures, costs related to the use of the stoves, time savings and productivity gains). Data will be collected using project expenses reports, interviews with project coordinators, and household surveys. Analysis will take into account QALYs from the primary and secondary health outcomes.

13.8. Implementation Fidelity

Measures of implementation fidelity will include whether: (1) data quality is assured according to protocol; (2) recruitment takes place and the target sample size is achieved and retention maintained as scheduled at each project site; (3) stove use monitoring and personal and household air quality monitoring are carried out per protocol; (4) household visits are happening with the intended frequency; (5) exposure and outcome assessments occur according to common protocol. Once a behavior change framework and behavior change interventions have been developed in accordance with formative research findings, appropriate measures of fidelity will be defined and collected for them as well. The primary data source for these fidelity measurements will be project surveillance data, supplemented with review of minutes from monthly Trial Coordinating Center and Steering Committee conference calls and the Steering Committee annual meeting. Additional data will come from key informant interviews with local research team members, direct observations of visits to participant households, and questionnaires to be applied annually among a small sample of participant households. A detailed implementation research protocol will be developed based on the stages defined by Steckler.

13.9. Collection of additional contact information at the end of the trial

Since the participants in this trial form such a valuable cohort, there may be interest in and opportunities to acquire additional research funding towards the end of the trial or even years later. We will therefore ask participants for additional contact information when they exit the study, including the contact information of up to 3 persons who are most likely to know how to contact the participant in 5 to 10 years' time.

14. Data Analysis Plan—Extended Follow Up Period

14.1 Statistical analysis plan for Aim 1 (Intention-to-Treat Analysis)

MDAT: The ITT analysis for the MDAT assessment will estimate intervention effects on mean longitudinal composite MDAT scores at ages 3, 4 and 5 using a random-intercept model with adjustments for age, randomization strata, and within-child clustering. For child *i* at age *j,* we assume $y_{ij} = \alpha_j + \beta \mathrm{X_i} + \sum_p \gamma_p Z_{i,p} + \theta_{0i} + \epsilon_{ij}$, where y_{ij} is the MDAT score, α_j is the average score at age *j* for the control group, $\mathrm{X_i}$ is an indicator for intervention versus control, $Z_{i,p}$ is the indicator for the pth stratum (Guatemala, Rwanda and 2 sites in India), θ_{oi} denotes child-specific Normal random intercept and ϵ_{ij} is the residual error. Hence, the ITT effect β represents a constant change in mean score averaged across the years. Secondary analyses will include (1) effect modification by infant sex and timing of LPG intervention (<14 weeks, >=14 weeks gestation), (2) stratification by study site, (3) analysis of specific domains, and (4) evaluation of age-specific ITT effect.

Growth: The ITT analysis for child growth will focus on linear growth (height-for-age z-score, HAZ) through age 5 (at birth and ages 1-5), with an interaction between study arm and age, allowing us to utilize the comprehensive longitudinal data to explore patterns over time that would not be captured by focusing on height at one age. We will also evaluate HAZ at age 5. The repeated measures approach will estimate intervention effects on linear growth rate of six annual length/height-for-age z-score (HAZ) between birth to age 5. For child *i* at age *j,* we assume $y_{ij} = \alpha_0 + \theta_{oi} + \alpha_1 X_i + (\alpha_2 + \theta_{1i} + \beta X_i)\,\mathrm{Age_i} + \sum_p \gamma_p Z_{i,p} + \epsilon_{ij}$, where y_{ij} is the HAZ value calculated using the WHO child growth standards reference median for age and sex, α_0 and α_2 are the intercept and slope for the overall linear trend, θ_{oi} and θ_{1i} are child-specific random intercepts and slopes. Hence the parameter of interest, β, describes the difference in average linear slope between intervention and control groups. Secondary analyses will include (1) effect modifications by infant sex and timing of LPG intervention (<14 weeks, >=14 weeks gestation), (2) stratification by site, and (3) consideration for non-linear trajectory as a function of age (e.g., piecewise linear splines or polynomial functions), as for child growth above. [84], [85], [86]. Given the randomized design and balance between study arms we do not plan to adjust for confounders for the ITT analyses in Aim 1; however, in secondary analyses we will evaluate the need to adjust for confounders.

14.2 Aim 2 Statistical Approach

The ITT analysis for Aim 2, evaluating the longer-term effect of the LPG intervention on personal exposure to $PM_{2.5}$ and black carbon, will estimate the intervention effects on mean exposure measurements (mean of exposure measurements at ages 3, 4, and 5 years) using a 2-sample t-test, log transforming as necessary. Secondary analyses will include (1) stratification by study site and (2) evaluation of age-specific ITT effect.

14.3 Exposure-Response Analysis (Aim 3)

We will use the same outcome definitions for MDAT and child growth as in the ITT analyses. A random-intercept model will be used to evaluate longitudinal associations between mean 24-hour $PM_{2.5}$ and black carbon exposures across critical periods of child development and growth. For exposure, we will evaluate: 1) gestational exposure - mother's 24-hour average personal exposure to $PM_{2.5}$ and black carbon averaged across pregnancy (3 measurements during pregnancy – pre-intervention, 18- 26 weeks gestation, 26-30 weeks gestation), and 2) child exposure after birth averaged through the outcome measurement (e.g., for height at age 5 we will average child exposure measurements at baseline 3 months, 6 months, and at ages 1-5 years). We will evaluate $PM_{2.5}$ and black carbon in separate models.

Each critical window of exposure will be examined separately. Expanding upon the models from Aim 1, we assume $y_{ij} = \alpha_0 + \theta_{oi} + (\alpha_2 + \theta_{1i})\,\mathrm{Age_i} + \beta\,\mathrm{PM}_{ij} + \sum_p \gamma_p Z_{i,p} + \sum_r \kappa_r W_{ij,r} + \epsilon_{ij}$, where PM_{ij} represent time-varying exposure (i.e., age-specific average exposure or cumulative average exposure). PM_{ij} can also be replaced with PM_i to represent exposure over a fixed window (i.e., gestational average and first year of life). The model also includes a set of confounders $W_{i,r}$. Potential confounders will be identified from the literature, as well as using directed acyclic graphs (DAGs). Potential confounders from the literature (see C.7) include infant sex, maternal height, socioeconomic status (maternal education, number of people in house, marital status), nulliparity, breastfeeding, dietary diversity, and household food insecurity. For the MDAT outcomes, we will also include the family care indicators and maternal stress. Gestational age at birth, birth weight, respiratory and diarrheal infections are likely mediators and will not be adjusted for, but we will collect these outcomes as well. Secondary analyses will include (1) consideration for nonlinear associations via regression splines, (2) interactions between $PM_{2.5}$ exposures and linear trend in growth, and (3) joint analysis of gestation and child exposures using methods that can accommodate temporal autocorrelation in exposures[87]. We will evaluate whether missing data patterns are differential between intervention and control groups, as well as the temporal pattern using Kaplan-Meier curves. For ITT and exposure-response analyses, the use of random effect models is protected against data that are missing at random, which specifies that the complete data distribution can be modeled using the observed data. We will also consider handling missing data using multiple imputation to incorporate imputation uncertainty and address potential not missing at random by estimating the survival average causal effect (SACE) that accounts for outcome-dependent attribution.

14.4. Power Considerations

Table 10 gives minimum detectable effects with 80% power for the primary analyses in Aims 1 and 3 using random effect models. Power was calculated using standard errors of regression coefficients derived from asymptotic theory and meta-regression formulation of random-slope models. The loss to-follow-up rate was assumed to be of 5% per year starting at year 2 (using actual numbers through age 1), with at most 3 repeated measurements for MDAT (ages 3, 4, 5), and 6 measurements for length/height-for-age Z score (HAZ) within each child (birth, ages 1-5). MDAT and HAZ values at each time point were assumed to have a residual variance of 1 after adjustment for randomization stratification, and a high intraclass correlation of 0.50. For modeling MDAT and HAZ linear trajectory, we further assume a heterogeneity standard deviation of linear random slopes of 0.01, which corresponds to 25% of the detectable main effect. For the exposure-response analyses, we assumed $PM_{2.5}$ exposures are time-invariant and follow a normal distribution. We further accounted for the presence of confounders in the model using the variance inflation method, assuming that confounders will explain R^2= 50% of the variation in $PM_{2.5}$ exposures, which is a conservative estimate. Although we have little to compare in terms of expected estimates for Aims 1 and 3, the minimum detectable estimates are for quite small shifts in these outcomes and, therefore, we have reasonable power for the proposed work. The ITT effect sizes here are similar to other large trials, on the order of 0.1 to 0.2, such as the Bangladesh trial of water and sanitation[89] and recent publications in household air pollution.

Table 10: Minimum Detectable Effect with 80% power and α=0.05

	Effect	All Sites	Single Site
MDAT[3]	ITT mean diff (ages 3, 4, 5)	0.102	0.176
	E-R slope[2] (ages 3, 4, 5)	0.073	0.132
HAZ[3]			
	ITT mean diff at age 5	0.130	0.226
	ITT diff in time trend[1] (ages 0 to 5)	0.029	0.050
	ITT diff in time trend[1] (ages 0 to 2)	0.072	0.125
	E-R slope[2] (ages 0 to 5)	0.057	0.100

[1] Linear time trend slope is per year in age; [2]E-R defined as change in score per standard deviation change in $PM_{2.5}$ exposure. [3]All outcomes are standardized.

15. Timeline

15.1 Trial period

As shown in **Table 9**, our study began with a planning and formative research phase in Year 1. Enrollment began in Year 2.

Table 9. Study timeline for randomized field intervention trial.

Tasks	Study Year																			
	1 *09/01/16-06/30/17*			**2** *07/01/17-06/30/18*				**3** *07/01/18-06/30/19*				**4** *07/01/19-06/30/20*				**5** *07/01/20-06/30/21*			
Formative research		X	X	X	X															
Training in SOPs, certification		X	X	X				X				X				X				
Annual study meetings	X				X				X				X				X			
Parallel pneumonia surveillance							X	X	X	X	X	X	X	X	X	X	X	X	X	X
Recruitment/enrollment							X	X	X	X	X	X	X	X						
Randomization							X	X	X	X	X	X	X	X						
Intervention delivery (*Int arm)*							X	X	X	X	X	X	X	X	X	X	X	X	X	X
Data collection (incl. QA/QC)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Data entry/mgt (incl. QA/QC)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ancillary study proposals								X	X	X	X	X	X	X	X					
Presentations and publications									X	X	X	X	X	X	X	X	X	X	X	X

15.2 Extended follow up period

As shown in **Table 10**, we will continue to follow HAPIN participants in Guatemala, India and Rwanda until the child's fifth birthday.

Table 10. Study timeline for extended follow-up period

Tasks	**R56** *07/01/21-09/15/22*				**Year 1** *09/16/22-06/30/23*				**Year 2** *07/01/23-06/30/24*				**Year 3** *07/01/24-06/30/25*				**Year 4** *07/01/25-06/30/26*			
IRB approvals, SOP development and training	X				X				X				X				X			
Develop / pilot MDAT materials	X	X	X	X	X															
24-month visits	X	X	X	X																
36-month visits				X	X	X	X	X												
48-month visits					X	X	X	X	X	X	X	X								
60-month visits										X	X	X	X	X	X	X				
Data management and cleaning	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Data analysis								X	X	X	X	X	X	X	X	X	X	X	X	X
Presentations and publications													X	X	X	X	X	X	X	X

References

ACOG Practice Bulletin Number 145: Antepartum Fetal Surveillance. Obstet Gynecol. 2014; 124:182–92

Ahmed S, Mitra DK, Nair H, Cunningham S, Khan AM, Islam AA, et al. Digital auscultation as a novel childhood pneumonia diagnostic tool for community clinics in Sylhet, Bangladesh: protocol for a cross-sectional study. BMJ Open. febrero de 2022;12(2):e059630.

Bedetti G, Gargani L, Corbisiero a, et al. Evaluation of ultrasound lung comets by hand-held echocardiography. Cardiovasc Ultrasound. 2006;4:34.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J. Health Econ. 2002;21(2):271-292.

Broor S, Parveen S, Bharaj P, Prasad VS, Srinivasulu KN, Sumanth KM, Kapoor SK, Fowler K, Sullender WM. A prospective three-year cohort study of the epidemiology and virology of acute respiratory infections of children in rural India. PLoS One. 2007 Jun 6;2(6):e491.

Bruce N, Weber M, Arana B, Diaz A, Jenny A, Thompson L, McCracken J, Dherani M, Juarez D, Ordonez S, Klein R, Smith KR. Pneumonia case-finding in the RESPIRE Guatemala indoor air pollution trial: standardizing methods for resource-poor settings. Bull World Health Organ. 2007 Jul;85(7):535-44.

Bruce NG, Dherani MK, Das JK, Balakrishnan K, Adair-Rohani H, Bhutta ZA, Pope D. Control of household air pollution for child survival: estimates for intervention impacts. BMC Public Health. 2013;13 Suppl3:S8. Doi: 10.1186/1471-2458-13-S3-S8.

Chavez M, Shams N, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. Respir Res. 2014;15(1):5

Childhood Acute Illness and Nutrition Network. Childhood Acute Illness and Nutrition (CHAIN) Network: a protocol for a multi-site prospective cohort study to identify modifiable risk factors for mortality among acutely ill children in Africa and Asia. *BMJ Open*. 2019;9(5):e028454. doi:10.1136/bmjopen-2018-028454

Clark et al. A Baseline Evaluation of Traditional Cook Stove Smoke Exposures and Indicators of Cardiovascular and Respiratory Health among Nicaraguan Women. Int J Occup Environ Health 2011;17:113–21.

Corretti MC, Anderson TJ, Benjamin EJ, et al. Guidelines for the ultrasound assessment of endothelial-dependent flow-mediated vasodilation of the brachial artery: a report of the International Brachial Artery Reactivity Task Force. J Am Coll Cardiol. 2002;39(2):257-265.

Crouse JR, Craven TE, Hagaman AP, Bond MG. Association of coronary disease with segment-specific intimal-medial thickening of the extracranial carotid artery. Circulation. 1995;92(5):1141-1147.

Danaei G, Andrews KG, Sudfeld CR, Fink G, McCoy DC, Peet E, Sania A, Smith Fawzi MC, Ezzati M, Fawzi WW (2016). Risk Factors for Childhood Stunting in 137 Developing Countries: A Comparative Risk Assessment Analysis at Global, Regional, and Country Levels. PLoS Med. Nov 1;13(11):e1002164.

Day LT, Ruysen H, Gordeev VS, et al. “Every Newborn-BIRTH” protocol: observational study validating indicators for coverage and quality of maternal and newborn health care in Bangladesh, Nepal and Tanzania. *J Glob Health*. 2019;9(1):010902. doi:10.7189/jogh.09.01902

de las Fuentes L, Gu CC, Mathews SJ, et al. Osteopontin promoter polymorphism is associated with increased carotid intima-media thickness. J. Am. Soc. Echocardiogr. 2008;21(8):954-960.

de las Fuentes L, Waggoner AD, Mohammed BS, et al. Effect of moderate diet-induced weight loss and weight regain on cardiovascular structure and function. J. Am. Coll. Cardiol. 2009;54(25):2376-2381.

Espeland MA, Tang R, Terry JG, Davis DH, Mercuri M, Crouse JR. Associations of risk factors with segment-specific intimal-medial thickness of the extracranial carotid artery. Stroke. 1999;30(5):1047-1055.

Espeland MA, Craven TE, Riley WA, Corson J, Romont A, Furberg CD. Reliability of longitudinal ultrasonographic measurements of carotid intimal-medial thicknesses. Stroke. 1996;27(3):480-485.

Farooqui H, Jit M, Heymann DL, Zodpey S. Burden of Severe Pneumonia, Pneumococcal Pneumonia and Pneumonia Deaths in Indian States: Modelling Based Estimates. PLoS One. 2015 Jun 18;10(6):e0129191.

Frison L, Pocock SJ. Repeated measures in clinical trials : analysis using mean summary statistics and its implications for design. Stat Med. 1992 ;11(13):1685-704.

Gu C, Flores HR, de las Fuentes L, Dávila-Román VG. Enhanced detection of genetic association of hypertensive heart disease by analysis of latent phenotypes. Genet. Epidemiol. 2008;32(6):528-538.

Gupta M, Kumar R, Deb AK, Bhattacharya SK, Bose A, John J, Balraj V, Ganguly NK, Kant L, Kapoor AN, Watt J, Shearer J, Santosham M; Hib study working group. Multi-center surveillance for pneumonia & meningitis among children (<2 yr) for Hib vaccine probe trial preparation in India. Indian J Med Res. 2010 May;131:649-58.

Jha P, Gajalakshmi V, Gupta PC, Kumar R, Mony P, Dhingra N, Peto R, RGI-CGHR Prospective Study Collaborators. Prospective study of one million deaths in India: rationale, design and validation results. PLOS Med. 2005. doi: 10.1371/journal.pmed.0030018

Johnson MA, Chiang RA. Quantitative Guidance for Stove Usage and Performance to Achieve Health and Environmental Targets. Environ Health Perspect. 2015 Aug;123(8):820-6.

Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation: the St. George's Respiratory Questionnaire. Am. Rev. Respir. Dis. 1992;145(6):1321-1327.

Lam NL, Chen YJ, Weyant C, et al. Household Light Makes Global Heat: High Black Carbon Emissions From Kerosene Wick Lamps. Environ. Sci. Technol. 2012;46(24):13531-13538.

Larson et al. 2017. A Cross-Sectional Survey in Rural Bihar, India, Indicates That Nutritional Status, Diet, and Stimulation Are Associated with Motor and Mental Development in Young Children. J Nutrition. doi: 10.3945/jn.117.251231

Mackenzie GA, Hill PC, Sahito SM, Jeffries DJ, Hossain I, Bottomley C, Uchendu U, Ameh D, Ndiaye M, Osuorah CD, Adeyemi O, Pathirana J, Olatunji Y, Abatan B, Ahameefula E, Muhammad BS, Fombah AE, Saha D, Mackenzie R, Plumb I, Akano A, Ebruke B, Ideh RC, Kuti B, Githua P, Olutunde E, Ofordile O, Green E, Usuf E, Badji H, Ikumapayi UNA, Manjang A, Salaudeen R, Nsekpong ED, Jarju S, Antonio M, Sambou S,

Ceesay L, Lowe-Jallow Y, Sowe D, Jasseh M, Mulholland K, Knoll M,Levine OS, Howie SR, Adegbola RA, Greenwood BM, Corrah T. Impact of the introduction of pneumococcal conjugate vaccination on pneumonia in The Gambia: population-based surveillance and case-control studies. Lancet Infect Dis. 2017 Jun 7. pii: S1473-3099(17)30321-3.

Mahomed, N et al. Pediatr Radiol (2017) 47:1399–1404.

McCollum ED, Park DE, Watson NL, Fancourt NSS, Focht C, Baggett HC, et al. Digital auscultation in PERCH: Associations with chest radiography and pneumonia mortality in children. Pediatr Pulmonol. noviembre de 2020;55(11):3197-208.

McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, Fawzi W, Fink G. Development and validation of an early childhood development scale for use in low-resourced settings. Popul Health Metr. 2017 Dec;15(1):3.

McCracken JP, Smith KR, Diaz A, Mittleman MA, Schwartz J (2007). Chimney stove intervention to reduce long-term wood smoke exposure lowers blood pressure among Guatemalan women. Environ Health Perspect. 115(7):996-1001. PMCID: PMC1913602.

Miranda JJ, Bernabe-Ortiz A, Smeeth L, Gilman RH, Checkley W, Group CCS. Addressing geographical variation in the progression of non-communicable diseases in Peru: the CRONICAS cohort study protocol. BMJ open. 2012;2(1):e000610.

Mortimer K, Ndamala CB, Naunje AW, Malava J, Katundu C, Weston W, Havens D, Pope D, Bruce NG, Nyirenda M, Wang D, Crampin A, Grigg J, Balmes J, Gordon SB. A cleaner burning biomass-fuelled cookstove intervention to prevent pneumonia in children under 5 years old in rural Malawi (the Cooking and Pneumonia Study): a cluster randomised controlled trial. Lancet. 2017 Jan 14;389(10065):167-175. doi: 10.1016/S0140-6736(16)32507-7

Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, Jaffer YA, Bertino E, Gravett MG, Purwar M, Noble JA, Pang R, Victora CG, Barros FC, Carvalho M, Salomon LJ, Bhutta ZA, Kennedy SH, Villar J; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet. 2014a,Sep 6;384(9946):869-79.

Papageorghiou AT, et al., International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester of pregnancy. Ultrasound Obstet Gynecol, 2014b. 44(6): p. 641-8.

Pereda MA, Chavez MA, Hooper-Miele C, et al. Lung Ultrasound for Diagnosis of Pneumonia in Children: A Meta-Analysis. Pediatrics. 2015;135(4):714-722.

Prado EL, Abubakar AA, Abbeddou S, Jimenez EY, Somé JW, Ouédraogo JB.Extending the Developmental Milestones Checklist for use in a different context in Sub-Saharan Africa. Acta Paediatr. 2014 Apr;103(4):447-54.

Ram J, Pineda-Cely J, Calhoun WJ. Forced Oscillometry: A New Tool for Assessing Airway Function-Is It Ready for Prime Time? J Allergy Clin Immunol Pract. 2019;7(8):2861-2862.

Reddy, U.M., et al., Fetal imaging: executive summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, American Institute of Ultrasound

in Medicine, American College of Obstetricians and Gynecologists, American College of Radiology, Society for Pediatric Radiology, and Society of Radiologists in Ultrasound Fetal Imaging workshop. Obstet Gynecol, 2014. 123(5): p. 1070-82.

Rudan I, Boschi-Pinto C, Biloglav Z, Mulholland K, Campbell H (2008). Epidemiology and etiology of childhood pneumonia. Bulletin of the World Health Organization , 86: 408-416.

Shah S, Noble VE, Umulisa I, et al. Development of an ultrasound training curriculum in a limited resource international setting: successes and challenges of ultrasound training in rural Rwanda. Int J Emerg Med. 2008;1(3):193-196.

Shah SP, Epino H, Bukhman G, et al. Impact of the introduction of ultrasound services in a limited resource setting: rural Rwanda 2008. BMC Int Health Hum Rights. 2009;9:4.

Sherpa CT, LeClerq SL, Singh S, et al. Validation of the St. George's Respiratory Questionnaire in Nepal. Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation 2015:2(4):281-289.

Simkovich SM, Underhill LJ, Kirby MA, et al. Design and conduct of facility-based surveillance for severe childhood pneumonia in the Household Air Pollution Intervention Network (HAPIN) trial. ERJ Open Res. 2020;6(1):00308-2019.

Smith KR, McCracken JP, Weber MW, et al. Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial. Lancet. 2011;378(9804):1717-1726.

Stein J, Korcarz C, Hurst R, et al. American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine. J. Am. Soc. Echocardiogr. 2008;21(2):93-111.

Tchetgen EJ, Phiri K, Shapiro R. A simple regression-based approach to account for survival bias in birth outcomes research. Epidemiology. 2015 Jul 1;26(4):473-80.

Thompson LM, Bruce N, Eskenazi B, Diaz A, Pope D, Smith KR (2011). Impact of reduced maternal exposures to wood smoke from an introduced chimney stove on newborn birth weight in rural Guatemala. Environ. Health Perspect 119(10):1489-1494.

WHO 2014. Handbook: IMCI integrated management of childhood illness. Geneva: World Health Organization.

WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl. 2006a Apr;450:76-85.

Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104. doi:10.1111/j.1524-4733.2005.04054.x

Appendix 1

Summary of Formative Research

1. Scoping Phase (Phase I)

1.1 Goals

The goals to be achieved in the Scoping Phase, and the specific way in which these will inform the Main Trial, are outlined in Table 1 below.

Table 1. Summary of goals and plans for synthesis/analysis of results for the Scoping Phase (Phase I).

Scoping Phase (Phase I)	**Goal:** The goals of the Scoping Phase (Phase I) are two-fold: 1. To select specific sites within each country where the Pilot Intervention Trial (Phase II) will be conducted. 2. To develop and/or adapt contextually appropriate behavior change interventions to promote exclusive LPG cookstove use and dissuade households from using solid fuel cookstoves.
	Synthesis/analysis: With regards to Goal #1, specific sites within each country will be selected based on the following four criteria: 1. Low rates of LPG penetration. 2. High uptake of LPG not anticipated over next 5 years. 3. Low anticipated cross-contamination from neighboring households and other ambient sources (e.g. waste incineration sites, brick kilns, etc.). 4. Optimal implementation logistics: setting up field offices and ease of transport from field offices to main study office; and engagement with primary health care centers, village leaders, and government departments. The scoping phase air pollution assessment approach was designed in conjunction with the IRC leads to gain information about relative air pollution concentration differences among potential sites within country. Although limited in nature, the air pollution exposure assessment approach was designed specifically to provide information important to scenarios identified by the IRC leads. These data, used in conjunction with the other information described above, will provide valuable information for finalizing the site selection for the Pilot Intervention Phase (Phase II). With regards to Goal #2, the behavior change interventions developed and/or adapted will be tested during the following phase.

1.2 Activities

Though the goals of this phase are the same for all IRCs, due to variation across countries in current cookstove practices as well as the information deemed necessary by the IRC leads to finalize specific site selection, Scoping Phase (Phase I) activities vary across IRCs. These activities are summarized in Table 2 below.

Table 2. Summary of Scoping Phase (Phase I) activities planned for Formative Research at each Intervention Research Center.

India	Guatemala	Rwanda	Peru
Rapid Assessment Survey: • India: up to 400 households in 2 districts • Guatemala: up to 600 households in 3 areas • Rwanda: up to 400 households in 8 districts Pregnancy census (primary health care center records)			Most information necessary to achieve objectives of Scoping Phase (Phase I) to be drawn from ongoing trial (see **Section 3** for details)
1-3 Focus Group Discussions (n=5-8 participants per group)			
In-depth interviews and direct observations with 10 women and 10 men (LPG users and non-users)		In-depth interviews, direct observations, and cooking demonstrations (15 LPG users and 15 non-users)	In-depth interviews and direct observations (up to 15 LPG users and 15 non-users)
Exposure assessment in 20 women >18 years of age with solid fuel cookstoves and 20 with LPG cookstoves	Grab sampling of indoor and outdoor $PM_{2.5}$ in 120 households	Exposure assessment in 50 women >18 years of age with solid fuel cookstoves before and after food provision/asking not to cook with solid fuel cookstove	

1.3 Results on Site Selection

Phase I of the Formative Research was completed in all of the IRCs by July 2017. As a result, we were able to identify study sites that met the prescribed criteria for the trial and allowed us to move into Phase II. We also were able to complete the other major goal for Phase I, which was to develop strategies to encourage near-exclusive use of the LPG stove that we subsequently piloted in Phase II (Section 1.4 below).

To identify and assess the suitability of potential study sites, we reviewed data from government and other sources, met with government officials and health care providers, visited clinics and hospitals and explored sources of LPG supplies. We then conducted rapid household surveys and collected and analyzed air quality samples among promising sites. The surveys were designed primarily to address household characteristics (including kitchen configurations), healthcare seeing activities during the antenatal period, LPG penetration and LPG stove usage, cooking practices, and other potential sources of exposure to smoke. We also collected and analyzed air quality samples from inside the home of a random sample of households.

As each IRC completed this Scoping Phase, it completed report for the Steering Committee that summarized its main results. Highlights of these reports follow:

India: Scoping phase activities at the India IRC were conducted between March and June 2017 after nearly 4 months of exploratory site selection activities, to shortlist potential districts/blocks for the conduct of the HAPIN Trial. Six blocks of Villupuram district and 4 blocks of Nagapattinam district in the state of Tamil Nadu were identified as potential sites. Potential for exposure contrast is high given the high prevalence of indoor cooking, low prevalence of outdoor cooking, absence of major point sources, (exposure measurements described in the next sections corroborate this expectation). The health care seeking behavioral indices are in alignment with what was expected based on demographic surveillance (NFHS data) (although district level differences in access to public vs. private health care facilities may be important for location of Ultrasound

machines). LPG access is widespread, although LPG use is not. Exposure assessment/biomarker protocols may need to address smoking and use of kerosene lamps while profiling participant exposures. The overall observed exposure contrast between biomass and LPG users is likely to be in the steepest part of known exposure-response (E-R) curves of $PM_{2.5}$ for major health conditions (24 to 85 µg/m3). BEC messaging would need to address the need to discontinue the use of biomass stoves in addition to promoting active LPG use in the intervention arm. In particular, LPG use in intervention would need to support use of heating bath water.

Rwanda: Based on census and other data, candidate districts were identified Eastern (Nyamagabe and Gisagara) and Southern (Kayonza and Kirehe) provinces. After conducting other scoping activities, the rapid assessment was completed among 400 households selected randomly from in each of the four potential districts. The survey revealed potentially important differences in the likelihood to cook indoors, to use wood and charcoal for cooking, and to use kerosene as source of lighting. Overall, use of LPG was extremely low (<5%) in all candidate sites. Exposure monitoring was conducted in Gisagara and Koyonza. Kitchen air pollution (KAP) levels were similar in both districts in the traditional cooking period and resulted in similar levels of improvement during the non-cooking period, however these were still higher than the WHO interim target 1. Personal monitoring of $PM_{2.5}$ revealed relatively high and were slightly higher in Kayonza. However, Kayonza showed a greater reduction in personal exposure during the non-cooking period, though it was nevertheless still quite high (>100 ug/m3). Based on these overall results, the Rwanda team elected to conduct Phase II in Koyonza.

Guatemala: Following slightly different procedures prescribed by protocol and described in the DSMB report on Formative Research, the Guatemala team interviewed 292 pregnant women at antenatal clinics in Jalapa District, the area identified as other published and field sources to be potentially meet prescribed site criteria. About half met the proposed trial criteria for age and gestational age of their fetuses. Household visits to these women revealed low levels of exclusive LPG stove use (9%) though about a third (32%) reported occasional use of LPG to supplement cooking on solid biomass. Valid 15-minute ambient air quality samples were obtained from 242 of these households, and overnight samples were obtained from 51 of these households. From these measurements, it appears that the open fire stove will generate very high level of $PM_{2.5}$,, reaching over 500 ug/m3 during 15-minute indoor monitoring. On the other hand, households using LPG stoves (even with some fuel stacking) were found to have relatively low levels of $PM_{2.5}$, with a mean of 58 ug/m3. For firewood plancha stove with chimney, indoor PM2.5 level is still very high, reaching over 200 ug/m3 during indoor and overnight measurements. The highest $PM_{2.5}$ concentration observed was during the open fire or firewood stoves were in use, could reach as high as 600 ug/m3. Based on these results, the team concluded that Jalapa District was likely to meet the trial criteria.

Peru: As described in the protocol, the aims of the Scoping Phase were accomplished in Peru in the context of the CHAP trial, a pilot LPG intervention trial enrolling 180 non-pregnant women (25-64 years) in rural communities of the province of Puno. For HAPIN, the Peru IRC expanded the study area to also include rural communities in the provinces of Huancane, Chichuito, and El Collao, all of which are also located in the Department of Puno. The area has a high prevalence of daily biomass fuel use and women who match eligibility criteria. Mean distance between households is 101 meters, and exposure monitoring showed little evidence of cross contamination from neighboring activities. While about of third of households also report cooking with gas, these households nevertheless had high kitchen level concentrations of $PM_{2.5}$ (669 ug/m3). Baseline personal level exposures in the CHAP study (used for scoping and pilot phases) were 103 to 128 ug/m3), and the intervention appears to be reducing those levels by around 10 fold (to 16.7 ug/m3). Based on these assessments, the proposed sites in Puno appear satisfactory, though the widespread presence of LPG stoves (though sporadic use) required a modification of the eligibility criteria for the Main Trial to allow households that have an LPG stove provided they are daily users of biomass.

1.4 Results on Behavior Change Strategies

Phase I also included extensive activities designed to assess the potential barriers and drivers to exclusive use of the LPG stove in the intervention arm of the Main Trial. These activities, which were led by the Behavior and Economics Core (BEC), began with the review and analysis of a growing body of research on adoption and use of improved cookstoves. This identified central barriers (e.g., cost, fear of explosions, inability to cook certain foods, preference for certain foods cooked with biomass) that were common to most settings. Each IRC then undertook a series of in depth interviews, focus group discussions and other research methods to identify the extent to which these and other issues may result in "stacking" (use of traditional as well as intervention stove) and specific strategies and approaches that could be used to overcome these barriers (e.g., stove demonstrations, taste tests, promotional activities, educational materials, etc.). These were then tested and reiteratively revised by each of the IRCs in subsequent qualitative research activities as well as in the Pilot Intervention Phase (Phase II). The result is an overall strategy with a goal of achieving near exclusive use of the LPG stove by intervention households. Actual use will be closely monitored using stove use monitors (SUMs) that provide objective data on continued use of traditional stoves (as well LPG stoves in intervention households). Throughout the Main Trial, these data will be used to trigger additional promotional activities if stacking is observed.

2. Pilot Intervention Phase (Phase II)

2.1 Goals

The goals to be achieved in the Pilot Intervention Trial, and the specific way in which these will inform the Main Trial, are outlined in Table 3 below.

Table 3. Summary of goals & plans for synthesis/analysis of results for the Pilot Intervention Trial (Phase II).

Pilot Intervention Trial (Phase II)	**Goals:** There are several goals to be achieved by the Pilot Intervention Trial (Phase II). With regards to piloting assessment procedures in pregnant women, the goals include, for India, Guatemala, and Rwanda: 1. To pilot the recruitment strategies to identify pregnant women (<20 weeks gestation). 2. To pilot the standardized operating procedures for: surveys, anthropometry, blood pressure, fetal ultrasound, and urine/dried blood spot collection, including transport of specimens from field offices to the main study office. *Note:* *For Peru, where a Pilot Intervention Trial in non-pregnant women is currently being conducted as part of another study, Goals #1-2 will be achieved in Phase III.* With regards to piloting intervention delivery, the goal is (for all four IRCs): 3. To pilot standardized operating procedures for LPG cookstove and fuel distribution. With regards to exposure assessment, the goals include (for all four IRCs): 4. To estimate anticipated baseline $PM_{2.5}$ and black carbon levels via assessment of a small sample of households determined by convenience. 5. To contrast exposure from pre- to post-intervention, and, therefore, estimate anticipated final and change in concentrations for the Main Trial. 6. To compare changes in $PM_{2.5}$ levels from pre- to post-intervention in kitchen and ambient samples in order to quantify anticipated cross-contamination from neighboring households and other ambient sources. 7. Finally, with regards to the behavior change interventions, the goals include (for all four IRCs): • To test the effectiveness of the behavior change interventions. • To assess the acceptability of the behavior change interventions. *Note:* *For Guatemala, where households with pregnant women **and** households with children <1 year old will be enrolled in Phase II along with any non-pregnant adult women and children < 2 years meeting inclusion criteria and also living in the households, all goals of Phase III will be achieved during Phase II.*
	Synthesis/analysis: With regards to Goals #1-2, we will evaluate the success of the execution and any challenges / barriers encountered across the IRCs. Based on this evaluation, we will make any necessary changes to implement in the Main Trial. We will adapt the behavior change interventions (Goal #7) as necessary given what we learn regarding air pollution exposure and stove use in this Phase. With regards to Goals #4-6 (air pollution assessment), the Pilot Intervention Trial (Phase II) will provide information on feasible post-intervention $PM_{2.5}$ exposure levels within each IRC. Our goals are to maximize exposure reductions beyond those previously demonstrated through our comprehensive fuel distribution and behavior change protocols. We will synthesize exposure data for each IRC and compare it to the exposure-response results and other data on major HAP-related health outcomes from previous and ongoing

	research to determine if we are within the range of exposure reductions that are likely to achieve meaningful health gains. Members of our Exposure Core are currently updating our reviews and analyses of these other studies to ensure that they reflect the best possible evidence. We will also use the field results to confirm issues around site selection, levels of stove adoption, risks and methods of minimizing cross contamination, and statistical power. This approach will allow us to make informed decisions about how to move forward with the Main Trial in a diligent and responsible manner. This decision will be made by the Steering Committee and reviewed by the External Advisory Board and the DSMB following the completion of the formative research stage and before moving forward with the Main Trial.

2.2 Progress and Results

India: Phase II was conducted August – October 2017. For the Phase II activities, the team in India conducted a pilot intervention in 40 pregnant women, 22 in Villupuram and 18 in Nagapattinam (see Appendix 2, Figure 4). For all participants in Phase II, after screening and enrollment the baseline exposure measurements were conducted and the LPG stove and fuel were provided. Follow-up exposure measurements (48-hour personal and kitchen measurements using the Enhanced Children's MicroPEM (ECM) and Ultrasonic Personal Aerosol Sampler (UPAS)—the two devices we are considering for personal exposure monitoring—for fine particulate matter and the Lascar for carbon monoxide were conducted at 1-month and 2-months post-intervention; we currently have data from the 1-month follow-up. Both the kitchen and area concentrations look promising for the goals of Phase II. For kitchen $PM_{2.5}$, the mean 48-hour average for baseline was 301ug/m^3 and for 1-month follow-up was 21ug/m^3. For personal $PM_{2.5}$, the mean 48-hour average for baseline was 125 ug/m^3 and for 1-month follow-up was 28 ug/m^3. Ambient concentrations were 8 ug/m^3 at baseline and 7 ug/m^3 at the 1-month follow-up. There were no issues with LPG connection, installation, delivery or support logistics. All of the households reported exclusive use of the LPG stove with no use of a secondary stove; in many cases, the women voluntarily destroyed the old biomass stove. The team has conducted 25 in-depth interviews regarding behavior change and messaging. They have developed a 1-page messaging document that will be delivered to households and will leave two pamphlets (one of LPG safety and one discouraging use of biomass) with households. They have held several focus groups to develop a plan for control compensation. The pilot intervention also included dried blood spot and urine collection for biomarker measurements and analysis. The team successfully collected urine samples from all participants at the appropriate times, and obtained 4-5 correct dried blood spots for 37/40 participants.

Rwanda: For the Phase II activities, the team in Rwanda conducted a pilot intervention among 40 women in the selected district of Kyonza (see Appendix 2, Figure 7). Participants were enrolled in September and October 2017. Exposure assessment (24-hour personal $PM_{2.5}$ using ECMs) was conducted at baseline and at 1 week after the LPG stove intervention. Due to problems with the ECMs, baseline data were available for only 8 participants. The first round of 1-week follow-up measurements are available for 31 participants. Initial results from this round look promising with regard to the goals of the Phase II trial. Mean 24-hour personal $PM_{2.5}$ (n=8) at baseline was 186 ug/m^3 and at the one-week follow-up was 44 ug/m^3. The team also obtained 24-hour average kitchen concentrations in 16 houses at follow-up; the mean concentration was 28 ug/m^3. The second round of follow-up in the pilot households was completed on November 10. Results will be available to the DSMB for the November 29 meeting. There have been no issues so far with LPG installation or delivery. The team in Rwanda has conducted multiple focus groups and in-depth interviews and is subsequently developing educational materials that will be left with the intervention households.

Guatemala: For Phase II activities, the team in Guatemala enrolled 40 households with pregnant women (see Appendix 2, Figure 14) less than 20 weeks gestation, confirmed by ultrasound, who used open fires (n=38) or deteriorated plancha stoves (n=2; no chimneys; high emissions). They measured baseline 24-hour average kitchen and personal gravimetric $PM_{2.5}$ prior to LPG stove installation, and measured $PM_{2.5}$ 1 month and 2 months after providing the LPG stove. They also monitored stove use, using i-buttons. For this report we have

data for baseline and for the 1-month follow-up. Both the kitchen and area concentrations look promising for the goals of Phase II. For kitchen $PM_{2.5}$, the mean 24-hour average for baseline was 350ug/m^3 and for 1-month follow-up was 19ug/m^3. For personal $PM_{2.5}$, the mean 24-hour average for baseline was 271 ug/m^3 and for 1-month follow-up was 30 ug/m^3. There have been no issues reported with LPG stove installation or LPG fuel delivery. While only 2 households reported continued use of the traditional wood stove at the 1-month follow-up, the stove use monitoring detected continued use of traditional wood stoves in about 30% of households. This stove stacking did not influence the reported kitchen and personal $PM_{2.5}$ measurements because the traditional stoves were not used on exposure assessment days. Training and educational material designed by the LPG company was provided as part of the pilot intervention and completed 18 key informant interviews, 6 focus groups, and 37 kitchen/cooking observations. The team also held focus groups to discuss approaches to control compensation and to get feedback drawings and key messages for calendars to reinforce behaviors during main trial. They also developed a video recording of behavior change messages to be used on iPads to reinforce behaviors during main trial. Starting in November 2017, the team will conduct pilot Phase II-B and will enroll 20 women between ages 35 and 79 who live in a household with a child < 1 year of age to conduct exposure assessment and pilot biomarker collection and processing.

Peru: In Peru we are using an ongoing LPG trial, the CHAP trial, among older women in the same region as the HAPIN trial to accomplish goals 3-7 of Phase II (the other goals will be addressed in Phase II). The CHAP trial has enrolled 170 women to date (see Appendix 2, Figure 10); we have exposure assessment (24-hour kitchen and personal $PM_{2.5}$ from ECMs) for a subset of those enrolled. The available data for $PM_{2.5}$ look promising for the HAPIN trial. For the women who received the LPG stove intervention the mean baseline 24-hour average kitchen $PM_{2.5}$ was 768ug/m^3; for 1 month after the LPG intervention it was 67ug/m^3 and 3 months after the intervention it was 40 ug/m^3. For personal 24-hour $PM_{2.5}$ the baseline mean was 104 ug/m^3, the mean 1 month after the LPG intervention was 17 ug/m^3, and the mean 3 months after the intervention was 20 ug/m^3. There have been no issues with LPG stove installation or LPG fuel delivery. The team in Peru has conducted 7 in-depth interviews and 1 focus group to discuss the LPG stove intervention, education approaches, and control compensation. The team has developed educational materials to be left with intervention households.

Updated Exposure Data from Phase II

HAPIN Phase II pilot intervention results for $PM_{2.5}$ assessment			
India; n=40 pregnant women; ECM and UPAS for $PM_{2.5}$			
	Baseline	**1 month post-intervention**	**2 months post-intervention**
	Mean / Median/ Range / N	**Mean / Median / Range / N**	**Mean / N**
48-hour average kitchen (ug/m^3)	301 / 171 / 44 – 1830 / n=34	21 / 18 / 1.1 – 74 / n=37	23 / n=37
48-hour average personal (ug/m^3)	125 / 75 / 17 – 991 /n= 40	28 / 23 / 7 – 138 / n=40	22 / n=40

Rwanda; n=40 pregnant women; Gravimetric and ECM for $PM_{2.5}$			
	Baseline	**1.5 weeks post-intervention**	**1.5 months post-intervention**
	Mean/ SD / Median / IQR / N	**Mean/ SD /Median/ IQR / N**	**Mean / SD / Median / IQR / N**
24-hour average kitchen (ug/m³) - grav	No measurements available	33 / 16 / 28 / 21-41 / n=21	30 / 18 / 26 / 15-43 / n=17
24-hour average personal (ug/m³) - grav	210 / 132 / 176 / 112 – 297 / n=35	120 / 127 / 87 / 39 – 159 / n=31	65 / 60 / 46 / 26-77 / n=36
24-hour average kitchen (ug/m³) - ECM	No measurements available	24 / 10 / 20 / 17-28 / n=12	Not yet available
24-hour average personal (ug/m³) - ECM	219 / 137 / 210 / 91-307 / n=7	43 / 22 / 37 / 29-54 / n=24	Not yet available
Guatemala; n=40 pregnant women; ECM for $PM_{2.5}$			
	Baseline	**1 month post-intervention**	**2 months post-intervention**
	Mean / SD / N	**Mean / SD / N**	**Mean / SD / N**
24-hour average kitchen (ug/m³)	432 / 343 / n=36	37 / 34 / n=39	83 / 203 / n=14
24-hour average personal (ug/m³)	251 / 237 / n=30	50 / 43 / n=32	41 / 24 / n=12
Peru; n=170 older women enrolled in ongoing CHAP LPG trial; ECM for $PM_{2.5}$			
	Baseline	**1 month post-intervention**	**3 months post-intervention**
	Mean / SD / Median / N	**Mean / SD / Median / N**	**Mean / SD / Median / N**
24-hour average kitchen (ug/m³)	923 / 699 / 835 / n=40	94 / 176 / 30 / n=13	53 / 61 / 39 / n=10
24-hour average personal (ug/m³)	120 / 139 / 66 / n=36	25 / 34 / 13 / n=19	26 / 34 / 14 / n=15

3. Pilot Assessment Procedures Phase (Phase III)

2.3 Goals

The goals to be achieved in the Pilot Assessment Phase, and the specific way in which these will inform the Main Trial, are outlined in Table 4 below.

Table 4. Summary of goals and plans for synthesis/analysis of results for the Pilot Assessment Procedures Phase (Phase III).

Pilot Assessment Procedures Phase (Phase III)	**Goals:** The goals of the Pilot Assessment Procedures Phase (Phase III) are three-fold: 1. To pilot the standardized operating procedures for non-pregnant, older adult women: surveys, anthropometry, blood pressure, vascular ultrasounds, urine/dried blood spot/venous blood collection, and personal exposure ($PM_{2.5}$ levels). 2. To pilot the standardized operating procedures for children <2 years of age: surveys, anthropometry, lung ultrasound, urine/dried blood spot collection, and personal exposure ($PM_{2.5}$ levels). 3. To validate biomarkers assessed in dried blood spots versus venous blood samples. In Peru, this Phase includes the following additional goals: 4. To pilot the recruitment strategies to identify pregnant women (<20 weeks gestation). 5. To pilot the standardized operating procedures for: surveys, anthropometry, blood pressure, fetal ultrasound, and urine/dried blood spot collection, including transport of specimens from field offices to the main study office. *Note: Goals #4-5 for Peru are identical to Goals #1-2 of the Pilot Intervention Trial (Phase II) for India, Guatemala, and Rwanda.*
	Synthesis/analysis: We will evaluate the success of the execution and any challenges / barriers encountered across the IRCs. Based on this evaluation, we will make any necessary changes to implement in the Main Trial.

3.2. Results

India: With respect to goals #1 and #2 of Phase III, field staff CRF training was completed in October and the piloting of standardized operating procedures among older adult women and children <2 years of age began the first week of November at both Villupuram and Nagapattinam. The field team plans to enroll households with both older adult women and children meeting inclusion criteria to improve efficiency of data collection for this Phase. Data collection is expected to be completed by mid-December, though two considerations will influence the timeline in India: (1) the design and testing of baby vests to hold the Beacon monitor and (2) the clearance from Indian customs and registration of the Sonosite ultrasound machines. We intend to pilot all assessment procedures using the Sonosites in registered clinics per existing rules and regulations in India.

Guatemala: Goals #1 and #2 of Phase III were achieved as part of Phase II in Guatemala, where households with pregnant women and households with children <1 year old were enrolled in the pilot intervention, and any non-pregnant, older adult women and/or children 1-2 years old who also lived in these households were enrolled in order to pilot the standardized operating procedures for this sample population. In order to reach the target sample size of n=40 non-pregnant, older adult women, they may also recruit n=20 women from clinics

where Phase IV is being conducted in order to pilot all procedures except exposure assessment in this sample population.

Rwanda: With respect to goals #1 and #2 of Phase III, CRF translation is currently underway and the piloting of standardized operating procedures among non-pregnant, older adult women and children <2 years of age will begin mid-November.

Peru: With respect to goal #4 of Phase III, the Peru team has held three successful health fairs in coordination with the Ministry of Health, in three communities in the HAPIN trial catchment area. However, though hundreds of community members attended each fair, the number of pregnancies detected was fairly low (n=3) compared to the effort required. Therefore, while the Peru team will continue to participate in health fairs that the Ministry of Health hosts, this will not be pursued as the main recruitment strategy. The team is currently ramping up coordination with each community health center, and with the existing network of community health promoters called 'agentes'. The promoters already perform house visits as part of their routine care, and the HAPIN team is training them in recruitment strategies.

With respect to goal #5 of Phase III, the standardized operating procedures for CRFs, urine/dried blood spot/venous blood collection, anthropometry, blood pressure, and imaging are currently underway. For biosamples and exposure measurements, the team has enrolled both pregnant women (n=20) and older adult women (n=20). Exposure measurements have been performed on all. Biomarkers have been collected from 4 pregnant women (urine and DBS) and 4 older women (urine, DBS, and venous blood).

Biomarker Core: The goal for Phase III is to collect 80 DBS and venous blood samples in Guatemala and Peru (n=40 per IRC). Similarly, 40 DBS and venous blood samples will be collected in India for local method development. We have done some validation with matched DBS and serum from biobanked samples at the Emory laboratory (from non-HAPIN participants). In addition, we have analyzed samples from Honduras and Peru (CHAP study) provided by HAPIN Co-Investigators to validate that the methods are working, but these studies have no matched blood. We have found that ICAM and VCAM have excellent agreement, but CRP does not. However, we have internal validity meaning that changes between an individual's DBS can be easily deciphered, but DBS CRP cannot be compared to blood CRP. In Phase III samples, we plan to validate ICAM, VCAM, CRP, e-selectin, endothelin, IL6, F2 isoprostanes, myeloperoxidase, CPY450 induction, metals, and HbA1c.

4. Respiratory Rate/Pulse Oximetry Assessment Phase (Phase IV)

4.1 Goals

The goals to be achieved in the Respiratory Rate/Pulse Oximetry Assessment Phase, and the specific way in which these will inform the Main Trial, are outlined in **Table 5** below.

Table 5. Summary of goals and plans for synthesis/analysis of results for the Respiratory Rate/Pulse Oximetry Assessment (Phase IV).

Respiratory Rate/Pulse Oximetry Assessment (Phase IV)	**Goals:** The goals of the Respiratory Rate/Pulse Oximetry Assessment (Phase IV) are two-fold: 1. To define tachypnea thresholds, defined as the upper limit of normal for respiratory rate in healthy children, for children 0 – 23 months of age across four IRCs located at various altitudes (sea level to 3,825 m above sea level). 2. To define oxyhemoglobin saturation thresholds, defined as the lower limit of normal for healthy children 0 – 23 months of age across four IRCs located at various altitudes (sea level to 3,825 m above sea level).
	Synthesis/analysis: One of the primary outcomes of the Main Trial is to assess severe pneumonia incidence in children < 2 years. The case definition for this outcome includes respiratory rates and oxyhemoglobin that are outside the normal range. Since the normal range depends largely on local altitudes and climates, it is necessary to develop standards at each IRC from children that are not suffering from respiratory symptoms. These assessments will help define respiratory rate and oxyhemoglobin saturation cut-offs for each IRC, to ensure that our pneumonia definition is appropriate for the IRCs of varying altitudes and climates.

4.2 Accomplishments

1. Obtained 13 Masimo Rad-97 devices (Masimo, Irvine, CA) for the assessment of automated respiratory rate and oxyhemoglobin saturation data. The Masimo Rad-97 device is a brand new product by Masimo (http://www.masimo.com/en-GB/home/rainbow-pulse-co-oximetry/rainbow-monitors/rad-97/) that can measured respiratory rate acoustically and photoplethysmographically.
2. Developed a plan and Standard Operating Procedure to validate the automated assessment of respiratory rate with standard manual respiratory rate counting.
3. Developed the Phase IV C4 form (titled Phase IV Piloting Form) which is now available in English and translated into Spanish, Kiyarwanda and Tamil.
4. Piloted the C4 form with field staff at IRCs.
5. Implemented C4 form into REDCap. We have drafted an instructional sheet to aid field staff in data collection.
6. Finalized Standard Operating Procedures for Phase IV study in English and Spanish.
7. Obtained U.S. (Emory University, Johns Hopkins University) and local IRB (AB PRISMA, SRU, LSHTM, Universidad del Valle de Guatemala) approvals across all IRCs for data collection.

4.3 Planned timeline for Phase IV

1. Planned start dates across all IRCs are in late November/mid-December with data collection planned for 2-3 working weeks.
2. Validation of automated respiratory rate will be available mid-December using data from all sites (n=160).

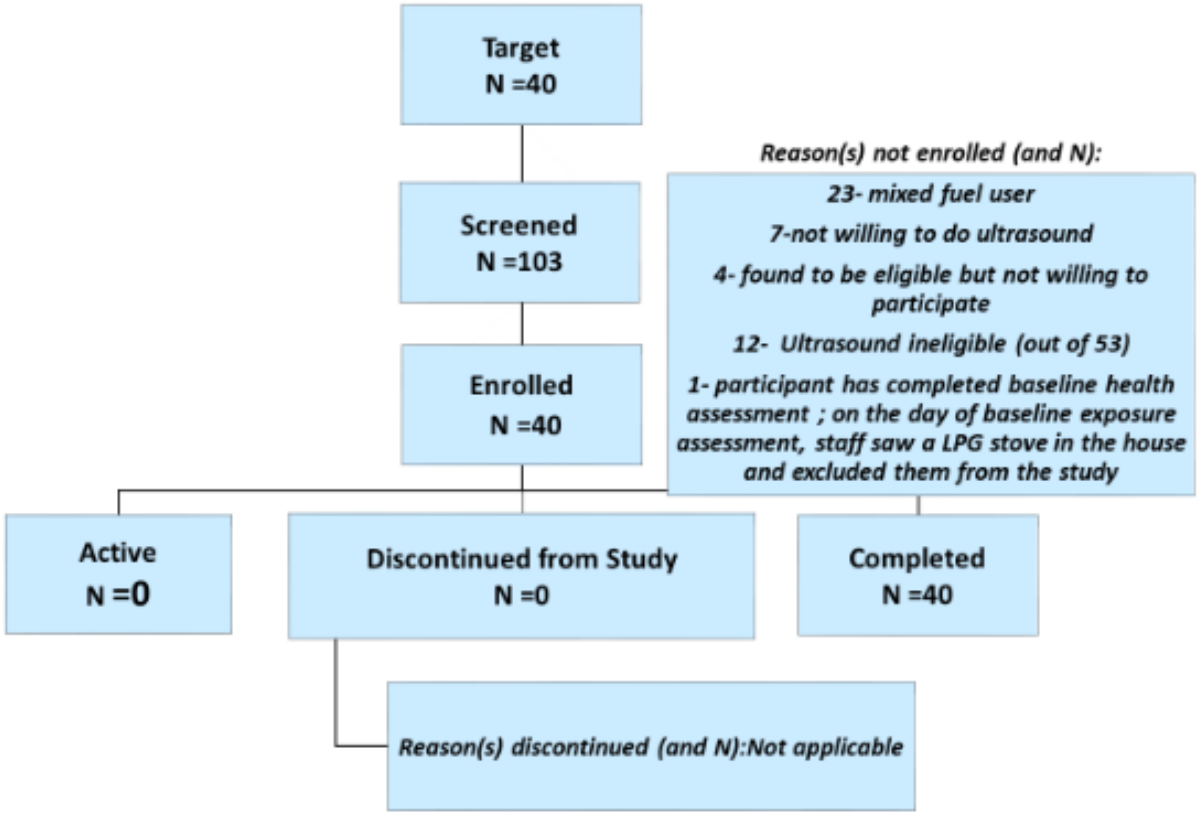
3. Analysis of data will occur in mid-January and thresholds for respiratory rate and oxyhemoglobin saturation (n=1600).
4. Report and publication available by the end of January.

Appendix 2: Pilot Study CONSORT Diagrams

(figure numbers not consecutive because figures are taken from December 2017 DSMB Report)

Figure 4: India IRC – Pilot Study Status – Pregnant Women

Date as of: 9 November 2017
Recruitment Start Date: 18 July 2017



We would like to seek consent from all the 40 pilot intervention participants for 3 more months of fortnightly follow-up for SUMS data. We will not give free LPG refills, but just observe their stove use pattern during this period (proposed amendment)

Figure 7: Rwanda IRC – Pilot Study Status – Pregnant Women

Date as of: 10 November,2017
Recruitment Start Date: 2 August 2017

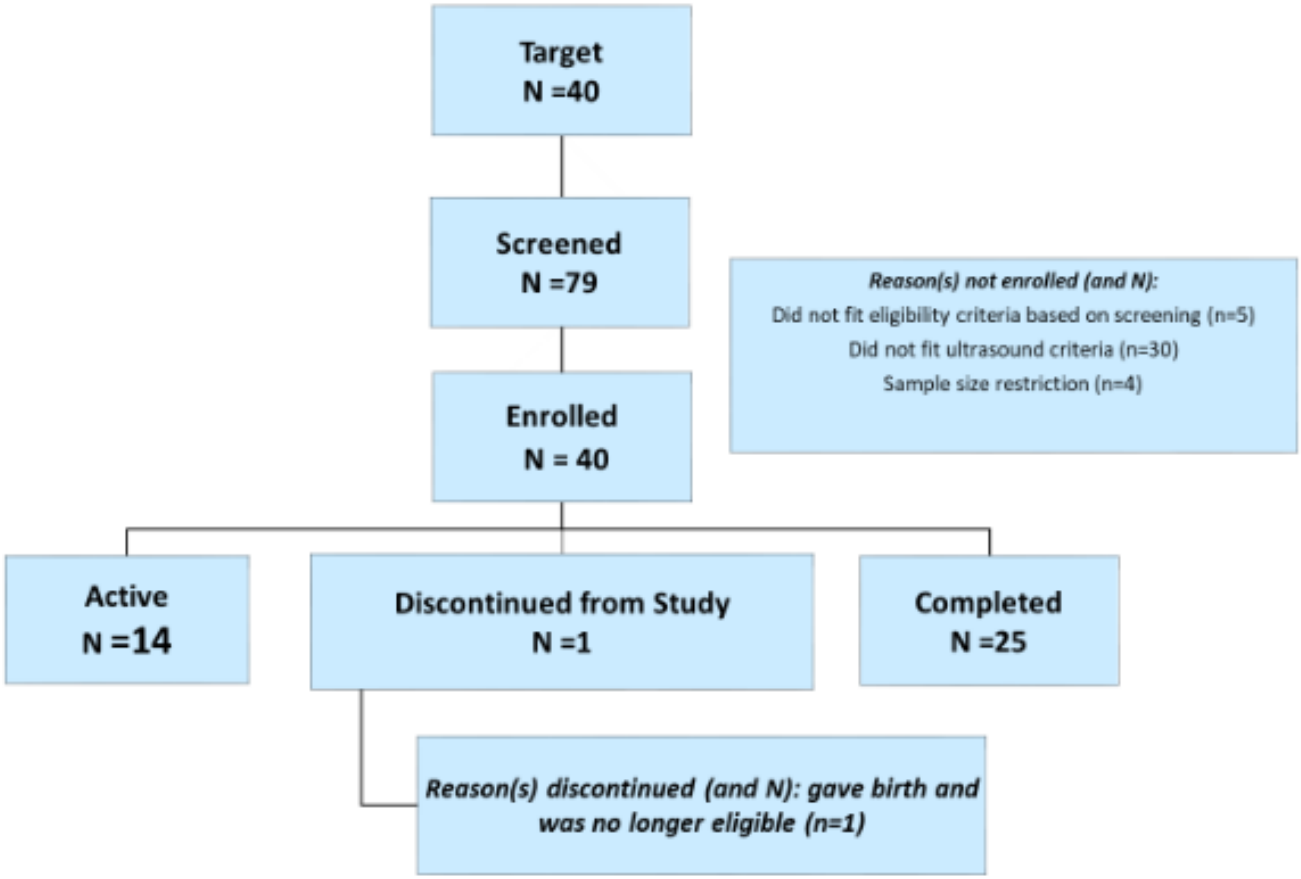
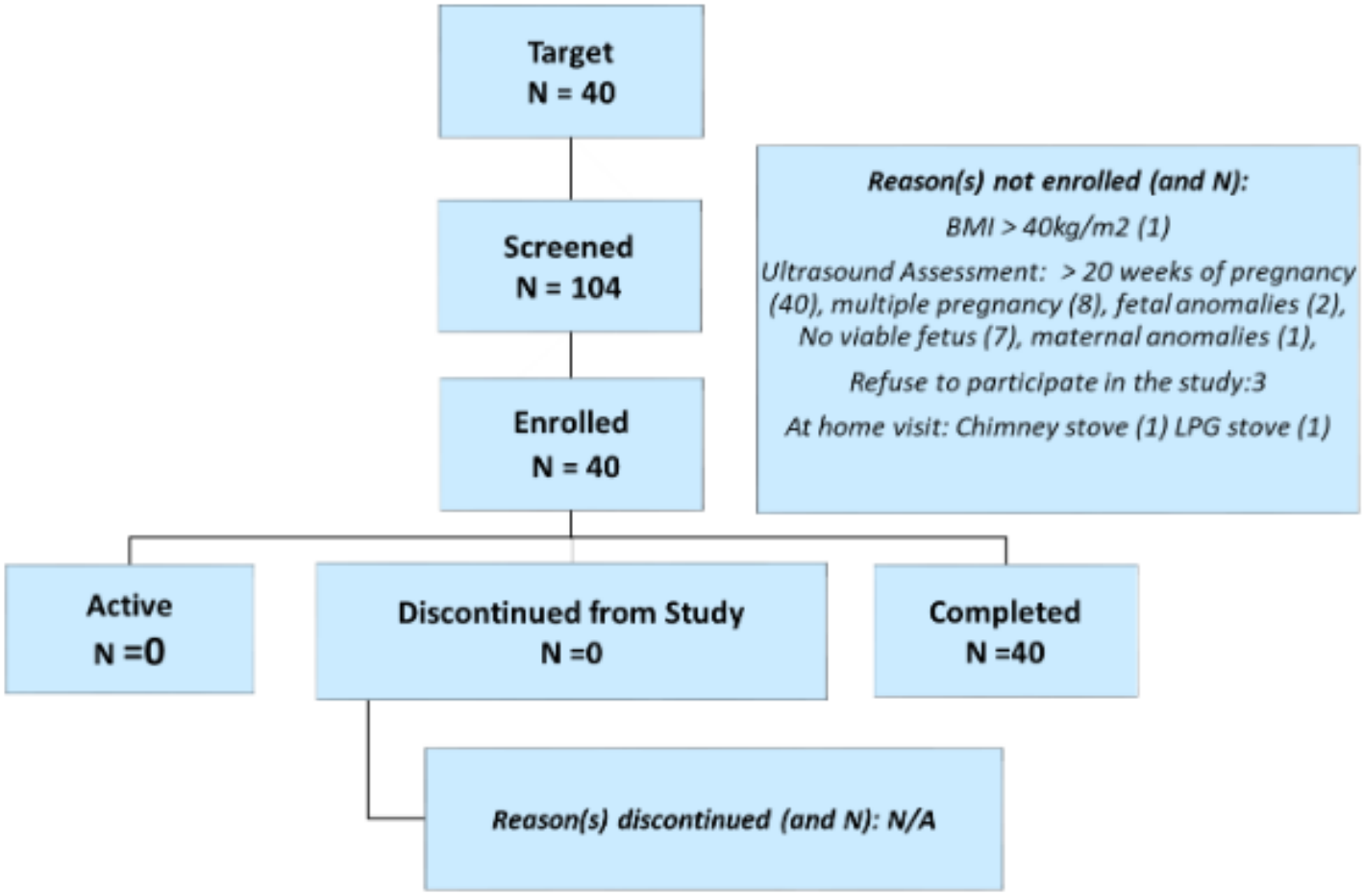


Figure 14: Guatemala IRC – Pilot Study Status – Pregnant Women

Date as of: 3 November 2017
Recruitment Start Date: 31 July 2017



Appendix 3: Statistical Analysis Plans for Primary Outcomes

3.A. Intention-to-Treat and Exposure-Response Analysis for Birth Weight

Version 1.6
September 13, 2021

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).

APPENDIX 3: STATISTICAL ANALYSIS PLANS

3.B. Intention-to-Treat Analysis for Stunting in 1-year-old Children

Version 1.2
August 18, 2022

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):

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

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:

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

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).

5. 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.
6. A table summarizing maternal characteristics at baseline (overall and by IRC).
7. The set of outcomes (primary and secondary) and subgroup analysis to be replicated.
8. 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:

5. Replicate summary statistics (e.g., mean, standard deviation, percentages, proportion missing) in the baseline characteristic table.
6. Replicate intention-to-treat analyses for primary and secondary outcomes according to models specified in Section 5.3.

Replicate results from interaction analyses (intention-to-treat only).

APPENDIX 3: STATISTICAL ANALYSIS PLANS

3.C. Intention-to-Treat Analysis for HAPIN Pneumonia

Version 1
August 31, 2022

1. INTRODUCTION

This document contains the statistical analysis plan (SAP) for severe pneumonia of the HAPIN Study. Severe pneumonia 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 at the time of consent. 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

The power for the hypothesis test for relative risk of pneumonia is 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 p_1 is the risk of pneumonia among the control group, p_2 is the risk of pneumonia among the intervention group, RR is the true relative risk (i.e. p_2/p_1), and n is the common sample size for each of the treatment group and the control group. For pneumonia, power for Poisson regression is also approximated by the equation for relative risk above assuming that follow-up times are the same for all participants.

The following assumptions, based on previous studies, are made for power calculations.

- For pneumonia, we assume a 1-year cumulative incidence rate among the control group as $p_1 = 0.06$ [a weighted average from Mackenzie et al. 2014, Gupta et al. 2010, Mortimer et al. 2007, Broor et al. 2017, Farooqi et al. 2015]. The actual incidence rate accounting for multiple events over the entire follow-up period is expected to be higher.

We assume a 10% attribution over the 1-year follow-up period, resulting in an effective sample size of $n = 1440$ per treatment arm for pneumonia.

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.

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 severe pneumonia follow-up is through the first year of life.

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 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 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 severe pneumonia 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 for whether the child did or did not have one or more illnesses that met criteria for the pneumonia endpoint

(*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, and withdrawal from study *prior to birth*. 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 prior to birth and after birth
    - Withdrawn by study team prior to birth and after birth
    - Moved away prior to birth and after birth
    - Pregnancy loss (termination/miscarriage/stillbirth)

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
- Child 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 restricted to liveborn children will be summarized by intervention versus control arms, separately by each IRC as defined by Table 1. A separate Table 2 will report additional key descriptive variables that may not be available at baseline. We will also report on characteristics of pneumonia cases as defined by Table 3. 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 of liveborn children

Variables	Type	Definition/Assessment Methods
International Research Center	Categorical	Guatemala, India, Peru, Rwanda
Gestational age (weeks) at intervention	Continuous	Calculated as the date of LPG installation minus the date of screening ultrasound plus gestational age at screening
Child sex	Categorical	Male/Female
Number of siblings in the house	Continuous	
Second-hand smoking	Binary	Whether someone other than the pregnant woman in household smokes (smoking of the pregnant mother was an exclusion criteria) (yes/no/missing)
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
Household assets	Categorical	• Color TV • Radio • Mobile phone • Bicycle

		• Bank account

Table 2. Other key characteristics of study population of liveborn children		
Variables	**Type**	**Definition/Assessment Methods**
Maternal variables		
Mother's age (years) at baseline	Continuous	Calculated as the date at baseline minus the date of birth. Date at baseline is assigned by the date of visit if not 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
Gestational age (weeks) at baseline	Continuous	Calculated as the date at baseline minus the date of screening ultrasound plus gestational age at screening, and then divided by 7
Child variables		
Birth weight-for-age z score	Continuous	
Breastfeeding	Binary	Breastfeeding during the first six months of life (yes/no/missing)
Up-to-date vaccination status	Binary	Number of respiratory vaccine doses by vaccine type by one year of age, including number of doses of pneumococcal conjugate vaccine (PCV) (*note*: PCV not available in India site), number of doses of haemophilus influenzae (HiB) vaccine, dose of measles vaccine Categorized as age <8 weeks & 1 dose PCV & HiB; age 8 to <12 weeks & 2 doses of PCV & HiB, age 12 to 32 weeks & 3 doses of PCV & HiB, age >32 weeks & 3 doses of PCV & HiB and 1 dose measles Yes/no

Table 3. Pneumonia case characteristics to be reported		
Variables	**Type**	**Definition/Assessment Methods**
Demographic information		
International Research Center, n (%)	Categorical	Guatemala, India, Peru, Rwanda
Child's age (months)	Continuous	Calculated as the date at diagnosis minus the date of birth. Date at diagnosis is assigned by the date of visit if not missing. Categorized as <2, 3-5, 6-11
Child sex, n (%)	Categorical	Male/Female
Weight (kg)	Continuous	
Severe acute malnutrition, n (%)	Binary	Weight-for-length or weight-for-age (if length unavailable) z score > -3 (use stata zscore06 at https://www.stata.com/statalist/archive/2011-04/msg01386.html)

		Yes/no
Up-to-date vaccination status, n (%)	Binary	Number of respiratory vaccine doses by vaccine type by age at diagnosis, including number of doses of PCV (*note*: PCV not available in India site), number of doses of HiB vaccine, dose of measles vaccine Categorized as age <8 weeks & 1 dose PCV & HiB; age 8 to <12 weeks & 2 doses of PCV & HiB, age 12 to 32 weeks & 3 doses of PCV & HiB, age >32 weeks & 3 doses of PCV & HiB and 1 dose measles Yes/no
$PM_{2.5}$, CO, BC	Continuous	
Antenatal $PM_{2.5}$, CO, BC	Continuous	
Postnatal $PM_{2.5}$, CO, BC	Continuous	
Clinical signs		
Temperature >38 C, n (%)	Binary	
Average Heart rate	Continuous	Average of measurements when multiple measurements available Beats per minute
Average Respiratory rate	Continuous	Average of measurements when multiple measurements available Breaths per minute
Average SpO_2	Continuous	Average of measurements when multiple measurements available
Hypoxemia, n (%)	Binary	Categorized as SpO_2 <93% for Guatemala, India, and/or Rwanda. <87% for Peru. Yes/No
Wheeze +/- crackles, n (%)	Binary	Yes/No
At least one respiratory danger sign, n (%)	Binary	Any of the following: chest indrawing, severe chest indrawing, head nodding, persistent nasal flaring, grunting, stridor when calm, audible wheeze, tracheal tugging, intercostal recessions Yes/No
Chest indrawing, n (%)	Binary	Yes/No
Head nodding, n (%)	Binary	Yes/No
Persistent nasal flaring, n (%)	Binary	Yes/No
Grunting, n (%)	Binary	Yes/No
Stridor when calm, n (%)	Binary	Yes/No
Audible wheeze, n (%)	Binary	Yes/No
Tracheal tugging, n (%)	Binary	Yes/No
Intercostal retractions, n (%)	Binary	Yes/No

At least one general danger sign, n (%)	Binary	Any of the following: unable to drink or breastfeed, vomiting everything, convulsions, lethargy or unconscious, unable to feed, not moving at all or moves with stimulation only Yes/No
Unable to drink or breastfeed, n (%)	Binary	Yes/No
Vomiting everything, n (%)	Binary	Yes/No
Convulsions, n (%)	Binary	Yes/No
Lethargy or unconscious, n (%)	Binary	Yes/No
At least one neonatal danger sign, n (%)	Binary	Any of the following: unable to feed well, not moving at all or moves with stimulation only, grunting, severe chest wall indrawing Yes/No
Unable to feed well, n (%)	Binary	Applies to <2 months only. Yes/No
Not moving at all or moves only, n (%)	Binary	Applies to <2 months only. Yes/No
Grunting, n (%)	Binary	Applies to <2 months only. Yes/No
Severe chest indrawing, n (%)	Binary	Applies to <2 months only. Yes/No
Lung Imaging		
Lung ultrasound, n (%)	Binary	Primary endpoint pneumonia Yes/No
Chest radiography, n (%)	Binary	Primary endpoint pneumonia Yes/No
Clinical care		
Hospitalized, n (%)	Binary	Yes/No
Oxygen treatment, n (%)	Binary	Yes/No
Advanced respiratory supportive care, n (%)	Binary	Any of the following: High flow, NIV (CPAP/BiPAP), mechanical ventilation Yes/No
Outcome		
Mortality, n (%)	Binary	Death <30 days since diagnosis and/or Verbal Autopsy positive Yes/No

We plan to include additional supplemental tables reporting more detailed imaging results and verbal autopsy results that will be included in the publication online supplemental information.

5. DATA ANALYSIS

In this section we provide the analysis approach for the intentional to treat aim. The primary outcome is HAPIN pneumonia. We present the primary analysis along with effect modification and secondary analyses (alternative model specifications, secondary outcomes).

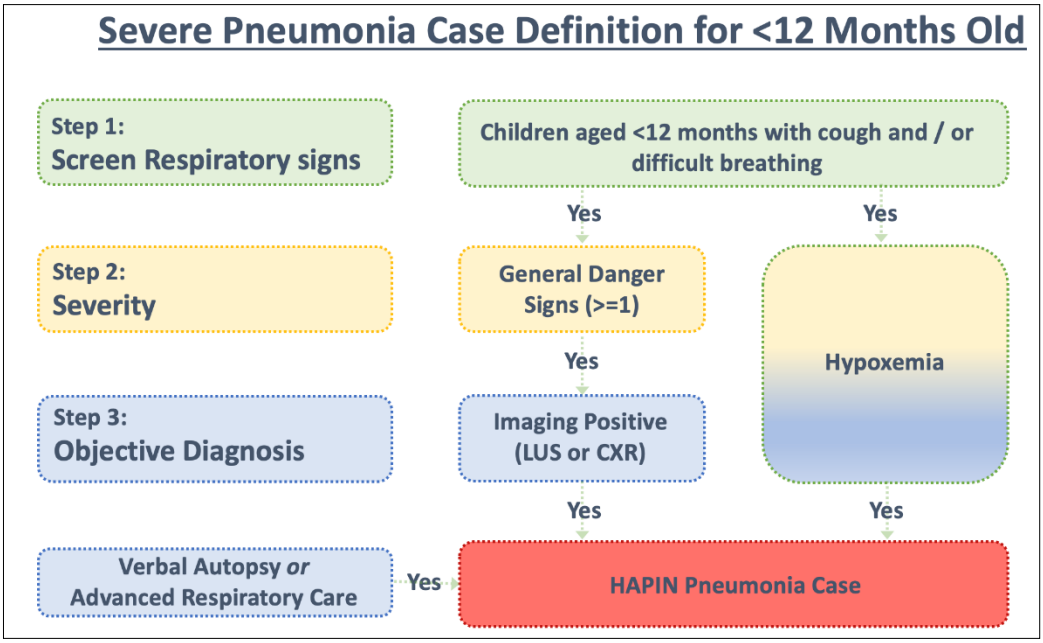
5.1. Outcome Definitions

This section describes each primary and secondary outcomes, including data collection approaches and calculations for derived outcomes.

The primary outcome is HAPIN pneumonia.

HAPIN pneumonia Case Definition (see Simkovich SM, et al. ERJ Open Res 2020;6(1))
We will define HAPIN pneumonia as a case of severe pneumonia adapted from the revised WHO classification of childhood pneumonia (World Health Organization, 2014), which includes three algorithms for the case definition: (1) the presence of cough and/or difficult breathing and at least one general danger sign and primary endpoint pneumonia on a lung ultrasound or chest x-ray, or (2) the presence of cough and/or difficult breathing and hypoxemia (measured via pulse oximetry, SpO_2), or (3) children who died and their death is attributed to pneumonia by verbal autopsy. General danger signs for all participants include any of the 5 signs of: (1) unable to drink or breastfeed, (2) vomiting everything, (3) convulsions, (4) lethargy or unconscious, (5) stridor at rest. There are 4 additional general danger signs for participants <2 months of age (neonatal danger signs): (1) unable to feed, (2) grunting, (3) not moving at all or moves with stimulation only, (4) severe chest indrawing. Severe acute malnutrition is also considered a danger sign for all participants, and is defined as a calculated weight-for-length z score < -3 or weight-for-age z score < -3 (if length unavailable or age <60 days).

Lung ultrasound will be the imaging modality obtained unless there are logistical or regulatory reasons this cannot happen and then we will obtain a chest x-ray. We will define tachypnea based on current World Health Organization thresholds ($\geq$60 breaths/minute for 0-2 month olds and $\geq$50 breaths/minute for 3-11 month olds). We will determine hypoxemia based on the physiologic threshold of $\leq$92% for altitudes <2,500 meters above sea level and $\leq$86% for altitudes $\geq$2,500 meters above sea level. Children are also considered hypoxemic, regardless of oxygen saturation, and assumed to have cough and/or difficult breathing if they are receiving advanced respiratory support at the time of evaluation, which includes any of the following: intubation and mechanical ventilation, non-invasive ventilation with continuous positive airway pressure support (CPAP), non-invasive ventilation with bi-level positive airway pressure support (BIPAP), or high-flow nasal cannula oxygen. Children who die but are determined to have had pneumonia or symptoms and/or signs of a respiratory illness consistent with pneumonia by verbal autopsy will also be considered a case of severe pneumonia. Children are permitted to meet the above case definition for severe pneumonia multiple times if the repeat event occurs either >14 days after the hospital discharge date or, if the hospital discharge data is not available or the child was not hospitalized, >30 days after the date the child met criteria for severe pneumonia. Implausible values and outliers identified from the variables that comprise the case definition will be considered missing.



Secondary Outcomes are given in Table 4.

Table 4. Secondary Outcomes

Parameter	Definition
IMCI (Integrated Management of Child Illnesses) WHO non-severe	Cough and/or difficult breathing and • chest indrawing or fast breathing and • no general danger sign (unable to drink or breastfeed, vomiting everything, convulsions, lethargic or unconscious) and • no stridor at rest and • no severe acute malnutrition and • no HIV infection or exposure and • no hypoxemia (SpO_2 <90%)
IMCI WHO severe	Cough and/or difficult breathing and • any general danger sign (unable to drink or breastfeed, vomiting everything, convulsions, lethargic or unconscious) or • stridor at rest or • severe acute malnutrition or • HIV infection or exposure (if chest indrawing also) or • Hypoxemia (SpO_2 <90%)
Pocketbook WHO non-severe	Cough and/or difficult breathing and • chest indrawing or fast breathing and • no sign of Pocketbook WHO severe pneumonia
Pocketbook WHO severe	Cough and/or difficult breathing and • central cyanosis or SpO_2<90% or • severe respiratory distress (grunting, very severe chest indrawing) or • sign of pneumonia (fast breathing, chest indrawing, or lung auscultation signs of decreased or bronchial breath sounds, crackles, abnormal vocal resonance, pleural rub) and at least

	one general danger sign (unable to drink or breastfeed, vomiting everything, convulsions, lethargic or unconscious) or other high-risk condition (severe acute malnutrition or HIV infection or exposure)
Hospitalization for respiratory illness	Hospitalization for respiratory illness at any time during the follow-up period
Lung ultrasound or chest radiograph pneumonia	Primary endpoint pneumonia Yes/No
Hypoxemic pneumonia	Categorized as SpO_2 <93% for Guatemala, India, and/or Rwanda. <87% for Peru.

5.3. Intention-to-Treat Analysis

For the primary outcome of the incidence of HAPIN pneumonia episodes (see Section 5.2) over a one-year follow-up period, we will use a Poisson regression model based on child-days at risk of follow up, to derive an incidence rate ratio for the intervention vs. control groups. The use of Poisson regression and rate ratios, rather than a relative risk of first occurrence of pneumonia, is motivated by observing that the occurrence of multiple severe pneumonias episodes in the study children, which occurs often enough to warrant taking the multiple events into account. We will account for multiple severe pneumonias episodes by estimating a dispersion parameter using quasi-likelihood estimation.

Specifically, the intention-to-treat analysis will be modelled as follows:

$$\log E(y_i) = \beta_0 + \beta_1 \times I(intervention_i) + \sum_{1}^{9} \beta_{k+1} I_{ik} + \log n_i ,$$

where y_i is the number of HAPIN pneumonia episodes for the i^{th}child during the first year of follow-up and $y \sim Poisson(\mu, \phi)$ with a variance function $var[y_i] = \phi \times E[y_i]$ where ϕ represents the dispersion parameter, $I(intervention_i)$ is an indicator variable that is 0 if the i^{th}child was assigned to the control group and 1 if assigned to intervention, $\hat{\beta}_2$ through $\hat{\beta}_{10}$ are the coefficients for the indicator variables ($I_{i.}$) representing nine strata (with one stratum is reference, for a total of 10 strata), and n_i represents the number of child-days at risk for the i^{th}child. For this generalized linear model, the canonical link is the logarithm function. The estimated parameter $\hat{\beta}_1$is the incidence rate ratio of HAPIN pneumonia between the intervention and control group.

The same Poisson regression model will be used for subgroup analyses and secondary outcomes.

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. A joint statistical test will be conducted to detect effect modification at a type I error rate of 0.05. Potential effect modifiers to be assessed are listed in Table 5.

Table 5. Subgroup analysis variables of liveborn children (effect modifiers)

Variables	Type	Definition/Assessment Methods
International Research Center and randomization stratification	Categorical	Randomization strata within Guatemala, India, Peru, Rwanda
Mother's age (years) at baseline	Continuous and 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. Will also include age categories.

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
Gestational age (weeks) at intervention	Continuous and Categorical	Calculated as the date at baseline minus the date of screening ultrasound plus gestational age at screening. Will also include gestational age categories.
Breastfeeding	Binary	Breastfeeding during the first six months of life, non-exclusive (yes/no/missing)
Up-to-date vaccination status	Binary	Receipt of 3 respiratory vaccines by one year of life, including 3 doses of PCV (*note*: PCV not available in India site), 3 doses of Hib vaccine, one dose of measles vaccine (yes/no/missing)

Secondary analyses. Secondary analyses include time-to-event analysis for estimating the intervention effect on the time to first pneumonia incidence and analysis of secondary outcome definitions as listed in Table 4 (section 5.2).

Additional Analysis. If imbalance between control and intervention groups for a baseline covariate (Section 4.4, Table 1) suggests problems with randomization, and the covariate is a potential confounder, covariate-adjusted effects will be evaluated as a sensitivity analysis.

We will also plan to assess the influence of (1) a marked *increase* of pneumonia cases in Rwanda from November 2019 due to surveillance strengthening at outpatient clinics, and (2) the COVID-19 pandemic, which started in April 2020 and may have contributed to an observed *decrease* in cases in all study sites. We may do this by following the analytical approach prescribed for the primary analysis with two indicator variables to time periods, with pre-November 1, 2019 as referent, November 1, 2019 to March 31, 2020 as the first indicator, and from April 1, 2020 as the second indicator, and/or evaluate using another appropriate modelling approach. We plan to include interaction terms with treatment arm and these time periods to assess whether the effect of intervention changed over time.

Incomplete Outcome Data. Our primary approach to incomplete outcome data that precludes assessment of a possible HAPIN pneumonia event will be to conservatively assume that such events are non-outcome events. For example, if the child's oxygen saturation and/or image is missing we assume that child is not a case, but keep the child in the denominator-at-risk rather than censor them. It is anticipated that incomplete outcome data will be rare and when present balanced between intervention arms. We will conduct additional analyses to evaluate the potential influence of any missing outcome data.

5.4. Analysis Replication Plan

Selected components of the intention-to-treat 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).

9. A cleaned analytic dataset where exclusions have been applied following the CONSORT diagram. The dataset will also include baseline variables and covariates for subgroup analysis.
10. A table summarizing baseline characteristics (overall and by IRC).
11. The set of outcomes (primary and secondary) and subgroup analysis to be replicated.

Specific replication tasks include:

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

APPENDIX 3: STATISTICAL ANALYSIS PLANS

3.D. Exposure-Response Analysis for Blood Pressure in Older Adult Women

Version 1.0
May 25, 2022

1. INTRODUCTION

This document contains the statistical analysis plan (SAP) for blood pressure in the older adult women, one of four primary outcomes of the HAPIN trial. This SAP will be posted with the trial registration.

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 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 household air pollution 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 LMIC settings (Tamil Nadu, India; Jalapa, Guatemala; Puno, Peru; and Kayonza, 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 enroll 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 are 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 year).

2.2. Randomization

To ensure 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, 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

The sample size and power for the trial were based on the intention to treat analyses for the primary outcomes.

2.4. Trial Framework

HAPIN is a superiority trial. The intention-to-treat analysis and exposure-response analysis (described here) will be conducted 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 is 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 participating older adult woman in a household the follow up time is approximately 18 months, including baseline measurements (prior to randomization and intervention), when the pregnant woman is at 24-28 and 32-36 weeks of gestation, and when the child is 3 months, 6 months, and 12 months of age (up to 6 total visits).

3. STATISTICAL PRINCIPLES

3.1. Confidence Intervals and P-Values

All confidence intervals will be presented at 95% confidence. Analysis of air pollution exposure-response associations will use an α-level of 0.05 to identify statistical significance (including effect modification). If the effect modifiers 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) on their traditional stoves, as well as a subset of approximately 80 homes in the control arm of each IRC. 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.

3.3. Analysis Populations

For each outcome, the analysis will include all 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 study participant and withdrawal from study prior to measurement.

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.

An **older adult woman** in the same household will be eligible to participate in the study if they fulfill the following inclusion and exclusion criteria at screening:

Inclusion criteria:

- Aged 40 to <80 years (via self-report)

Exclusion criteria:

- Currently smokes cigarettes or other tobacco products
- Pregnant (via self-report)
- Plans to move out of her current household in the next 12 months
- Takes blood pressure medication at enrollment and/or any point during follow-up (will be included in sensitivity analysis, below)

If two or more older adult women live in the same household and are interested in participating, one woman will be randomly selected to participate (the one with the next birthday[month and day]).

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
- Participants determined to be ineligible after randomization
- Reasons for exits after randomization
  - Voluntary withdrawal
  - Withdrawn by study team
  - Moved away
- Reasons for exclusion due to missing data

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
- 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.

5. DATA ANALYSIS

In this section we provide the analysis approach for the exposure-response aim. The primary outcome is systolic blood pressure. We present the primary analysis along with secondary outcomes, subgroup (effect modification), sensitivity, and additional analyses.

5.1. Outcome Definitions

This section describes the outcomes, including data collection approaches and calculations for derived outcomes.

Following the 24-hour exposure assessment period, a nurse or trained field worker measured resting blood pressure in the right arm in triplicate (with at least 2 minutes between measurements) using an automatic monitor (model HEM-907XL; Omron®) at the participant's home. Before starting the measurement, the participant was instructed to sit on a chair in a quiet room for 5 min with legs uncrossed, their back supported by the chair, and their arm supported on a table. The participant also confirmed that she had not smoked, consumed alcohol, or caffeinated beverages (coffee, tea, or Coca-Cola), or cooked using biomass in the past 30 minutes. If she had done any of those activities in the 30 minutes prior to the measurement, she was asked to refrain from doing these activities for 30 minutes before proceeding with the measurements.

A participant with a measured systolic blood pressure $\geq$ 140 mmHg and/or a diastolic blood pressure $\geq$ 90 mmHg was checked again during the same visit. If the same result was observed on two measurements, the participant was referred to the nearest health center or hospital to receive age-appropriate treatment. If a participant had systolic blood pressure < 80 mmHg or diastolic blood pressure < 40 mmHg, she was also referred to the nearest health center or hospital. In analyses, the average of all three blood pressure measurements will be used. Systolic blood pressure values less than 70 mmHg and diastolic blood pressure values less than 35 mmHg will be excluded as implausible.

Pulse pressure is the difference between systolic blood pressure (SBP) and diastolic blood pressure (DBP) and mean arterial pressure is calculated as $DBP + (SBP - DBP)/3$.

5.2. Exposure-Response Analysis

For each pollutant (PM2.5, black carbon and CO), time-weighted average exposures will be estimated using 24-hr personal measurements at baseline, and al follow-up visits (up to 5). For participants in the control group, an average will be calculated from all available measurements. For participants in the intervention group, baseline exposure levels will be weighted by days prior to LPG installation, while post baseline measurements will be weighted by the days after LPG installation. For the intervention group, if the baseline measurement is missing the mother will be excluded from the analysis.

The following two approaches will be used for both primary and secondary blood pressure outcomes. After excluding any women who reported using blood pressure medication at any visit, we will estimate the association between long-term exposure and the blood pressure outcome using the regression model given by

$$E[d_i] = \beta_0 + f(X_i) + \boldsymbol{\gamma Z}_i$$

where d is the change in outcome between the 6th follow-up measurement (B4 visit) and the baseline measurement, $f(X_i)$ is a function that uses the average exposure of interest across the follow-up period, defined above, and $\boldsymbol{Z}_i$ is the vector of confounders. This model evaluates the change in BP from baseline to end of follow-up, assessing the cumulative effect of exposure over the whole period.

Confounders and covariates to be included in the model are listed in Table 2, chosen a priori.

We will also use the following random-intercept model to analyze repeated measurements of blood pressure (separately in relation to each of our three pollutants) given by

$$E[Y_{ij}] = \beta_0 + \theta_i + \beta_1 X_i + \beta_2 (X_{ij} - X_i) + \boldsymbol{\gamma Z}_i$$

where y_{ij} denotes the outcome for woman i at time point j, θ_i is the woman-specific random intercept, and $(X_{ij} - X_i)$ is the woman-specific deviation for exposure at time point j. Here parameter β_1 reflects the 'long term' or average effect of exposure across the study period. The exposure variable X_i will be the average of all measurements as described above. The parameter β_2 describes the within-woman short-term association between exposure and outcome, adjusted for impacts of long-term exposures (both such effects have been found in studies of ambient $PM_{2.5}$) This model contains two special cases: (1) when $\beta_1 = \beta_2$, the model reduces to a random-intercept model that only uses the time-varying exposure directly, and (2) when $\beta_2 = 0$, the model reduces to a random-intercept model with only average exposure.

In the exposure-response analyses, all models will be adjusted for the covariates listed in Table 2. Confounder selection are based on conceptual directed acyclic graphs and from previous studies. Additional covariates included to explain variance in the outcome (e.g., age) to increase power.

Table 2. A priori covariate adjustments in exposure-response analyses

Parameter	Type	Subgroup Definitions
International Research Center	Categorical	Guatemala, India, Peru, Rwanda
Baseline: Age at baseline (years)	Continuous	Calculated as the date at baseline minus the date of birth. Date at baseline is assigned by the date of visit if not missing.
Baseline: 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
Baseline: Body mass index (BMI)	Continuous	BMI calculated as the average weight (kg) divided by the average height squared (m^2)
Baseline: 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
Baseline: Minimum diet diversity	Categorical	Categories (corresponding diet diversity score): • Low (< 4)

		• Medium (4-5) • High (>5) • Missing
Time of the blood pressure measurement; time-varying	Categorical	AM or PM

We will evaluate the shape of the relationship between blood pressure and exposures. Non-linear associations between blood pressure and exposure will be evaluated via (1) log transformation of the exposure, (2) exposure categories based on quartiles, (3) regression splines of varying order and number of internal knots (we will start with a knot selection based on quartiles), and (4) penalized smoothing splines. Model comparison will be based on traditional goodness-of-fit (e.g., plotting observed and predicted values, use of residual plots and added variable plots, use of R^2) and information criteria to measure prediction error (i.e., AIC) to identify the most parsimonious model.

Subgroup Analysis. Subgroup (effect modification) analyses will be conducted using interaction terms between each of our three pollution variables and the effect modifiers. The list of pre-specified subgroup analyses for the exposure-response analysis is given in Table 3.

Table 3: Definition for variables for subgroup analysis for exposure-response analysis

Parameter	**Subgroup Definitions**
Age	Continuous and dichotomous (using median age)
International Research Center	Guatemala, India, Peru, Rwanda
BMI	Continuous and categorical (underweight, less than 18.5 and normal, 18.5-24.9, combined; overweight/obese, 25.0 and above). And also using 3 categories (underweight, normal, overweight/obese).
Study Arm	Categorical; intervention vs. control

Secondary Outcomes: We will evaluate diastolic blood pressure, pulse pressure, and mean arterial pressure as secondary outcomes.

Sensitivity Analysis. We will conduct an analysis including women taking blood pressure medication. We will also evaluate other potential confounders (physical activity, reported smokers in household, recent caffeine intake) by including them in the model.

Additional Analysis: We will consider extend the random intercept model to (1) include a linear time trend and its interactions with average exposure (X_i) for estimating associations between exposure and outcome trajectory, and (2) consider time-varying exposures defined as cumulative averages.

Missing Data. Our primary approach to missing outcome data will be a complete-case analysis by excluding participants with a missing baseline blood pressure measurement. All follow-up measurements will be used in the ITT analysis. It is anticipated that missing blood pressure will be balanced between intervention arms.

5.3. Analysis Replication Plan

Selected components of the exposure-response analyses will be replicated by an independent analyst. Sensitivity analyses will not be replicated.

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

12. A cleaned analytic dataset where exclusions have been applied following the CONSORT diagram. The dataset will also include covariates for subgroup analysis and covariates to include in the exposure-response analyses.
13. The set of outcomes (primary and secondary) and subgroup analysis to be replicated.
14. The list of pre-specified covariates to be included in the regression models and forms of the exposure-response function.

Specific replication tasks include:

10. Replicate summary statistics (e.g., mean, standard deviation, percentages, proportion missing) in the baseline characteristic table.
11. Replicate exposure-response analyses for primary and secondary outcomes according to models specified in Section 5.2

APPENDIX 3: STATISTICAL ANALYSIS PLANS

3.E. Intention-to-Treat Analysis for Blood Pressure in Older Adult Women

1. INTRODUCTION

This document contains the intention-to-treat (ITT) statistical analysis plan (SAP) for blood pressure in the older adult women, one of four primary outcomes of the HAPIN trial. This SAP will be posted with the trial registration.

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 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 household air pollution 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 LMIC settings (Tamil Nadu, India; Jalapa, Guatemala; Puno, Peru; and Kayonza, 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 enroll 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 are 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 year).

2.2. Randomization

To ensure 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, 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 power calculation for blood pressure, we assume the number of post-randomization to be $m = 5$ and an intraclass correction of $p = 0.76$ for all individuals. The power for detecting a difference in average blood pressure (β_1) in an analysis-of-covariance analysis with baseline blood pressure adjustment for n individuals per arm can be approximated by Frison and Pocock (1992):

$$\Phi\left(-Z_{1-\alpha/2} + \frac{|\beta_1|}{\sqrt{\frac{2\sigma^2}{nm}[1 + (m-1)p - mp^2]}} \right).$$

Table 1. Minimum Detectable Effects with 80% Power and α = 0.0125.*

	Parameter	Sample Size Per Arm	-20%	-10%	Original	+10%	+20%	Previous Studies Estimate (95% CI)
Change in mean blood pressure (mm Hg)	σ^2	200	2.30	2.44	2.58	2.70	2.82	3.7 (-0.6, 8.1)[a]

*We examined the impact of power when the key population parameter (residual variance of blood pressure $\sigma^2 = 12^2$ are set to ±10% and ±20% of their values. Calculations assume a 10% attrition during follow-up.
[a]McCracken et al. (2007) is a study of 120 older women (>38) from the same randomized trial (RESPIRE), which adjusted for confounders.

2.4. Trial Framework
HAPIN is a superiority trial. The intention-to-treat analysis is a test to evaluate whether the outcome data are consistent with the assumption of there being no difference between the intervention and control arms.

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 is 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 participating older adult woman in a household the follow up time is approximately 18 months, including baseline measurements (prior to randomization and intervention), when the pregnant woman is at 24-28 and 32-36 weeks of gestation, and when the child is 3 months, 6 months, and 12 months of age (up to 6 total visits).

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 blood pressure outcome (systolic blood pressure) will utilize repeated measures to examine differences in mean blood pressure post-randomization while adjusting for baseline systolic blood pressure; the statistical inference test will be a Wald 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.

Analyses of secondary blood pressure outcomes and subgroup analysis will use an α-level of 0.05 to identify statistically significance. If the effect modifiers 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) on their traditional stoves, as well as a subset of approximately 80 homes in the control arm of each IRC. 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 will not consider adherence.

3.3. Analysis Populations

For each outcome, the analysis will include all 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 study participant and withdrawal from study prior to measurement. 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.

An **older adult woman** in the same household will be eligible to participate in the study if they fulfill the following inclusion and exclusion criteria at screening:

Inclusion criteria:

- Aged 40 to <80 years (via self-report)

Exclusion criteria:

- Currently smokes cigarettes or other tobacco products
- Pregnant (via self-report)
- Plans to move out of her current household in the next 12 months
- Takes blood pressure medication at enrollment and/or point during follow-up (will be included in sensitivity analysis, below)

If two or more older adult women live in the same household and are interested in participating, one woman will be randomly selected to participate (the one with the next birthday[month and day]).

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
- Participants determined to be ineligible after randomization
- Reasons for exits after randomization
  - Voluntary withdrawal
  - Withdrawn by study team
  - Moved away
- Reasons for exclusion due to missing data

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
- 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 2. 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 2. Baseline characteristics of the older adult women and the household to be reported (including participants randomized and not taking blood pressure medication at any time during follow-up)		
Variables	**Type**	**Definition/Assessment Methods**
Age (years)	Continuous	Calculated as the date at baseline minus the date of birth. Date at baseline is assigned by the date of visit if not missing.
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
Body mass index (BMI)	Continuous	BMI calculated as the average weight (kg) divided by the average height squared (m^2)
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
Minimum diet diversity	Categorical	Categories (corresponding diet diversity score): • Low (< 4) • Medium (4-5) • High (>5) • Missing
Number of people who sleep in this house	Continuous	Number of people reported to sleep in the house
Second-hand smoking	Categorical	Whether someone in the household other than the older adult woman smokes (smoking of the older adult woman 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 ITT analysis. The primary outcome is systolic blood pressure. We present the primary analysis along with secondary outcomes, subgroup (effect modification), sensitivity, and additional analyses.

5.2. Outcome Definitions

This section describes the outcomes, including data collection approaches and calculations for derived outcomes.

Following the 24-hour exposure assessment period, a nurse or trained field worker measured resting blood pressure in the right arm in triplicate (with at least 2 minutes between measurements) using an automatic monitor (model HEM-907XL; Omron®) at the participant's home. Before starting the measurement, the participant was instructed to sit on a chair in a quiet room for 5 min with legs uncrossed, their back supported by the chair, and their arm supported on a table. The participant also confirmed that she had not smoked, consumed alcohol, or caffeinated beverages (coffee, tea, or Coca-Cola), or cooked using biomass in the past 30 minutes. If she had

done any of those activities in the 30 minutes prior to the measurement, she was asked to refrain from doing these activities for 30 minutes before proceeding with the measurements.

A participant with a measured systolic blood pressure ≥ 140 mmHg and/or a diastolic blood pressure ≥ 90 mmHg was checked again during the same visit. If the same result was observed on two measurements, the participant was referred to the nearest health center or hospital to receive age-appropriate treatment. If a participant had systolic blood pressure < 80 mmHg or diastolic blood pressure < 40 mmHg, she was also referred to the nearest health center or hospital. In analyses, the average of all three blood pressure measurements will be used. Systolic blood pressure values less than 70 mmHg and diastolic blood pressure values less than 35 mmHg will be excluded as implausible.

Pulse pressure is the difference between systolic blood pressure (SBP) and diastolic blood pressure (DBP) and mean arterial pressure is calculated as $DBP + (SBP - DBP)/3$.

5.3. Intention-to-Treat Analysis

The following approach will be used for both primary and secondary blood pressure outcomes. After excluding any women who reported using blood pressure medication at enrollment and/or at any visit, repeated measurements of blood pressure of the older adult woman will be analyzed using a model to examine differences in mean blood pressure post-randomization, while adjusting for baseline BP. Specifically, let y_i denote the blood pressure for woman i. The regression model is given by

$$\bar{y}_i^{Post} = \beta_0 + \beta_1 X_i + \beta_2\left(y_i^{Base} - \bar{y}^{Base}\right) + \varepsilon_i$$

where for individual i, $\bar{y}_i^{Post}$ is the average post-randomization blood pressure measurement, X_i is an indicator variable (0 for control and 1 for intervention), $\left(y_i^{Base} - \bar{y}^{Base}\right)$ is the centered baseline measurement of blood pressure, and $\varepsilon_i \sim N(0, \tau_i^2)$ represents independent Normal error. The parameter of interest β_1 captures differences in average blood pressure due to the intervention, accounting for baseline blood pressure A Wald's test will be conducted under the null hypothesis of $\beta_1 = 0$.

The variance τ_i^2 is a function of the number of post-randomization measurements for individual i and the between-measurement correlations. Specifically, τ_i^2 is given by

$$\tau_i^2 = \frac{\sigma^2}{m_i}\left[1 + (m_i - 1)p - m_i p^2\right]$$

where m_i is the number of post-randomization measurements, σ^2 is the cross-sectional between-individual variance of blood pressure at any time point, p is the intra-class correlation of blood pressure among both baseline and post-randomization measurements. Our primary analysis will assume τ_i^2 to be constant for all participants. In sensitivity analyses, we will estimate p from a random-intercept model and use weighted least-square to account for heterogenous residual variance τ_i^2 due to the different number of follow-up measurements across individuals.

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 3. For each categorized subgroup variable a joint (simultaneous) statistical test (Wald chi-square test for categorical variables; t test for continuous variables) will be conducted to evaluate effect modification.

Table 3. Definition for variables for subgroup analysis for intention-to-treat analysis

Parameter	Subgroup Definitions
Age	Continuous and dichotomous (using median age)
International Research Center	Guatemala, India, Peru, Rwanda
BMI	Continuous and categorical (underweight, less than 18.5 and normal, 18.5-24.9, combined; overweight/obese, 25.0 and above); And also with 3 categories (underweight, normal, overweight/obese).

Secondary Outcomes: We will evaluate diastolic blood pressure, pulse pressure, mean arterial pressure, and change between baseline and last follow-up measurement (~18 months after baseline, when the child is 1 year old) as secondary outcomes.

Sensitivity Analysis. If imbalance between control and intervention groups for a baseline covariate (Table 2) suggests problems with randomization, and the covariate is a potential confounder, covariate-adjusted effects will be evaluated as a sensitivity analysis. We will also conduct an analysis including women taking blood pressure medication.

Additional Analysis: We will also consider mixed effect models to evaluate changes in blood pressure trajectories with participant-specific random intercepts and random slopes. We will also analyze the difference between baseline and final BP ('change score'). This additional does not include baseline BP as a covariate, and will be of particular importance if there is an imbalance in baseline BP between arms, as in some instances including baseline measurements can introduce bias

Missing Data. Our primary approach to missing outcome data will be a complete-case analysis by excluding participants with a missing baseline blood pressure measurement. All follow-up measurements will be used in the ITT analysis. It is anticipated that missing blood pressure will be balanced between intervention arms.

5.4. Analysis Replication Plan

Selected components of the intention-to-treat analysis will be replicated by an independent analyst. Sensitivity analyses will not be replicated.

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

15. A cleaned analytic dataset where exclusions have been applied following the CONSORT diagram. The dataset will also include characteristics at baseline and covariates for subgroup analysis.
16. A table summarizing characteristics at baseline (overall and by IRC).
17. The set of outcomes (primary and secondary) and subgroup analysis to be replicated.

Specific replication tasks include:

12. Replicate summary statistics (e.g., mean, standard deviation, percentages, proportion missing) in the baseline characteristic table.

13. Replicate intention-to-treat analyses for primary and secondary outcomes according to models specified in Section 5.3.
14. Replicate results from subgroup analyses (effect modification) for primary and secondary outcomes