Reducing Childhood Lead Exposure in Ghana Pre-analysis Plan*

Lee Crawfurd[†] Carl Stephen Osei[‡] Usamatu Salifu[§]
Justin Sandefur[¶] Hannah Timmis[‖]

August 13, 2026

Abstract

This document presents a pre-analysis plan for a randomised controlled trial evaluating the contribution of contaminated cookware and traditional eyeliner to child lead exposure in Ghana's Northern Region. We randomise 300 households to receive lead-free replacements for cookware, eyeliner, both, or neither, and estimate the causal effects of contaminated product use on children's blood lead levels using instrumental variables. Should one or more interventions prove effective, a second phase of the study will evaluate the causal effects of reduced lead exposure on children's cognitive development. The findings will improve understanding of the contribution of household consumer products to child lead exposure and inform the design of larger-scale lead abatement interventions.

Keywords: Lead exposure, child health, environmental health
JEL Codes: I18, Q53, C26

*We are deeply grateful to the teams at Pure Earth and the Ghana Health Service for their partnership in implementing this research, and for the expertise, commitment, and insights they have contributed throughout the project. We also acknowledge the generous financial support of Coefficient Giving, Founders Pledge, and Pure Earth. The research was approved by the Oxford Tropical Research Ethics Committee at the University of Oxford (Ref: 722127) and the Ghana Health Service Ethics Review Committee (Ref: 034/05/25).

[†]Center for Global Development, lcrawfurd@cgdev.org
[‡]Ghana Health Service, carl.osei@ghs.gov.gh
[§]Kwame Nkrumah University of Science and Technology, usalifu@st.knust.edu.gh
[¶]Coefficient Giving, justin.sandefur@coefficientgiving.org
[‖]University of Oxford, hannah.timmis@bsg.ox.ac.uk (**Corresponding author**)

1 Introduction

Around half of children in developing countries have elevated blood lead levels (BLLs), defined as at or above 5 micrograms per decilitre (μg/dL) (Crawfurd et al., 2024). Lead is a highly toxic metal, and exposure during childhood is associated with impaired cognitive development, reduced educational attainment, and adverse health outcomes (Larsen and Sánchez-Triana, 2023; Crawfurd et al., 2024). Because these damages are largely irreversible, effective policy must focus on limiting children's exposure.

Doing so is made challenging by two evidence gaps. First, we know relatively little about the sources of lead exposure across contexts, and thus which products, processes, or behaviours should be targeted by policymakers. Second, there is also limited evidence on the *causal* relationship between children's BLLs and developmental outcomes, making it difficult to compare the social returns to lead abatement with those of other health and education interventions (Bonnifield and Todd, 2024a).

This study addresses both gaps by experimentally evaluating the effects of removing potential lead sources in children's homes on their BLLs and cognitive performance. We implement a two-phase randomised controlled trial (RCT) in Ghana's Northern Region, in partnership with Pure Earth, an international NGO specialising in lead exposure reduction, and the Ghana Health Service (GHS), the country's public health agency. In Phase I, we randomise 300 households to receive lead-free replacements for two commonly contaminated products—cookware and cosmetics—and measure changes in children's BLLs using blood samples. This phase estimates the share of child exposure attributable to these home-based sources in the study sample. In Phase II, we scale the most effective intervention(s) to a larger sample and use the resulting exogenous variation in BLLs to estimate the causal effects of exposure on children's cognition over 12—18 months.

This document outlines our pre-analysis plan for the Phase I RCT, including the study context (Section 2), experimental design (Section 3), measurement approach (Section 4),

and empirical strategy (Section 5). Should one or more of the Phase I interventions prove effective, we will submit a supplementary pre-analysis plan for Phase II.

2 Context

Our study takes place in Tolon and Yendi districts in Ghana's Northern Region. These districts, among the poorest in the country,[1] were identified as lead exposure hotspots in a recent large-scale BLL survey conducted by Pure Earth, GHS, and UNICEF (Nash et al., 2025). The survey tested 3,277 children aged 1-5 years across nine selected district capitals using a portable LeadCare II device;[2] it found that 77% and 79% of children in Tolon and Yendi towns, respectively, had elevated BLLs, compared with an overall average of 54%.

Where this lead is coming from is unclear. To investigate potential exposure pathways, Pure Earth et al. conducted detailed home-based assessments (HBAs) for a random subset of 288 surveyed children. The assessments measured lead concentrations in soil, dust, paint, drinking water, cookware, cosmetics, cooking spices, and other items found in children's homes, and collected survey data on household exposure risks and child health. While most items showed low contamination rates, soil, dust, metal cookware, and traditional eyeliner ("chilo") frequently had elevated lead concentrations.[3] The extent to which these sources contribute to children's lead exposure remains uncertain, however. The HBAs found that contaminated soil and chilo use were strongly associated with elevated child BLLs,[4] whereas contaminated dust and contaminated cookware were not. Yet these associations should be

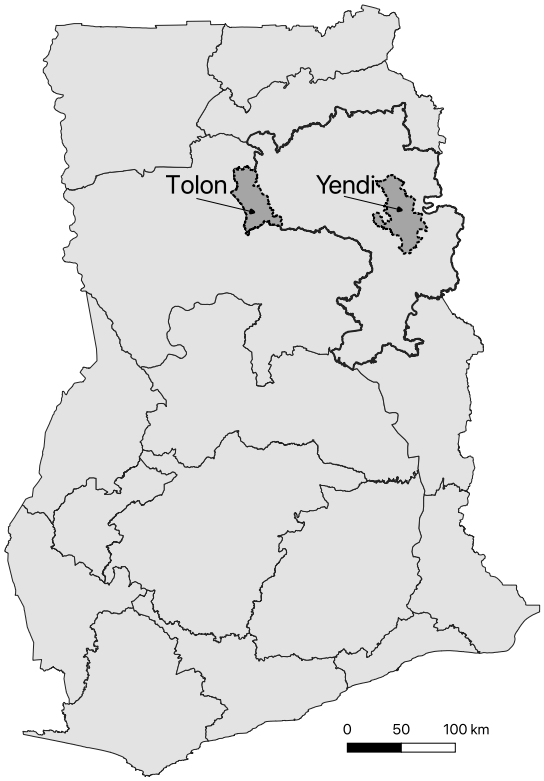
[1] Tolon and Yendi rank in the top quartile of districts on Ghana's Multidimensional Poverty Index (MPI), which measures deprivations in health, education, and living standards; 58% and 39% of their populations are classified as multidimensionally poor, respectively (Ghana Statistical Service, 2021).

[2] LeadCare II is a point-of-care device that uses anodic stripping voltammetry to measure blood lead concentrations from capillary blood samples. It is widely used for large-scale lead screening in field settings.

[3] Chilo is commonly applied to young children in Northern Ghana for religious, aesthetic, and perceived health-related reasons.

[4] Because few households had eyeliner samples that were suitable for testing, the study could not examine the association between total lead concentrations and child BLL for this product. Instead, it reported the association between BLL and whether households reported regularly applying chilo to the child.

Figure 1: Study districts in Ghana



Notes: The map shows Ghana's 16 administrative regions, with the Northern Region outlined in bold and the study districts of Tolon and Yendi highlighted.

interpreted with caution. Chilo use was heavily concentrated in the Northern Region, raising the possibility that the observed association is confounded by unobserved geographic factors. Moreover, a review of the available evidence on traditional eyeliner use and child BLL has reached mixed conclusions (Bonnifield and Todd, 2024a). For cookware, laboratory studies suggest that total lead concentrations may be a poor proxy for realised exposure because the amount of lead that leaches into food depends on cooking practices, cookware condition, and other factors (Binkhorst et al., 2025).

This study therefore experimentally evaluates the contribution of contaminated metal cookware and traditional eyeliner to children's lead exposure. We focus on these products because, unlike soil and dust, they can be readily removed from children's homes, making

them well-suited to a randomised intervention. Both are also widely used beyond Ghana. A recent study of consumer goods covering 25 low- and middle-income countries found that 51% of metal foodware samples had lead levels above the reference value, with contaminated items identified in every country (Sargsyan et al., 2024). The same study also found that, although only 12% of cosmetic samples were contaminated, traditional eyeliners widely used across West and North Africa, the Middle East, and South Asia sometimes contained extremely high lead concentrations.

3 Experimental Design

3.1 Sampling

Our study population comprises children aged 4–6 years living in rural areas of Tolon and Yendi districts. Eligible children are required to have BLLs of at least 3.5 μg/dL,[5] reside in households that use metal cookware and apply eyeliner to the child regularly (at least once per month), and have no history of bleeding disorders or specific eye illnesses.[6] We focus on children aged 4–6 because lead exposure is particularly harmful during early childhood, and children in this age range are old enough to complete the cognitive assessments planned for Phase II of the RCT, described above.

We sampled 300 eligible children using a three-step process:

1. **Site selection:** Tolon and Yendi were divided into 513 enumeration areas (EAs), each containing approximately 150 households, for the 2021 Ghana Population Census. We excluded 167 urban EAs and a further 75 large, sparsely populated EAs ($> 2.5\text{km}^2$), where geographic dispersion made BLL screening operationally infeasible. We ran-

[5] Although much of the literature defines elevated BLL as ≥ 5 μg/dL, we use a threshold of 3.5 μg/dL to align with the US Centers for Disease Control and Prevention's (CDC) latest blood lead reference value.

[6] Children with evidence of severe allergic conjunctivitis or corneal ulcers were excluded since this could be exacerbated by eyeliner application.

domly selected 20 EAs from the remaining sampling frame, stratifying by district. Early screening indicated substantially higher rates of both eyeliner use and elevated BLLs in Tolon than in Yendi. To achieve the target sample size within available resources, we therefore replaced three Yendi EAs with ten additional Tolon EAs drawn from localities where screening had revealed the highest prevalence of elevated BLLs.[7] The final sample comprised 27 EAs across 15 localities: 10 in Tolon and five in Yendi.

2. **BLL screening:** We conducted a household census in the selected EAs to identify potentially eligible children. Households containing children aged 4–6 years who reported using metal cookware and applying eyeliner were invited to participate in BLL screening. One child per household (the youngest in the eligible age range) was tested using LeadCare II, and the results were shared with caregivers, together with information on their local district health directorate for further advice and support. In total, 750 children were screened across the 27 EAs, of whom 454 (61%) had BLLs above 3.5 μg/dL.

3. **Sample selection:** The 400 households whose children had the highest BLLs completed a modified home-based assessment (HBA),[8] to confirm eligibility and collect pre-trial information on household lead exposure risks and child health. This process identified 315 children who met all the inclusion criteria.[9] The 300 children with the highest BLLs were invited to participate in the study, while the remaining 15 were assigned to a reserve sample to be enrolled if a selected child later proved ineligible or declined to participate during the baseline visit.

[7] Localities are population clusters defined by the Ghana Statistical Service that comprise one or more EAs.

[8] The modified HBA was a shortened version of the assessment described in Section 2, retaining only dust, cookware, and eyeliner testing, plus the household survey. Items shown to have low contamination rates in the earlier study were not tested. Dust wipes were collected from all households; cookware and eyeliner tests were discontinued, as discussed below.

[9] Most exclusions occurred because households that had previously reported applying eyeliner during the census either were no longer using it or applied it less frequently than required for eligibility. A small number of households could not be re-contacted.

The modified HBA originally included X-ray fluorescence (XRF) testing of households' cookware and eyeliner, following the protocol described in Section 4.1, to construct baseline measures of contaminated product use. However, this was discontinued after 130 households due to concerns that revealing products' lead concentrations could alter participant behaviour prior to the intervention and reduce statistical power. Consequently, baseline measures of contaminated product use are available only for a non-random subset of the study sample; Section 5.1 discusses implications for the analysis.

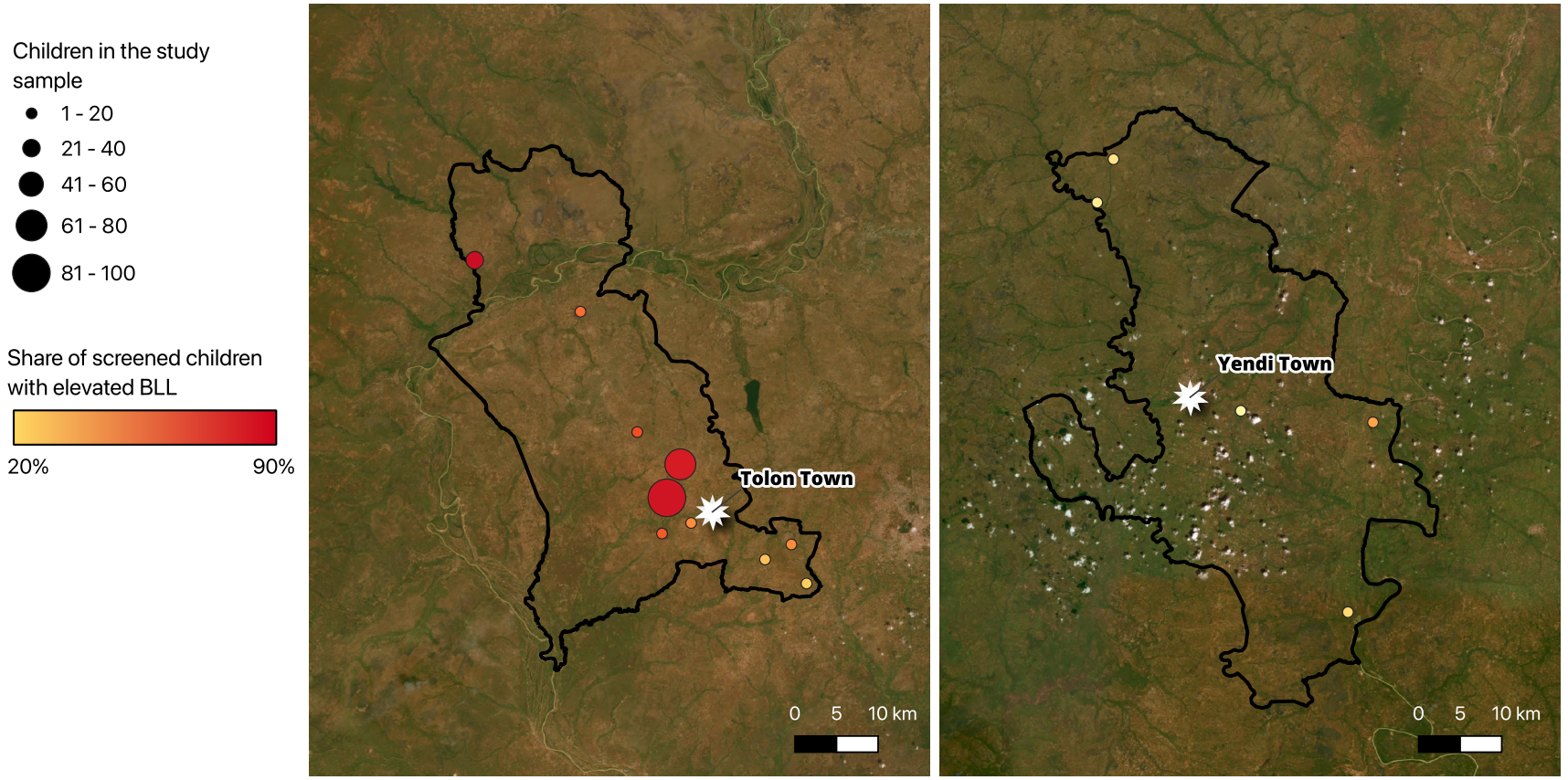
Figure 2 maps the geographic distribution of the final study sample and the share of screened children with elevated BLLs by locality.

3.2 Randomisation and interventions

The 300 sampled children were randomised into four study arms: control ($N = 101$), cookware replacement ($N = 102$), eyeliner replacement ($N = 49$), and cookware and eyeliner replacement ($N = 48$). More children were assigned to the cookware replacement arm because Pure Earth et al.'s previous survey found a weaker association between child BLLs and contaminated cookware use than between child BLLs and contaminated eyeliner use (Nash et al., 2025). We therefore assumed smaller treatment effect sizes for this arm and allocated the sample accordingly. Appendix A provides further details on this assumption and the resulting power calculations.

The randomisation was stratified by locality and screening BLL to help ensure balance on our primary outcome of interest. Within each locality, children were ranked by their LeadCare II result, divided into blocks of six, and randomised in a 2:2:1:1 ratio across the four groups. Where the number of children in a locality was not divisible by six, the remaining children were randomised independently using the same 2:2:1:1 allocation probabilities to preserve the intended treatment shares overall. Table **??** reports a preliminary balance check for key child and household characteristics collected during the census, BLL screening,

Figure 2: Map of study localities in Tolon and Yendi districts



Notes: Each point represents one of the 15 sampled localities. Colour indicates the share of screened children with elevated BLL ($\geq$3.5 μg/dL) in that locality; circle size indicates the number of study children in that locality, excluding the 15 reserve children. Screening BLLs were measured using a LeadCare II point-of-care device.

and modified HBA.

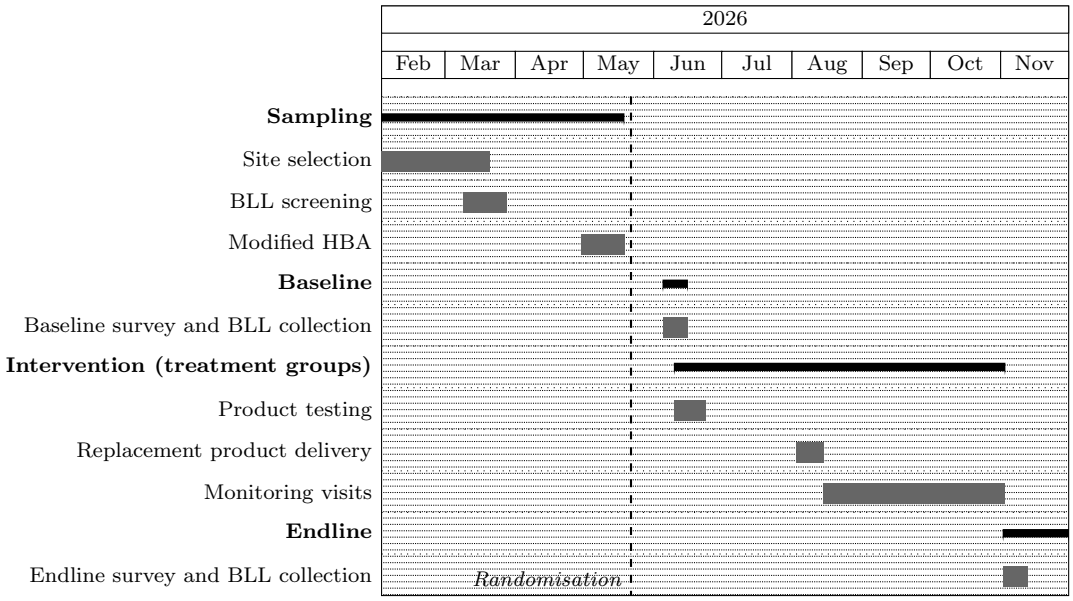
The households of children assigned to a treatment arm will be offered lead-free replacements for all contaminated cookware and/or eyeliner items. Treatment households whose products were *not* XRF-tested during the modified HBA were visited immediately after baseline data collection was completed in their locality, and their cookware and/or eyeliner were tested for lead following the protocol described in Section 4.1. Households whose products had already been tested were not re-visited; the original measurements were used instead to identify contaminated items. All treatment households will be informed of the XRF results and offered replacements for items with average lead concentrations above established reference levels: 100 parts per million (ppm) for cookware and 2 ppm for eyeliner. Replacement cookware will be of a similar size to the original item where possible. Households that accept replacement items will have their contaminated products removed and will receive monthly monitoring visits to assess compliance and determine whether replacements have been sold, lost, depleted, or transferred to other households.

The interventions were piloted prior to this study to test delivery procedures and gather household feedback on the replacement products. Twenty households in Tolon and Yendi that had been identified as using contaminated cookware and applying chilo during Pure Earth et al.'s previous survey received lead-free replacements and fortnightly monitoring visits. Follow-up surveys conducted two months later indicated that pilot households had used the replacements and had overwhelmingly positive feedback, with most respondents describing the new products as high quality and reporting no regrets about participating.

3.3 Data collection and timeline

Following randomisation, baseline data collection was conducted sequentially across localities for the 300 study households. This round collected children's BLLs using ICP-MS (see Section 4.2), providing the pre-treatment measure of our primary outcome, together with

Figure 3: Timeline of study activities



Notes: The figure is organised by field activity. Product testing was conducted only for treatment households whose cookware and eyeliner were not tested during the modified HBA.

a short survey capturing household covariates that were not collected during the census and modified HBA. To ensure households' treatment assignment did not bias the baseline measurements, surveyors were blinded to treatment assignment, and product testing for the treatment groups began only after the baseline was completed in each locality.

The study timeline is presented in Figure 3. Sampling activities, including site selection, BLL screening, and the modified HBA, were conducted between February and May 2026. The final study sample was selected and randomised on 21 May. Baseline data collection and product testing for treatment households were conducted in June. Replacement cookware and eyeliner are expected to be distributed during the first half of August.[10] Endline data collection, which will measure the first- and second-stage outcomes described in Sections 4.1 and 4.2, will begin approximately three months after product distribution, in November.

The three-month follow-up period was chosen because blood lead has a relatively short half-life of approximately 20–30 days, meaning that any intervention-induced changes in lead exposure should be fully reflected in children's BLLs within 75–100 days (U.S. EPA, 2024).

[10]The gap between product testing and delivery reflects a short delay in procuring the replacement items.

4 Measurement

4.1 First-stage outcomes: Contaminated product use

The first-stage outcomes are measures of contaminated product use in the sampled child's household. During the endline survey, we will collect the following information for every metal cookware and eyeliner item currently in the child's household, including original products and study-provided replacements:

- **Lead concentration (ppm)**: Lead concentrations of original household products will be measured in situ using a portable Thermo Scientific Niton XL3T XRF analyser. Cookware will be tested in Test All mode using three measurements (two from the interior cooking surface and one from the exterior) to account for within-item heterogeneity, while eyeliner will be tested in Soil mode using two measurements. The mean concentration for each item will be used in the analysis. Study-provided replacement products were tested and confirmed to contain no lead prior to distribution and will therefore be assigned a lead concentration of zero.
- **Frequency of use**: The child's caregiver will be asked (a) how many meals the sampled child consumed at home during the previous seven days and, for each cookware item, how many of those meals were prepared using that item; and (b) how many times each eyeliner item was applied to the child during the same period.

Using these data, we construct two complementary measures of contaminated product use. The first is a binary indicator for the presence of any contaminated product in the child's household:

$$I_i^C = \mathbf{1}\left\{\exists\, j : Lead_{ij}^C > 100\right\}, \qquad I_i^E = \mathbf{1}\left\{\exists\, j : Lead_{ij}^E > 2\right\},$$

where $Lead_{ij}^C$ is the mean lead concentration of cookware item j in ppm, and $Lead_{ij}^E$ is defined

analogously for eyeliner. The second measure constructs an approximate lead-dose index by weighting each product's lead concentration by its frequency of use:

$$U_i^C = \frac{1}{1000} \sum_j Lead_{ij}^C \times Meals_{ij}^C, \qquad U_i^E = \frac{1}{1000} \sum_j Lead_{ij}^E \times Applications_{ij}^E.$$

Here, $Meals_{ij}^C$ is the number of meals prepared using cookware item j for the sampled child during the previous seven days, and $Applications_{ij}^E$ is the number of times eyeliner item j was applied during the same period. We divide each index by 1,000, so that one unit corresponds to 1,000 ppm-meals of cookware use or 1,000 ppm-applications of eyeliner use.

The two measures are used to address related questions: how children's blood lead responds to the presence of contaminated products in the home, and how it responds to the intensity of their use.

4.2 Second-stage outcome: Blood lead levels

The primary second-stage outcome is the sampled child's BLL, measured in μg/dL. BLL is measured at both baseline and endline using venous blood samples analysed by inductively coupled plasma mass spectrometry (ICP-MS), the gold-standard method for blood lead analysis (Bonnifield and Todd, 2024b). Blood samples are collected by trained GHS personnel in trace-element-free tubes and transported under cold-storage conditions to the Metropolis Global Reference Laboratory, an ISO 15189-accredited laboratory, in Mumbai, India.[11] Results are shared by the laboratory approximately 21 days after sample collection.

We selected ICP-MS for BLL measurement in place of LeadCare II, the point-of-care test used for BLL screening, because it offers substantially greater precision and a lower limit of detection (< 0.1 μg/dL, compared with 3.3 μg/dL for LeadCare II). In addition, venous blood collection reduces the risk of sample contamination from lead present on the

[11] Analysis will occur at the Metropolis Global Reference Laboratory because ICP-MS testing capacity is limited within Ghana.

surface of the skin. These features reduce measurement error and increase statistical power, while the lower detection limit allows treatment effects to be accurately measured even if the interventions reduce children's BLLs to very low levels. However, ICP-MS is substantially more costly and less rapid than LeadCare II, making the latter better suited to large-scale screening.

As secondary outcomes, we construct binary indicators for whether the child's BLL is at or above (i) 3.5 μg/dL, the current CDC blood lead reference value, and (ii) 5 μg/dL, a widely used threshold in the literature. These outcomes speaks more directly to the policy question of how many children the intervention moves below the level at which action is recommended, at the cost of reduced sensitivity to changes in BLL among children who remain above the threshold.

4.3 Heterogeneity

We will conduct exploratory analyses of treatment effect heterogeneity across four pre-treatment child and household characteristics:

1. **Child baseline BLL (mean-centred):** We do not have a strong prior regarding whether treatment effects will be larger or smaller among children with higher baseline BLLs, as measured using ICP-MS. Higher baseline BLLs may reflect greater exposure from cookware and eyeliner, greater exposure from alternative lead sources, or greater physiological absorption of lead, each of which has different implications for the effectiveness of contaminated product replacement.

2. **Child age in months (mean-centred):** We hypothesise that treatment effects will be larger for younger children, who are evidenced to absorb lead more readily and may also exhibit behaviours that increase exposure, such as more frequent hand-to-mouth activity.

3. **Household dust lead concentration (mean-centred):** A dust sample was collected from all households during the modified HBA. Dust wipes were used to collect dust from a 25×25 cm area of the floor immediately outside the room where the child sleeps, and samples were analysed for lead concentration using an XRF analyser. We hypothesise that treatment effects may be smaller among children living in households with higher dust lead concentrations, as these children are likely to experience greater exposure from environmental lead sources that are unaffected by the intervention.

4. **Alternative lead exposure risk index:** We will construct an Anderson (2008) index summarising alternative household lead exposure risks measured during the modified HBA. The index will comprise indicators for: (i) household members working in occupations associated with lead exposure (paint removal, building renovation, demolition, plumbing, vehicle repair, smelting, welding, foundry work, paint manufacturing, scrapyard work, electronics repair, jewellery making, chemical plants, glass manufacturing, pottery, and battery manufacturing or recycling); (ii) take-home exposure pathways (wearing work clothes home, bringing work clothes home, or washing work clothes at home); and (iii) household activities involving potential lead exposure (home renovation, paint removal from furniture or vehicles, soldering, vehicle repair, battery recycling, melting lead, and glazing). We similarly hypothesise that treatment effects may be smaller among households with higher alternative exposure risk.

4.4 Covariates

The robustness specifications will include the following pre-treatment covariates, collected during the census, BLL screening, modified HBA, and baseline survey (see Section 3.3):

- **Child age:** Child's age (in months) at baseline.
- **Child sex:** Indicator for whether the child is female.

- **Household size:** Total number of household members, defined as individuals who share the same housekeeping arrangements, recognise one household head, and live in the same compound.
- **Caregiver education:** Highest level of education completed by the child's primary caregiver.
- **Household multidimensional poverty index (MPI):** Composite measure of household socioeconomic status based on deprivations in living conditions, health, education, and employment, adapted from the Ghana National MPI. Our measure omits the child nutrition indicator (whether at least one child under five is underweight, stunted, or wasted), which was not collected. Further details of the methodology are available from the Oxford Poverty and Human Development Initiative: https://ophi.org.uk/national-mpi-directory/ghana-mpi.
- **Household dust lead concentration:** See Section 4.3.
- **Alternative lead exposure risk index:** See Section 4.3.

5 Empirical Strategy

5.1 Estimation and inference

Our primary estimands are the causal effects of contaminated cookware and eyeliner use on children's BLLs. Contaminated product use is endogenous: treatment assignment determines whether households are offered lead-free replacements but not their actual product use, which may be correlated with unobserved determinants of children's BLLs. We therefore estimate these causal effects using two-stage least squares (2SLS), instrumenting the endogenous regressors of interest with households' random assignment.

We estimate two specifications: one using the binary indicators of contaminated product presence and one using the lead-dose indices, as defined in Section 4.1. Let $X_i^C \in \{I_i^C, U_i^C\}$ and $X_i^E \in \{I_i^E, U_i^E\}$ denote the cookware and eyeliner first-stage outcomes, respectively. The first-stage regressions are:

$$X_i^C = \pi_0^C + \pi_C^C Z_i^C + \pi_E^C Z_i^E + \pi_{CE}^C (Z_i^C \times Z_i^E) + \gamma^C BLL_i^{base} + \delta_s + u_i \quad (1)$$

$$X_i^E = \pi_0^E + \pi_C^E Z_i^C + \pi_E^E Z_i^E + \pi_{CE}^E (Z_i^C \times Z_i^E) + \gamma^E BLL_i^{base} + \delta_s + v_i \quad (2)$$

where Z_i^C and Z_i^E indicate assignment to receive lead-free cookware and eyeliner, BLL_i^{base} denotes the child's baseline BLL, and δ_s denotes randomisation-stratum fixed effects. The second-stage regression is:

$$BLL_i^{end} = \beta_0 + \beta_C \hat{X}_i^C + \beta_E \hat{X}_i^E + \gamma BLL_i^{base} + \delta_s + \epsilon_i \quad (3)$$

where BLL_i^{end} denotes the child's endline BLL, and $\hat{X}_i^C$ and $\hat{X}_i^E$ are the fitted values from Equations 1 and 2. Baseline BLL is included in the second-stage regression as an ANCOVA control to improve precision and also enters the first-stage regressions, as is standard in 2SLS. As secondary analyses, we estimate Equation 3 replacing BLL_i^{base} and BLL_i^{end} with indicators for whether the child's BLL was elevated at baseline and endline, using the thresholds described in Section 4.2.

The interpretation of β_C and β_E depends on the first-stage outcome used. In the binary specification, they give the average effect on child BLL of the presence of contaminated cookware and eyeliner in the household. In the lead-dose specification, they give the average effect of increasing a child's lead dose by one unit — 1,000 ppm-meals of cookware use or 1,000 ppm-applications of eyeliner use. Both are local average treatment effects, defined over households whose product use responds to treatment assignment.

We examine treatment effect heterogeneity using the following specification:

$$BLL_i^{end} = \beta_0 + \beta_C \hat{X}_i^C + \beta_E \hat{X}_i^E + \theta_C(\widehat{X_i^C \times H_i}) + \theta_E(\widehat{X_i^E \times H_i}) + \eta H_i + \gamma BLL_i^{base} + \delta_s + \varepsilon_i \quad (4)$$

where H_i denotes the pre-specified child or household characteristic of interest (see Section 4.3). The coefficients θ_C and θ_E give the change in the average effect of contaminated cookware and eyeliner associated with a one-unit increase in H_i. Because $(X_i^C \times H_i)$ and $(X_i^E \times H_i)$ are endogenous, these interaction terms are instrumented using interactions between the treatment-assignment instruments and H_i. When H_i denotes mean-centred baseline BLL, BLL_i^{base} is omitted.

Finally, we report intention-to-treat (ITT) estimates, obtained by regressing endline BLL directly on treatment assignment:

$$BLL_i^{end} = \tau_0 + \tau_C Z_i^C + \tau_E Z_i^E + \tau_{CE}(Z_i^C \times Z_i^E) + \gamma BLL_i^{base} + \delta_s + \xi_i \quad (5)$$

Here, τ_C and τ_E give the average effects on child BLL of being *offered* replacement cookware and eyeliner, irrespective of take-up, while τ_{CE} gives the interaction between the two treatments (i.e., whether the effect of receiving both differs from the sum of their individual effects). We use τ_{CE} to assess the additivity assumption underlying the primary 2SLS specification; a non-zero interaction could reflect either differences in product-use responses when the treatments are offered jointly or a non-additive biological response of BLL to contaminated cookware and eyeliner use. If there is evidence of the latter, we will additionally estimate a 2SLS specification including an interaction between cookware and eyeliner use.

All specifications are estimated using heteroskedasticity-robust standard errors. Because treatment was assigned at the household level and inference is for the study sample rather than a broader target population, we do not cluster standard errors (Abadie et al., 2023). For

our primary 2SLS and ITT specifications, which comprise a small number of pre-specified hypotheses, we report conventional two-sided p-values only; for the heterogeneity analyses, we additionally report false discovery rate-adjusted q-values, treating all interaction coefficients θ_C and θ_E across moderators H_i as a single family separately for the binary and lead-dose specifications.

We assess the robustness of the specifications in Equations 3 and 5 in the following ways:

- Adding the pre-specified baseline covariates described in Section 4.4;
- Adding an indicator variable for households whose products were XRF tested during the modified HBA. We omit this indicator from the primary specification because pre-trial XRF testing was conducted sequentially across localities before being discontinued, so stratum fixed effects should absorb most of the resulting variation;
- Estimating the second-stage regression using the natural logarithms of endline and baseline BLL, to assess the sensitivity of the results to an alternative functional form of the outcome; and
- Winsorising the lead-dose indices at the 95th percentile to assess sensitivity to extreme values. This check applies only to the lead-dose specification.

5.2 Identification assumptions and diagnostics

The causal interpretation of the 2SLS estimates relies on the standard instrumental variable assumptions of relevance, exclusion, independence, and monotonicity. This section summarises the design decisions taken to support these identification (ID) assumptions, together with the diagnostic tests that will be used to assess their plausibility.

Relevance: The treatment assignments must generate sufficient independent variation in contaminated cookware and eyeliner use to identify the causal effect of each on BLL. In prac-

tice, this requires that (a) the offer of replacement cookware (eyeliner) substantially reduces contaminated cookware (eyeliner) use—i.e., high compliance with treatment, and (b) behavioural spillovers that induce control households to reduce their use of contaminated items remain limited. The latter is a particular concern because treatment was assigned at the household rather than community level, creating more opportunities for control households to learn about the potential risks associated with cookware and eyeliner. Such spillovers would weaken the first stage and change the composition of the complier population, although they do not invalidate identification of our primary estimand (i.e., the treatment effect on compliers) provided the remaining ID assumptions hold.

To reduce these risks, we discontinued pre-trial product testing (Section 3.1), so that control households would be less likely to alter their behaviour after learning their products' lead concentrations. We will also conduct monthly visits to treatment households (Section 3.2) to monitor compliance and replace missing or depleted lead-free products. We will assess the relevance assumption using the first-stage estimates from Equations 1 and 2 and the corresponding Sanderson–Windmeijer conditional F-statistics, which test whether the excluded instruments provide sufficient independent variation to identify each endogenous regressor, conditional on the other (Sanderson and Windmeijer, 2016).

Exclusion: Assignment to receive replacement cookware and/or eyeliner must affect children's BLLs *only* via its effect on contaminated cookware and eyeliner use. This assumption is violated if the intervention encourages treatment households to reduce lead exposure through alternative pathways, such as adopting safer occupational practices, improving hygiene or nutrition, or avoiding other contaminated products.

We will mitigate this risk by ensuring that, aside from information on product contamination and the opportunity to receive replacement products, treatment and control households receive the same information about lead exposure risks throughout the study. In particular, during product replacement and subsequent monitoring visits, field staff will be trained

not to provide additional advice on alternative lead exposure reduction strategies. We will assess the plausibility of the exclusion restriction by testing for treatment effects on endline measures of the alternative lead exposure index described in Section 4.3. Because the model is overidentified—three instruments for two endogenous regressors—we will also report the Hansen J statistic, which tests whether the instruments are jointly consistent with the estimated model (Hansen, 1982; Baum et al., 2003). Failure to reject the null is consistent with the exclusion restriction holding, though the test detects only violations that cause the instruments to disagree; rejection may indicate either invalid instruments or misspecification, so is likewise not conclusive.

Independence: Independence requires that treatment assignment is independent of potential outcomes. This assumption is supported by the randomisation procedure described in Section 3.2 and the preliminary balance check presented in Table **??**. We will additionally assess balance in the analysis sample using baseline ICP-MS BLL and other characteristics collected during the baseline survey. Finally, we will test for differential attrition across treatment arms by regressing an indicator for endline missingness on treatment status and its interactions with baseline BLL. If the joint F-tests indicate differential attrition across treatment arms, we will report Lee bounds as an additional robustness check.

Monotonicity: Monotonicity requires that assignment to receive replacement cookware (eyeliner) does not increase contaminated cookware (eyeliner) use for any household. We consider violations unlikely: replacement products are confirmed lead-free before distribution, contaminated items are removed when replacements are accepted, and households that decline are unaffected. The principal risk is that a household dissatisfied with a replacement acquires a further contaminated item to substitute for it, though the pilot findings indicate that dissatisfaction is uncommon (Section 3.2). In any case, our first-stage outcome measures will capture whether treatment households increase their contaminated product use.

5.3 Source apportionment

Finally, we use the 2SLS estimates to quantify the share of children's lead exposure attributable to contaminated cookware and eyeliner in our sample. For each product, we estimate the reduction in mean BLL that would result from eliminating all contaminated products from children's households and divide this by mean endline BLL in the control group:

$$A^k = \frac{\hat{\beta}_k \, \bar{I}^k}{\overline{BLL}^{end}_{ctrl}}, \qquad k \in \{C, E\}, \tag{6}$$

where $\hat{\beta}_k$ is the estimated effect of contaminated product presence on BLL among compliers, $\bar{I}^k$ is the pre-replacement share of treatment households with contaminated product k, which under random assignment estimates its pre-intervention prevalence in the study sample, and $\overline{BLL}^{end}_{ctrl}$ is mean endline BLL among control children. We use the control-group endline mean rather than the baseline mean so that numerator and denominator refer to the same point in time; we report the baseline-denominator version alongside as a robustness check. This calculation assumes that the average effect of product removal is the same for compliers and non-compliers. We obtain standard errors for A^k by the delta method.

We estimate the share attributable jointly to cookware and eyeliner by summing their individual contributions, A^C and A^E, assuming additive effects; we assess this assumption using the treatment interaction in the ITT specification, as described above.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Baum, C. F., Schaffer, M. E., and Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, 3(1):1–31.

Binkhorst, G., Jones, B., Sargsyan, A., van Geen, A., and Wagner, K. (2025). Potential lead exposure from aluminum cooking pots in lower and middle-income countries. *Journal of Hazardous Materials*, 492:138134.

Bonnifield, R. and Todd, R. (2024a). Starting from the Source: Methods and Priorities for Explaining Lead Exposure. Policy Paper 333, Center for Global Development, Washington, DC.

Bonnifield, R. S. and Todd, R. (2024b). Tools for Measuring Human Lead Exposure: A Review of Methods and Implications for Future Research and Practice. Policy Paper 329, Center for Global Development, Washington, DC.

Crawfurd, L., Todd, R., Hares, S., Sandefur, J., and Bonnifield, R. S. (2024). The Effect of Lead Exposure on Children's Learning in the Developing World: A Meta-Analysis. *The World Bank Research Observer*.

Ghana Statistical Service (2021). Multidimensional Poverty Metrics by Locality, District, and Region. PHC 2021 StatsBank Ghana.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(3):1029–1054.

Larsen, B. and Sánchez-Triana, E. (2023). Global health burden and cost of lead exposure in children and adults: a health impact and economic modelling analysis. *The Lancet Planetary Health*, 7(10):e831–e840.

Nash, E., Baylor, C., Gbadago, B. E., Baaku, W., Quansah, E. W., Berger, U., and Bose-O'Reilly, S. (2025). Industry and consumer products as lead exposure sources among children across 3 regions in Ghana. *Environmental Research*, 287:123013.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Sargsyan, A., Nash, E., Binkhorst, G., Forsyth, J. E., Jones, B., Sanchez Ibarra, G., Berg, S., McCartor, A., Fuller, R., and Bose-O'Reilly, S. (2024). Rapid Market Screening to assess lead concentrations in consumer products across 25 low- and middle-income countries. *Scientific Reports*, 14:9713.

U.S. EPA (2024). Integrated Science Assessment (ISA) for Lead. Technical Report EPA/600/R-24/042, U.S. Environmental Protection Agency, Washington, DC.

A Power Calculations

A.1 Ex ante calculations

The study's sample and treatment group sizes were calculated ex ante to ensure sufficient power to detect the causal effects of contaminated cookware and eyeliner use on child BLL, as estimated by the 2SLS specification described in Section 5.1. The calculations used parameters estimated from data collected by Pure Earth, GHS, and UNICEF from households in Tolon and Yendi towns in a previous survey (see Section 2). They assumed a 5% significance level, 80% power, and individual-level randomisation. Based on the previous survey, control-group BLL was assumed to have a mean of 10.1 μg/dL, a standard deviation of 4.4 μg/dL, and a baseline–endline correlation of 0.6, implying a residual standard deviation of 3.52 μg/dL under the ANCOVA specification.

We further assumed that the presence of contaminated cookware and eyeliner in the household would increase child BLL by $\beta_C = 2.0$ and $\beta_E = 4.7$ μg/dL, respectively. These are naive estimates based on the observed associations between contaminated product presence and BLL in the previous survey. Finally, we assumed that all study households would have contaminated products, as this was an inclusion criterion under the initial research design; that 70% of treatment households would fully adopt the lead-free replacements offered, which the pilot findings suggest is conservative (see Section 3.2); and that there would be no cross-treatment effects ($\pi_E^C = \pi_C^E = 0$) and no complementarity between the two treatments ($\tau_{CE} = 0$). Together, these assumptions imply first-stage effects of $\pi_C^C = \pi_E^E = -0.7$ and ITT effects of $\tau_C = (-0.7)(2.0) = -1.4$ μg/dL and $\tau_E = (-0.7)(4.7) = -3.3$ μg/dL.

Table A1 reports the resulting sample size requirements. While the calculated minimum sizes for the eyeliner and combined arms were 18 and 9 households, we allocated around 50 to each, given the risk that the naive association used to estimate the causal effect of eyeliner presence on BLL may overstate the true effect (see Section 2).

Table A1: Ex ante assumed effect sizes and sample size calculations

Treatment arm	First-stage effect, π	Second-stage effect, β	ITT effect, τ	Required N	Allocated N
Cookware	−0.7	2.0	−1.4	99	102
Eyeliner	−0.7	4.7	−3.3	18	49
Both	−0.7	6.7	−4.7	9	48

Notes: Calculations assume a 5% significance level, 80% power, a control-group BLL mean of 10.1 μg/dL and standard deviation of 4.4 μg/dL, and a baseline–endline correlation of 0.6. The row for the combined arm reports the implied effect of receiving both replacements under the assumption of no complementarity; the pre-specified model contains no separate parameter for this arm. The control arm was allocated 101 households to provide the required power for the cookware comparison.

The calculations for cookware and eyeliner treat each as a separate IV model with one endogenous regressor and one instrument; they therefore approximate rather than exactly reproduce power for the multivariate 2SLS specification. While the assumptions above are expressed in terms of the binary first-stage outcome, we note that expressing them in terms of the lead-dose index would not change the calculated sample sizes, since power ultimately depends on the reduced-form effect of treatment assignment on BLL.

A.2 Updated calculations

We update the ex ante power calculations using data collected from the study sample to refine two key assumptions. First, we replace the assumed control-group BLL distribution with that observed at baseline using our primary ICP-MS outcome measure. Mean baseline BLL is 4.26 μg/dL, with a standard deviation of 1.8 μg/dL, compared with the ex ante assumptions of 10.1 and 4.4 μg/dL, respectively. We retain the assumed baseline–endline correlation of 0.6.[12]

Second, we update the first-stage assumptions using observed rates of contaminated product prevalence. Pre-replacement product testing found that 99% of households had contaminated cookware and 87% had contaminated eyeliner. Assuming as previously that 70% of

[12]The ex ante parameters were estimated from BLL data collected from a different study population using LeadCare II rather than ICP-MS, which likely accounts for the differences in the BLL distributions.

treatment households fully adopt the replacement products, this implies first-stage effects of $\pi_C^C = -(0.99)(0.70) = -0.69$ and $\pi_E^E = -(0.87)(0.70) = -0.61$. Given the high adoption observed in the pilot, we also report calculations assuming 90% of treatment households adopt the replacements.

Table A2 reports the resulting minimum detectable effects (MDEs) using the realised treatment-group sizes, maintaining the other assumptions of the ex ante analysis. The second-stage MDEs are well below the naive effect size estimates assumed ex ante: at 70% adoption, the study can detect effects of 0.82 μg/dL for cookware and 1.15 μg/dL for eyeliner, against assumed effects of 2.0 and 4.7 μg/dL. This mainly reflects the lower variance of BLL in the study sample than was assumed ex ante, and is reassuring given that the naive associations may overstate the true causal effects.

Table A2: Minimum detectable effects using observed baseline parameters

Treatment arm	First-stage effect, π	Second-stage MDE	ITT MDE
Panel A: 70% adoption			
Cookware	−0.69	0.82	0.57
Eyeliner	−0.61	1.15	0.70
Both	–	–	0.71
Panel B: 90% adoption			
Cookware	−0.89	0.64	0.57
Eyeliner	−0.78	0.90	0.70
Both	–	–	0.71

Notes: Minimum detectable effects (MDEs) are in μg/dL, at a 5% significance level and 80% power, using the realised arm sizes reported in Section 3.2. The second-stage MDE is the ITT MDE divided by the assumed first stage. The combined arm has no separate second-stage parameter, so only its ITT MDE is reported.