Assessing Changes in Blood Lead Levels:
Study Design Considerations

PURE EARTH

GAHP
GLOBAL ALLIANCE ON HEALTH AND POLLUTION

August 2021

Acknowledgements

This document was written by Dr. Jenna Forsyth, Stanford University. The following people reviewed and provided comments on the document:

Gordon Binkhorst, VP Global Programs, Pure Earth

Mary Jean Brown, US Centers for Disease Control and Prevention (retired)

Daniel Kass, Senior Vice President, Environmental Health, Vital Strategies

Emily Nash, Program Director for South Asia, Pure Earth

Laura Schaefli, Program Manager, GAHP

Support: GAHP and Pure Earth gratefully acknowledge the financial contribution of the Clarios Foundation for the development of this document.

Note to the reader:

This guide was developed to provide a high-level overview of study design considerations, especially for studies aiming to test a hypothesis or test the effectiveness of an intervention to reduce blood lead levels (BLLs). Each topic covered in this guide could be a course or textbook on its own. The purpose of this overview is to provoke thought and identify topic areas where a researcher may benefit from further training or guidance.

Before developing a study design, it is important to form a research question or hypothesis. A research question or hypothesis forms the backbone of any study design. These will indicate the anticipated outcome or change in the outcome before and after an intervention, as well as the study population, location, and timeframe and set the stage for all study-related decisions. Although initial drafts of a working hypothesis may start as a general statement, ultimately it will be as detailed as possible and informed by prior scientific studies. For example, “We hypothesize that a soil remediation intervention will reduce median BLLs by 15% after 14 months among children under 13 years of age in Kathgora, Bangladesh.”

Table of Contents

PART 1. Methods and best practices for identifying exposed and control populations for blood lead level sampling

1.1 Decide if a control group is appropriate

1.1.1 What is a control group?

A control group is a group of individuals who are similar in demographic characteristics to the exposed group of individuals but have not been exposed to the source of concern (or intervention if testing an intervention).

1.1.2 Why is a control group important?

Having a control group is necessary to make causal inference (e.g., to conclude that an intervention caused the observed change in BLLs). This is because pre- and post-intervention BLLs can be compared within an exposed group and also compared between an exposed and control group at similar time points to determine the "difference-in-differences" between the pre- and post- BLL measures (Figure 1). By contrast, an uncontrolled study would be comparing the pre- and post-intervention BLLs within the exposed group alone (Figure 1). If the exposed and control groups are similar, then the change in BLLs in the control group between two time points can be assumed to represent what would have happened to the exposed group over the same timeframe if the contaminated site (or other exposure) was not present. For example, BLLs in the control group could increase over a certain time period if another source of lead appeared in the region. Conversely, BLLs in the control group could decrease if a regional source of lead was removed due to effective policies or other intervention. Without a control group, it would be impossible to know about changes to BLLs unrelated to the focal exposure/intervention of study.

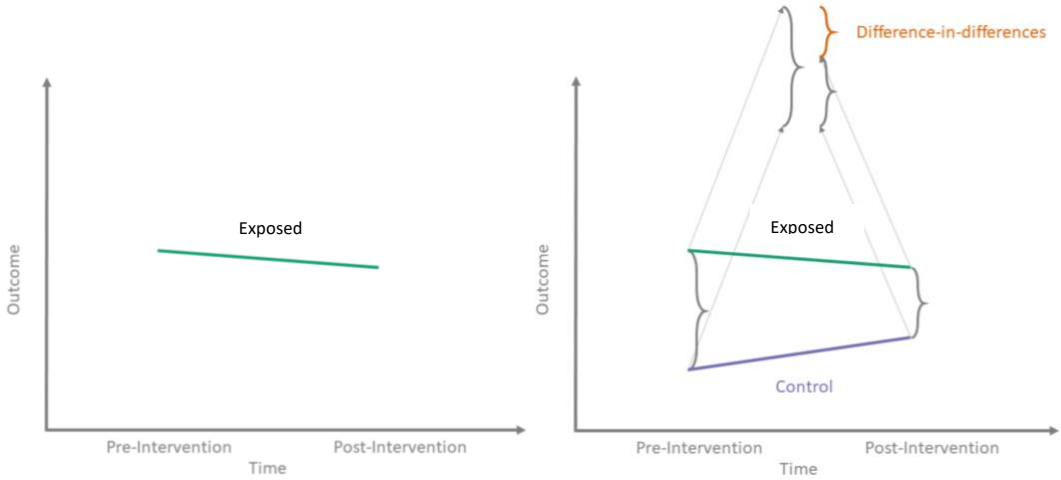

Figure 1. Comparison of a before-after design with no control group (left) versus a difference-indifferences design with an exposed and control group (right). Adapted from https://diff.healthpolicydatascience.org.

As an example, consider an exposed group in Bangladesh living near a used lead acid battery recycling site and another control group living far from the contaminated site. Both populations may still be exposed to other non-battery sources of lead like spices or cosmetics. Spice lead concentrations may vary over time equally for the exposed and control groups and they may cause the control group's BLLs to be slightly elevated, but not so elevated as the exposed group. If conducting an intervention to reduce contamination at the battery recycling site, pre- and post-intervention BLLs should be assessed among both the control and exposed groups to determine how much of the BLL reduction in the exposed group is likely attributable to the remediation and how much exposure from other sources remains. It is worth noting, however, that BLLs may change with age among children (generally decreasing as a child ages due to behavioral and developmental changes).

1.1.3 How to select a control group

To adequately compare a control and exposed group, a control group needs to be carefully identified and selected. A good control group is one that is identical to the exposed group in as many ways as possible except being exposed to the contaminated site (or is not a beneficiary of the intervention). As long as the proportion of individuals with key characteristics are similar between control and exposed groups, then comparisons should be reliable.

The most important characteristics to match control and exposed groups can vary based on the context, but, as a rule of thumb, there should be a match on characteristics that are known to impact lead metabolism, lead uptake, or possible lead exposure unrelated to the focal exposure from the contaminated site.

Characteristics worth considering include:

- **Socioeconomic status:** Children of low socioeconomic status tend to have poorer nutrition and a higher likelihood of micronutrient deficiencies. Individuals low in calcium or iron absorb more lead. In the United States, children of low socioeconomic status also tend to live in older housing where legacy lead paint or lead soldered pipes might be a source of exposure. Similarly, people of lower socioeconomic status may live in homes with dirt floors which may be a continued source of lead exposure and are more difficult to clean than homes with concrete or other floor materials. Moreover, family members may work at low-paying jobs where they are exposed to lead and can expose others by bringing lead home on clothes and shoes.
- **Age:** BLLs often exhibit an inverse relationship with age, especially if the exposure is from soil or dust, because younger children absorb more lead than older children and have higher rates of hand-to-mouth activity.[3] Children under 5 absorb up to 50% of ingested lead compared to less than 10% among older children and adults.[1] Young children also unintentionally ingest more lead from soil and dust in their

environments because they touch their mouths more often than older children and adults.[2] For young children, age matters, especially if the suspected lead source is food, because very young children may be exclusively breastfeeding.[5] As a general rule, children under 5 years of age undergo rapid development and behavioral changes, so matching on a more precise age range (e.g., 12-24 months old, or 24-48 months old) would be more important than it would be for children over 5 years old. Overall, having a precise age range reduces variability in BLLs and thus requires a smaller sample size. However, depending on the environmental media of exposure and local context, age may be a more or less important consideration.

- **Gender:** In some cultures and countries, male children are given more freedom to play and explore outside the home than female children who may be restricted to the indoor environment. On the other hand, female children may be more likely to be exposed to lead in cosmetics. Such gender-based behavioral differences would impact exposure to environmental contamination and should be assessed based on the local context.
- **Ethnicity/religion:** Individuals of different ethnicities or religions may use different products known to be contaminated with lead due to cultural or dietary preferences (e.g., spices, ayurvedics, cultural powders, etc.).
- **Geographic characteristics like the degree of urbanization or industrialization:** rural vs. urban areas (and minimally industrial vs. highly industrial areas) typically differ in terms of lead exposure and contamination because of the numerous industrial sources of lead. To the extent possible, individuals from the exposed and control groups would hail from neighborhoods with similar levels of urbanization and industrialization.

In addition to those listed above, it is worth considering how transient or permanent a population is. Ideally, control and exposed group participants would remain at the same place of residence for the duration of the study and would have lived there for at least 6 months prior to the first blood collection. Limiting the study population to children who have spent the last 6 months in the same location will ensure that local environmental lead exposures are relevant to the child's measured BLL. Families who have not moved recently may also be less likely to move during the study period.

1.1.4 How to determine whether a control group is necessary

It is important to pause to consider if a control group is even appropriate by asking the following questions:

- Would your budget allow for a control group, which would essentially double the cost?

- In your study region, would it be possible to find an unexposed group of individuals with the same characteristics as the exposed group that do not live near the contaminated site that will receive the intervention?
- Is it possible to define the bounds of contamination or is it widespread? Lead-polluting industries typically have definable bounds of contamination that spread from the main site. However, during the time of leaded gasoline usage, lead pollution was widespread and the assessment of the reduction in BLLs had to be done in the absence of a control group.
- Does the intervention you aim to assess have localized impacts only or would the impact be far-reaching? Some interventions have widespread impacts (e.g., a national spice intervention) and therefore it would not be appropriate to use a control group since all individuals in a country would be exposed and would benefit from the intervention. Choosing a control group from a different country would be unlikely to be similar enough in important characteristics as defined above. In these situations, a before-after study design would be the most robust method of assessing BLLs. This design would either measure repeat BLLs among the same population, or if the intervention time is long enough, would measure pre- and post-intervention BLLs among similar (but not the same) populations to maintain similar-aged cohorts. In a before-after study design without a separate control group, measuring BLLs of similar-aged cohorts at two points in time help to control for age-related behavioral aspects of exposure, helping to isolate the impact of changes in exposure due to interventions. This is especially important if studying young children but less critical if studying adults.

1.2 Evaluate the exposure pathway and catchment area for exposed and unexposed groups

1.2.1 What is an exposure pathway?

As defined by the U.S. Environmental Protection Agency (EPA), an exposure pathway is the way a person encounters a contaminant or hazardous substance (in this case, lead). The U.S. Agency for Toxic Substances and Disease Registry (ATSDR) breaks down exposure pathways into five elements:

1) The contaminant source or release (e.g., airborne lead from a smelter);
2) Environmental fate and transport (e.g., once lead is released into the air, how does it move through different media like soil or water);
3) Exposure point or area (these are specific locations where people come into contact with the contaminated medium like playgrounds where children may play in contaminated soil);
4) Exposure route (how exactly someone comes into contact with the contaminant, e.g., via inhalation, ingestion, or direct contact), and ;
5) Potentially exposed population (e.g., this may be residents living within a certain distance, children playing on playgrounds, workers working at the site, etc.).

An exposure pathway typically has multiple contaminant sources, affects different environmental media, has several exposure points, and more than one exposure route depending on the exposed population. See Figure 2 below for a conceptual model that provides an example of the multitude of impacts from a hazardous waste drum.

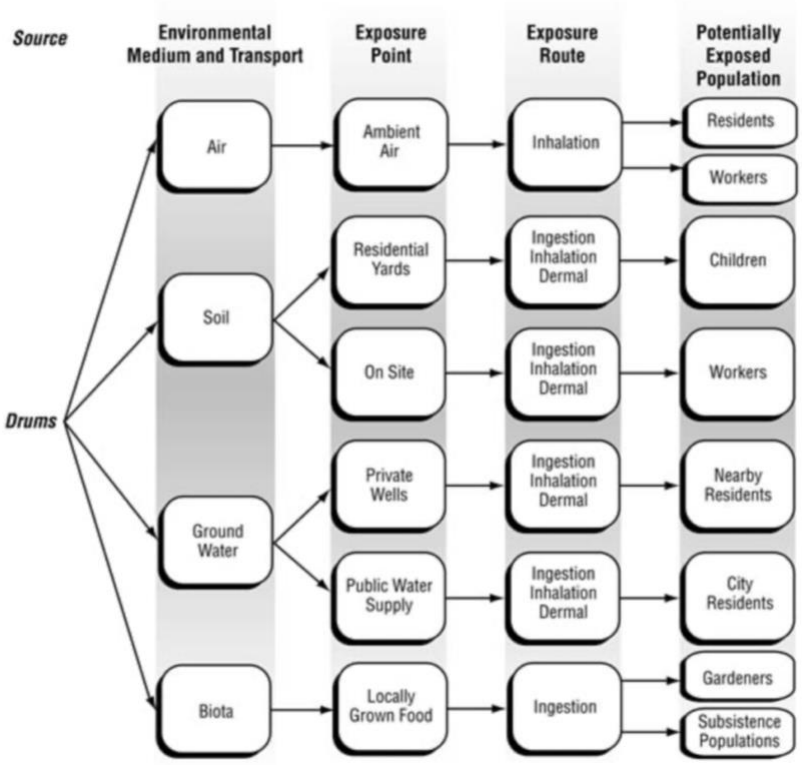

Figure 2. Conceptual model of the 5 elements of the exposure pathway from hazardous waste drums (Source: ASTDR https://www.atsdr.cdc.gov/hac/phamanual/ch6.html).

1.2.2 What is a catchment area?

In this case, a catchment area refers to the specific area (region, district, city, village, neighborhoods) in which the study population lives and from which the study sample will be drawn. More information on these terms can be found in section 1.3. The catchment area should have clearly defined boundaries that could be mapped for study purposes.

1.2.3 How to define the exposed group catchment area

Step 1: Review prior studies to determine the extent to which the five elements of the exposure pathway have been defined for the specific contaminant source. Of particular interest when defining the bounds of the catchment area would be how far the contamination is likely to travel from the focal site of pollution generation (e.g. on the order of meters or kilometers?), and how far people are likely to travel to the exposure points from their place of residence. This information will help to define the geographic bounds that demarcate the catchment area from

which an “exposed” population should be drawn. A distance from a contaminated site should be specified, for example, a 200-meter radius from the site of pollution generation. For informal used lead acid battery recycling with informal smelters that do not have tall chimneys, a threshold distance of 50-200 meters is reasonable as the contamination is likely to be fairly localized.[11] However, for a more formal industrial-scale smelting operation with a tall chimney, the threshold distance may be much longer.

Step 2: Whenever possible, and especially if no prior studies have been conducted to inform decision-making, it is advantageous to assess and map environmental contamination to define the geographic bounds of the contaminated area (Figure 3, taken from [11]). For example, if the primary contaminated media is soil, the researcher could measure soil lead levels using a portable x-ray fluorescence analyzer along a radial transect, every X meters (X to be defined based on exposure source but could be 50 meters to start). Conduct these radial measurements in all cardinal directions north, east, south, and west from the central point of contamination. Using a GPS unit, record the latitude and longitude of each soil measurement and map the data to determine where contamination appears to taper off. Soil lead levels above the EPA reference level of 400 ppm can be used as lead levels of concern for soil. Additional measurements can be made on a more granular scale between the initial measurements to determine the pattern of contamination.

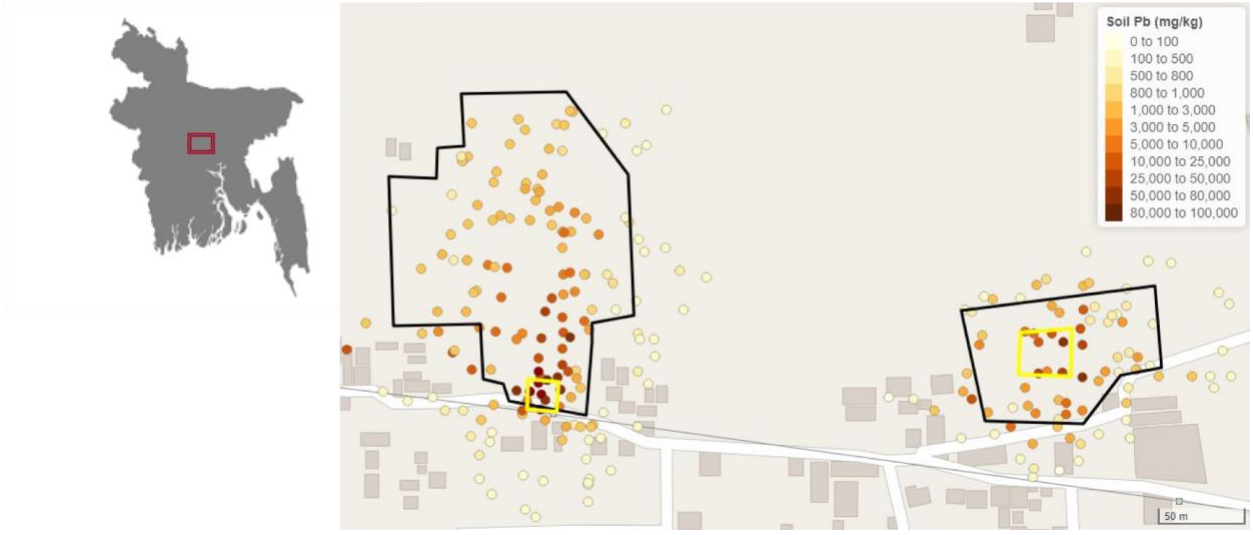

Figure 3. Map of study site and soil lead (Pb) concentrations before remediation (not to scale). The two impacted areas are marked with a black border. The yellow border indicates the two smelting zones. Figure from a study conducted in Bangladesh by researchers from Pure Earth, Dhaka University, the International Centre for Diarrheal Diseases Research, and Stanford University.[11]

Step 3: Conduct some formative research in the proposed catchment area to understand how far potential study subjects travel in the area and their likelihood of encountering exposure points that are outside of their residences.

1.2.4 Defining a control catchment area

The control group's catchment area should be similar in size to the exposed group's catchment area and should be far enough away from the exposed group's catchment area to be free from contamination. It is worth considering not only the extent of the environmental contamination to ensure that the control catchment area is not contaminated, but also the movement of residents in the area. For example, how far do children travel for school or play from the control catchment area and would they be likely to spend their days in the contaminated environment of the exposed catchment area? If the answer to the second question is yes, then the control catchment area should be further from the border of the exposed catchment area to remove the risk of spillover. The characteristics of the control catchment and the individuals living in it should be similar in other respects, as described above.

Follow the same steps described in section 1.2.3 for the control catchment area to ensure that there are no point sources of lead that may be contributing to BLLs (e.g., another battery recycling site).

1.3 Define who will be sampled

1.3.1 What is a target population?

The target population is also known as the theoretical population to which you hope to generalize your findings.

1.3.2 What is a study population?

The study population is also known as accessible population from which you will draw your sample.

1.3.3 What is a study sample?

The study sample is the group of individuals who you collect data from. More information on these terms can be found in any statistical reference material.[12] If funds were unlimited, a researcher could sample from every person in the target population (conduct a census). However, conducting a census is not usually an option due to logistical or financial constraints.

1.3.4 Characteristics of the study sample

The first step is to decide on target population characteristics and make this as specific as possible. Then, identify the study population you will be able to get access to. Key characteristics of your sample will likely be those defined in section 1.1.3 that have an impact on lead metabolism/exposure: socioeconomic status, age, gender, housing characteristics, ethnicity, nutritional status, degree of neighborhood industrialization.

1.4 Define the sampling strategy

1.4.1 What is a sampling strategy?

Once you know the study population from which you will be selecting your sample, you need to define the process by which you will draw your sample, or the sampling strategy. Sampling strategies can either be probability-based (for example, random sampling) whereby every individual in the study population has the same probability of being selected, or non-probability-based (for example, convenience-based sampling). It is best to avoid non-probability-based sampling when conducting hypothesis-testing studies because these sampling strategies introduce bias. Section 1.4.2 describes different probability-based sampling strategies that are recommended.

1.4.2 Factors to consider when choosing a sampling strategy

The best sampling strategy for drawing a representative sample and for generalizing findings to the target population is a random sample. However, drawing a random sample is not always feasible. In an ideal scenario, you would have a list of all residents of the study population in the catchment area, you could number each resident, then generate random numbers to select participants randomly. However, it is rare to have access to a list of all residents, especially when doing research in low-and middle-income countries. A random representative sample can also be drawn by generating random GPS coordinates within the specified catchment area and identifying the nearest resident that meets the study inclusion criteria (e.g., age, gender, etc.). This is robust but can be logistically expensive. If needed, a catchment area can be subdivided into clusters.

In a two-stage cluster random sampling approach, clusters are randomly selected from the entire pool of available clusters (e.g., census tracts) and then households are enrolled. Cluster selection is probabilistic, proportional to the total population in the catchment area, and based on the study outcomes (e.g., households with children 12-24 months living in the residence). Typically, a minimum of 30 clusters is recommended. Once clusters are identified, a simple random sample of housing units can be taken or a systematic sample, such that each household has an equal probability of selection. If taking a systematic sample, the first step would be to identify a random starting point within the catchment area, then to move systematically from household to household traveling in a pre-specified direction to enrol subjects based on a certain rule. One rule could be to visit every n^{th} house, where “n” is defined based on the overall study population and the desired total sample size. Households are visited until the total sample of families with the population of interest (children 12-24 months) are identified and enrolled in the study. More information about typical two stage cluster random sampling can be found at https://www.cdc.gov/nceh/casper/sampling-methodology.htm. Note that using a

cluster random sampling strategy instead of a simple random sampling strategy will have implications for sample size (not covered in this document).

PART 2. Methods and best practices for testing blood lead levels before and after an intervention

2.1 Timing of blood collection and intervention impacts

2.1.1 Why timing of blood collection is important

Ensuring that an appropriate amount of time has elapsed between the pre- and post-intervention BLL measurements is critical for drawing inference about the intervention effect. A BLL is a time-integrated measure of lead exposure. In heavily exposed populations, internal sources of lead (e.g., bones and tissues) may keep BLLs high for months to years. In less heavily exposed populations, BLLs represent recent exposure within the past several months since absorbed lead circulates throughout the body in the blood, with a mean half-life of 21-28 days (compared to a mean half-life in the bones of decades).[10] It is important to ensure that enough time has passed for the intervention to have impacted BLLs. If post-intervention BLLs are measured too soon after the pre-intervention measurement, then a researcher would falsely conclude that the intervention had no effect. For children, pre- and post-intervention BLLs should be drawn about a year apart. For occupationally exposed adults, the timeframe may be much longer.

2.1.2 Factors to consider in defining the timing of blood collection

Pre- and post-intervention blood collection should be timed after careful consideration of the context and intervention.

Depending on the exposure pathway characteristics, climate of the region and human behavior, BLLs may vary seasonally. Seasonal variation in BLLs has been measured when soil and dust are the contaminated environmental media.[13] BLL variability may be a result of the climate (e.g., hot dry windy seasons increase the suspension of soil and dust in the air whereas cold wet seasons reduce suspension) and/or human behavior (e.g., children spend more time playing outside in the soil in dry seasons than wet seasons, and/or in summer time versus winter time). Evaluate the likelihood of seasonal BLL trends in the catchment areas and determine if pre- and post-blood collection should fall within the same season (e.g., in different years) or if seasonal impacts are unlikely.

How quickly the intervention is likely to impact BLLs should also be considered. Interventions may be quick to implement and have short-term impacts on BLLs or be prolonged to implement and have delayed impacts depending on the intensity and duration of exposure. Generally, the amount of time needed for BLLs drop to homeostatic levels after an intervention varies based on several factors including starting BLLs, time exposed, age, and nutritional status.[13] This is in part because lead stored in the bones from long-term exposure can be remobilized into the blood based on age and nutritional status and so a post-intervention BLL could appear elevated

even if the exposure has been removed. Lead deposited in soft tissue or bone has a mean half-life up to 19 years.[10]

On average, prior studies indicate that 1 year should be sufficient for BLLs to decline from more than 10 µg/dL to less than 10 µg/dL but other studies indicate that up to 4 years could be needed if exposure time was very long.[14–16] In the United States, governmental agencies assessing child lead poisoning cases typically re-measure BLLs after a period of 6 months has elapsed after the suspected source of lead has been removed from the child's home.

If the time elapsed between the pre- and post-intervention BLL measurement is on the order of years, it would be helpful to do another environmental assessment and/or human behavior assessment to collect information about potential exposures in both control and exposed catchment areas after the intervention. This would help to ensure that no new sources of lead emerged during the study period or would identify any changes in the exposure characteristics of the control and exposed groups that should be considered.

2.2 Blood collection and analytical method

2.2.1 Best practices on collecting blood samples

The World Health Organization (WHO) has developed a detailed guide to blood collection and analysis.

Below is a summary of the key information:

Avoid contamination of the blood sample: Because lead contamination is widespread, care should be taken to avoid contamination during sample collection, storage, transport, and manipulation. Prior to collecting blood, an individual's skin surface must be cleaned to ensure that ambient lead-containing dust does not interfere with and overestimate the BLL. The use of venous samples will decrease the risk of sample contamination.

Utilize high-quality, metal-free needles and tubes: If possible, utilize certified lead-free materials. If this is not possible, have the needles and tubes assessed to determine lead content.

Trained phlebotomist should collect blood: All precautions to prevent the transmission of bloodborne pathogens must be followed including the use of gloves. More information about this can be found in the WHO guide to phlebotomy.

Blood is ideally stored in a coolers, refrigerator, or freezer prior to analysis to prevent spoilage prior to analysis. However, note that lead levels are not impacted by blood storage temperature or spoilage, this is more for the comfort of the research team and ensuring the sample stays hydrated.

2.2.2 Deciding between field and laboratory BLL measurement methods

There are two primary methods for assessing BLLs: field-portable rapid measurement methods and laboratory methods. The most common field portable measurement method uses anodic stripping voltammetry (ASV) and the Magellan Industries LeadCare II kit is the industry leader. Laboratory methods fall under two broad categories: atomic absorption spectrometry (AAS), or inductively-coupled plasma mass spectrometry (ICP-MS). Among the AAS methods, there are several to choose from: electrothermal atomic absorption spectrometry (ETAAS), flame atomic absorption spectrometry (FAAS), and graphite furnace atomic absorption spectrometry (GFAAS). Table 1, below, provides a summary of the strengths and limitations of each method.

Table 1. Comparison of the strengths and limitations of different BLL measurement methods. Note an additional limitation of ASV is that it cannot be used at altitudes greater than 2,440 meters (8,000 feet). Source: WHO, Brief guide to analytical methods for measuring lead in blood

Method	Strengths	Limitations
Flame atomic absorption spectrometry (FAAS) *(11, 12, 14, 15)*	• Short analysis time (seconds) • Relatively easy to use • Relatively few interferences • Relatively low capital and running costs	• Large sample size usually needed • Relatively high detection limit (5 µg/dL) • Cannot be left unattended (flammable gas)
Electrothermal atomic absorption spectrometry (ETAAS) *(8, 11, 14, 15)*	• Low detection limit (< 1 µg/dL) • Can analyse small samples (50–100 µL) • Can be fitted with autosampler so multiple samples can be processed • Well documented applications • May be left unattended • No need for sample preparation	• Limited analytical working range • Requires some laboratory expertise • Longer analysis time so low sample throughput
Inductively coupled plasma mass spectrometry (ICPMS) *(10, 11, 14, 16)*	• Very low limit of detection (0.02 µg/dL) • Can analyse small samples (50–100 µL) • Very fast analysis time (< 1 minute) • Wide analytical working range • Multi-element capabilities and can be economical if used for large sample runs • Potential to perform isotopic ratio analyses with some forms of ICP-MS, which may help to identify the source of the lead	• High purchase and running costs • Requires highly skilled laboratory staff • Analysis of large number of samples is cheaper than ETAAS
Portable ASV *(1, 17)*	• Small sample size (50 µL) • Can be used at non-laboratory sites • Uses finger prick (capillary sample), though venous samples can also be used • Simple to use, does not require skilled laboratory personnel • Low purchase and running costs • Rapid results • Has comparable accuracy with laboratory-based methods	• Limited analytical working range • Levels above 5 µg/dL should be confirmed by a high-complexity laboratory method • High risk of sample contamination • Risk of low-biased results on venous blood collected with certain evacuated blood tubes *(20)*

The decision to use one analytical method for determining BLLs versus another depends on budget, accessibility of a local laboratory, expected BLL values, desired lower limit of detection, and willingness of individuals in the catchment area to consent to venous blood draw versus a simple finger prick.

ASV Field method: This approach works well if BLLs are known to fall within LeadCare II unit's range in detection: between 3.3-65 µg/dL (note that if there is access to materials for dilution, the upper limit of detection can be much higher). This approach works well if individuals are unlikely to consent to providing venous blood since only capillary blood, a finger prick of blood, is needed. Sometimes parents are resistant to having blood drawn from very small children so a finger prick may be more appropriate. This method does not require a laboratory and is field portable. The cost is relatively low once a LeadCare II analyzer is purchased as it is just the cost of consumables, typically less than $10 per sample. More information can be found here:

https://www.magellandx.com/leadcare-products/leadcare-ii/. To ensure that the BLL results are accurate a 10% randomly selected sub sample of BLLs should be sent to a clinical laboratory.

Laboratory methods: These methods are more robust than LeadCare II and have a much lower limits of detection, typically 0.5 µg/dL but can be as low as 0.01 µg/dL, and there is no upper bound on detection. These approaches require venous blood samples, 0.1-0.5 mL of blood, a clean laboratory for conducting analyses, and highly skilled technicians. The cost per sample varies from region to region but would typically cost $15-40 per sample. As noted in the table above, some methods like ICP-MS can have a high throughput and therefore, there can be some reduction in the per sample cost if many samples are analyzed. Laboratories involved in BLL testing should be enrolled in a proficiency testing program. The US Centers for Disease Control and Prevention (CDC) maintains the Lead and Multi-Element Proficiency Program (LAMP) to the help laboratories ensure consistent, high-quality blood lead measurements. LAMP is a voluntary, no cost program that focuses on assuring the quality of multi-element analyses in whole blood including lead. LAMP is not an accreditation or certification program; however, the program does improve the precision and accuracy of blood lead, cadmium, and mercury measurements.

PART 3. Methods and best practices for determining statistically significant reductions in blood lead levels

3.1 Measuring Change in the Outcome Before and After Intervention

3.1.1 Types of data distributions

First, it is important to briefly review the different types of data distributions and how they affect the outcomes of interest. Normal or symmetrical distributions are those where all central tendency values are basically the same: the median (middle value), arithmetic mean (average value), and mode (most common value) (Figure 4). These are symmetric distributions because 50% of the data are above the central tendency and 50% are below. Normal distributions are common for physical characteristics like height and weight. If you were to plot a histogram of adult men's heights in a given country, it would look normally distributed or like a bell curve: there would be lots of individuals with average height and fewer who are extremely tall or extremely short. For this normal symmetrical distribution, the average (arithmetic mean) height would be very similar to the median height and the arithmetic mean and standard deviation would adequately describe the distribution.

For BLLs of a population, however, distributions tend to be asymmetrical (right-skewed and not normal, as in Figure 4). In a right-skewed BLL distribution, the arithmetic mean BLL would be higher than the median BLL and not be a good measure of central tendency because of the few individuals with high BLLs (outliers). For right-skewed asymmetrical distributions, a median value can be used or a geometric mean can be calculated to remove the weighting of the high BLL outliers. Unlike an arithmetic mean calculation, which uses the sum of the values, the geometric mean calculation uses the product of the values. As a result, the geometric mean would be more similar to the median value than the arithmetic mean. Either a geometric mean and geometric standard deviation, or a median value (50^{th} percentile) and interquartile range (25^{th}-75^{th} percentiles) would be better metrics for describing the asymmetric distribution of BLLs.

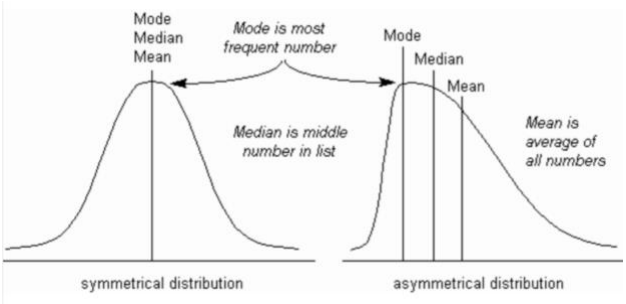

Figure 4. Representation of symmetrical (normal) and asymmetrical (right-skewed) distributions. Source: www.statisticsanswered.com.

3.1.2 Possible outcomes to measure

It is important to decide what will be compared before and after the intervention. This should have been defined within the hypothesis or research question. Once the primary outcome measure has been defined, review prior studies to determine an appropriate expected change in that outcome for a given intervention.

Among many possible outcomes of interest, here are a few:

- Measure of central tendency (e.g., median or geometric mean since an arithmetic mean is not advisable for asymmetrical distributions like BLLs, see below for more information)
- Measure of the highly exposed (e.g., the BLL value that corresponds to the 90^{th} percentile or the geometric mean of 75^{th} and 90^{th} percentiles. Assessing the BLL reduction before and after intervention for those with the highest BLLs might be important since this is the group of highest concern and also the group that might be most targeted by the intervention. We may expect the median (50^{th}) percentiles to remain unchanged, but the 75^{th} and 90^{th} percentiles to be reduced)
- Measure of prevalence of a BLL of clinical importance (e.g., % of individuals with BLLs exceeding 5 micrograms/deciliter)

3.1.3 What is a statistical test?

Although statistical tests could be the topic of entire statistics courses, in brief, a statistical test is an approach to evaluating the data against a pre-specified hypothesis. In statistical tests, two hypotheses are defined. One is the null hypothesis (H_0), which is essentially a statement of equality. The second is the alternative hypothesis (H_A or H_1). The alternative hypothesis is the statement of difference which is typically what a researcher wants to demonstrate. As an example, in the Kathgora, Bangladesh study recall that the hypothesis was "We hypothesize that a soil remediation intervention will reduce median blood lead levels (BLLs) by 15% after 14 months among children under 13 years of age in Kathgora, Bangladesh."

So, in this study, the general form of the null and alternative hypotheses would be as follows:

H_0: Median BLL Pre-intervention = Median BLL Post-intervention

H_A: (1-0.15) x Median BLL Pre-intervention > Median BLL Post-intervention

The results of statistical tests are described in terms of "rejecting the null hypothesis" or "failing to reject the null hypothesis." One would reject the null hypothesis if the test statistic, known as the p-value, is less than a pre-specified significance level (alpha), indicating that the observed reduction in BLLs can be reported as being statistically significant. The statistical power is the probability that the test correctly rejects the null hypothesis. See more on levels of significance, confidence levels, and statistical power in section 3.2 or consult a statistical reference book.

3.1.4 Factors to consider in deciding on statistical tests

The choice of which statistical test to employ depends on the research question and the distribution of the data. If conducting a BLL comparison among a single group of people before and after an intervention, use a matched analytical method like a Wilcoxon Signed-Rank Test (if an asymmetric distribution) or a paired t-test (if a symmetric distribution. If conducting a BLL comparison among two different groups of people before and after an intervention (e.g., not a matched design), use a Mann-Whitney U Test (if an asymmetric distribution) or an independent sample t-test (if a symmetric distribution). If comparing BLLs among control and exposed groups before and after an intervention, use a difference-in-differences analytical approach. Engaging a statistician would be the best way to identify the most appropriate analytical approach for your hypothesis and research question.

3.2 Sample size determination and power analysis

3.2.1 Components of a sample size calculation

The following are needed to determine an appropriate sample size: the desired level of significance, the level of statistical power, the expected effect size (change in primary outcome due to the intervention accounting for variability in the data, e.g., standard deviation or spread of the data), and the statistical test that will be used to analyze the data. An adequate amount of prior research and a well-informed hypothesis is critical for an adequate sample size calculation.

Level of significance and statistical power: Appropriate sample sizes can be calculated for a given level of significance (alpha/type 1 error) and for a given level of statistical power. The level of significance is the odds that the observed result is due to chance (and not due to the intervention, for example). The standard level of significance is 0.05 or 5% odds that resulting BLL declines are due to chance and "1-alpha" is the confidence level, typically at 95%. Statistical power (1-beta), typically set at a minimum of 0.8 or 80% is the odds that you will observe a treatment effect when it occurs. These can be relatively complicated statistical decisions and a researcher may decide to increase or decrease the level of significance or statistical power depending on the research question, hypothesis, and logistical considerations. Generally, the lower the alpha and the higher the statistical power, the larger sample size will be needed. More information can be found here or in any statistical reference book.

Effect size: One important consideration is that the estimated effect size used in sample size and power calculations considers not only the central tendency of the data (e.g., geomean or median for BLLs) but importantly it also considers the variability of the data (e.g., standard deviation). So, taking an example of a decline from a median of 23 µg/dL to 15 µg/dL observed in the Kathgora, Bangladesh study referenced in Figure 2,[11] if the standard deviation of the data were quite large (it is not uncommon in BLL studies for the standard deviation to be equal to the mean), then the drop in BLLs from 23 to 15 ug/dL (with standard deviation = mean) would

be equivalent in effect size and desired sample size to the drop from 7 to 4.5 µg/dL (with standard deviation = mean) because the effect size would be the same.

Statistical test: Sample size calculations depend on the choice of statistical test and would be different depending on the study design.

3.2.2 Resources to determine sample size and power

Once these have been decided, the researcher or a statistician can use software that calculates sample size. It can be helpful to do a "power analysis" which means creating a graph or table that shows the tradeoff between sample size and statistical power. Such an analysis can also be performed assessing the impact of varying effect sizes (or other parameters) on the needed sample size.

There are numerous pre-made sample size calculators that will work well for an individual with minimal training. At the time of writing, recommended resources include the G*Power application (free download, see Figure 5). This application can easily be used to compare how sample size might change when varying effect size, significance level, or power. Other resources include online sample size calculators which can be selected depending on the need. These typically are more suited to simple sample size calculations and not as suited for making tables or graphs to compare variations in parameters. This website provides a range of sample size calculators but these tools are always changing and a quick google search will likely return up-to-date options: https://sample-size.net/all-calculators-on-this-site/.

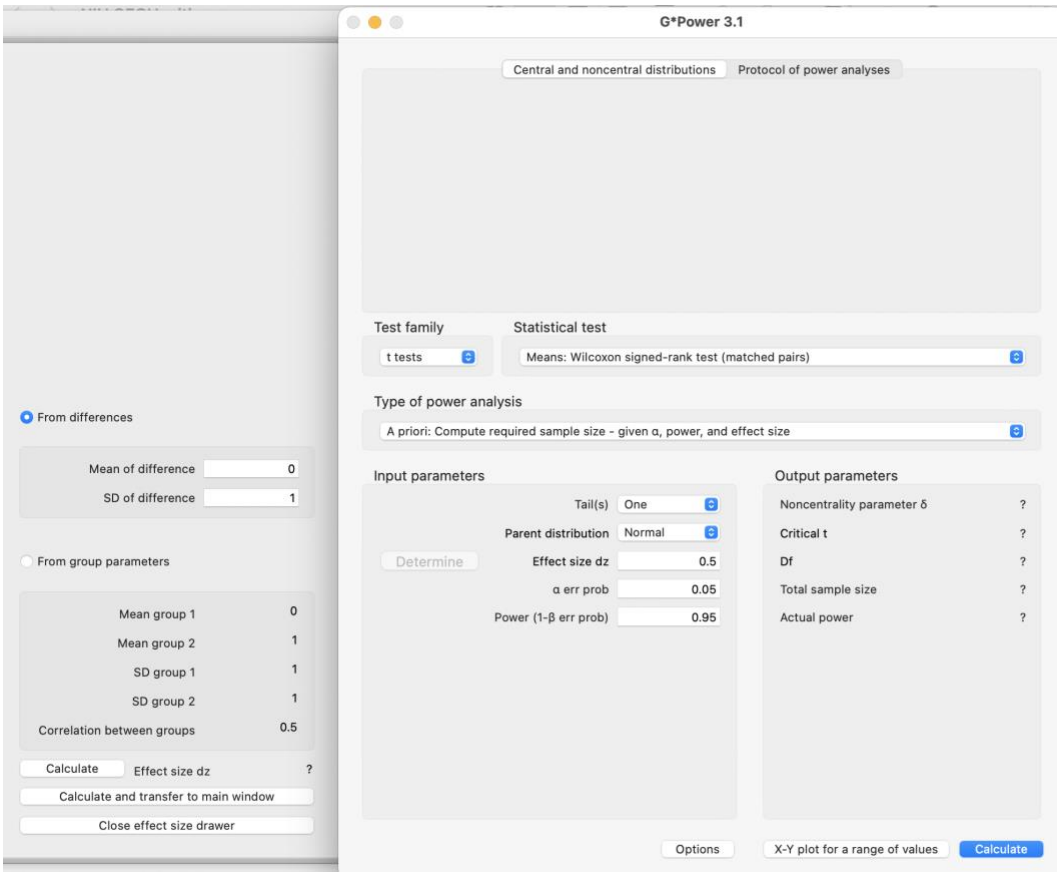

Figure 5. Screenshot of the sample size calculator from the G*Power app for a comparison of two means using a Wilcoxon Signed-Rank analytical method. Note that the pop-out window on the left side allows for the calculation of effect size after inputting the estimated mean and standard deviation for the two groups. The main window asks the user to input the desired significance level (alpha) and statistical power (1-beta).

REFERENCES

1. Laidlaw MAS, Filippelli GM, Brown S, et al. Case studies and evidence-based approaches to addressing urban soil lead contamination. *Applied Geochemistry*. 2017;83:14-30. doi:https://doi.org/10.1016/j.apgeochem.2017.02.015

2. Xue J, Zartarian V, Moya J, et al. A meta-analysis of children's hand-to-mouth frequency data for estimating nondietary ingestion exposure. *Risk analysis : an official publication of the Society for Risk Analysis*. 2007;27(2):411-420. doi:10.1111/j.1539-6924.2007.00893.x

3. Rodrigues EG, Bellinger DC, Valeri L, et al. Neurodevelopmental outcomes among 2- to 3-year-old children in Bangladesh with elevated blood lead and exposure to arsenic and manganese in drinking water. *Environmental Health*. 2016;15:44. doi:10.1186/s12940-016-0127-y

4. Bellinger DC. Prenatal Exposures to Environmental Chemicals and Children's Neurodevelopment: An Update. *Safety and Health at Work*. 2013;4(1):1-11. doi:http://dx.doi.org/10.5491/SHAW.2013.4.1.1

5. Silbergeld EK. Lead in bone: implications for toxicology during pregnancy and lactation. *Environmental Health Perspectives*. 1991;91:63-70.

6. Davidsson L, Jamil KA, Sarker SA, Zeder C, Fuchs G, Hurrell R. Human milk as a source of ascorbic acid: no enhancing effect on iron bioavailability from a traditional complementary food consumed by Bangladeshi infants and young children. *The American journal of clinical nutrition*. 2004;79(6):1073-1077.

7. Kwong LH, Ercumen A, Pickering AJ, Unicomb L, Davis J, Luby SP. Age-related changes to environmental exposure: variation in the frequency that young children place hands and objects in their mouths. *Journal of exposure science & environmental epidemiology*. 2020;30(1):205-216.

8. Das M, Elsey H, Shawon RA, et al. Protocol to develop sustainable day care for children aged 1-4 years in disadvantaged urban communities in Dhaka, Bangladesh. *BMJ Open*. 2018;8(7):e024101-e024101. doi:10.1136/bmjopen-2018-024101

9. UNICEF. *BANGLADESH MICS 2019 REPORT: KEY FINDINGS*.; 2020.

10. Rabinowitz MB, Wetherill GW, Kopple JD. Kinetic analysis of lead metabolism in healthy humans. *Journal of Clinical Investigation*. 1976;58(2):260.

11. Chowdhury KIA, Nurunnahar S, Kabir ML, et al. Child lead exposure near abandoned lead acid battery recycling sites in a residential community in Bangladesh: Risk factors and the impact of soil remediation on blood lead levels. *Environmental Research*. 2021;194:110689.

12. Banerjee A, Chaudhury S. Statistics without tears: Populations and samples. *Ind Psychiatry J*. 2010;19(1):60-65. doi:10.4103/0972-6748.77642

13. Laidlaw MA, Filippelli GM, Sadler RC, Gonzales CR, Ball AS, Mielke HW. Children's blood lead seasonality in flint, Michigan (USA), and soil-sourced lead hazard risks. *International journal of environmental research and public health*. 2016;13(4):358.

14. Dignam TA, Lojo J, Meyer PA, Norman E, Sayre A, Flanders WD. Reduction of elevated blood lead levels in children in North Carolina and Vermont, 1996–1999. *Environmental health perspectives*. 2008;116(7):981-985.

15. Manton WI, Angle CR, Stanek KL, Reese YR, Kuehnemann TJ. Acquisition and Retention of Lead by Young Children. *Environmental Research*. 2000;82(1):60-80. doi:http://dx.doi.org/10.1006/enrs.1999.4003

16. Roberts JR, Roberts J, Reigart JR, Ebeling M, Hulsey TC. Time required for blood lead levels to decline in nonchelated children. *Journal of Toxicology: Clinical Toxicology*. 2001;39(2):153-160.

17. Faul F, Erdfelder E, Lang A, Buchner A. G* Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. Behavioral Research Methods, 39 (2), 175-191. Published online 2007.

18. Lin CG, Schaider LA, Brabander DJ, Woolf AD. Pediatric lead exposure from imported Indian spices and cultural powders. *Pediatrics*. 2010;125(4):e828-35. doi:10.1542/peds.2009-1396

19. Prihartono NA, Djuwita R, Mahmud PB, et al. Prevalence of blood lead among children living in battery recycling communities in Greater Jakarta, Indonesia. *International Journal of Environmental Research and Public Health*. 2019;16(7):1276.

20. Haefliger P, Mathieu-Nolf M, Lociciro S, et al. Mass Lead Intoxication from Informal Used Lead-Acid Battery Recycling in Dakar, Senegal. *Environ Health Perspect*. 2009;117. doi:10.1289/ehp.0900696

APPENDICES: Examples of sample size projections for different sources and study designs

Appendix 1. Sample size calculation for children in Bangladesh – a before-after study powered on exposure to a used lead acid battery recycling site

For the before-after study design assessing child lead exposure near a used lead acid battery recycling site in Kathgora, Bangladesh,[11] we calculated that we would need a minimum sample size of 50 to detect a 15% decrease in BLLs from a median of 20 to 17 µg/dL with 90% power and an alpha of 0.05. This calculation was based on a Wilcoxon Signed-Rank test to compare median BLLs in the same population of children before and after a soil remediation intervention.

Below is an example of a sample size projection for this population with a fixed significance level (0.05) and varying statistical power (Figure 6).

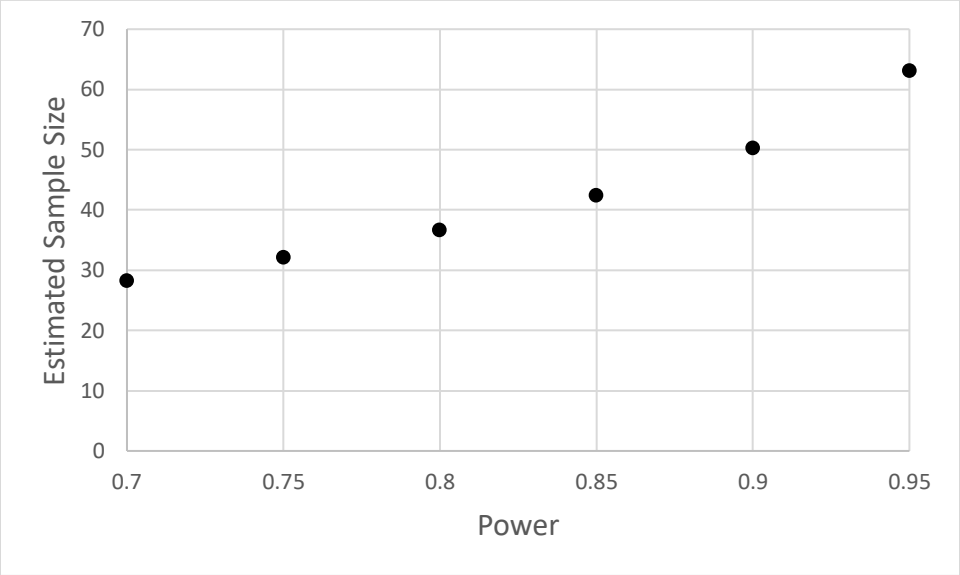

Figure 6. Sample size projections with an alpha of 0.05, and an estimated 15% decrease in median BLLs after a soil remediation (from 20 to 17 µg/dL) with varying statistical power.

Appendix 2. Sample size calculation for Bangladeshi children – a population-based cross-sectional study powered on exposure to spices

Below is an example of a sample size projection for different ratios of exposure versus non-exposure to adulterated spices for a given significance level (0.05), varying statistical power, and using a regression or independent t-test to assess differences in the mean BLLs among exposed and unexposed children in a population (Figure 7).

With a sample size of 500 children, we would have at least 80% power to detect a 1 µg/dL increase in BLL attributable to a source, at a 95% level of confidence. We based the sample size calculation on a simplified analysis: a t-test for difference in mean BLLs among "exposed" and "unexposed" children.[17] With a null hypothesis that the mean BLL of exposed children is 1 µg/dL greater than the mean BLL of unexposed children, we conducted a one-tailed test and modeled estimated sample sizes over a range in power and different ratios of exposure versus non-exposure in the population.

We would have 80% power to detect a 1 µg/dL increase in BLLs if exposure to adulterated spices were uncommon (1:5 ratio of exposure versus non-exposure) and at least 90% power if exposure were more common (1:3 or 1:1 ratio of exposure versus non-exposure).

We consider the 1 µg/dL BLL increment an appropriate minimum difference to detect. The contribution of any exposure to BLLs depends on dose. We consider the 1 µg/dL increment conservative, and we anticipate that lead-tainted turmeric is likely to generate a smaller increase in BLLs than ingestion of contaminated soil from the battery industry or other sources. A study of South Asians in North America suggested that typical consumption of adulterated spices increased mean BLLs by nearly 1 µg/dL from 3.2 to 4.1 µg/dL.[18] Since turmeric is likely to contain more lead in Bangladesh than North America, a 1 µg/dL increase in BLLs from turmeric is a reasonable minimum. Exposure to lead in soil from battery recycling has been estimated to increase child BLLs by much more than 1 µg/dL.[19,20] Moreover, guided by preliminary data in rural Bangladesh, we assumed that the BLLs vary more among children in the exposed group, with standard deviations twice as high in exposed versus unexposed children. This increased variability stems from exposure to more than one lead source or more intense exposure to a single source. Overall, a sample size of 500 children balances scientific rigor and practicalities, enabling us to conduct robust assessments.

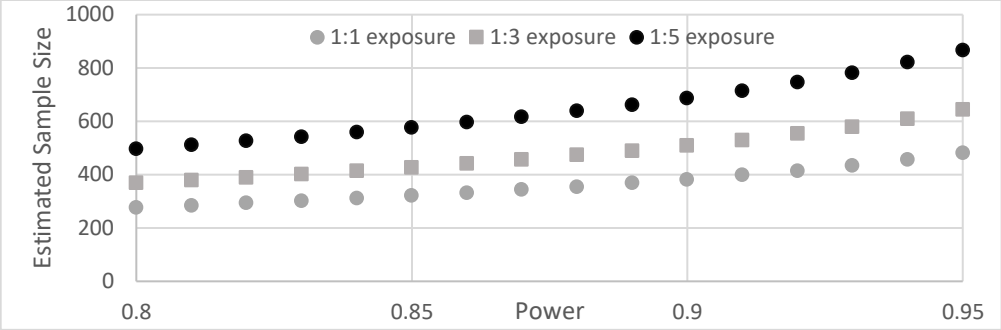

Figure 7. Sample size projections with an alpha of 0.05, and a 1 µg/dL higher mean BLL among exposed versus unexposed children based on different ratios of exposure versus non-exposure in the population (1:1, 1:3, and 1:5).

Appendix 3. Sample size calculation for a case-control study of children in the Republic of Georgia – powered on adulterated spice exposure

Below is an example of a case-control study design that aimed to identify sources of exposure among children with high and low BLLs, assuming spices were the primary source. Although this study design, hypothesis, and research question differ from what has been described in this guide, it might provide another useful reference for how to think about sample size calculations.

The projections in Figure 8 use a fixed significance level (0.05), a fixed statistical power (0.80) and vary the effect size/outcome measure – in this case an odds ratio. The planned analytical approach was to use logistic regression analysis to estimate unadjusted and adjusted odds ratios. Odds ratios approximate relative risk as described and assess how much more likely exposure to a certain source is among cases compared to controls.

Enrolling a total of 335 children (112 controls and 223 cases) would enable us to detect an odds ratio of 2.0 with 80% power and an alpha of 0.05. This means that if 33% of children in the control group were exposed to lead-tainted spices, we would be able to say case and control groups are statistically different (i.e. not due to random effects) if the prevalence of exposure to lead-tainted spices in the case group was 50%. Below I describe the assumptions and information used to make this calculation, as well as how to interpret this.

Table 2 and Figure 8 show the sample sizes needed for different proportions of cases versus controls with lead exposure. For all estimates, 80% statistical power and an alpha of 0.05 are used. These values for power (1-beta) and alpha are standard for epidemiologic studies.

An alpha of 0.05 means that there is a 5% chance we would detect a difference in exposure sources between children with elevated and low BLLs when there was actually no difference. On the other hand, having 80% power means that we have an 80% chance of detecting a difference when there is one.

The sample size calculation in Table 2 and graphed in Figure 8 assumes spices are the primary lead exposure sources. The calculation uses an assumption that ~50% of cases are exposed to lead-tainted spices and varies the percent of controls exposed to lead-tainted spices from 22-40% (equivalent odd ratios of 1.5-3.5). This choice is informed by exposure data from the 2019 Pure Earth NCDC study of 25 children. Although the study had a small sample size, it is the best information about the percent of exposure among case versus control children. In that study, 10 of the 19 children with elevated BLLs also had elevated lead in spices (53%). By comparison, only 2 of the 6 children with low BLLs had elevated lead in spices (33%). This ratio of exposure among “cases” versus “controls” would equate to an odds ratio of 2.2. The interpretation of the odds ratio is that a case child with elevated BLLs is 2.2 times more likely to have elevated lead in spices.

Table 2. Table of sample size for different proportions of controls and cases exposed to a lead source (spices), with 80% power and an alpha of 0.05.

Odds Ratio	Total Sample Size	# of Controls	# of Cases	Percent of Controls Exposed	Percent of Cases Exposed
1.5	918	306	612	40	50
1.55	792	264	528	39.2	50
1.6	693	231	462	38.5	50
1.7	551	184	367	37	50
1.8	455	152	303	35.7	50
2	335	112	223	33.3	50
2.25	252	84	168	30.8	50
2.5	204	68	136	28.6	50
2.75	171	57	114	26.7	50
3	150	50	100	25	50
3.25	135	45	90	23.5	50
3.5	122	41	81	22.2	50

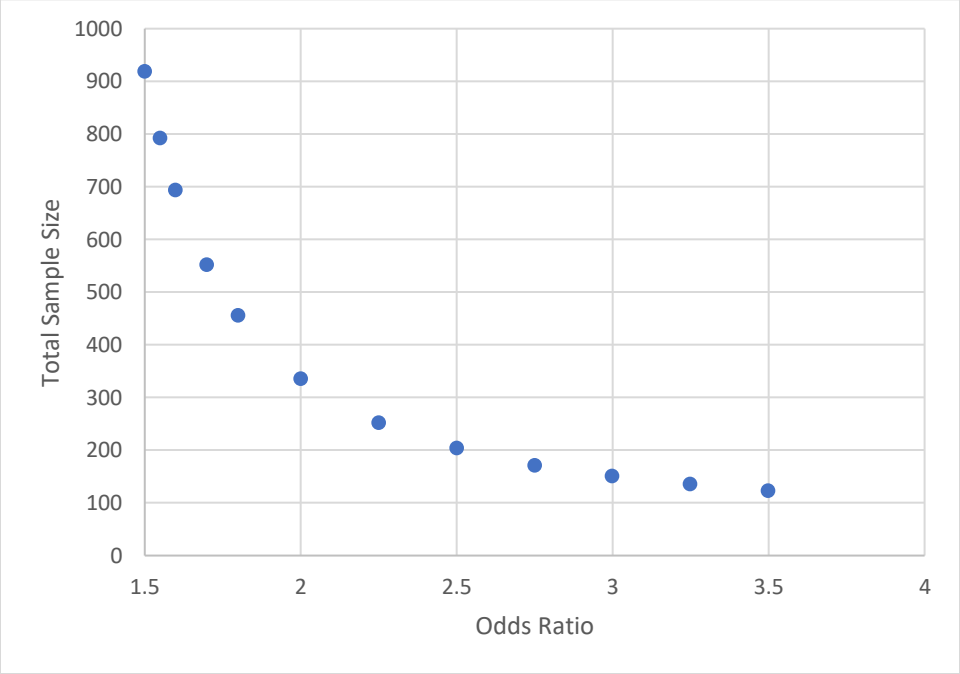

Figure 8. This is a graph of the data presented in Table 2. It shows the total sample size needed (with a ratio of 2:1 for the number of cases: controls enrolled) for different odds ratios, with 80% power and an alpha of 0.05.