Study Objective 1: Assessing the Impact of Environmental Exposures on Child Nutrition and Growth Patterns in Rural Populations

Outcome Variable (Dependent Variable):

* Child Nutrition/Growth (e.g., stunting, underweight, growth velocity)
  + Growth measures: Height-for-age z-scores (HAZ), weight-for-age z-scores (WAZ), mid-upper arm circumference (MUAC), or growth velocity (change in weight/height over time).

Independent Variables (Factors that could influence growth)

1. Socioeconomic Status:
   * Household income, education levels of parents, access to social services, parental occupation.
2. Nutritional Intake:
   * Frequency and diversity of food intake, protein, fat, and micronutrient consumption (especially iron, zinc, vitamin A).
3. Health Interventions:
   * Immunization status, deworming treatments, iron supplementation.
4. Access to Healthcare:
   * Proximity to health facilities, frequency of health visits.

Exposure Variables (Environmental factors influencing growth)

1. Water Quality:
   * Contamination levels (e.g., E. coli, heavy metals, fluoride, etc.), access to safe drinking water, sanitation facilities (toilets, waste disposal).
2. Air Pollution:
   * Indoor air pollution (from biomass cooking, use of coal or wood), outdoor pollution levels (PM2.5, NO2, ozone).
3. Housing Quality:
   * Ventilation, presence of mold or damp conditions, floor materials (dirt vs. cement floors), overcrowding.
4. Climate/Weather:
   * Temperature, rainfall, seasonal variations affecting disease transmission (e.g., diarrheal diseases).

Mediator Variables (Variables that mediate the relationship between exposure and growth outcomes)

1. Gut Microbiome Diversity:
   * Richness and diversity of gut bacteria (e.g., Firmicutes, Bacteroidetes).
   * Dysbiosis indices (imbalance in gut microbiota).
   * Short-chain fatty acids (SCFAs): Products of fermentation of fiber in the gut, which may influence nutrient absorption.
2. Nutrient Absorption Efficiency:
   * Biomarkers of nutritional status (e.g., serum levels of vitamins, iron, zinc) that mediate the relationship between diet and growth outcomes.
3. Immune System Activation:
   * Immune markers (e.g., C-reactive protein, cytokine profiles) as mediators linking environmental exposures (e.g., poor sanitation) to inflammation and growth restriction.
4. Intestinal Permeability:
   * Markers of gut barrier integrity (e.g., fecal calprotectin) may mediate how environmental stressors like waterborne diseases or pollution affect growth.

Study Objective 2: The Role of Iron Deficiency, Malnutrition, and Environmental Factors on Anemia Prevalence

Outcome Variable (Dependent Variable):

* Anemia/Hemoglobin Levels (e.g., hemoglobin concentration, ferritin, serum iron levels, transferrin saturation)
  + Prevalence of iron deficiency anemia (IDA), microcytic hypochromic anemia, and overall anemia prevalence.

Independent Variables (Factors influencing anemia)

1. Socioeconomic Status:
   * Household income, maternal education, and access to food (diversity and quantity).
2. Nutritional Status:
   * Intake of key nutrients like iron, folic acid, and vitamin A, as well as the diversity of the diet.
3. Health Interventions:
   * Iron supplementation, deworming, and vaccination status.
4. Micronutrient Deficiencies:
   * Zinc, folate, vitamin A deficiency (potential confounders).

Exposure Variables (Environmental factors contributing to anemia)

1. Water and Sanitation Quality:
   * Availability of clean drinking water, sanitation facilities, and exposure to waterborne pathogens.
2. Pollution Exposure:
   * Exposure to indoor and outdoor air pollution, which can affect iron absorption or contribute to systemic inflammation (e.g., PM2.5, carbon monoxide).
3. Dietary Contaminants:
   * Presence of dietary inhibitors like phytates or tannins, which may reduce iron absorption, and presence of toxic substances (e.g., lead).
4. Climate and Seasonal Variations:
   * Seasonal food availability and agricultural practices (e.g., iron-rich food availability during dry season vs. rainy season).

Mediator Variables (Mediating factors between exposure and anemia)

1. Gut Microbiome Function:
   * Gut flora's role in iron absorption and metabolism. For example, Lactobacillus species might help improve bioavailability of iron.
2. Intestinal Inflammation:
   * Levels of fecal calprotectin or other biomarkers of gut inflammation that may impair nutrient absorption and contribute to anemia.
3. Inflammation and Immune Response:
   * Elevated levels of C-reactive protein (CRP) or other immune markers due to environmental stressors (e.g., air pollution) that could impair iron status.
4. Helminth Infection:
   * Worm load or parasitic infections that mediate iron deficiency through blood loss or nutrient malabsorption.

Study Objective 3: Investigating the Impact of Environmental Sanitation and Microbiome Changes on Diarrheal and Respiratory Illness Incidence in Children

Outcome Variable (Dependent Variable):

* Diarrhea and Respiratory Illness Incidence (e.g., frequency, severity, and duration of diarrheal episodes, number of respiratory infections or ARI hospitalizations)

Independent Variables (Factors contributing to diarrhea and respiratory illness)

1. Socioeconomic and Demographic Factors:
   * Household income, parental education, access to healthcare services, child’s age, sex, and vaccination status.
2. Nutritional Factors:
   * Nutritional status, specifically undernutrition (low weight-for-height) or deficiencies in key micronutrients like zinc and vitamin A.
3. Health Interventions:
   * Immunization rates, access to health services, deworming, and vitamin supplementation.

Exposure Variables (Environmental factors influencing disease incidence)

1. Sanitation and Hygiene:
   * Access to clean water, waste disposal, sanitation systems, and hygiene practices (e.g., handwashing).
2. Indoor Air Pollution:
   * Use of solid fuels (e.g., wood, coal) for cooking or heating, household exposure to smoke.
3. Climate and Seasonality:
   * Temperature, rainfall, and humidity influencing the spread of waterborne diseases or respiratory infections (e.g., flu season).
4. Pest Infestation:
   * Presence of rodents or insects (e.g., cockroaches) in the household, which could transmit pathogens.

Mediator Variables (Factors that mediate the relationship between exposure and disease outcomes)

1. Gut Microbiome Composition:
   * The composition of the gut microbiota influencing the immune response and gastrointestinal health (e.g., Firmicutes, Bacteroidetes).
2. Immune Function:
   * Cytokine levels, CRP, or other markers of systemic inflammation as mediators between environmental exposures and disease.
3. Microbial Dysbiosis:
   * Changes in the gut microbiota (e.g., decreased diversity or overgrowth of pathogenic bacteria) that influence susceptibility to infections.
4. Respiratory Microbiome:
   * Differences in nasal or lung microbiota that mediate respiratory diseases such as asthma or pneumonia in the context of environmental exposures.

Study Design Recommendations:

* Cross-Sectional Analysis: Ideal for examining associations between environmental factors (e.g., sanitation, pollution) and health outcomes (e.g., anemia, growth).
* Longitudinal Analysis: Suitable for tracking changes in outcomes (e.g., growth, anemia, disease incidence) over time and identifying causal relationships.
* Mediation Models: Use statistical techniques like mediation analysis or structural equation modeling (SEM) to test how microbiome changes mediate the relationship between environmental exposures and health outcomes.

Other Objectives:

A. Eco-epidemiology of parasitic infections

* Objective: To assess how land use change, water quality, and sanitation practices influence intestinal parasite prevalence among Indigenous Amazonian communities.
* Potential question: How does proximity to disturbed forest or gold mining sites predict infection risk?

B. Environmental determinants of anemia and nutritional status

* Objective: To evaluate associations between dietary diversity, micronutrient intake, water quality, and anemia prevalence.
* Potential question: How do dietary transition and environmental contamination (e.g., mercury) contribute to iron-deficiency anemia?

C. Longitudinal eco-social health transitions

* Objective: To examine how modernization (e.g., road access, market integration) affects health outcomes over time.
* Potential question: Do communities with greater exposure to market economies exhibit different patterns of infectious vs. chronic disease?

How to create a “road access” variable

You can derive *road access* using geospatial analysis:

| Method | Data Needed | Description |
| --- | --- | --- |
| Distance to nearest road | GIS shapefile of roads (from OpenStreetMap or national mapping agency) + GPS coordinates of each community | Compute Euclidean distance (km) from community center or households to nearest road. |
| Travel time to urban center | Same road data + elevation and river network | Model accessibility (in hours of travel) using friction surface or cost-distance modeling (e.g., in QGIS, Google Earth Engine, or R’s *gdistance* package). |
| Road density | Area-based measure | Useful if you’re working at regional or buffer-zone level. |

How to estimate “market integration”

“Market integration” is typically not a single variable — it’s a composite index that reflects how connected a community or household is to the regional economy.

You can build this variable from:

* Distance/time to nearest town or market (derived from spatial data)
* Frequency of selling or buying goods (if your questionnaire includes this)
* Diet composition: proportion of purchased vs. foraged/produced foods
* Occupation or income source: wage labor, cash cropping, etc.
* Possession of market goods: cell phones, radios, metal pots, plastic containers, etc.
* Participation in tourism, logging, or government programs

If those components exist in your dataset, you can:

* Assign numeric scores to each indicator
* Combine them via Principal Component Analysis (PCA) or factor analysis  
  → creating a *Market Integration Index (MII)*
* 4. Supplementary Data Sources for Deriving Access/Integration

| Variable | Source | Notes |
| --- | --- | --- |
| Road networks | OpenStreetMap | Extract shapefiles for Peru/Amazon region |
| Settlements and towns | WorldPop, GADM, or Peru’s GeoPortal | Identify nearest towns with permanent markets |
| Travel time surfaces | Malaria Atlas Project Accessibility Map | Global 1 km grid of travel time to nearest city |
| Economic activity | MODIS night-time lights (VIIRS) | Proxy for market intensity or infrastructure presence |

D. Microbiome and environmental exposure

* Objective: To explore how changes in water sources, diet, and antibiotic exposure shape gut microbiome diversity and its relation to parasitic load or inflammation.

E. Climate change and vector-borne disease risk

* Objective: To model how local climate variability (temperature, rainfall, humidity) correlates with parasite or vector prevalence.
* Potential question: How does seasonal rainfall influence transmission intensity of soil-transmitted helminths or protozoa?

F. One Health approach to ecosystem health and human infection

* Objective: To investigate linkages between wildlife biodiversity, livestock management, and zoonotic infections.
* Potential question: Does reduced biodiversity or increased domestic animal contact elevate zoonotic parasite risk?

G. Social determinants and community resilience

* Objective: To identify how education, traditional ecological knowledge, and community organization buffer environmental health risks.

📊 2. Complementary Variables You Can Integrate

To strengthen ecological and epidemiological inference, you can integrate these external variables:

| Domain | Example Variables | Purpose |
| --- | --- | --- |
| Land Use & Environment | Forest cover %, deforestation rate, NDVI (vegetation index), mining sites, distance to river/road | Environmental exposure, ecosystem integrity |
| Water Quality | pH, turbidity, E. coli, heavy metals (Hg, Pb, As), pesticide levels | Assess contamination and link to disease outcomes |
| Climate | Rainfall, temperature, humidity, drought index (SPI), flood frequency | Seasonal or climate-driven disease dynamics |
| Socioeconomic | Education level, market distance, household income, access to healthcare | Mediating factors for disease vulnerability |
| Nutrition | Food frequency, dietary diversity, caloric intake | Nutritional determinants of immune response and anemia |
| Biodiversity | Mammal/bird species richness, wildlife abundance, domestic animal density | One Health and zoonotic perspectives |
| Health Outcomes | Parasite species, anemia status, hemoglobin level, diarrhea incidence, anthropometrics | Epidemiological core outcomes |

🌎 3. Potential Open-Source Data Sources

| Data Source | What It Provides | URL / Access |
| --- | --- | --- |
| MODIS / Landsat (NASA) | Land cover, forest loss, NDVI, surface temperature | https://earthdata.nasa.gov |
| WorldClim / CHELSA | High-resolution climate data (temperature, rainfall, humidity) | https://worldclim.org |
| HydroSHEDS / HydroBASINS | River networks, watershed boundaries | https://hydrosheds.org |
| Global Surface Water Explorer (JRC) | Water occurrence and seasonality | https://global-surface-water.appspot.com |
| Global Forest Change (Hansen et al.) | Deforestation, tree cover loss | https://earthenginepartners.appspot.com/science-2013-global-forest |
| Demographic and Health Surveys (DHS) | Regional-level health and demographic data | https://dhsprogram.com |
| World Bank / UNDP | Socioeconomic and environmental indicators | https://data.worldbank.org |
| FAO AQUASTAT | Water and agricultural statistics | <https://www.fao.org/aquastat> |
| GBIF | Biodiversity records (species distribution) | https://www.gbif.org |

🧭 Possible Integrative Frameworks

You could situate your PhD around one of these theoretical or methodological frameworks:

* Ecohealth / One Health approaches
* Geo-epidemiology (spatial analysis of disease ecology)
* Exposome framework (integrating environmental and biological exposures)
* Socio-ecological resilience theory (linking traditional practices to health outcomes)
* Bayesian hierarchical modeling (for multi-level, spatial, or temporal data)