OECD and UNHCR Datathon 2025

Multidimensional Vulnerabilities and Child Health: Evidence from Displaced Households in Uganda

By: **Eyram Espoir TETSHIE, Master student at ENSAI, France**
& **William Kokou AMEDANOU, Master student at PSE, France**

Summary

Context and Refugee Situation in Uganda

Largest Refugee Population in Africa

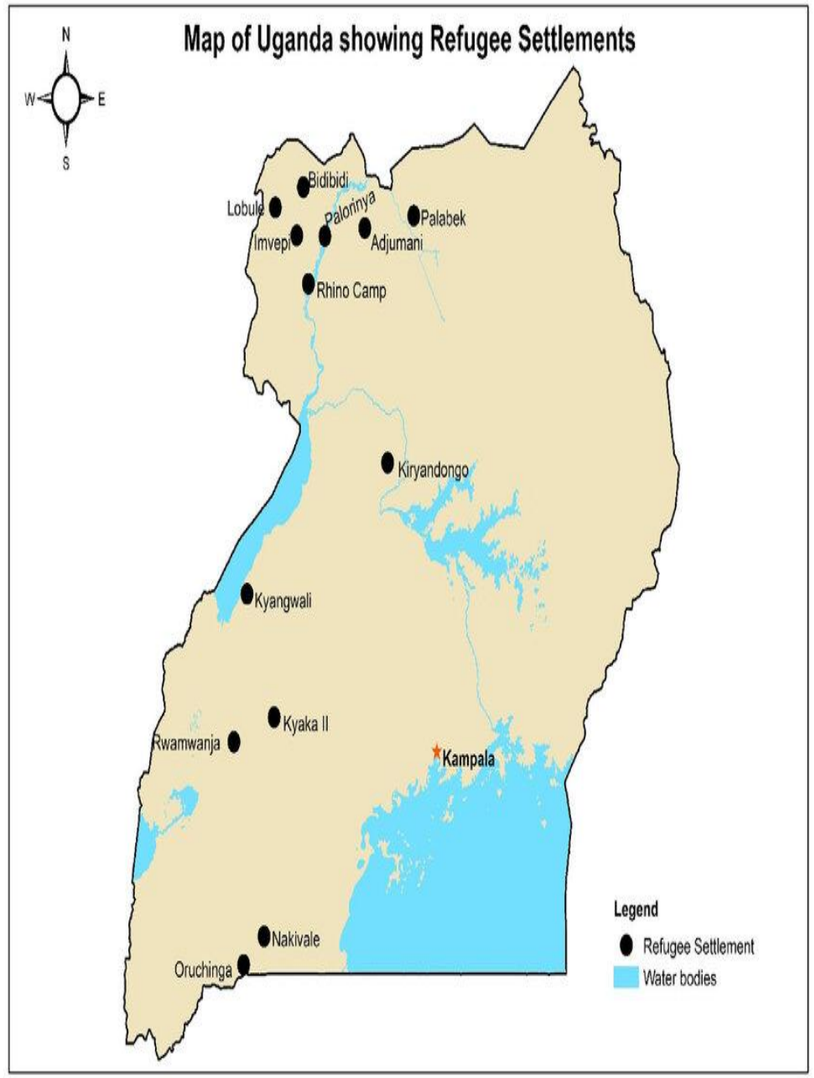
Uganda hosts over 1.8 million refugees and asylum seekers, making it the 6th largest refugee-hosting country globally. The majority come from South Sudan (54%) and the Democratic Republic of Congo (32%) in 2025.

Progressive Refugee Policy

Despite being known for its welcoming refugee policy, Uganda faces significant challenges as the steady influx strains its capacity to provide adequate services and support.

Multiple Challenges

The country faces natural disasters, disease outbreaks, worsening climate conditions, and funding cuts for humanitarian aid, reducing access to essential services and putting pressure on already under-resourced public services in refugee settlements.



Child Health Vulnerability and Study Objectives

Children are the most vulnerable group within refugee populations. Nothing is more distressing for parents than threats to their children's health. Ensuring their well-being is therefore a critical priority.

Challenges

- Poor sanitation and hygiene
- Unsafe shelter
- Limited healthcare access
- Food insecurity
- Lack of legal protection
- Extreme climate event shocks

Our Study

1- Understanding how this multiple dimensions: education, legal status, sanitation, and living conditions impact the health outcomes among displaced children.

2- Identify the interactions between those dimensions that hold the greatest potential impact on child health status for high-impact and cost-effective interventions

Using advanced analytical techniques, including Machine Learning models for complex relation identification, to provide data-driven insights for targeted humanitarian interventions.

Key Factors Affecting Child Health

WASH (Water, Sanitation, and Hygiene)

WASH interventions significantly impact child growth and development. Poor sanitation, unsafe water, and inadequate hygiene increase exposure to disease and environmental enteric dysfunction, limiting nutrient absorption and impeding development. (Britto, 2016)

Food Assistance and Nutrition

Proper nutrition is critical for optimal growth and development. Chronically hungry and malnourished children are more likely to experience adverse health outcomes that affect their lifelong wellbeing (Hannah Wagner, 2024).

Education and Early Childhood Programs

Education shapes health through multiple pathways, including health literacy, access to health programs, school feeding, and moral support. It directly impacts health indicators such as life expectancy and health-related behaviors.

Data and Methodology

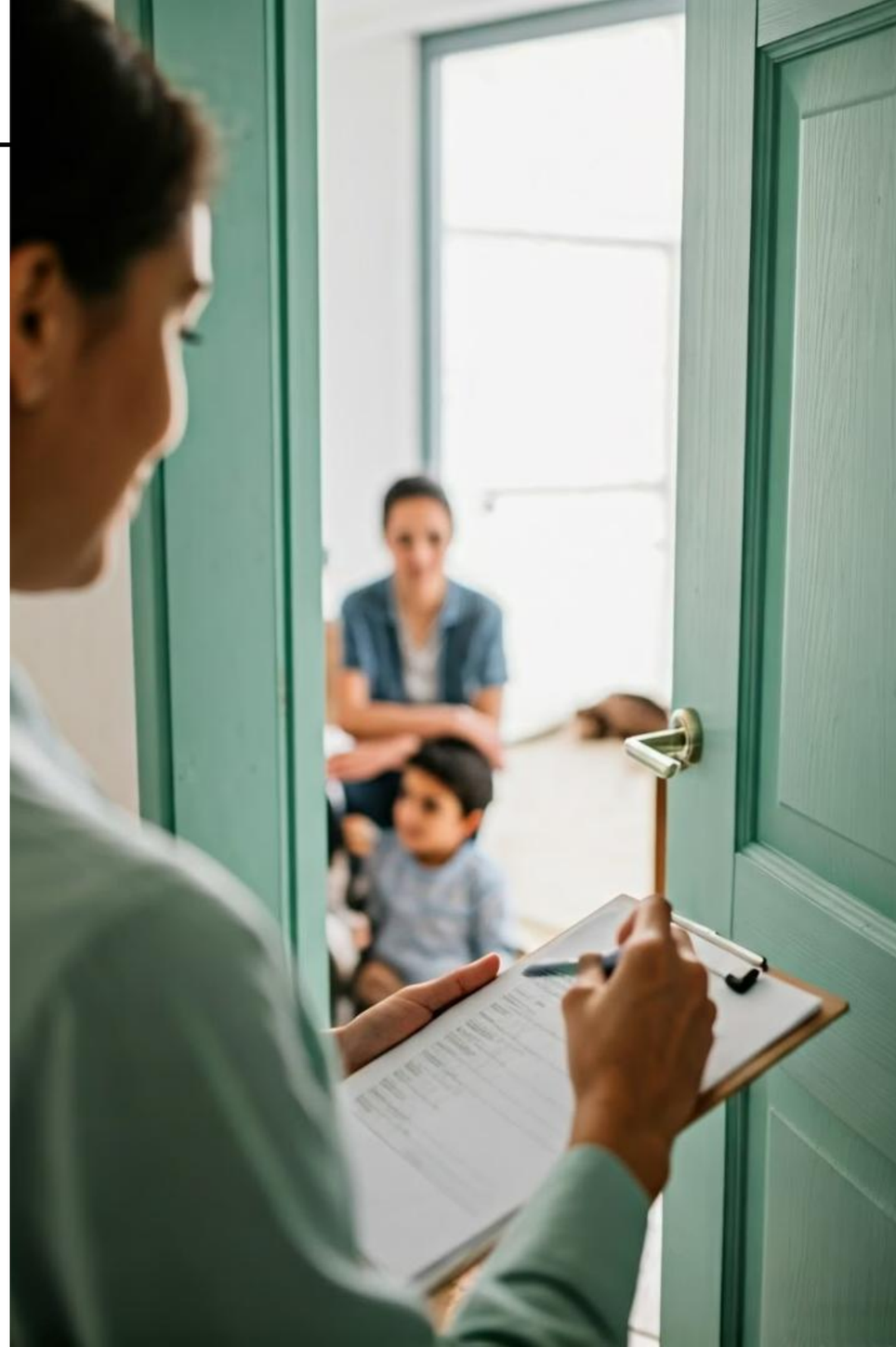
Data

Data Collection

The study uses data from the 2018 Uganda Joint Multi-Sector Needs Assessment (JMSNA), covering all 30 refugee settlements and 11 host community districts across Uganda's Midwest, Northwest, and Southwest regions.

Health Index Construction

A composite Health Index was created using six health-related indicators: recent diarrhoea in young and older children, polio vaccination, measles vaccination, vitamin A supplementation, and use of insecticide-treated mosquito nets.



Health Index Construction

Variable Recoding

Each health indicator was recoded: +1 for favorable outcomes, -1 for unfavorable outcomes, 0 for missing/no answers

Principal Component Analysis

PCA was applied to standardized indicators to reduce dimensionality and compute a weighted composite index

Index Scaling

The final Health Index was scaled to range from 0 to 1, with higher values indicating better health outcomes

$$\text{Index}_i = \frac{\lambda_1 I_{i1} + \lambda_2 I_{i2}}{\lambda_1 + \lambda_2} \qquad I_{ik} = \sum_{j=1}^{p} w_{jk} \cdot X_{ij} \qquad w_{jk} = \frac{|\phi_{jk}|}{\sum_{j=1}^{p} |\phi_{jk}|}$$

Where I_{ik} is the intermediate score of individual i on component k

Indexi : weighted average of the first two components with weights based on their explained variances λ1 and λ2

Summary of Explanatory Dimensions

The explanatory dimensions are constructed as an index ranging from 0 to 1, with higher values indicating better conditions. Control variables included household characteristics such as size, number of children, age, gender, and resources.

WASH Index

- Access to latrines by children
- Access to clean drinking water
- Use of soap for handwashing

Food Assistance Index

- Sources of food assistance
- Diversity of food consumed
- Presence of proteins, vegetables, fruits, and cereals

Environment Index

- Type and condition of shelter
- Availability of kitchen
- Exposure to environmental hazards

Other Variables

- School attendance
- Legal refugee registration status
- Household and children characteristics

Linear Regression methodology

Specification choice

- Causal inference via IV regression is infeasible due to weak/exogenous instrument requirements.
- 2019 data exclusion stems from non-comparable survey designs.

OLS

Consequently, we rely on a linear regression approach, including a comprehensive set of control variables in an attempt to minimize omitted variable bias as much as possible.

We include:

- **District fixed :** to control for unobserved heterogeneity across broader geographic
- **Cluster standard errors** at the settlement level to correct for intra-group correlation

$$HealthIndex_h = \beta_1 \text{WASH}_h + \beta_2 \text{Food}_h + \beta_3 \text{School}_h + \beta_4 \text{Environment}_{ih} + \beta_5 \text{LegalStatus}_h + \mathbf{X}_h\gamma + \delta_d + \epsilon_h$$

Clustering methodology

K-means

To identify distinct profiles of child health vulnerability among households, we performed an unsupervised classification using the K-means clustering algorithm.

Correspondence Analysis (MCA)

MCA allows us to project both households and variable categories (modalities) onto the same coordinate system, providing a visual and analytical representation of how different response categories cluster together

Interpretation

The clustering results can then be interpreted by analyzing which variable modalities are situated near each cluster in the MCA projection.

Random Forest methodology

Goal

To predict the health status of children in refugee households.

Advantages

- Identify the most influential variables, It works well with **nonlinear relationships** and complex interactions
- Test whether combinations (interactions) of factors matter more than individual ones,
- And explore whether we can *anticipate* vulnerability using available household data.

How does it work?

Ensemble Method: Combines multiple decision trees trained on bootstrapped data samples.

The Random Forest algorithm works by building multiple decision trees on bootstrap samples of the data and averaging their predictions.

At each node of the trees, a random subset of predictors is considered to split the data. Formally, the final prediction is the majority vote among all B trees:

$$\hat{y} = mode(T_1(x), T_2(x), ..., T_B(x))$$

RESULTS

- The health index distribution
- Clustering analysis
- Impact analysis
- Random Forest results

Health Status Distribution by Household Type

Refugee Health Distribution

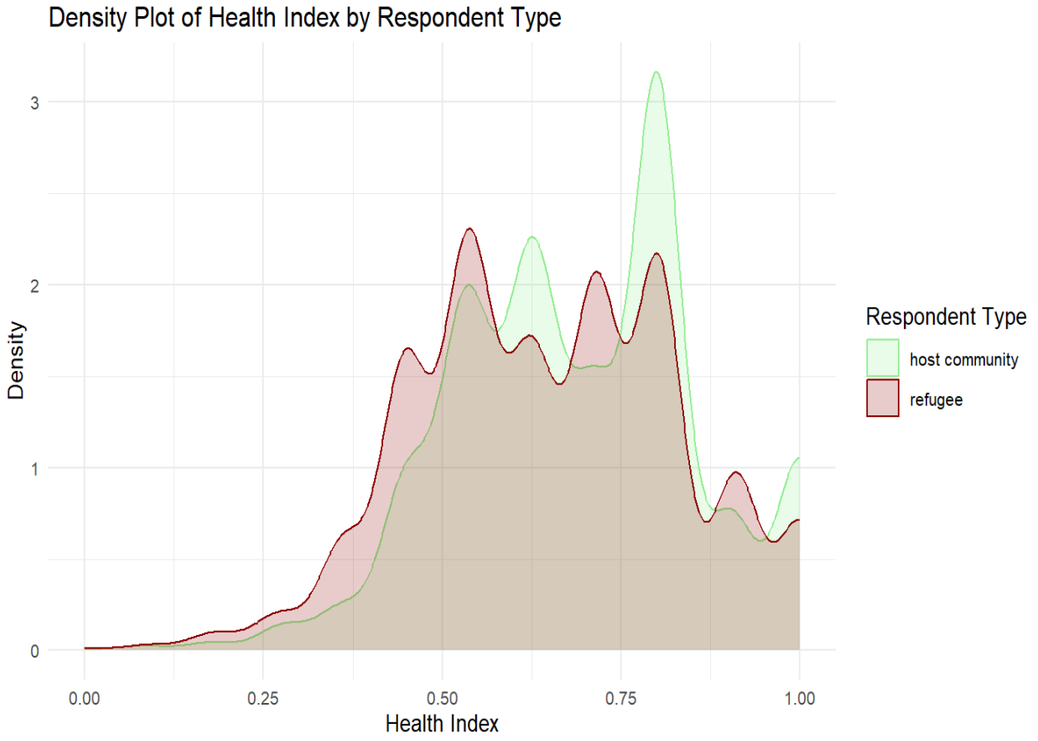
The refugee curve shows a slightly broader spread and higher density in the lower half of the index, indicating a greater share of children with poorer health outcomes.

Host Community Health

The host community distribution is more concentrated toward higher index values, suggesting relatively better overall health conditions for children in these households.

Vulnerability Impact

Being in a vulnerable group is not strongly associated with a lower Health Index, suggesting that vulnerability status alone does not significantly affect children's health outcomes.



The overlapping error bars across subgroups indicate that differences might not be statistically significant, suggesting a need for deeper analysis of the factors influencing health outcomes.

Overview of the clustering analysis

K-means Clustering

Applied K-means with K=2 clusters on the health index to distinguish between "Poor Health" (51.3%) and "Good Health" (48.7%) households

Clear Separation

"Poor Health" cluster: mean index of 0.52, maximum of 0.66; "Good Health" group: mean index of 0.82, minimum of 0.67

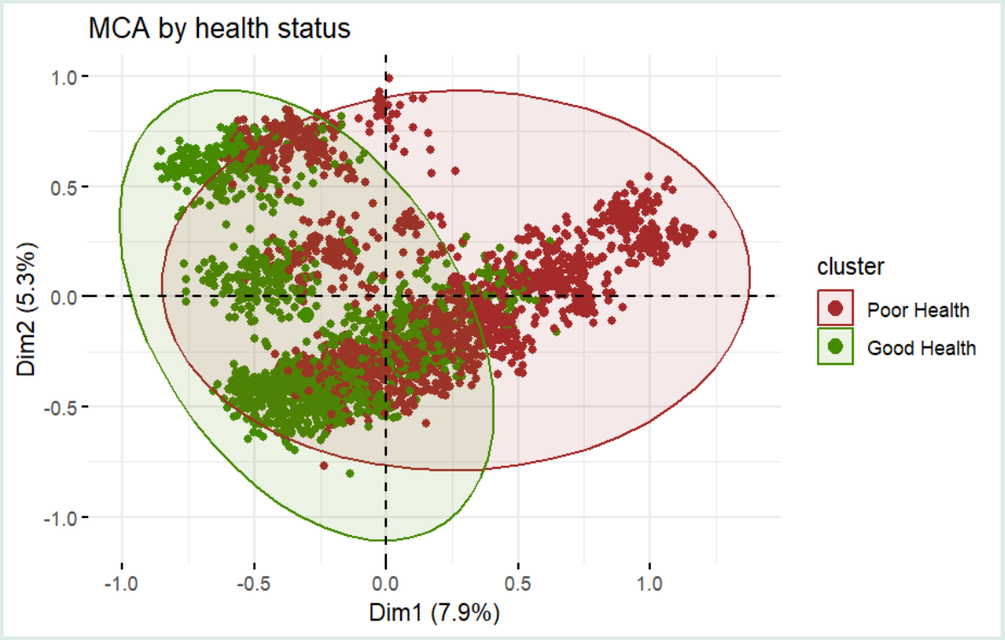
Multidimensional Vulnerabilities

Analysis confirmed convergence of vulnerabilities that likely exacerbate health risks for children in "Poor Health" households

Multiple Correspondence Analysis

MCA revealed "Poor Health" group associated with poor hygiene, poor diet, inadequate environment, and lack of legal documentation. The MCA thus reinforces the multidimensional and intersection alnature of child health and provides valuable insights for targeted interventions.

Clustering Analysis of Health Vulnerability



The MCA plot shows distinct spatial clustering of households by health status.

- "***Poor Health***" households group on the **right side** of the graph.
- "***Good Health***" households concentrate on the **left side**.

- ***Poor Health Group***: Strongly linked to poor hygiene, inadequate nutrition, unfavorable environmental conditions, and lack of legal documentation.
- ***Good Health Group***: Associated with better hygiene and dietary practices, indicating these factors may significantly improve child health outcomes.

Regression Analysis Results

Variables	*Model 1*	*Model 2*	*Model 3*
Wash index	0.039**	0.0178	0.0458***
Food index	0.066***	0.0762**	0.0483
Attending school	-0.010	-0.0355***	-0.0066
Environment index	0.059***	0.0684***	0.0611***
Legal registration	0.003	0.0229	-0.0029
Fixed Effect	No	Yes	Yes
Controls Added	Yes	No	Yes

The regression analysis highlighted the importance of environmental conditions and WASH conditions in improving children's health. The environment index showed consistent and robust relationships across all models, with coefficients ranging from 0.059 to 0.068, all significant at the 1% level.

Food also emerged as important for child health, with a statistically significant result of at 1% level in the simple regression and fixed effects model . Interestingly, school attendance showed mixed results, while legal registration showed no significant association with children's health status across all models.

Interaction Effects from Random Forest Model

Model Performance

Achieved 65% accuracy with 13% error rate, indicating meaningful predictive information

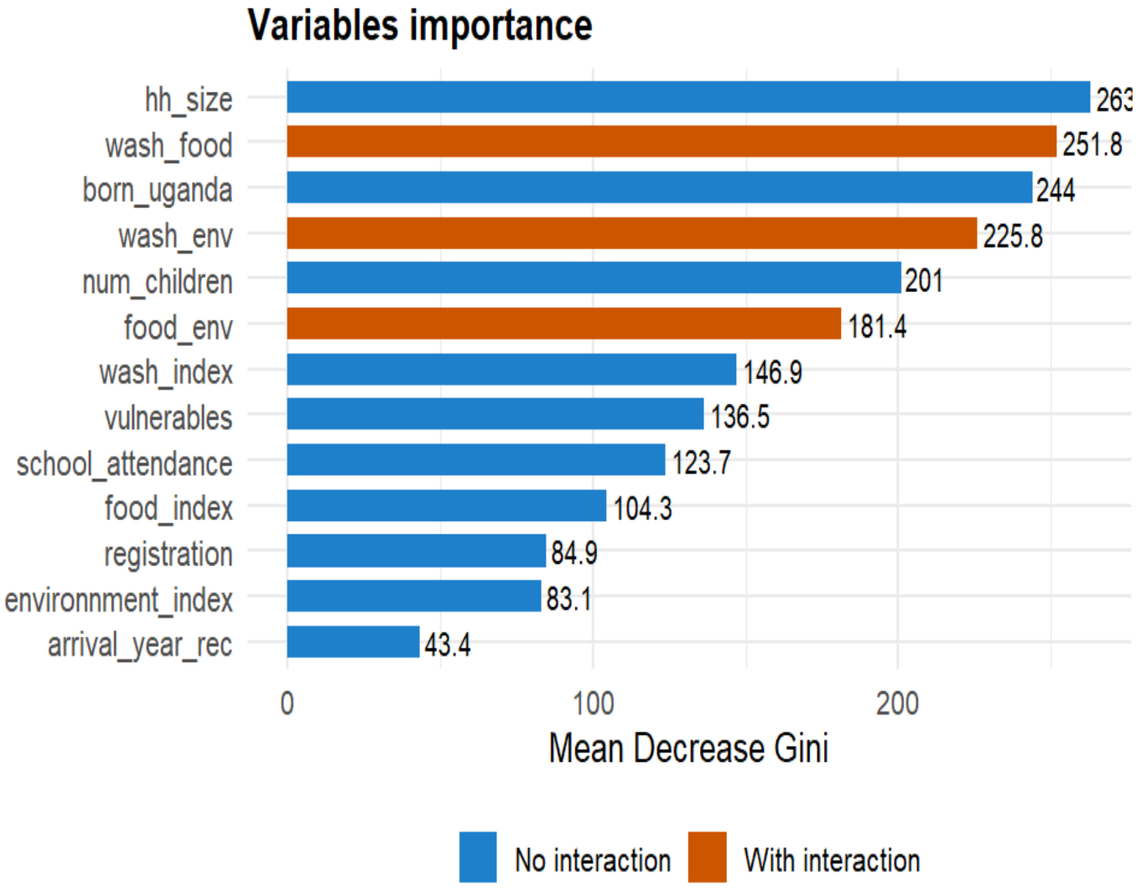
Key Finding

Interaction terms—especially **wash×food, wash×environment, and food×environment**—were among the most influential predictors

Poor hygiene alone may not be highly harmful, but combined with poor nutrition or unsafe environments, health risks increase.

The combination of food insecurity and poor environmental conditions undermines children's physical and emotional well-being.

These overlapping deprivations reinforce each other, reflecting real-world compounded vulnerabilities.



Conclusion

This project investigated the determinants of children's health among displaced and host communities in Uganda, using data from the 2018 Uganda Household Survey

Our findings reveal that poor WASH conditions, inadequate food diversity, and unsafe environments jointly contribute to deteriorated health outcomes

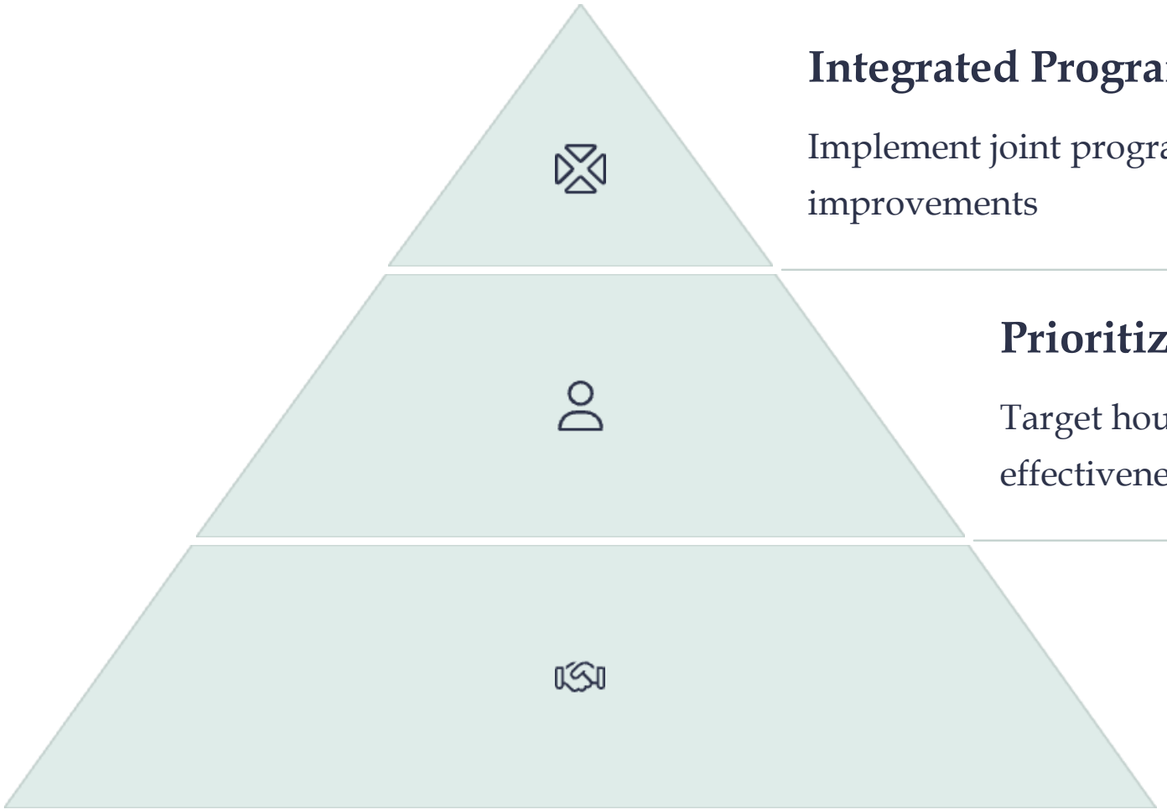
Recognize that the deterioration of hygiene and living conditions can disproportionately harm displaced children's health, highlighting the need for cross-sectoral interventions.

Limitation

The survey was household-focused rather than child-specific, which led us to use proxy variables to approximate child-level health.
Some indicators used in the health index, such as mosquito net use or vaccination, reflect access to services rather than direct health outcomes
Inconsistency between multiple survey waves making difficult to monitor evolution and panel analysis

Finally, although cross-country comparisons were encouraged, we focused solely on Uganda mainly due to time constraints.

Policy Recommendations



Integrated Programming

Implement joint programs combining food security, WASH, and environmental improvements

Prioritize Vulnerable Subgroups

Target households with compounded vulnerabilities for enhanced equity and effectiveness

Cross-Sectoral Collaboration

Foster partnerships between humanitarian and development actors

This project aligns strongly, but not exclusively with Sustainable Development Goal (SDG) 3 on Good Health and Well-being, as well as SDG Target 10.7, which calls for implementing well-managed migration policies. By focusing on displaced households in Uganda, the study shed slight on how multidimensional vulnerabilities affect children's health outcomes in fragile and often under-resourced contexts.

Thank you for attention