Child Health in Bangladesh

Harold Haugen
School of Data Science
The University of Virginia
waa4bq@virginia.edu

Aqsa Majeed
School of Data Science
The University of Virginia
zry8zr@virginia.edu

Wyatt Priddy
School of Data Science
The University of Virginia
tds8tv@virginia.edu

I. Introduction

Malnutrition remains one of the most pressing and multifaceted global health challenges, particularly in low- and middle-income countries (LMICs), where it continues to impede healthy development in early childhood. Among LMICs, Bangladesh has undergone significant economic growth and urbanization in recent decades, yet child undernutrition, particularly in urban slum settings remains alarmingly prevalent.

Mirpur, a densely populated and socioeconomically diverse urban community in Dhaka, Bangladesh, presents a unique opportunity to study the impacts of rapid urbanization and development on child health. Over the past 20 years, a series of prospective birth cohort studies conducted in this area have tracked maternal and early childhood health metrics, providing a rich longitudinal dataset.

This project aims to examine how child health, specifically growth-related outcomes like stunting, wasting, and being underweight has changed in Mirpur over time. We will identify the factors that best predict improvements in these outcomes. In doing so, we hope to uncover key opportunities for intervention and provide actionable insights that can guide public health strategies in similar urbanizing contexts.

II. Literary Review

Childhood malnutrition and growth faltering are persistent public health concerns in low- and middle-income countries (LMICs), especially during the first two years of life. A large pooled analysis of longitudinal studies from 33 LMICs, published in *Nature*, revealed that linear growth faltering (measured as length-for-age z-score, or LAZ) often begins within the first six months of life and worsens until about 24 months of age [7]. The study also found that once a child falls behind in growth, catch-up growth is rare in later childhood. This highlights the critical need for early interventions during pregnancy and the first 1,000 days of life.

In the context of Bangladesh, stunting remains a significant challenge despite improvements in economic development and healthcare access. According to a multilevel analysis by Sultana et al. [8], child stunting is closely associated with maternal education, household wealth, sanitation, and family size. The study emphasizes that not just nutritional intake, but also social and environmental conditions, play a crucial role in determining a child's growth trajectory. Children in poorer households, those with uneducated mothers, and those lacking access to clean water or hygienic sanitation were found to be at significantly higher risk of being stunted.

To track and compare child growth over time and across regions, researchers use z-scores standardized by the World Health Organization (WHO). These scores such as LAZ (length-for-age), WAZ (weight-for-age), and WHZ (weight-for-height) help identify children who are underdeveloped for their age and height. The WHO provides a detailed methodology for calculating these scores based on a child's sex, age, height, and weight [1]. These standardized measures are essential for monitoring global trends and evaluating the effectiveness of interventions.

Another recent study in *Nature* took a deeper look at when and how linear growth faltering occurs. It showed that not only does faltering start early, but it also tends to repeat, meaning some children experience multiple periods of slowed growth in the first two years. The findings underscore the importance of regular monitoring and integrated strategies that address nutrition, infections, and caregiving practices together [7].

Overall, the literature supports the idea that early childhood is a critical window for growth and development. Intervening during this period can prevent long-term consequences such as poor cognitive development, lower academic performance, and reduced economic productivity in adulthood. Our project builds on these insights by applying them to a specific urban context, Mirpur, Dhaka, where about 15 years of cohort data offer a unique opportunity to understand how urbanization, maternal health, and environmental factors influence child growth over time.

III. Project Proposal

A. Overview

The purpose of this project is to identify specific longitudinal trends associated with key child health outcomes in the communities of Mirpur and Dhaka Bangladesh. Using several field study data sets assembled by the sponsor from 2008 to 2022, our team is being asked to clean, wrangle, structure and analyze this data to develop visualizations and algorithmic inferences between independent and dependent features. These developed data outputs will be further reviewed and assessed to determine if strong correlations and potentially causation can be concluded between a range of factors (e.g., child ailments including diarrheal episodes, maternal health, socio-economic circumstances, etc.) and child health metrics during early life (e.g., birth weight, intrauterine growth restriction, growth rate (height and weight) across moments in child age, etc.)

B. Motivation

Our work will potentially help improve child health in Dhaka, Bangladesh and children in similar urban slum conditions by providing cleaned data, visualizations, and model results that can be used in practical research ways in the future. As seen in Figure 1, children in Bangladesh are below the curve when compared against World Health Organization standards for weight-for-height. Doctors and health care workers may use our findings to identify children at risk and improve nutrition programs. Policymakers and NGOs can use our analysis to decide where to grant resources, such as food and healthcare, to have the greatest impact. Researchers will also benefit from our organized data, which would allow them to build on our work without having to start from scratch. By focusing on these real-world applications, we ensure that our work addresses the right problems and leads to better child health policies.

This project is important in several ways that include practical resource support in working through the provided independent data sets to aggregate, clean and wrangle data into usable data bases/tables for algorithmic and visual analysis. This work will save time for the current sponsor and may be utilized by future research teams. Through statistical analysis, we hope to identify and determine instances of strong correlation and causation between investigated factors and growth outcomes that could expand our understanding of key factors that may cause child growth restrictions in similar types of urban slum environments.

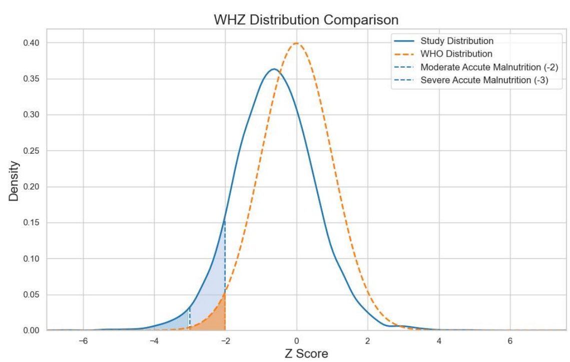

Fig. 1: WHZ of Study Participants Against WHO Standards

C. Data Collection

Data was provided to our team via the Capstone sponsor through secure file transfer that represented four distinct studies performed from 2008 through 2022. Study data was saved across a large array of individual text files structured in a tabular format.

IV. Methods

A. Data and Processing

As cited above, Data was provided through multiple sets of individual text files that required thorough assessment, cleaning, and transformation before being aggregated into a cohesive dataset suitable for visualization and modeling. Derived tables used for visualization and modeling needed to be structured with dependent response variables and independent features organized by SID and Study as the primary keys. Of note, we observed 1,779 total child IDs in the data set at the time of birth [Provide 700, NIH DBC: 629, PRKC: 200, and NIH Crypto 2000: 250]. As the data is longitudinal, the number of children remaining at each interval date was observed to decline over time.

1) Data Assessment and Cleaning: Due to the large number of individual data files, our team performed the following steps to assess and clean the data. This enabled us to more easily consolidate the information into usable tables for our analysis. These steps included:"

- Obtaining, aggregating and reviewing associated data dictionaries for the tables provided to identify where features across studies were located, as well as identifying where inconsistencies were present that required additional work.
- Macro Table Analysis: Developed a Python utility to assess each study's dataset by determining the number of records, feature columns, and null values. The results were visualized for our team's analysis.
- Standardization of column names, data types, and formats across tables.
- Throughout the range of text files, null values were represented in several different formats. By review of the data dictionaries and leveraging prepared EDA visualizations, our team identified where missing values appeared and how they were encoded.
  Nulls were represented by different combinations of the digits 8 and 9, which highlighted the importance of building a consistent approach to detect and transform them where needed. From this, we were able to compile a comprehensive list of the null indicators used by the field teams, which included a variety of integers, floats, and strings: [e.g., 88888, -9, -9.99, -9.999, -999, -999.99, 9, 99, 99.9, 99.99, 999, 999.9, 999.99, 9.9, '-9', '9', '-9.99', '-9.999', '-999', '-999.99', '99', '99.9', '99.99', '999', '999.9', '999.99', '9.9', '99:99', '1999-09-09', '9/9/99']

2) Developed Data Tables: Our team created several tables to support the project, consolidating information from the four studies, merging specific independent features from various source tables, and storing new variables generated through feature engineering.

The full dataset consisted of 243 individual tables sourced from the PROVIDE, PRKC, NIH Crypto 2000/3000, and NIH DBC studies. As shown in Figure 1, we used common keys such as Study and Participant SID to accurately

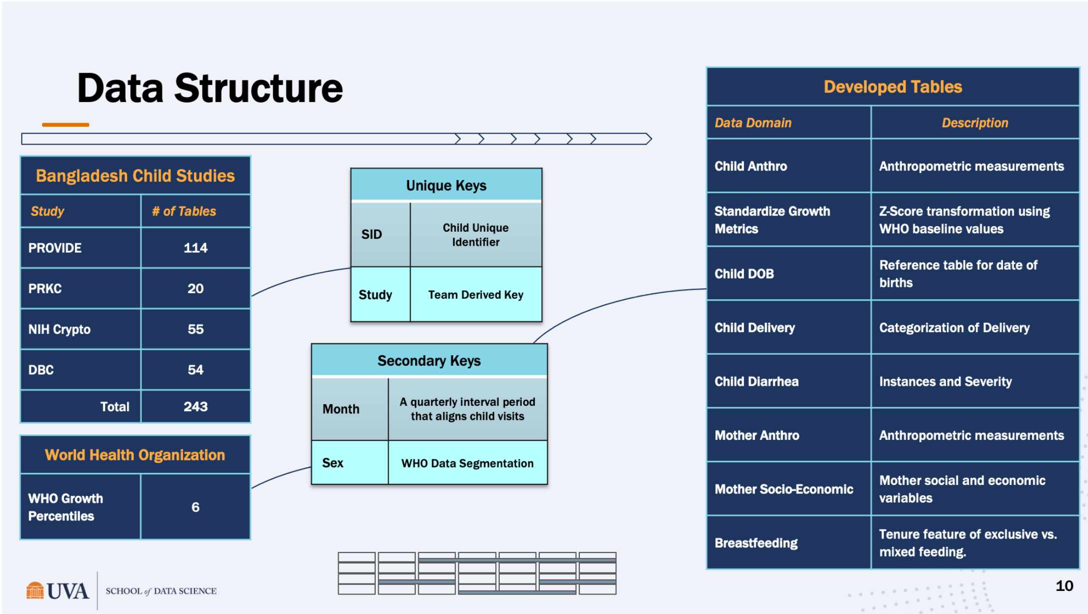

Fig. 2: Data Structure

merge information across studies into a unified structure. A critical early step involved reviewing raw data from these multiple sources to identify consistent features, align variable naming conventions, and standardize formats. Using Python, we aggregated, cleaned, and transformed the data into well-organized tables with a consistent schema. We also created a 'Month' variable to represent the quarterly intervals at which child data was recorded, a feature not explicitly available in the PROVIDE dataset and thus derived during our feature engineering process. Special attention was given to handling missing values, standardizing data types, and consolidating related variables. Ultimately, we constructed eight key consolidated tables that served as the foundation for our exploratory data analysis and growth modeling work.

3) Exploratory Data Analysis (EDA): We explored the data through various EDA mechanisms during the project to gain perspective on the distribution of numeric features, frequency of categorical levels, statistical metrics on numerical data, and correlation between features, etc. Some of the tools used include:

- Visualization via Histograms, Pairplots, and Bivariate plots (see Figure 3)
- Summary tables reviewing statistics and grouped frequencies
- Outlier review through Box-plot, and Z-score analysis, etc.

4) Feature Engineering: Feature engineering became a predominant portion of the project enabling the transformation of both independent and dependent variables into a more usable and consistent format for modeling and correlation analysis. Our feature engineering efforts focused on the following areas:

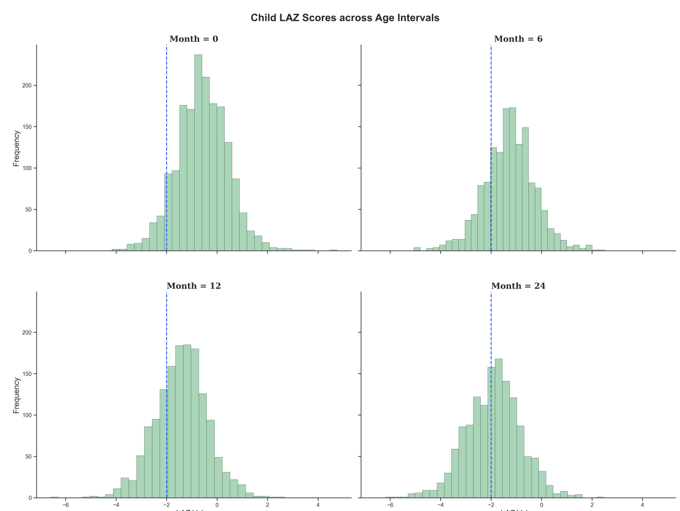

Fig. 3: EDA Example: Histogram of Child LAZ across by Child Age (Z-Score of -2 equals Child Stunting) [3]

- **Feature Creation**: Developing new features associated with both dependent and independent variables:
  - Calculation of new numerical features including: BMI (see further commentary on this below), incidents of diarrhea, growth change over time.
  - Development of informative categories (binning) representing a mother's Stature, Underweight, and BMI growth status at birth.
- **Time Features**: Standardization of numerical
  - Classification of research visits into quarterly intervals for longitudinal review.
  - Aggregation of instance over time (% of stunted or

wasting children by period)
  - Segmentation of study timelines into Time Period Groups for A/B regression testing.

- **Transformation**: Standardization of numerical variables:
  - Standardization of key growth metrics including LAZ, WHZ, and WAZ using WHO reference data for Z-Score calculation (see further commentary on this below).
  - Standardization of select numerical variables, such as income, was performed to account for the longitudinal nature of the data spanning 15 years. Values were standardized within each year to enhance the comparability of the feature across time.
- **Reclassification / Level Reduction**: To improve model performance (e.g., as a means to reduce overfitting) and reduce complexity, we consolidated categorical variables with a high number of unique levels. This process involved combining similar and/or infrequent category levels into broader, more meaningful groups. Reducing the number of levels improves model interpretability and ensures more stable training.
  - Modification of level definitions / labels for certain categorical features across studies for consistency.
  - Level reduction of multi-level categories to derive more informative features (e.g., 27 to 3 descriptive levels).

Growth Level Z-Score Standardization: Because our primary research goal was to identify trending and correlation of child associated factors to growth, we applied standardization to the child growth outcomes found in the study data (e.g. height/length and weight for age, as well as weight for height) at the different age points found in the data. We understand that this is the normal practice in this type of research, citing from a primary literary reference "a common practice in growth references is to model the raw data that is normally distributed, or to transform the raw data so that the transformed values are approximately normal." [Chapter 4, pg. 55 to 57] [2]. We calculated Z values for LAZ, WHZ and WAZ for all child IDs present in the data.

We used the LMS method to derive z-scores at each time period. The LMS method is a mathematical model that allows summarization of anthropometric data to obtain normalized percentile curves. LMS method summarizes the distribution of the variable of interest according to age, based on three parameters or curves: $L(\lambda)$, $M(\mu)$, and $S(\sigma)$. These three parameters indicate the power in the Box-Cox transformation for the skewness adjustment (L), the median (M), and the generalized coefficient of variation (S) for each annual measurement. Assuming the residuals follow a normal distribution and by using the three estimated parameters (L(x), M(x) and S(x)) for each age (x), axial length values can be converted into percentiles and Z-scores." [5]

$$Z_{ind} = \frac{\left(\frac{y}{M(t)}\right)^{L(t)} - 1}{S(t)L(t)} \tag{1}$$

[LMS formula reviewed in referenced book (Chapter 4) and WHO paper] [2], [3]

Body Mass Index (BMI): BMI is another growth index that is generated from one variable (weight) divided by another variable (height) and for this review was used to roughly assess the overall health and nutrition of the related mothers. We understand that the index itself is a ratio index where "The division serves to standardize the numerator for the denominator. In the studies of growth, ratio indices are usually used to represent body shape or proportionality" [Chapter 4, pg. 51] [2].

$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m)}^2} \tag{2}$$

B. Assumptions (Data and Study)

Through collaboration with our research sponsor and ongoing data analysis, we gained an understanding of the typical relationships between the key dependent and independent variables from the studies and operated under several key assumptions as follows:

- **Health Measures**: From our research, the following health categories are generally set at certain quantitative levels:
  - Adult women are considered Short Statured at a height below $< 145\,\text{cm}$ [6]
  - Adult women are considered Underweight at a BMI level below < 18.5 [4]
- **Interval Selection for PROVIDE**: Because child visits to the research team were not tagged with a quarterly 'annum' indicator, we needed to make an assumption on how to determine which individual dates would best reflect the quarterly time period (e.g., would 360 or 370 days from birth be considered 1 year of age). When selecting visit dates to represent quarter-end intervals within our Python scripts, we assumed that any visit falling within approximately ±10 days of the target interval would be eligible for inclusion. In cases where multiple dates qualified for a given interval, we prioritized selecting the visit date that was further from the date of birth to better align with the intended time point.
- **Crypto 3000**: Due to the location for this study being in a more rural location and distinctly separate from the other studies, we chose to remove Crypto 3000 from our data when fitting our regression models.
- **Child Participation over Study Timeline**: One assumption we observed within the data is that as the studies began, not all children who participated entered each study at the same time. This then would cause the span of the studies to vary in time. Secondly, participation from

children was not always consistent and as the age of each child progressed from date of birth to age 24 months, we noticed a decline in SID count across each of the studies.
- **Child SID / Record Loss**: Various factors caused the sample size to change due to the nature of the question being investigated in the models. The main one being children participating in the study leaving at varied periods due to factors such as consent being withdrawn, physical moving of the participant, or death of an individual. Metrics such as those derived as change in z-score from 0 - 12 months could not be calculated if the participant left before their one year check in. Another factor which caused participants to be excluded from the analysis was the inability to derive z-scores due to a participants seemingly abnormal metrics. For example, deriving WHZ necessitated a corresponding height within the WHO reference dataset. For females this threshold was 45cm - therefore participants below that threshold were unable to have a z-score derived.

C. Modeling

For assessing each of the proposed hypotheses, we used various regression models due to their simplicity, interpretability, and effectiveness in quantifying relationships between variables.

1) Ordinary Least Squares Linear Regression: Ordinary Least Squares (OLS) are used to determine the direction, magnitude, and significance of feature variables on a quantitative response variable. In multivariate linear regression, numerous features are implemented and are held constant against the response to isolate each feature's effect. To derive the importance of the coefficients, the residuals are "estimated in a way that minimizes the residual sum of squares" [Chapter 5] [2]. These coefficients can be interpreted as the magnitude the response changes based on a one-unit change in the predictor. OLS is heavily utilized due to its interpretability.

2) Generalized Linear Models (GLM): Generalized Linear Regressions - mainly logistic regression - were utilized for response variables with a binomial distribution. In logistic regression, we are estimating for coefficients that show a one unit different in the log odds of the probability the response variable will be positive. The model is represented below:

$$\ln\left(\frac{p}{1-p}\right) = a + b_1 x_1 + \ldots + b_p x_p$$

3) Quantile Linear Regression: Quantile Linear Regression (QLR), different from OLS and GLM, minimizes the sum of the weighted absolute residual values rather than minimizing the residual sum of squares. "Quantile regression is a nonparametric procedure to estimate a specific quantile in relation to covariates" [Chapter 5] [2] and allows the model to be less sensitive to outlier influence. Specifically, if the quantile under review is the 90th percentile, the model will adjust to ensure that 90% of the predicted values and therefore residuals fall below the prediction line. The loss function will weight residuals differently depending on whether the prediction is above or below the actual value. The model is:

Quantile selected = τ

$\hat{\beta}_\tau = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta)$

The check function or weight $\rho_\tau(u)$ assigns different weights to residuals. $\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases}$

V. Results

Our team performed exploratory data analysis, created visualizations, and developed regression models to evaluate several hypotheses and scenarios outlined by our research sponsor. Through our regression results, we aimed to investigate the following:

- The strength of feature influence on target growth outcomes, while controlling for other relevant factors
- Calculation of statistical significance of independent features
- Assess confidence intervals for notable features, and
- Quantify any potential variations in feature influence across different quantile distributions of target responses

A. Scenario I - Longitudinal Trending

Through exploratory data analysis (EDA) and visualization, we evaluated whether thematic longitudinal trends were present across a range of target responses over the study timeline from 2008 to 2022. These included LAZ and WHZ, as well as indicators of undernutrition and growth impairment, such as moderate acute malnutrition (MAM), among others. Table I below lists several scenarios we explored, along with identified outcomes.

Over the course of our EDA work, we noted several thematic observations including:

- Some outliers were observed beyond the 2.5th and 97.5th percentiles.
- There did appear to be some level of upward trajectory in longitudinal trending over the timeline of the studies.

	Scenario	Outcome
1	WZH Distribution (Box-Plots) at Birth (by Quarter) at Birth, 12 and 24 Months	Slight / progressive upward trajectory in z-score over time for children at birth. This becomes more pronounced at 12 Months. By 24 Months, the trajectory becomes flat.
2	WHZ Rate of Growth Change Between Interval Periods	There was no discernible change in rate over time for WHZ for each interval period (e.g., 0-3, 0-6, 0-12, 0-24).
3	WHZ Rate of Growth Change by Sex	No significant difference between male and female, as it relates to growth change across the different interval periods over time.
4	LAZ Distribution (Box-Plots) at Birth (by Quarter) at Birth, 12 and 24 Months	Trajectory in z-score is relatively flat at birth; however gradually moves upward at 12 and 24 Months.
5	LAZ Rate of Growth Change Between Interval Periods	Rate of growth for children between birth and 3 months of age had a progressive upward trajectory over the study timeline. Other interval periods (e.g., 0-3, 0-6, 0-12, 0-24) did not present a clear change over time.
6	LAZ Rate of Growth Change by Sex	No discernible difference by sex over time.
7	Percentage Children with Growth Impairments at Birth	Instances of growth conditions have decreased since the initial study despite a slight uptick in the past years. Variance is likely due to sample sizing.

TABLE I: Longitudinal Scenarios with Observed Outcomes

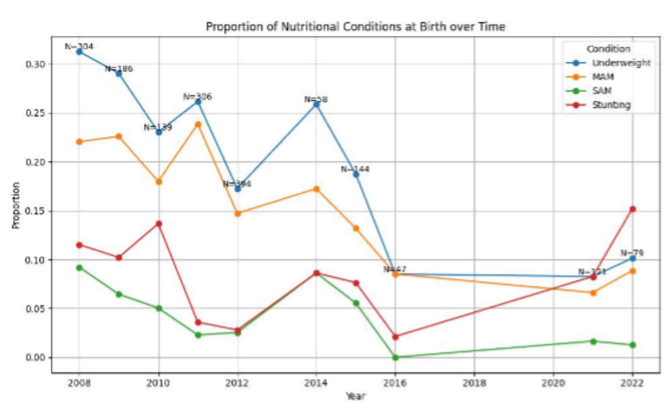

Fig. 4: EDA Example: Longitudinal Line Plot of % Percentage Growth Impairments Over Time

B. Scenario I - Growth over Time

The unique nature of our multi-study longitudinal analysis allowed for analysis to be cultivated on children's anthropometric features over time. As seen in Figure 5, while the LAZ score in later years tends to be above the median, the trend is not definitive.

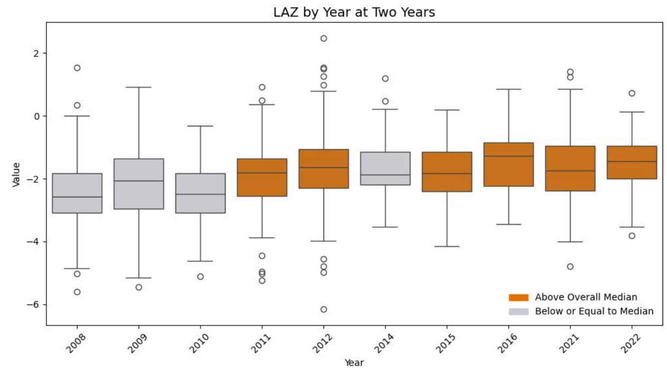

Fig. 5: LAZ Distribution at Age 2

To assess whether time had a statistically significant impact on z-scores, a linear regression was used to derive the following results:

Z-score	Coefficient	p-value
WHZ	0.0179	0.0148
LAZ	0.0484	0
WAZ	0.0377	0

TABLE II: Regression results on z-score changes over time

Each z-score has increased at a statistically significant level with LAZ seeing the largest increase with approximately a 0.05 increase in z-score per year. Over the span of 15 years, that accounts for an increased median z-score of 0.75 which equates to about an added 2.29cm of height. While time shows a significant impact on growth, the model had a low R^2 with year only accounting for 4% of the variability in z-score. Overall, it is shown that z-scores are definitively trending upwards across the years.

C. Scenario II - Assessing the Influence of Maternal Growth on Child Stunting

We investigated how maternal growth status categorized as short statured or underweight affected the odds of a child being considered stunted (LAZ $<$ -2). We approached this analysis through the use of GLM with a logistic regression framework. We also evaluated whether maternal categorical status remained predictive after accounting for numeric growth indicators, assessing the added value of this categorical information to the model. Our baseline model and results showed the following:

Initial GLM Model with Mother Categorical Features

$$\log\left(\frac{p(\text{Child Stunted})}{1 - p(\text{Child Stunted})}\right) = \beta_0 + \beta_1 \cdot \text{Maternal Stature} + \beta_2 \cdot \text{Maternal Weight Category}$$

Variable	Log Odds	Odds	Marginal Prob.	P-Value	Odds Ratio (CI)
Short Stature	0.876	2.401	0.706	1.96e-07	1.73 to 3.34
Underweight	0.513	1.671	0.626	1.62e-03	1.21 to 2.30

TABLE III: GLM Outcomes when Regressing with Child Stunting vs Mother Growth Status

After incorporating and controlling the model with mother's numeric growth indicators, the influence of the categorical growth status noticeably diminished as noted in Table IV. Here we can see the numerical features for a mother's height and weight are the significant factors.

Variable	Log Odds	Odds	Marginal Prob.	P-Value	Odds Ratio (CI)
mhtcm	-0.067	0.935	0.483	3.4e-05	0.91 to 0.97
mwtkg	-0.041	0.959	0.490	.05e-05	0.94 to 0.98
Short Stature	-0.021	Ins.	Ins.	9.24e-01	Ins.
Underweight	0.016	Ins.	Ins.	9.34e-01	Ins.

TABLE IV: GLM Outcomes from Regressing Child Stunting vs Mother Growth Metrics and Mother Growth Status. Insignificant results noted as Ins.

1) Multicolinearity: We were concerned about the potential for multicolinearity given the relationship between a mother's growth metrics and their growth categories (e.g., Stature and Weight per BMI), therefore we assessed several combinations of features through the GLM and OLS models with Variance Inflation Factor (VIF). We found that the only scenario where multicolinearity developed was when our model included both the mother's BMI numeric value and their height and weight numerical metrics. We also included the following socio-economic features within this analysis and found their VIF values to all be below 1.1, indicating a lack of multicolinearity; Education, Occupation, Food Availability, Sex of Child, and Home variables including: Floor, Roof, Electricity, and TV. Results did not differ when running VIF with OLS.

Variable	Adj. VIF	Adj. VIF w/ BMI Value
mhtcm	1.388	6.115
mwtkg	1.322	15.127
bmi_val	NA	14.008
mom_short_cat	1.296	1.298
mom_underweight	1.212	1.210

TABLE V: VIF Results with GLM covering with and without Mother BMI.

D. Scenario III - Linear Relationship with Child LAZ

To further examine how maternal health and socio-economic factors influenced child growth, we fitted an OLS regression model using the child's standardized height (LAZ) as the outcome variable.

1) Main Effects: Overall, we quickly observed through EDA that a linear relationship between a mother's height and weight towards the response LAZ was present.

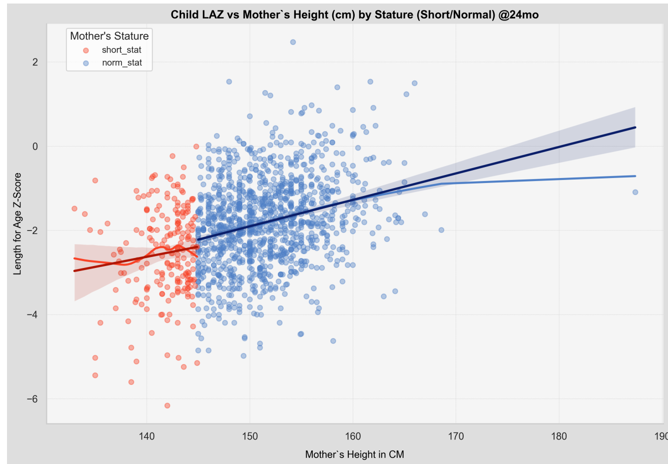

Fig. 6: Regression Plot: Child LAZ vs Mothers Height w/ Interaction from Mother's Stature)

Below are results from our multi-variate OLS model where we fitted the following model: Child LAZ $\sim$ Mother's Health Metrics $+$ Mother's Health Categories $+$ Socio-Economic Factors.

Variable	Est. Coeff.	Error	P-Value
mhtcm	0.043	0.007	4.84e-11
mwtkg	0.014	0.004	9.26e-05
Stature Normal to Short	Ins.	Ins.	Ins.
Underweight Normal to Under.	-0.198	0.085	0.019
TV: No to Yes	0.235	0.062	1.50e-04
Educ. High to Min.	-0.261	0.068	1.20e-04
Educ. High to None	-0.396	0.077	3.24e-07
Occup. Housewife to Prof.	Ins.	Ins.	Ins.
Occup. Housewife to Unskilled	Ins.	Ins.	Ins.
Floor: Cement to Earth/Wood	-0.247	0.099	1.26e-02
Electricity: No to Yes	Ins.	Ins.	Ins.
Latrine to Toilet	Ins.	Ins.	Ins.
Roof Finished to Natural	Ins.	Ins.	Ins.
Roof Finished to Tin	-0.277	0.066	2.42e-05
Food Avail. Deficit to Avg.	Ins.	Ins.	Ins.
Food Avail. Deficit to Surplus	0.247	0.083	3.16e-03
Child Sex	Ins.	Ins.	Ins.

TABLE VI: Results for the multi-variate model between LAZ response and Mother related factors. Features where the associated P-Value was insignificant are noted as Ins.

From the resulting model, we continue to observe that a mother's height and weight are positively correlated to LAZ. Several controlling features that were added related to a mother's health and socio-economic circumstances did not remove the significnace of this relationship; however, we did note that there was some additional effect on the mean response of LAZ when taking these factors into account. Specifically, the higher impact factors were Education, when moving from High to No education status (decrease by .396) and housing related factors when moving from assumed higher standards to lower (approximate decrease by .25).

2) Interaction Effects: We also investigated the potential for interaction effects between the socio-economic factors and key maternal health characteristics. In order to manage model complexity, we ran a focused model where each variable only interacted with two primary features; a mother's numerical height and weight. We used the following model to explore this scenario: Child LAZ $\sim$ mhtcm + mwtkg + Maternal Health and Socio-Economic Features + mhtcm:Maternal Health and Socio-Economic Features + mwtkg:Maternal Health and Socio-Economic Features:

Variable / Reference	mhtcm	mwtkg	ANOVA
Stature Normal to Short	Ins.	Ins.	Ins
Underweight Normal to Under.	0.041	0.071	0.047
TV: No to Yes	-0.032	Ins.	0.033
Education	Ins	Ins	Ins
Occupation	Ins.	Ins.	Ins.
Floor Cement to Earth	0.053	Ins	0.029
Electricity	Ins.	Ins.	Ins.
Toilet	Ins.	Ins.	Ins.
Roof Finished to Tin	0.091	Ins.	Ins.
Food Availabilty	Ins.	Ins.	Ins.
Child Sex	Ins.	Ins.	Ins.

TABLE VII: Interaction Effects towards Mother Height / Weight by Variable. Insignificant results noed as Ins.

Based on the above results, we found that certain factors did have an influence on the slope of the regression trajectory for mother's height and weight. Among these, Weight Status of the mother, TV, and Floor (e.g., of the home) showed significant ANOVA p-values when selecting the full model with interaction terms. This indicated for weight status and tv a strengthening relationship between mother's height and weight to child LAZ, while TV noted a moderating effect. Separately, while Roof was significant on its own and shows a localized effect, its inclusion as an interaction term did not improve the model.

E. Scenario IV - Impact of Growth Faltering on Health Growth

One assumption made was that early-age growth faltering, or a decline in weight-for-age z-score of one or greater over a period of three months, has a negative impact on a child's growth from birth to age two. To examine this assumption, we utilized a mixed-effects linear model, incorporating year and gender as grouping variables to account for longitudinal trends and gender difference. Growth faltering was assessed at changes from birth to three months and from three months to six months.

Variable	Est. Coeff.	Std. Error	p-value
Intercept	-0.170	0.093	0.068
Growth Faltering 0 - 3 Month	-0.509	0.086	**0.000
Growth Faltering 3 - 6 Month	-0.405	0.154	**0.009
Access to TV	0.177	0.054	**0.001
Mother Educated	0.080	0.057	0.164
Mother Highly Educated	0.258	0.068	**0.000
Food Deficit	-0.172	0.069	**0.012
Food Surplus	0.051	0.056	0.355
Mother's Height	0.103	0.023	**0.000
Mother's BMI	0.030	0.024	0.202
Group Var	0.081		

TABLE VIII: Mixed Linear Model Regression Results

Accounting for maternal influence and proxies for socio-economic status, it was determined that growth faltering at these early periods is attributed to a significant decline in the LAZ score delta from birth to two years. The 0 - 3 month period had the largest effect on LAZ change at two years with an associated decline of half a standard deviation. An instance of growth faltering from the 3 - 6 month period showed a decline in LAZ score of 0.4. Factors positively associated with growth were proxies for socio-economic status such as access to television and a highly educated mother. A reported deficit of food was associated with a significant negative decline in LAZ of 0.172.

F. Scenario V - Maternal and Socio-Economic Impacts Across the Outcome Distribution Using Quantile Regression

Our analysis explored whether the impact of maternal and socio-economic factors differed across the distribution of LAZ outcomes among children. We used quantile regression to perform this, so that we could isolate the effects and any differences in coefficient values depending on whether the child was at the 5th percentile (i.e, the child is lower on the LAZ distribution) or the Median which is a close approximation to OLS.

Fig. 7: Quantile Regression between LAZ and Mother's Weight (kg), displaying multiple percentile regressions for example.

We performed a quantile regression with the full model used in Scenario III between LAZ and the mother related factors and noted the following results.

Variable	5th Percentile Coeff.	Median Coeff.
Underweight Status: Normal to Underweight	-.52	-.13
TV: No to Yes	0.33	0.21
Educ. High to Min.	-0.33	-0.18
Educ. High to None	-0.51	-0.37
Occup. Housewife to Professional	0.30	0.10
Latrine to Toilet	0.16	0.03

TABLE IX: Factors with Statistically Significant Coefficient Changes ($\Delta > 0.05$) between 5th and 50th Percentiles.

Based on the comparison between the 5th percentile and the median, several factors showed noticeable and statistically significant shifts in their influence on LAZ. The largest factor movements of note include a mother's weight status at birth, where if the mother is considered "Underweight", it impacted the lower distribution of children more severely by a reduction of 0.39. Another notable shift was seen with maternal occupation. As occupational status increased to the professional level, reflecting higher assumed economic standing, there was an associated uplift of 0.20 in LAZ at the lower end of the distribution.

G. Scenario VI - Diarrhea and Child Growth Outcomes

Research Question:

Do more diarrhea episodes in the first year of life lead to more stunting or underweight outcomes later in childhood?

Summary Answer

While exploratory plots suggest that children with more diarrhea episodes in their first year tend to have lower LAZ (Length-for-Age Z-score) scores, statistical analysis reveals that the association is weak and not statistically significant. This suggests that diarrhea burden alone may not be a sufficient predictor of growth outcomes like stunting or underweight status.

Visual Findings

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.3728	0.032	-43.331	0.000	-1.435	-1.311
instances_of_diarrhea	-0.0109	0.009	-1.247	0.212	-0.028	0.006

Fig. 8: OLS regression coefficients show a very weak and statistically insignificant relationship between diarrhea episodes in the first year and LAZ at 12 month.

- **Box Plot:** As shown in Figure 9, the distribution of diarrhea episodes is slightly higher among stunted children compared to non-stunted ones. This suggests a possible link between higher illness burden and stunting, but the differences are subtle and not dramatic. The overlap in the interquartile ranges indicates that other factors may also contribute to stunting.

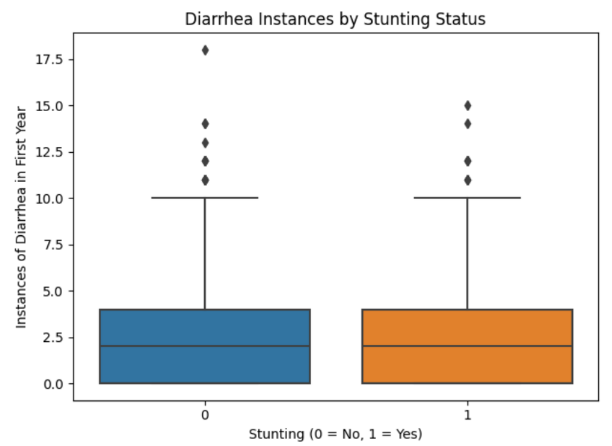

Fig. 9: There is no visible difference in diarrhea burden between stunted and non-stunted children.

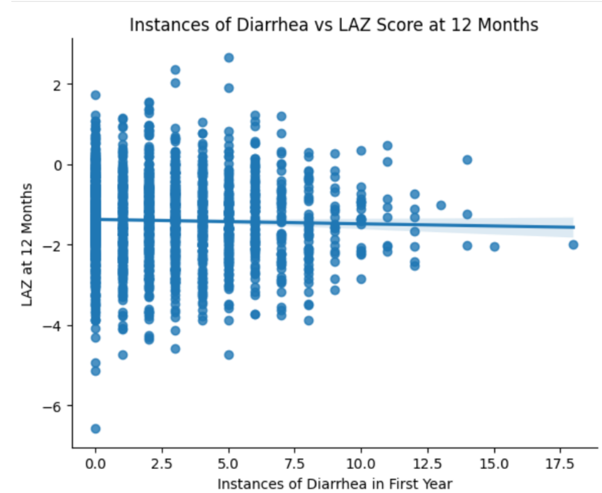

Fig. 10: Scatter plot is showing a slight negative trend between diarrhea episodes and LAZ scores; however, the relationship is weak.

- **Scatter Plot:** As shown in Figure 10, the scatter plot of total diarrhea episodes against LAZ scores shows a downward sloping regression line, implying a weak negative relationship between illness burden and child growth. However, the data shows high variability, with many children having the same number of episodes but very different LAZ scores.

OLS Regression Analysis

An Ordinary Least Squares (OLS) regression was performed to quantify the relationship between diarrhea episodes and LAZ scores.

- **Model:** LAZ $\sim$ Instances of Diarrhea

- **Coefficient for Diarrhea Episodes:** -0.0109
- **P-value:** 0.212 (Not statistically significant at $\alpha = 0.05$)
- **R-squared:** 0.001 (Only 0.1% of the variance in LAZ is explained by diarrhea)

Interpretation: The coefficient indicates that each additional diarrhea episode is associated with a 0.0109 decrease in LAZ score. However, the p-value (0.212) shows that this relationship is not statistically significant. Moreover, the extremely low R-squared value means that the model explains almost none of the variability in LAZ, reinforcing the idea that diarrhea alone is not a strong determinant of stunting.

Although visualizations provide some suggestive patterns of a negative association between diarrhea burden and child growth, the statistical analysis does not support a meaningful or significant relationship. Diarrhea may be one of many contributing factors, but other variables such as maternal health, nutrition, socioeconomic status, and sanitation likely play more substantial roles in determining long-term growth outcomes. Future research should explore these factors in combination to better understand the determinants of stunting and under-nutrition in early childhood.

VI. Conclusion

Child health in the Mirpur region of Dhaka in Bangladesh has improved markedly over the past 15 years. Weight, height, and weight-for-height ratios have improved as the region has gone through rapid development. Although general conditions have allowed for more substantial growth, specific factors also contribute to the trajectory of child development. Conditions such as growth faltering in early periods of life have detrimental effects to child development in the later years. Having a highly educated mother and access to luxuries such as television and improved home conditions, which may serve as proxies for socioeconomic status, have positive effects on child development. Surprisingly, this study finds that diarrheal incidences had no bearing on a child's growth outcomes.

A. Implications

This research shows that to enhance the development of children in LMICs, Non-governmental organizations and governments should focus on uplifting their population's education, the common standard of living, and intervene early when children show signs of faltering in their growth.

B. Limitations

Despite the promising results, some limitations remain in this study. Sample sizing was one limitation with only 1,779 children being included across the four studies. This number excludes those omitted in models due to leaving the study or not having qualifying metrics as previously asserted. A noteable limitation was the frequency of data collection - with data only being consistently collected across studies on participants at three month intervals. Access to more granular data would allow researchers to more closely monitor the growth of children in these studies. Additioanlly, consistency across study variables among the four data sets was observed that impaired the use of certain features (e.g., income as an example due to varying scope) data when their definitions or categorical levels did not align.

C. Future Development

Scenario VI analysis focused on the association between diarrhea episodes in the first year of life and growth outcomes (LAZ and stunting) at 12 months. However, because exposure (diarrhea) and outcome (growth) are measured concurrently, it's difficult to draw conclusions about the long term effects of early diarrhea burden.

In future work, this relationship should be re-evaluated by measuring growth at later time points such as 24 months to better capture delayed impacts on child development. Additionally, separating the analysis by severity of diarrhea may reveal whether more serious episodes have stronger associations with poor growth. Further exploration of confounding variables like maternal BMI, socioeconomic status, and access to healthcare could also help clarify the pathways between illness and stunting.

In addition, building on the above analysis of early child growth in relation to child health metrics and key maternal and socio-economic factors (e.g., LAZ), future research teams may further explore how additional biomedical, physiological, and clinical characteristics relate to child growth outcomes.

D. GitHub

Our team used the following repository to house our code base and outputs: https://github.com/Shakeri-Lab/Child-Heath-in-Bangladesh-Capstone

References

[1] WHO Child growth standards; https://www.who.int/tools/child-growthstandards/standards

[2] Statistical Analysis of Human Growth and Development, Cheung YB. Milton, UNITED KINGDOM: CRC Press LLC; 2013 [cited 2025 Feb 7]; http://ebookcentral.proquest.com/lib/uva/detail.action?docID=1377020

[3] WHO COMPUTATION OF CENTILES AND Z-SCORES FOR HEIGHT-FOR-AGE, WEIGHT-FOR-AGE AND BMI-FOR-AGE; https://cdn.who.int/media/docs/default-source/child-growth/growth-reference-5-19-years/computation.pdf

[4] Malnutrition in women; https://www.who.int/data/nutrition/nlis/info/malnutrition-in-women

[5] LMS parameters, percentile, and Z-score growth curves for axial length in Chinese schoolchildren in Wuhan; https://www.nature.com/articles/s41598-022-08907-5

[6] Short stature as an independent risk factor for cephalopelvic disproportion in a country of relatively small-sized mothers; https://pubmed.ncbi.nlm.nih.gov/22187064/

[7] Early-childhood linear growth faltering in low- and middle-income countries; https://www.nature.com/articles/s41586-023-06418-5

[8] Correlates of stunting among under-five children in Bangladesh: a multilevel approach; https://bmcnutr.biomedcentral.com/articles/10.1186/s40795-019-0304-9