UVA DATA SCIENCE

Identifying Key Opportunities for Intervention in Child Health and Development in Dhaka Bangladesh

Harold Haugen, Aqsa Majeed, Wyatt Priddy

5/1/2025

Introduction

- Malnutrition disproportionally effects children in Bangladesh
- 4 studies spanning ~15 years were conducted in Mirpur, a crowded area in the capital Dhaka.
- Exploratory Analysis of longitudinal trends in child growth methods.
- Prescriptive Analysis with interpretable models to determine factors of change longitudinally.

Key Hypothesis and Scenarios

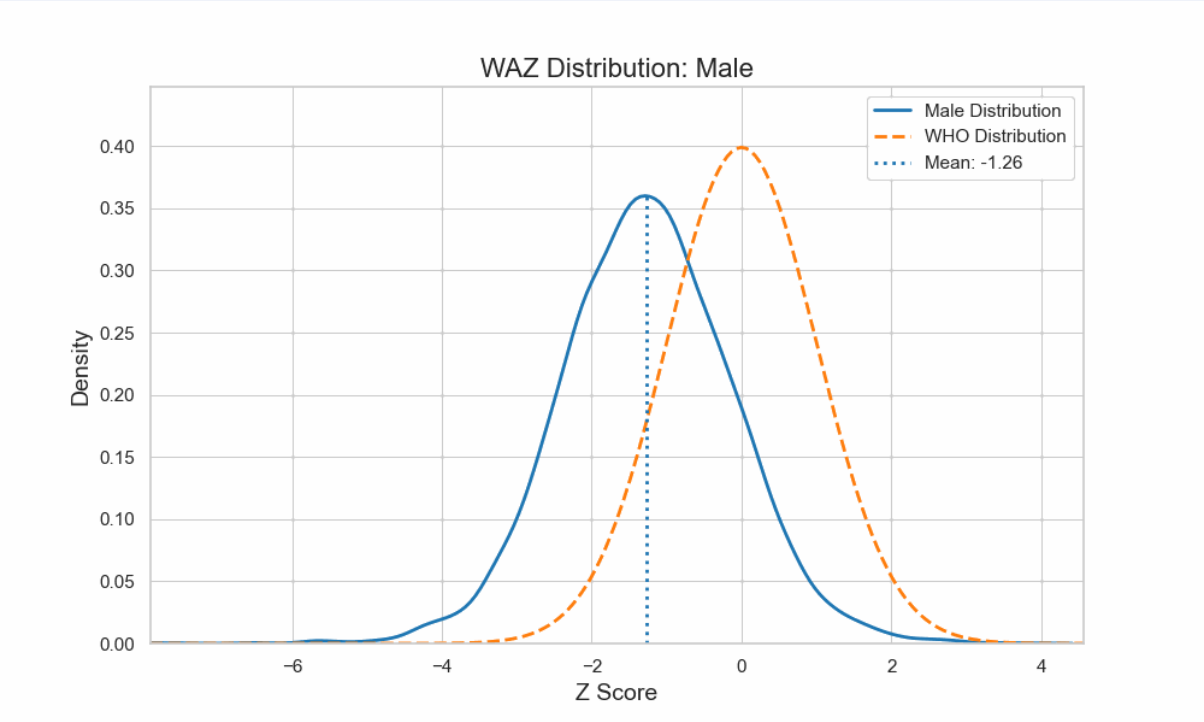
Key Hypothesis:

We believe that child health in Mirpur, Bangladesh has improved over the past 20 years, and that factors like maternal health, family standard of living, and instances of diarrhea are linked to growth outcomes in children.

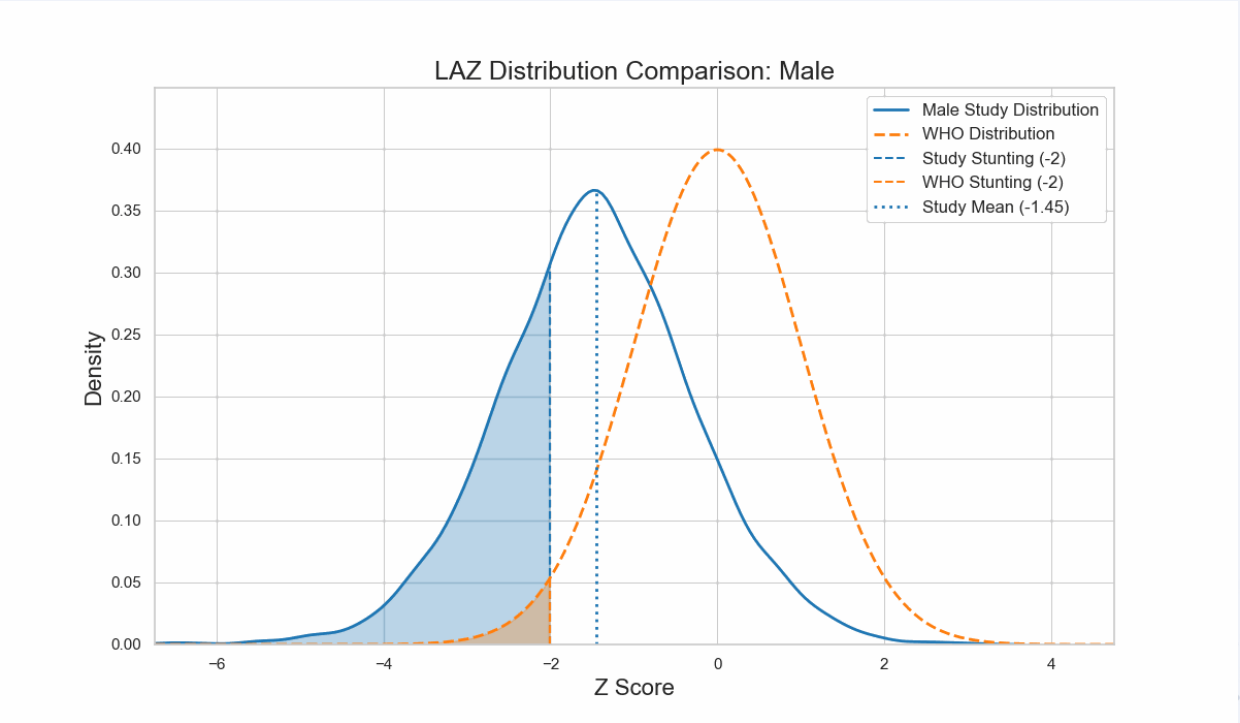
Scenarios:

- **Health Over Time:**
  Has the average child height and weight improved from 2007 to 2024?
- **Predictors of Healthy Growth:**
  Does growth faltering in early periods result in a decline in LAZ at two years?
- **Maternal Influence:**
  Do mothers who are healthier or have more education tend to have children who grow better?
- **Diarrhea and Growth:**
  Do more diarrhea episodes in the first year of life lead to more stunting or underweight outcomes later?
- **Socioeconomic Trends:**
  As income and urban development increased in Mirpur, did child health improve along with them?

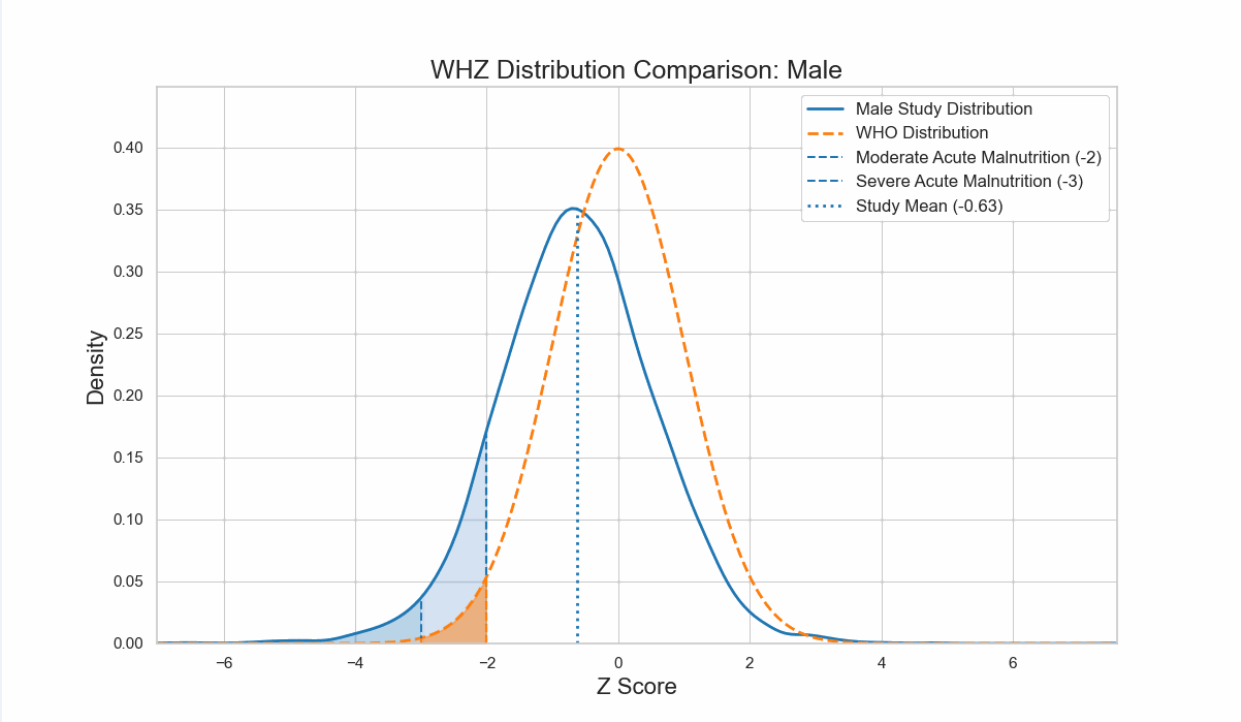
Introduction of Metrics: WAZ



Introduction of Metrics: LAZ



Introduction of Metrics: WHZ

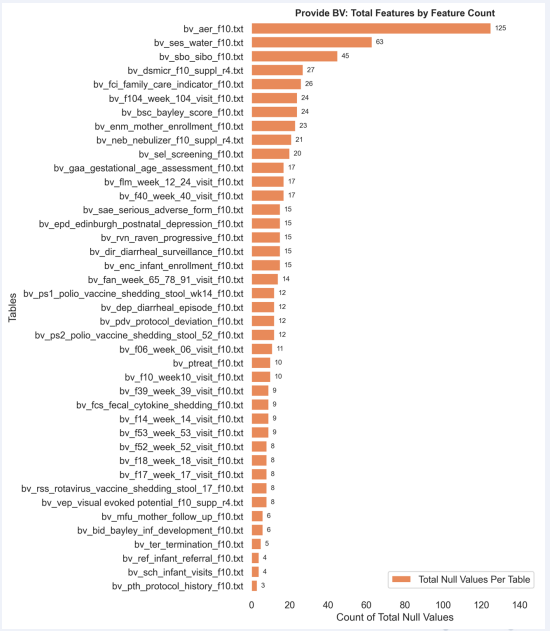


Data Structure & Feature Engineering

A significant allocation of time was invested to assess, understand, design and establish the future state data / table structure.

Data Overview

Example of features per Table from the PROVIDE Study 'BV' Data : [Range 3 to 125]

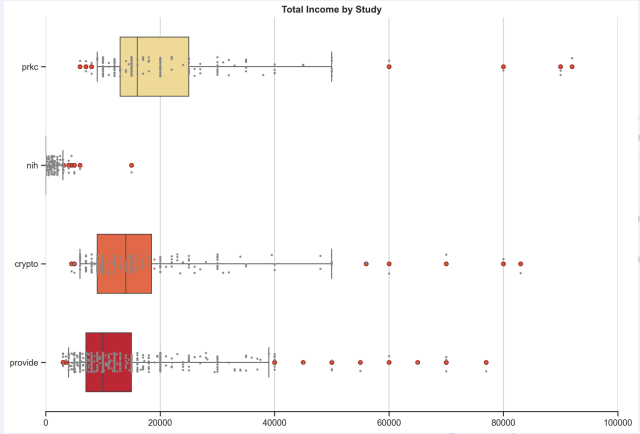


Integration and Preparation of Data for Statistical Analysis

Consolidating the data from four separate research periods into singular structured tables for statistical analysis was a highly time-intensive process.

Tasks involved:

- IRB Training by the Team to gain access to data on vulnerable subjects.
- Research and assessment of data dictionaries / tables to locate specific informational features across studies.
- EDA on specific scenarios.
- Altering table identifiers and category definitions for consistency.
- Cleaning of data for null values and updating data types.
- Feature engineering to create numeric and categorical features that align with study hypotheses and scenarios.
- Creating consolidated tables using unique keys across studies.



EDA of Income by Study: Noted Inconsistent Definitions and Null Values

Data Structure

Bangladesh Child Studies	
Study	*# of Tables*
PROVIDE	114
PRKC	20
NIH Crypto	55
DBC	54
Total	**243**

World Health Organization	
WHO Growth Percentiles	6

Unique Keys	
SID	Child Unique Identifier
Study	Team Derived Key

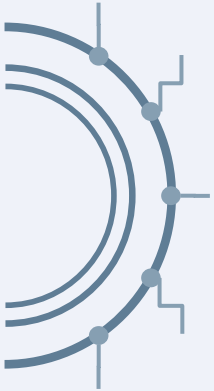
Secondary Keys	
Month	A quarterly interval period that aligns child visits
Sex	WHO Data Segmentation

Developed Tables	
Data Domain	*Description*
Child Anthro	Anthropometric measurements
Standardize Growth Metrics	Z-Score transformation using WHO baseline values
Child DOB	Reference table for date of births
Child Delivery	Categorization of Delivery
Child Diarrhea	Instances and Severity
Mother Anthro	Anthropometric measurements
Mother Socio-Economic	Mother social and economic variables
Breastfeeding	Tenure feature of exclusive vs. mixed feeding

Feature Engineering

Feature Creation

- *Calculation of values*
- *Binning of numerical features*



Transformation

- *Standardization to Z-Score*

Missing Values

- *Identification, assessment and removal*

Time Features

- *Segmentation into specific time periods*

Reclassification / Level Reduction

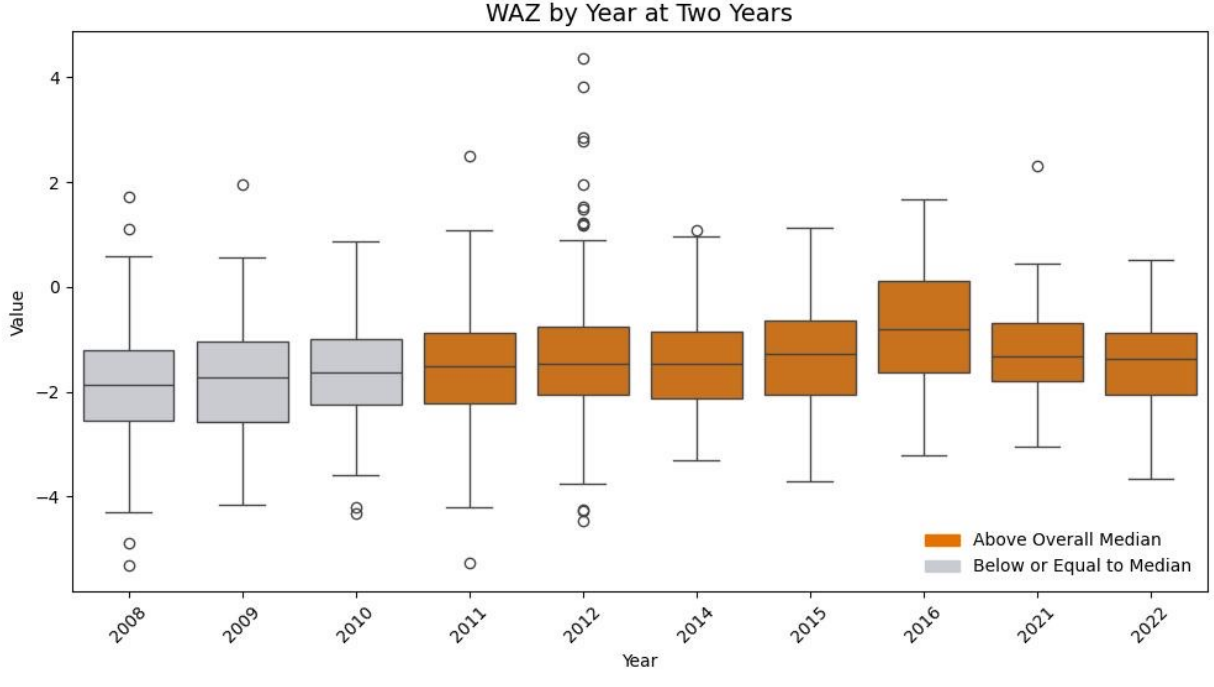
- *Categorical level consistency*
- *Level reduction (e.g., 27 > to 3 levels) for more informative categorical features*

Developed Tables	
Data Domain	*Impacts*
Child Anthro	Standardize Time intervals Across Studies.
Standardized Growth Metrics	Calculation of LAZ, WAZ, WHZ Z-Scores w/ WHO reference data, child growth categories .
Child DOB	Aggregate Child DOB Across Studies.
Child Delivery	Aggregate Delivery Modes of Children Across Studies.
Child Diarrhea	Creation of metrics around days and instances of diarrhea by severity.
Mother Anthro	BMI, height, weight, BMI category, mother stunted and underweight status, standardized anthro.
Mother Socio-Economic	Missing values, adjustments to category levels, level reduction, numeric standardization.
Breastfeeding	Creation of days exclusive and mixed breastfeeding across standard time intervals.

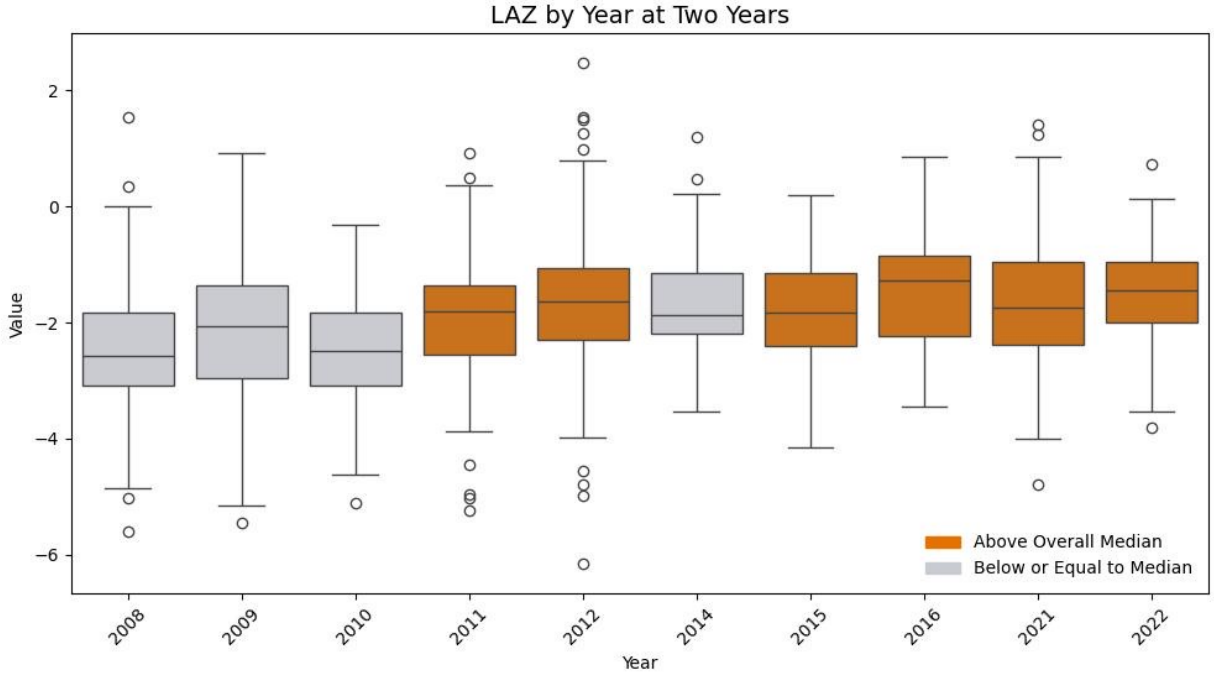
Models & Scenario Results

- Linear Regression to determine magnitude and significance of components
- Mixed Effect Regression to account for longitudinal effect
- Logistic Regression for binary outcomes
- Quantile Regression to account for outliers, and assess influence across different points (quantiles) of the outcome distribution

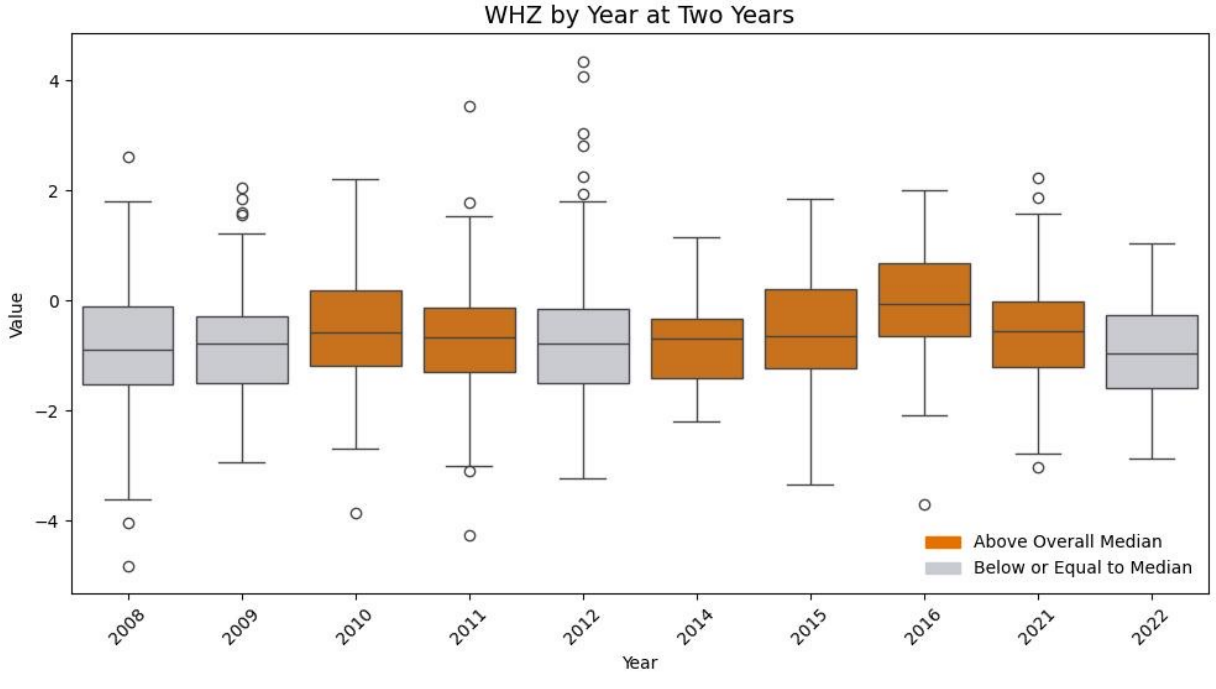
Health Over Time: Age Two Snapshot



Health Over Time: Age Two Snapshot

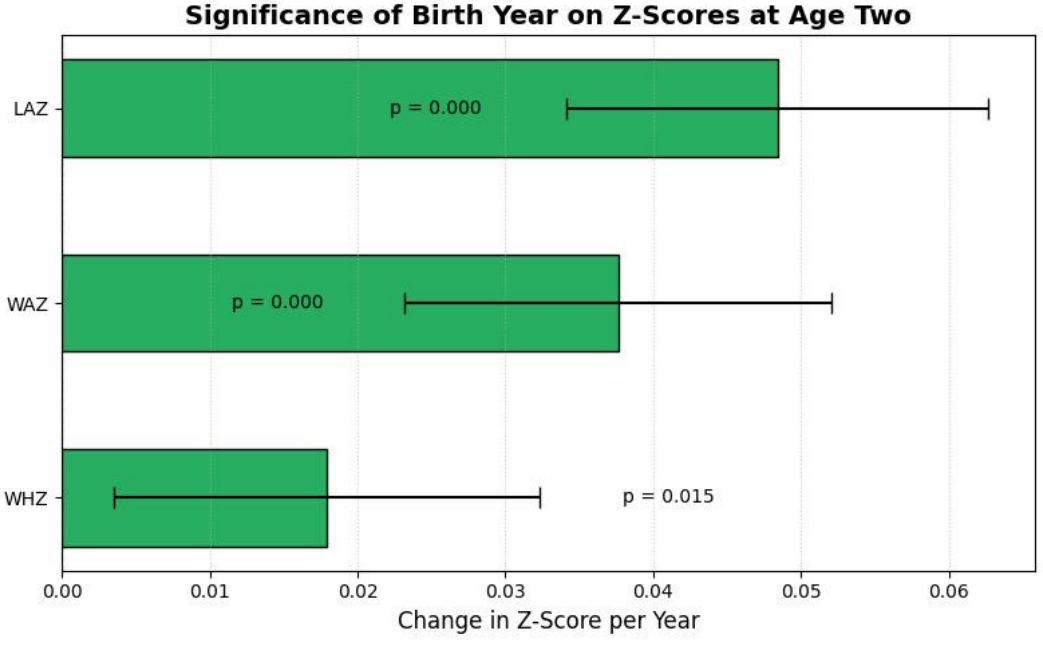


Health Over Time: Age Two Snapshot



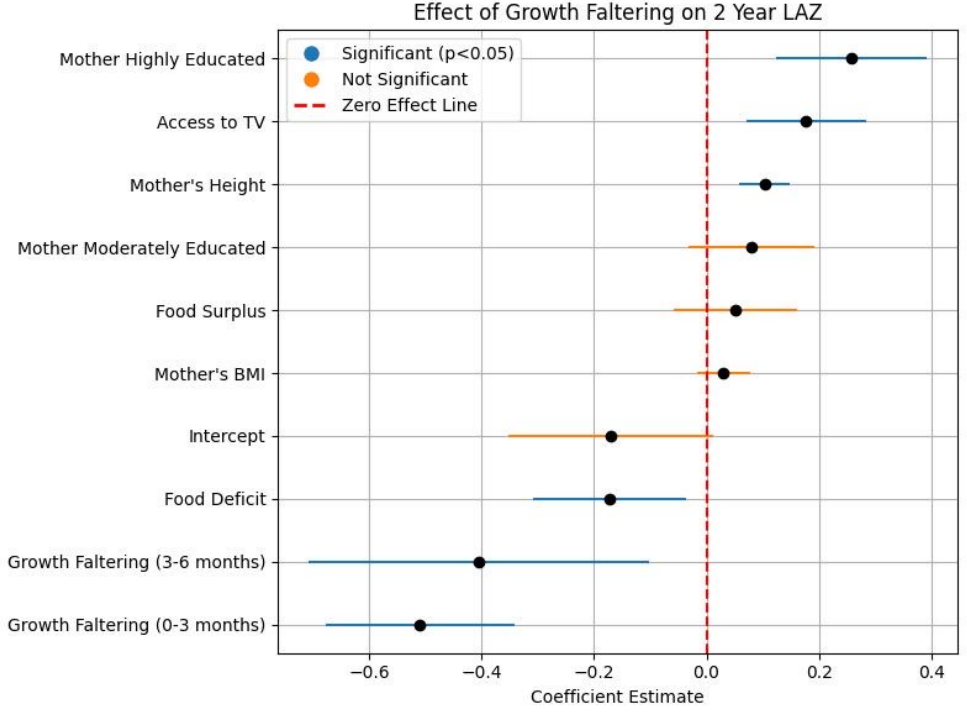
Health Over Time: Age Two Snapshot

- Every year, Z-scores increased at a statistically significant level
- A 1 unit increase in year resulted in an increase in LAZ of ~0.05 at age two
- Low R2 (~4%) indicates that while explanatory, longitudinal trending is not most explanatory factor



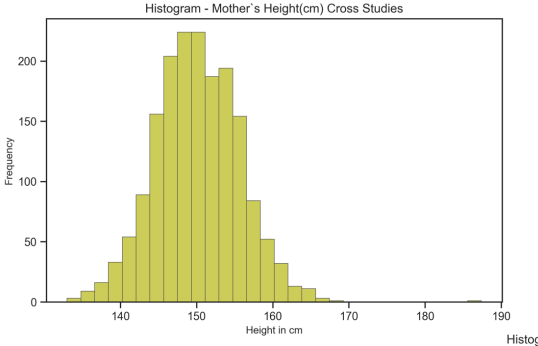
Predictors of Healthy Growth

- Growth faltering at the 0-3 month and 3–6-month period negatively impact growth of a child
- Educated mothers and access to TV positively impact growth
- Inadequate access to food hinders growth

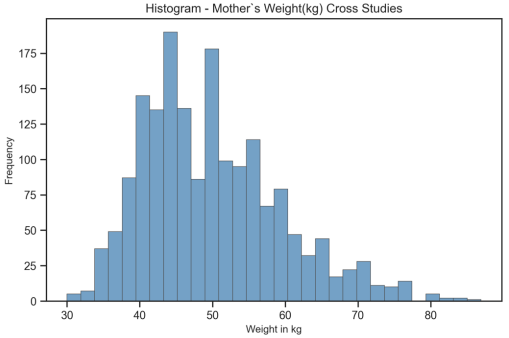


Impact of Mother's Health on Child Stunting

- Initial GLM indicated a strong relationship between Stunted Children (reference) vs Mother's Short Stature / Underweight build at birth.
- The addition of numerical features (Height/ Weight) reduced the influence of Stature/Underweight BMI.
- 1cm Increase ~ 6.5% Decrease in the Odds of Stunting.
- 1kg Increase ~ 4.1% Decrease in the Odds of Stunting



Assumptions:
a. Children at 24mo
b. Child is Stunted at LAZ < -2
c. Mothers < 145cm are Short Statured
d. Mothers < 18.5 BMI are Underweight



Probability of Child Being Stunted at Age 24mo

Stunted ~ Short Stature + Underweight

Variable	Log-Odds	Odds	Marg. Prob.	P_Value	Odds Ratio
Short Stature	0.876	2.401	0.706	1.962e-07	1.73 to 3.34
Underweight	0.513	1.671	0.626	1.620e-03	1.21 to 2.3

Stunted ~ mhtcm + mwtkg + Short Stature + Underweight

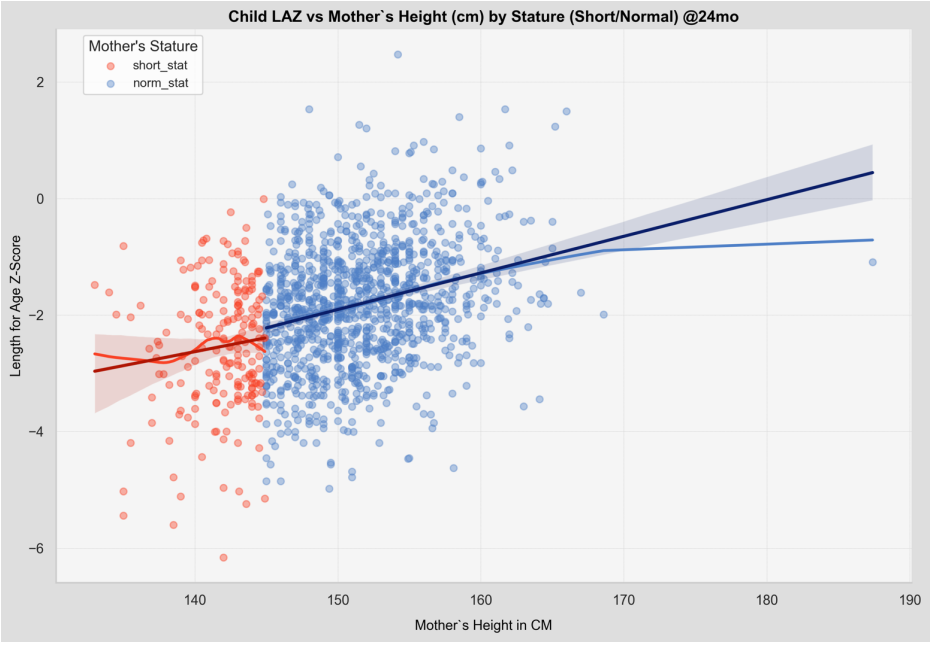
Variable	Log-Odds	Odds	Marg. Prob.	P_Value	Odds Ratio
mhtcm	-0.0674	0.935	0.483	3.4e-05	0.91 to 0.97
mwtkg	-0.0412	0.959	0.49	0.05e-05	0.94 to 0.98
Short Stature	-0.0210			0.9238	
Underweight	0.0164			0.9336	

Addition of Mother Numeric Anthro

Maternal Influence

Exploration of Linearity

- Overall, there is a strong linear trend among a mother's height and weight to a child's LAZ @ 24mo
- No significant impact to 'mean LAZ response' given a mother's growth category



LAZ vs Mother's Height (cm) by Stature

Linear Effects to LAZ

LAZ ~ mhtcm + Short Stature

Variable	Est.	Error	P_Value
mhtcm	0.0621	0.0066	<2e-16
Short Stature	-0.1295	0.1006	0.198

LAZ ~ mwtkg + Underweight

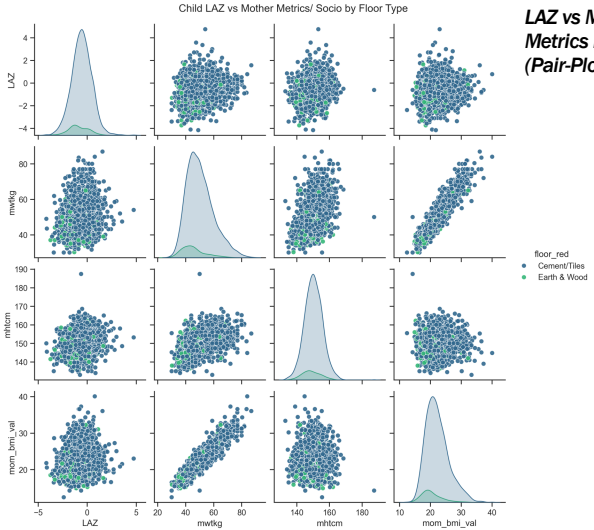
Variable	Est.	Error	P_Value
mwtkg	0.0399	0.0033	<2e-16
Underweight	-0.0115	0.0880	0.896

LAZ ~ bmi + Underweight

Variable	Est.	Error	P_Value
mom_bmi_val	0.0617	0.0088	4.4e-12
Underweight	-0.1831	0.09414	0.052

Controlling for Socio-Economic Factors vs LAZ

- Addition of Socio-Economic factors related to the Mother as a control variable.
- Although Socio-Economic factors are at play, Height and Weight are still quite significant.
- Leveraged automated parameter selection Step, Adj R^2, BIC, Mallows CP to derive meaningful features.



LAZ vs Mother Anthro Metrics by Floor Type (Pair-Plot)

Notes:
a. Grey = Insignificant
b. Step Function used the 'Both' parameter
c. Top two factors highlighted in both directions.

Variable	Reference	Est.	Error	P_Value	Step()	Adj. R^2
mhtcm	N/A	0.0434	0.0065	4.84e-11	Y	Y
mwtkg	N/A	0.0142	0.0036	9.26e-05	Y	Y
Short Stature	Normal - Short					Y
Underweight	Normal - Underweight	-0.1988	0.0849	0.0193	Y	Y
TV	No - Yes	0.2345	0.0618	1.50e-04	Y	Y
Education	Higher Edu – Min. to Moderate Edu	-0.2606	0.0676	1.20e-04	Y	Y
Education	Higher Edu – No Formal Edu	-0.3955	0.0770	3.24e-07	Y	Y
Occupation	Housewife_No_Occ – Professional					Y
Occupation	Housewife_No_Occ - Unskilled	-0.1138	0.0655		Y	Y
Floor	Cement/Tiles – Earth/ Wood	-0.2472	0.0989	1.26e-02	Y	Y
Electricity	No - Yes					
Toilet	Latrine - Toilet					
Roof	Finished – Natural Mat.					
Roof	Finished – Tin	-0.2773	.0655	2.42e-05	Y	
Food Avail.	Deficit – Neither Def./Surpl.					
Food Avail.	Deficit - Surplus	0.2466	0.0834	3.16e-03	Y	
Sex	Femail - Male					

Interaction Effects – Mother's Height / Weight

- Is the influence of Mother's Height and Weight modified by other Maternal Factors ?
- Model: LAZ ~ mhtcm + mwtkg + category etc. + mhtcm:category + mwtkg:category

LAZ vs Mother's Height (cm) by Underweight

Interaction Variable	Reference	mhtcm	mwtkg	ANOVA
Short Stature				0.7522
Underweight	Normal - Underweight	0.0408	0.0706	0.0471
TV	No - Yes	-0.0319		0.0327
Education				0.3015
Occupation				0.8785
Floor	Cement/Tiles – Earth/ Wood	0.0529		0.02998
Electricity				0.0909
Toilet				0.3814
Roof	Finished – Tin	0.09136		0.0914
Food Avail.				0.5412
Sex				0.4016

Notes:

a. A positive value/ effect to the targeted variable is a steepening of the slope (+ x1 to the slope); i.e., LAZ decreases at a steeper rate.

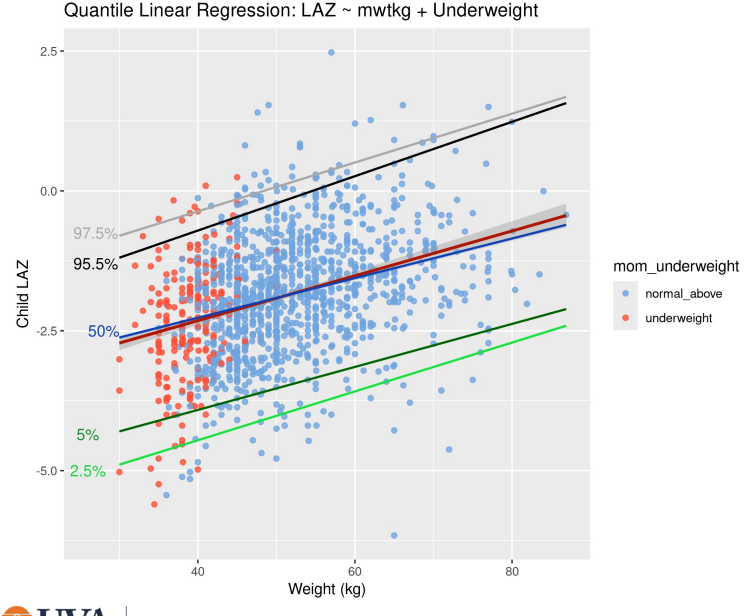
b. ANOVA is between the reduced and model w/ interactions.
   - Roof was significant, while the interaction model was not.

c. Grey = Insignificant

Quantile Regression

Maternal Influence

- Do predictors influence different subgroups of the distribution (e.g., other than the mean).



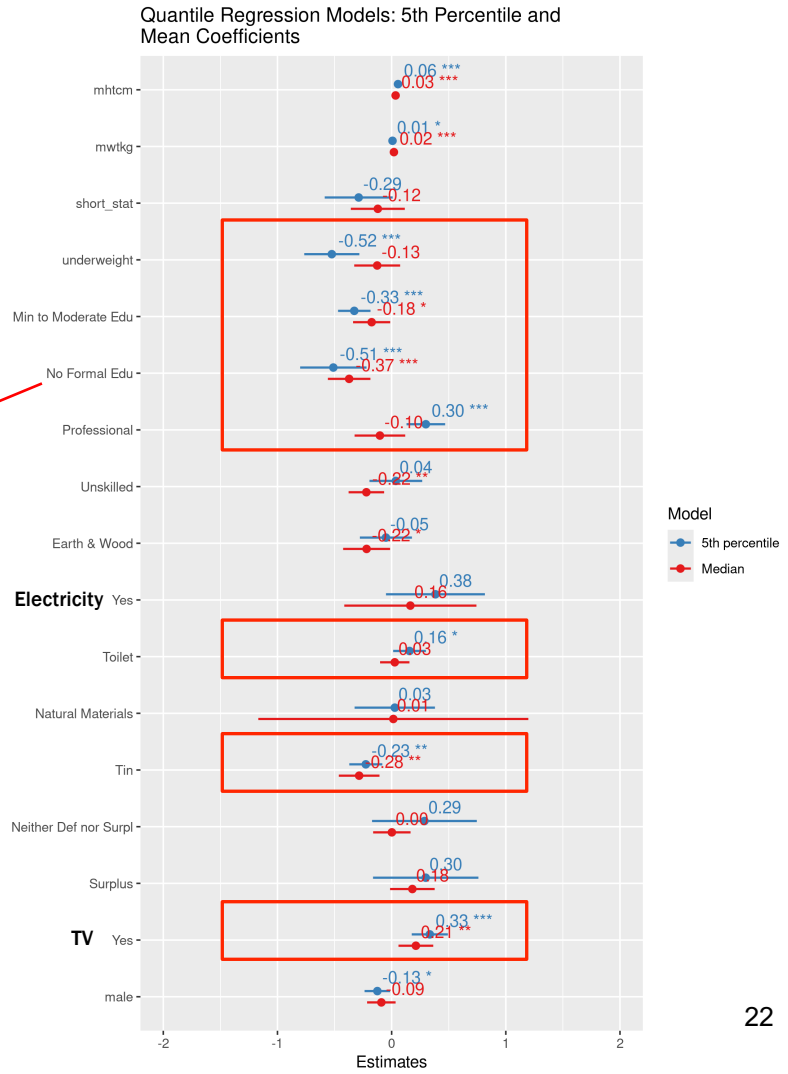
- In real-world cases, effects are not constant across the outcome distribution.
  - Are vulnerable groups impacted differently?
- QR is robust to outliers, which are present.
- Observation: As we move lower on the distribution, a mother's underweight category has more influence on the child response; however, we see the trajectory of mother's weight relatively static across the distribution.

Quantile Regression

Socio-Economic Factors

- Quantile Response movements across features:
  - Features w/ Significance & Movement
  - Example: Mother with No Formal Education vs Ref. Higher Education has a larger effect on the lower distribution of Child LAZ. From -.37 to -.51.

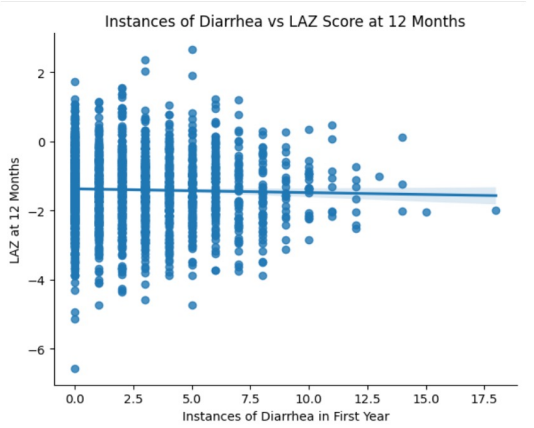
```
tau: [1] 0.05

Coefficients:
                                 Value     Std. Error t value   Pr(>|t|)
(Intercept)                    -12.59320   0.93058  -13.53264   0.00000
mhtcm                            0.05592   0.00614    9.11149   0.00000
mwtkg                            0.00771   0.00365    2.11140   0.03492
mom_short_catshort_stat         -0.28862   0.15180   -1.90134   0.05747
mom_underweightunderweight      -0.52401   0.12304   -4.25893   0.00002
edu_redMin to Moderate Edu      -0.32727   0.07246   -4.51641   0.00001
edu_redNo Formal Edu            -0.51061   0.14818   -3.44580   0.00059
occup_redProfessional            0.29960   0.08619    3.47612   0.00052
occup_redUnskilled               0.03679   0.11772    0.31252   0.75470
floor_redEarth & Wood           -0.05028   0.11627   -0.43248   0.66546
elecYes                          0.38371   0.22078    1.73801   0.08243
toiletToilet                     0.15684   0.07310    2.14562   0.03208
roof_redNatural Materials        0.02739   0.17936    0.15272   0.87864
roof_redTin                     -0.22622   0.07371   -3.06914   0.00219
food_avail_redNeither Def nor Surpl  0.28722  0.23386  1.22816  0.21960
food_avail_redSurplus            0.29855   0.23525    1.26905   0.20464
tvYes                            0.33427   0.08066    4.14435   0.00004
sexmale                         -0.12511   0.05711   -2.19071   0.02864
```



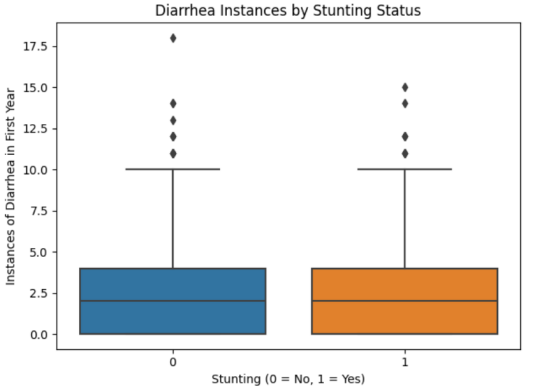
Diarrhea & Growth

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

- Children with more diarrhea instances in the first year tend to have slightly lower LAZ scores at 12 months, indicating worse growth outcomes.
- Although a negative trend between diarrhea burden and growth is suggested by the plots, the relationship is **weak and not statistically significant** based on the regression results (p = 0.212).
- **Regression:** The coefficient for instances of diarrhea is negative (-0.0109) but not statistically significant (p = 0.212), suggesting that while the visual trend shows a decline in LAZ with more diarrhea, the effect is small and could be due to chance.



Scatter Plot: More diarrhea instances are associated with lower LAZ scores, but the relationship is weak and highly variable.

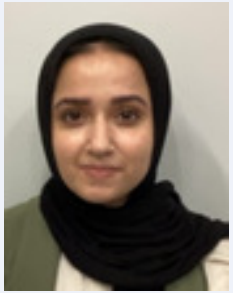


Box Plot: Stunted children show slightly more diarrhea episodes, suggesting a possible link between illness burden and stunting.

Capstone Team

Harold Haugen

Wyatt Priddy



Aqsa Majeed