Learning Lab 1a: Establishing nutrition surveillance system and analyzing longitudinal data

Centre for Non-Communicable Diseases and Nutrition (CNCDN)
BRAC James P Grant School of Public Health (BRAC JPGSPH)
BRAC University, Dhaka, Bangladesh

Facilitators: Md. Mokbul Hossain[1], Fahmida Akter[1], Mahbub Latif[1,2], Ali Ahsan[1], Malay Kanti Mridha[1]

[1]CNCDN, BRAC JPGSPH, BRAC University; [2]Institute of Statistical Research and Training, Dhaka University

Preamble

Facilitators

Md. Mokbul Hossain, Senior Analyst (Statistics)
Fahmida Akter, Assistant Scientist
Ali Ahsan, Deputy Research coordinator
Mahbub Latif, Scientist and Professor*
Malay Kanti Mridha, Professor and Director

All the facilitators are from the Center for Non-communicable and Nutrition
BRAC James P Grant School of Public Health BRAC University

*Also affiliated with Dhaka University

Session outline

- ❖ Importance of longitudinal data
- ❖ Introduction to the nationwide nutrition surveillance system in Bangladesh
- ❖ Longitudinal data analysis: a practical example

Learning objectives

1. Gain skills on how to establish a nutrition surveillance system

2. Get an orientation on analysis and interpretation of longitudinal nutrition data

Why do we need longitudinal studies in nutrition?

Why do we need longitudinal study in nutrition? (1)

1. Distinguishes between ‘within-person’ and ‘between-person’ change
2. Can predict within-person changes so that interventions can be targeted effectively
3. Can investigate the intergenerational cycle of malnutrition
4. Track changes in nutrition indicators over time
5. Better suited to examine how multiple factors (e.g., behavioral, environmental, socio-demographic factors) influence nutrition outcomes
6. Can capture fluctuations in nutritional indicators based on seasonality and cyclical trends
7. Can identify causal relationships with malnutrition and other health outcomes

Why do we need longitudinal study in nutrition? (2)

8. Can evaluate the outcomes and impact of nutrition-specific and nutrition-sensitive interventions (e.g., school meal program, nutrient supplementation)

9. Better suited to improve understanding on how early life exposures to malnutrition influence the long-term health and nutrition outcomes

10. Facilitates risk and protective factors identification

11. Improves statistical power because repeated measures of the same individuals reduce variability and improve the precision of estimates.

12. Reduces bias (e.g., recall bias)

13. Supports design of programs and development of policy

Some famous longitudinal studies

1. The Framingham Heart Study (1948, USA, n=5209)
2. The Whitehall Studies (1967, UK, n=10,308)
3. The Millennium Cohort Study (2000, UK, n=18,818)
4. The Malmo Diet and Cancer Study (1991, Sweden, n=28,449)
5. The Pelotas Birth Cohort Studies (1982, Brazil, n=4,275)
6. The Tohoku Medical Megabank Project (2011, Japan, n=150,000)
7. The European Prospective Investigation into Cancer and Nutrition (1992, Europe, n=519,978)
8. The PURE Study (Prospective Urban Rural Epidemiology) (2003, 25 countries, n=424,921)
9. The Growing Up in Singapore Towards Healthy Outcomes (GUSTO) Study (2009, Singapore, n=1,247)

Bangladesh Food Security and Nutrition Surveillance

Background

- First nutrition survey after independence at 1975-76 and second was in 1981-82
- IPHN conducted Food Security and Nutrition Surveillance (FSNS) from 1990 to 2007
- BRAC JPGSPH, HKI, and BBS conducted FSNS from 2008 to 2013
- The government of Bangladesh, development partners, and other stakeholders provided funding to continue the FSNS in 2015 and 2018
- The World Bank supported FSNS in 2023

Rationale

- ❖Bangladesh is passing through an epidemiological and demographic transition
- ❖Data are scarce about the nutritional status of men, adolescent boys, older adults
- ❖Planning and implementation of nutrition programs following life-cycle approach need better data for each life stage
- ❖Tracking of progress in necessary to understand the success and failures nutrition-sensitive and nutrition-specific programs
- ❖Local level planning and implementation of programs are needed to improve nutrition

Institutional partners

National Nutrition Services, Institute of Public Health Nutrition, Directorate General of Health Services

James P Grant School of Public Health, BRAC University

Bangladesh Bureau of Statistics

The World Bank

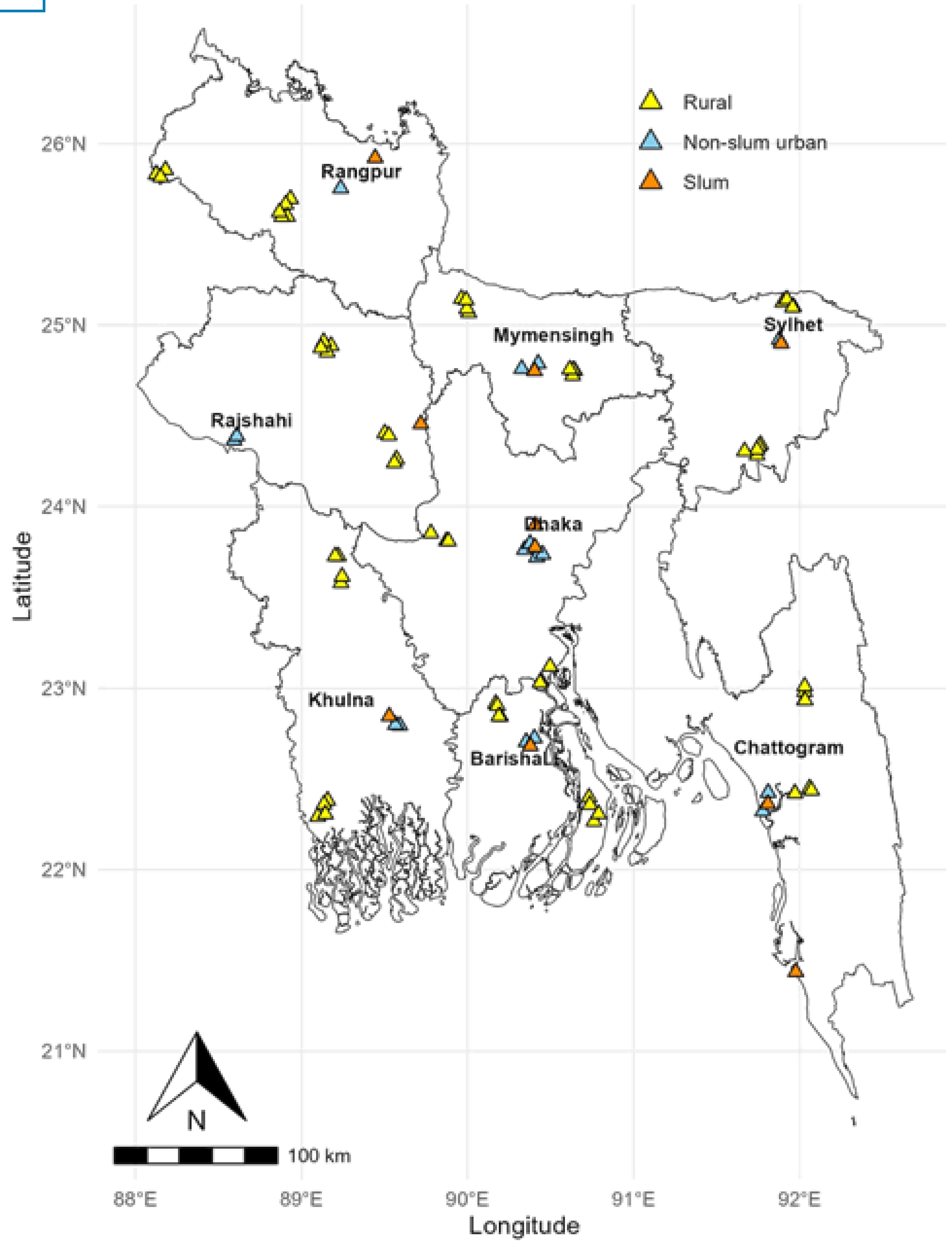
Objectives

- To assess household socio-economic status, food security and water, sanitation and hygiene status in rural, non-slum urban, and slum households of Bangladesh
- To assess dietary diversity, feeding practices and nutritional status of the under-five children and 6-9 years old children
- To assess dietary diversity, the burden of non-communicable diseases related risk factors and nutritional status of the adolescent boys and girls
- To explore dietary diversity, the burden of non-communicable diseases related risk factors, nutritional status of adult women and men
- To assess dietary diversity, the burden of non-communicable diseases and related risk factors, nutritional status, and quality of life of older adults

Design and Study Sites

Design: Longitudina, two waves so far: 2018-2019, 2023			
Sl#	**Rural**	**Urban**	**Slum**
1	Division: 08	Division: 08	Division: 08
2	District: 16	District: 08	District: 10
3	Upazila: 16		
4	Union: 32	Ward: 09	
5	Village/Mouza: 64	Mahalla: 16	Slum: 10

Study Sites Map



Sample Size

- ❖Outcome indicators ranges from 4% to 98%
- ❖Minimum required sample:

$$n = DEF \times \{z_{\alpha/2}^{2}(p)(1-p)\}/d^{2}$$

  - ➢DEF = the design effect
  - ➢p = apriori proportion of the relevant indicator
  - ➢$z_{\alpha/2}$= Standard normal quantile
  - ➢d = allowable margin of error
- ❖$\alpha = 0.05$
- ❖d = 0.05 for >10%; d/2 for 10% or less
- ❖DE = 1.61 considering ICC = 0.01
- ❖620 persons in each population group in each division, 3720 total for each of the 6 groups in each division
- ❖Another 3720 from 10 slums [Total sample 33,480]

Study population

Inclusion Criteria

- ❖ Living in the household for at least 3 months
- ❖ Sharing meal from the same cooking pot
- ❖ Age: 0-5, **6-9**, 10-19, 20-59, 60+
- ❖ Not more than 1 in each population group in one household

Exclusion Criteria

- ❖ Unable to give consent (e.g., major psychiatric illness)

Methods: Sampling
Rural areas: Five stage cluster sampling design

Stage 1	• Random selection of two districts from each eight divisions of Bangladesh (16 districts)
Stage 2	• One sub-district from each of the selected districts
Stage 3	• Two unions from each of the selected sub-districts
Stage 4	• Two segments/clusters from each of the unions
Stage 5	• For each age group, eighty (80) households with at least one member of that age group were randomly selected to interview 62 individuals in each age group from each cluster

Final rural sample was 11,790 households with 21,104 participants

Methods: Sampling
Non-slum urban: Three stage cluster sampling design

Stage 1
- 2 wards were selected from each of the divisional headquarters (Large city corporations)

Stage 2
- Two segments/clusters from each of the Wards

Stage 3
- For each age group, eighty (80) households with at least one member of that age group were randomly selected to interview 62 individuals in each age group from each cluster

Final non-slum urban sample was 3,368 households with 5,256 participants

Methods: Sampling
Slums: Three stage cluster sampling design

Stage 1

- 2 slums from Dhaka and Chittagong and 1 slum from other 6 divisional headquarters were randomly selected

Stage 2

- Two segments/clusters from each of the selected slums

Stage 3

- For each age group, eighty (80) households with at least one member of that age group were randomly selected to interview 62 individuals in each age group from each cluster

Final slum sample was 2,165 households with 3,645 enrolled participants

Study Tools

Respondents	Interviews	Measurement
Household head	Socio-economic status, remittance, food security, cooking oil, iodized salt, and water sanitation and hygiene practices	
<5 children	Age, sex, infant and young child feeding, dietary diversity, morbidity, nutrition status	Weight, height/length, MUAC
6-9 years old children	Age, sex, dietary diversity, morbidity, nutrition status	Weight, height
Adolescent boys	Age, dietary diversity, behavioral risk factors of non-communicable diseases, mental health (depression)	Weight, height
Adolescent girls	Age, dietary diversity, reproductive history, menstrual hygiene, marital status, behavioral risk factors of non-communicable diseases, mental health (depression)	Weight, height
20-59 years old men	Age, dietary diversity, behavioral risk factors of non-communicable diseases, self reported chronic disease	Weight, height, Waist circumference, % body fat, BP
20-59 years old women	Age, dietary diversity, reproductive history, menstruation/menopause/menstrual hygiene, behavioral risk factors of non-communicable diseases, self reported chronic disease, pregnancy	Weight, height, Waist circumference, % body fat, BP
60 years and above	Age, dietary diversity, behavioral risk factors of non-communicable diseases, self reported chronic disease, nutritional status, quality of life (EQ-5D-5L), mental health (depression)	Weight, height, Waist circumference, % body fat, BP

Measurement devices

SL#	Anthropometry/BP	Device
1	Weight [1-5 years, 6-9 years, adolescent boys, girls, 20-59 years old women and men, older adults]	TANITA UM-070
2	Weight (<1 years)	EB-522
3	Height/Recumbent length	Locally made height board
4	MUAC	ShorrTape
5	Waist circumference (Adult and elderly)	Local measuring tape
6	BP for adult and elderly male & female	Omron (HEM 7120)

Distribution of listed household members by population group and place of residence (2018-2019)

Variables	Population summary			
	Rural	non-slum Urban	Slum	Total/Overall
Households	16,403 (64.7%)	5,726 (22.6%)	3,242 (12.8%)	25,371
Total Population	70,762 (71.3%)	23,207 (23.4%)	13,691 (13.8%)	99,209
0-5 years old children	6,891 (67.0%)	1,931 (18.8%)	1,460 (14.2%)	10,282
6-9 years old children	6,426 (67.5%)	1,798 (18.9%)	1,296 (13.6%)	9,520
Adolescent girls	6,736 (64.0%)	2,318 (22.0%)	1,475 (14.0%)	10,529
Adolescent boys	6,654 (65.2%)	2,164 (21.2%)	1,393 (13.6%)	10,211
20-59 years old women	12,692 (63.4%)	4,683 (23.3%)	2,638 (13.2%)	20,013
20-59 years old men	17,717 (61.8%)	7,388 (25.8%)	3,546 (12.4%)	28,651
Older adults	7,341 (75.0%)	1,527 (15.6%)	914 (9.3%)	9,782

Distribution of enrolled respondents by age groups and place of residence (2018-2019)

Variables	Population summary			
	Rural	Urban	Slum	Total/Overall
Households	11,790 (68.1%)	3,368 (19.4%)	2,165 (12.5%)	17,323
Sentinel sites	57	15	10	82
Study population	21,104 (70.3%)	5,256 (17.5%)	3,645 (12.1%)	30,005
0-5 years old children	3,525 (70.0%)	887 (17.6%)	621 (12.3%)	5,033
Adolescent girls	3,490 (69.7%)	898 (17.9%)	622 (12.4%)	5,010
Adolescent boys	3,499 (69.9%)	889 (17.8%)	616 (12.3%)	5,004
20-59 years old women	3,565 (69.7%)	921 (18.0%)	626 (12.2%)	5,112
20-59 years old men	3,504 (70.8%)	840 (17.0%)	607 (12.3%)	4,951
Older adults	3,521 (71.9%)	821 (16.8%)	553 (11.3%)	4,895

Longitudinal data analysis: a practical example

Contents

- ❖ Introduction to longitudinal data
- ❖ Regression models for longitudinal data
  - Linear mixed effects and marginal models
- ❖ Analysis of nutrition surveillance data
  - Descriptive statistics and regression models
- ❖ Summary

Longitudinal study…1

- In a longitudinal study, individuals are measured repeatedly over time, whereas in cross-sectional studies, a single outcome is measured from each individual
- Longitudinal study requires special statistical techniques because observations from one individual are assumed to be correlated and from different individuals are assumed to be independent

Longitudinal study…2

- In longitudinal study, within-individual change in response can be captured in addition to between-individual change

- The main objective of a longitudinal study is to examine how these within-individual changes are associated with selected covariates

Longitudinal study…3

- ❖ Regression model for analysing independent responses, e.g., linear regression model, logistic regression model, proportional hazards model, can be extended for analysing longitudinal data
- ❖ There are two main approaches to analysing longitudinal data
  - Conditional model, e.g., linear and generalized linear mixed effects models, frailty models, etc.
  - Marginal model, e.g., generalized estimating equations (gee) method, which is an extension of generalized linear model (glm)

Linear mixed effects (LME) model…1

- ❖LME model accommodates between-individual variability as well as within individual variability over time
- ❖The fixed effect estimates obtained from an LME model for repeated continuous outcome measures describe the population average effects that can be obtained by marginal models
- ❖LME can be used to explore subject-specific prediction and to adjust for possible confounders
- ❖Model assumptions can be examined using residual analysis

Linear mixed effects model…2

- $\boldsymbol{Y}_i = (y_{i1}, \ldots, y_{id})' \rightarrow$ response vector corresponding to *ith* individual $(i = 1, \ldots, n)$

- Linear mixed effects model for the response vector $\boldsymbol{Y}_i$

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

  - X_i is the covariate matrix and $\boldsymbol{\beta}$ is the corresponding vector of fixed effects
  - Z_i is the covariate matrix and $\boldsymbol{b}_i$ is the corresponding vector of random effects, and $\boldsymbol{\epsilon}_i$ is the vector of random errors

Linear mixed effects model…3

❖ Assumptions regarding random effects and error terms

$$\boldsymbol{b}_i \sim N_d(\boldsymbol{0}, G) \quad and \quad \boldsymbol{\epsilon}_i \sim N_n(\boldsymbol{0}, \Sigma)$$

❖ Independence assumption

- Responses within a cluster are marginally correlated, but they are independent conditional on the random effects

❖ Marginal and conditional mean

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta} \quad and \quad E(\boldsymbol{Y}_i \mid \boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

Marginal models…1

- In marginal models, also known as population-averaged models, the mean function is defined for the response at each occasion as a function of covariates

$$E(Y_{ij} \mid X_{ij}) = X'_{ij}\boldsymbol{\beta}$$

  - Y_{ij} is the response of the ith individual at the jth occasion and X_{ij} is the corresponding p-dimensional covariate vector
- The mean function does not include any random effects, and no joint distributional assumptions are required in some marginal models, e.g., the generalized estimating equations (GEE) method

Marginal models…2

- To accommodate within-subject correlation in the analysis, the covariance matrix of the ith response $\boldsymbol{Y}_i$ is defined as

$$V_i = A_i^{1/2} R_i(\boldsymbol{\alpha}) A_i^{1/2}$$

- A_i is diagonal matrix with elements $Var(y_{ij})$ and $R_i(\cdot)$ is the working correlation matrix, which is specified by the user
- The selection of the working correlation matrix depends on the type of within-subject correlations, e.g., exchangeable, auto-regressive, unstructured, etc.

Analysis of Nutrition Surveillance Data

An Example of Longitudinal Data…1

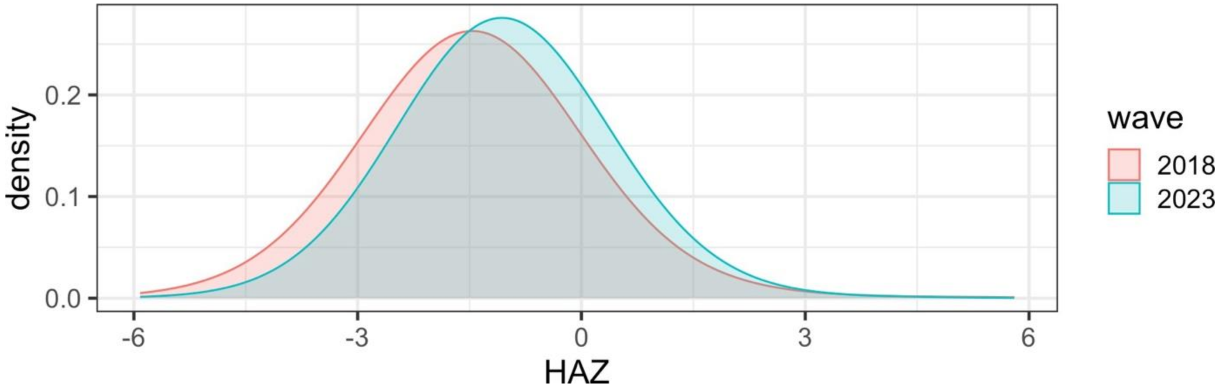
- For this example, the data were obtained from 90 surveillance sites
- The surveillance sites are distributed in the rural, non-slum urban, and slum urban areas of eight administrative divisions of Bangladesh
- Nutritional data were captured from 3,005 children in 2018 and 2023

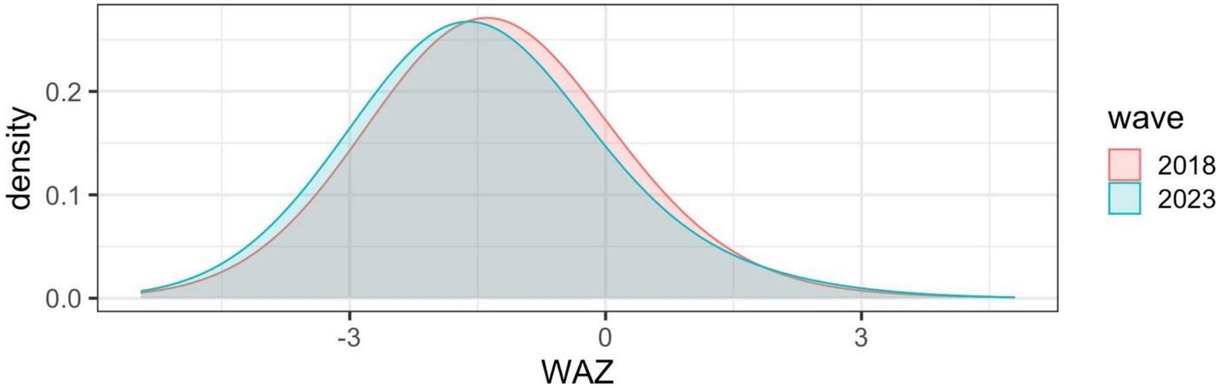
An Example of Longitudinal Data…2

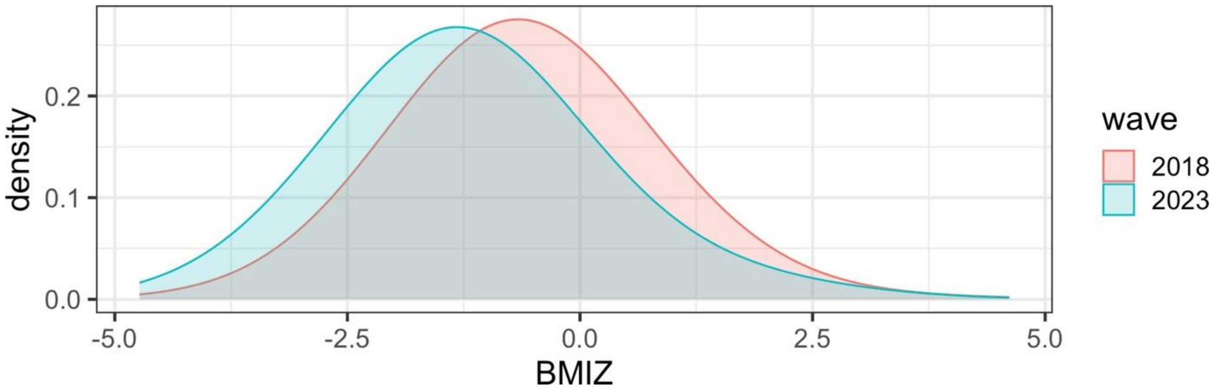
- Responses
  - Height-for-age *z*-score (HAZ), Weight-for-age *z*-score (WAZ), and Body mass index *z*-score (BMIZ)

- Predictors
  - Mother and father education, religion, division, place of residence, children's age and sex, wave
  - Access to improved sanitation, food security, consumption of processed food, household size, wealth index, dietary diversity

Distributions of responses







Characteristic	2018[1] N = 3,005	2023[1] N = 3,005
HAZ	-1.43 (1.27)	-1.02 (1.11)
WAZ	-1.30 (1.15)	-1.44 (1.23)
BMIZ	-0.59 (1.11)	-1.18 (1.23)
[1]Mean (SD)		

Descriptive Statistics of Responses

Stunting

	2023	
2018-19	**Not Stunted**	**Stunted**
Not Stunted	1941 (96.1%)	78 (3.9%)
Stunted	513 (54.9%)	422 (45.1%)

Underweight

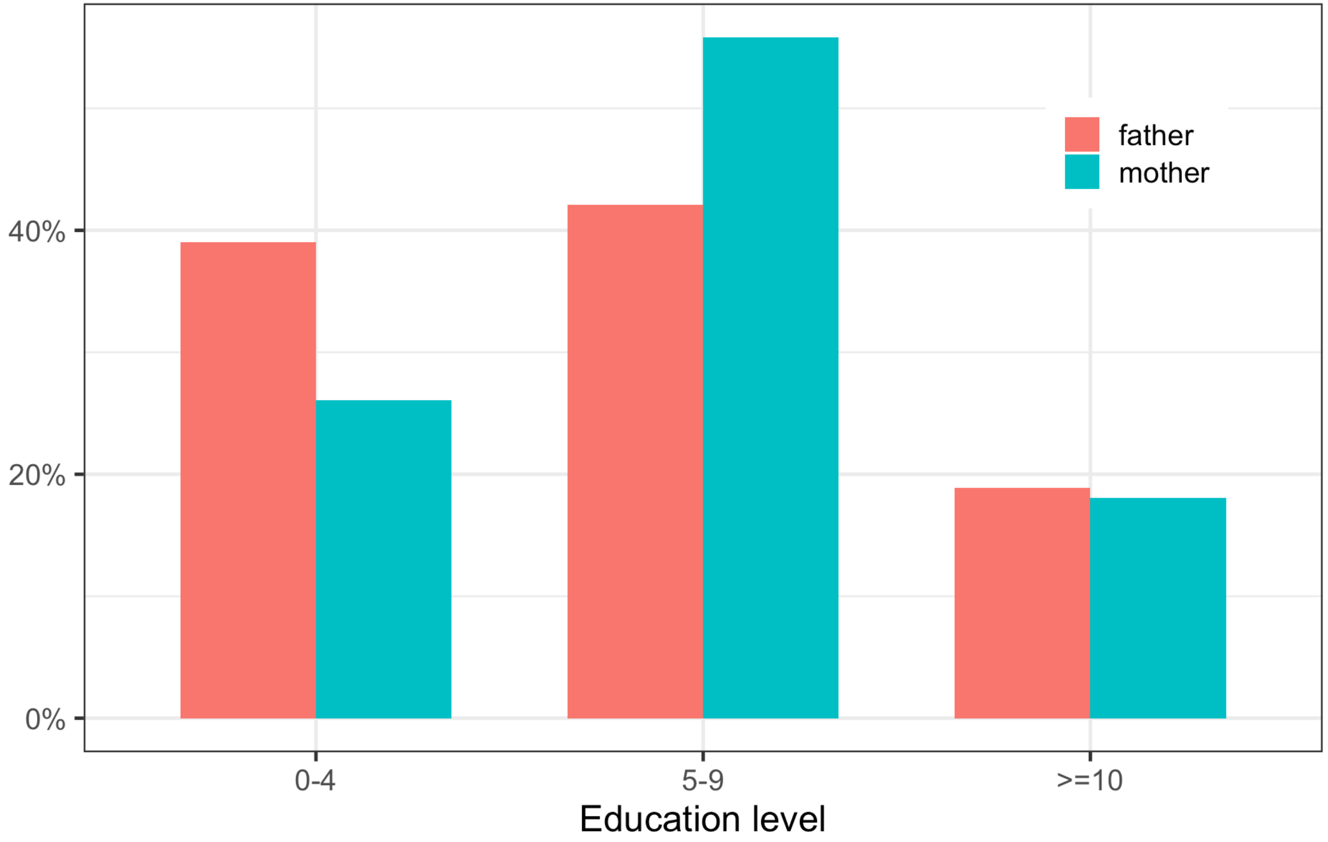
	2023	
2018-19	**Not Underweight**	**Underweight**
Not Underweight	1834 (83.7%)	78 (16.3%)
Underweight	179 (22.5%)	615 (77.5%)

BMI

	2023		
2018-19	**Underweight**	**Normal weight**	**Overweight + Obese**
Underweight	362 (77.5%)	101 (21.6%)	4 (0.9%)
Normal weight	671 (29.2%)	1511 (65.7%)	117 (5.1%)
Overweight + Obese	18 (7.5%)	164 (68.6%)	57 (23.8%)

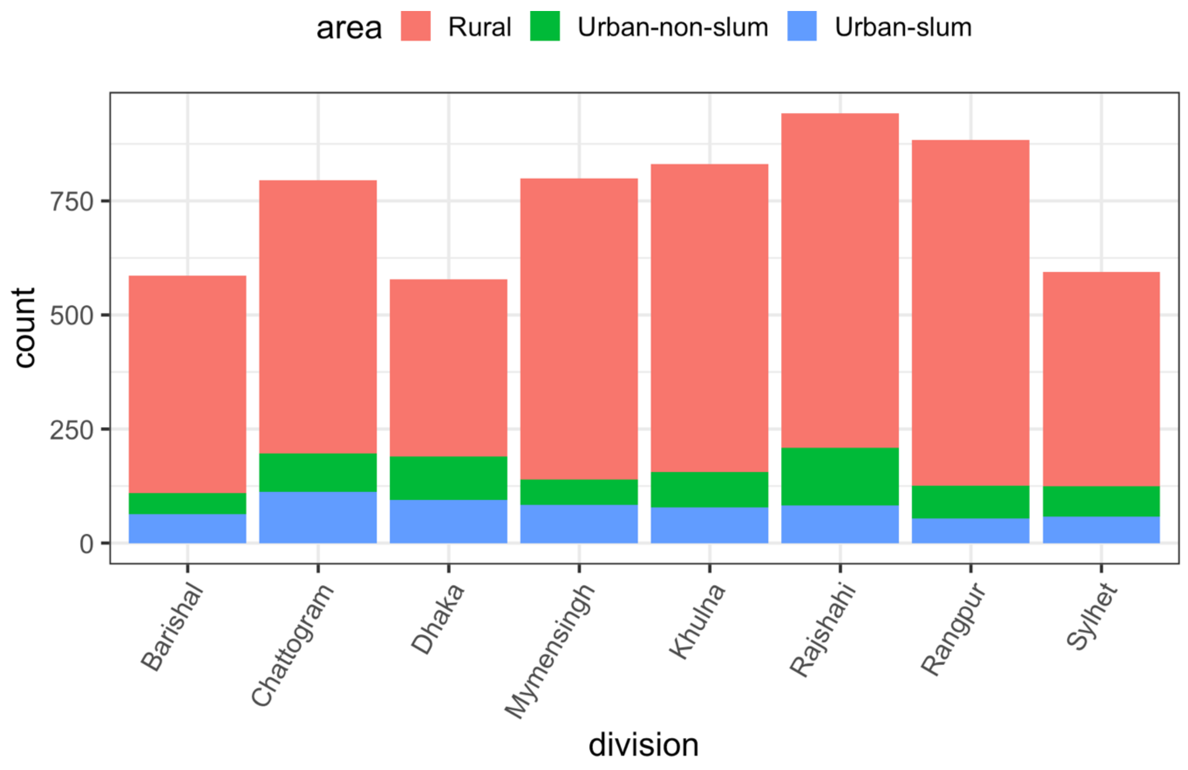
Descriptive Statistics of Selected Predictors…1

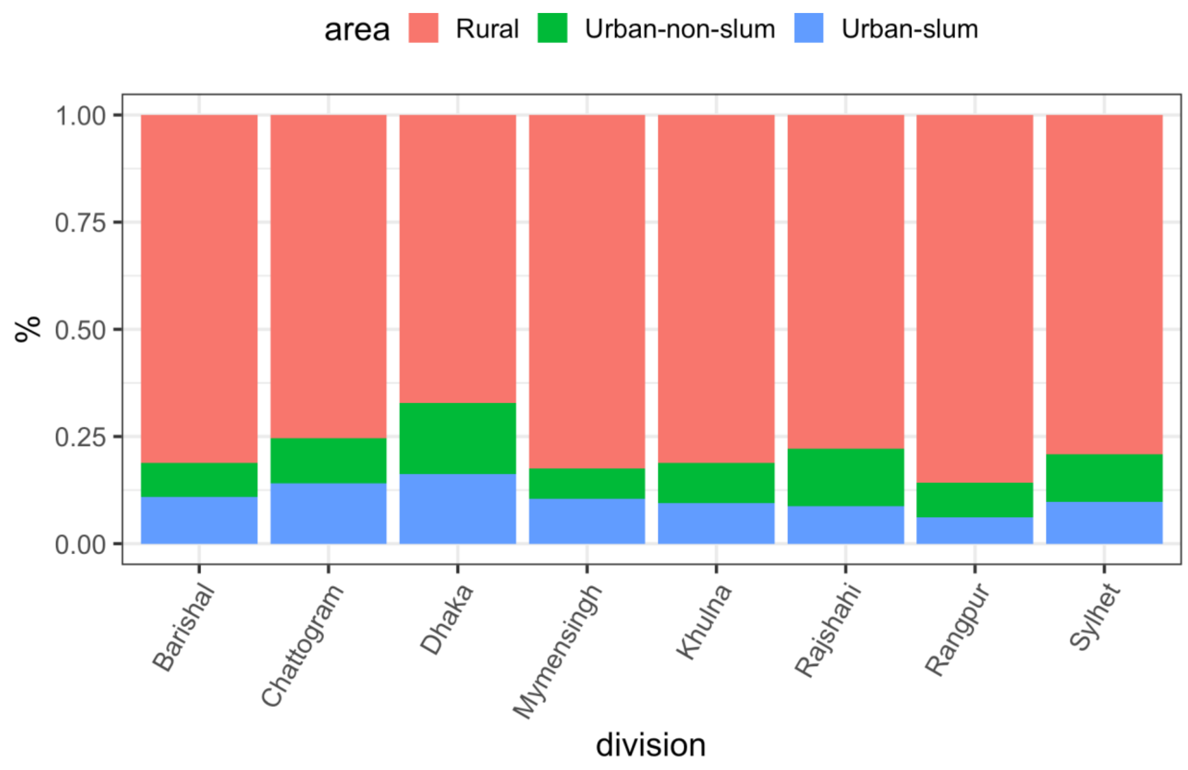
Distributions of parents' educational levels



Characteristic	N = 3,005[1]
Girl	1,557 (52%)
Child age (in yrs)	2.69 (1.28)
Area of residence	
Rural	2,380 (80%)
Urban-non-slum	312 (10%)
Urban-slum	313 (10%)
[1] n (%); Mean (SD)	

Descriptive Statistics of Selected Predictors…2





Descriptive statistics of selected predictors…3

Characteristic	2018 N = 3,005	2023 N = 3,005
Food security		
Secure	1,751 (58%)	1,800 (60%)
Mild-insecure	746 (25%)	727 (24%)
Mod+sev-insecure	508 (17%)	476 (16%)
Dietary diversity (>=5)	1,274 (43%)	1,483 (49%)
Processed food intake (any)	2,439 (82%)	2,220 (80%)
Any morbidity in last two weeks	1,835 (61%)	1,127 (38%)

[1] n (%)

Characteristic	2018 N = 3,005	2023 N = 3,005
Wealth index		
lowest	597 (20%)	598 (20%)
second	605 (20%)	600 (20%)
middle	605 (20%)	611 (20%)
fourth	609 (20%)	600 (20%)
highest	589 (20%)	594 (20%)
Access to improved sanitation	1,784 (59%)	2,055 (68%)

[1] n (%)

Time-dependent predictors…1

- ❖ In a longitudinal study, predictors could be either fixed or time-dependent depending on whether values of the variable can change over time or not
- ❖ Fixed predictors
  - Sex of the child, parent's educational level, place of residence, division, etc.
- ❖ Time-dependent predictors
  - Improved sanitation, wealth index, food insecurity, area of residence, etc.

Time-dependent predictors…2

❖ Distributions of households with Food security over the study period

	2023		
2018-19	**Secure**	**Mild-insecure**	**Moderate to severe insecure**
Secure	1207 (69.0%)	365 (20.9%)	178 (10.2%)
Mild-insecure	377 (50.6%)	223 (29.9%)	145 (19.5%)
Moderate to severe insecure	216 (42.5%)	139 (27.4%)	153 (30.1%)

```
## # A tibble: 6,010 × 5
##    unique_id wave     HAZ edu_m Food_s
##    <chr>     <fct>  <dbl> <fct> <fct>
##  1 10101     2018  -1.94  5-9   Secure
##  2 10101     2023  -2.04  5-9   Mild-insecure
##  3 10103     2018  -0.343 0-4   Mild-insecure
##  4 10103     2023  -1.29  0-4   Mild-insecure
##  5 10107     2018  -2.11  >=10  Secure
##  6 10107     2023   0.307 >=10  Secure
##  7 10108     2018  -1.86  >=10  Secure
##  8 10108     2023  -0.247 0-4   Secure
##  9 10120     2018  NA     5-9   Secure
## 10 10120     2023   2.72  5-9   Secure
## # ℹ 6,000 more rows
```

Long and Wide Format Data

Long and Wide format data

- In a longitudinal study, more than one observations are collected on each subject, and data can be represented in either long or wide format

- In wide format data, there is one row for observations of each unit, where as in long format data, there is more than one row for observations of each unit

Wide format data

```
> Wdat
```

```
## # A tibble: 3,005 × 5
##    unique_id y_2018 y_2023 edu_m sex
##    <chr>      <dbl>  <dbl> <fct> <fct>
##  1 10101     -1.94  -2.04  5-9   girl
##  2 11101     -3.28  -2.12  5-9   girl
##  3 13101     -0.426 -0.681 5-9   girl
##  4 21101     -1.68  -1.53  0-4   girl
##  5 23101     -1.65  -0.862 0-4   girl
##  6 25101     -2.56  -0.972 5-9   boy
##  7 26101     -2.06  -1.17  >=10  boy
##  8 27101     -0.152 -0.569 5-9   boy
##  9 28101     -0.458  0.817 0-4   boy
## 10 30101     -0.459 -0.044 >=10  girl
## #  2,995 more rows
```

Long format data

```
> Ldat
```

```
## # A tibble: 6,010 × 5
##    unique_id edu_m sex   time      HAZ
##    <chr>     <fct> <fct> <chr>   <dbl>
##  1 10101     5-9   girl  y_2018 -1.94
##  2 10101     5-9   girl  y_2023 -2.04
##  3 11101     5-9   girl  y_2018 -3.28
##  4 11101     5-9   girl  y_2023 -2.12
##  5 13101     5-9   girl  y_2018 -0.426
##  6 13101     5-9   girl  y_2023 -0.681
##  7 21101     0-4   girl  y_2018 -1.68
##  8 21101     0-4   girl  y_2023 -1.53
##  9 23101     0-4   girl  y_2018 -1.65
## 10 23101     0-4   girl  y_2023 -0.862
## #  6,000 more rows
```

Wide → Long

```
pivot_longer(
  data = Wdat,
  cols = c(y_2018, y_2023),
  names_to = "time",
  values_to = "HAZ"
  )
```

```
## # A tibble: 6,010 × 5
##    unique_id edu_m sex   time     HAZ
##    <chr>     <fct> <fct> <chr>  <dbl>
##  1 10101     5-9   girl  y_2018 -1.94
##  2 10101     5-9   girl  y_2023 -2.04
##  3 11101     5-9   girl  y_2018 -3.28
##  4 11101     5-9   girl  y_2023 -2.12
##  5 13101     5-9   girl  y_2018 -0.426
##  6 13101     5-9   girl  y_2023 -0.681
##  7 21101     0-4   girl  y_2018 -1.68
##  8 21101     0-4   girl  y_2023 -1.53
##  9 23101     0-4   girl  y_2018 -1.65
## 10 23101     0-4   girl  y_2023 -0.862
## # ℹ 6,000 more rows
```

Long → Wide

```
pivot_wider(
  data = Ldat,
  id_cols = unique_id,
  names_from = time,
  values_from = HAZ
  )
```

```
## # A tibble: 3,005 × 3
##    unique_id y_2018 y_2023
##    <chr>      <dbl>  <dbl>
##  1 10101     -1.94  -2.04
##  2 11101     -3.28  -2.12
##  3 13101     -0.426 -0.681
##  4 21101     -1.68  -1.53
##  5 23101     -1.65  -0.862
##  6 25101     -2.56  -0.972
##  7 26101     -2.06  -1.17
##  8 27101     -0.152 -0.569
##  9 28101     -0.458  0.817
## 10 30101     -0.459 -0.044
## #  2,995 more rows
```

Regression Models for Longitudinal Data

Linear mixed effects model…1

- Considered a linear mixed effects mode for HAZ with random intercept

- Wave, age at the baseline (centered at 2 years), sex of the child, and area of residence are included in the model

- The R function lme4::lmer is used to fit the linear mixed effects model

Linear mixed effects model...2

```
> library(lme4)
> mod1 <- lmer(formula = HAZ ~ wave + age2018c + Girl + area +
  (1|unique_id), REML = TRUE, data = dat_f)
```

- The `formula` argument contains both the fixed and random effects
- The term `(1|unique_id)` indicates a random intercept model is considered for the HAZ score
- Since we have only two responses for each child, we cannot consider a model with a random slope

LME: Estimates of model parameters

```
## # A tibble: 6 × 4
##   var                   Estimate      se p_value
##   <chr>                    <dbl>   <dbl> <chr>
## 1 (Intercept)             -1.51  -38.3   <0.001
## 2 wave2023                 0.402  23.6   <0.001
## 3 age2018c                 0.009   1.07  0.142
## 4 Girl                     0.035   0.875 0.191
## 5 areaUrban-non-slum       0.546   8.30  <0.001
## 6 areaUrban-slum          -0.098  -1.50  0.067
```

❖ The fixed effects parameters have population-averaged interpretations

- Average HAZ score increases by 0.402 over 2018 to 2023
- The average HAZ score of children from non-slum Urban areas is 0.035 unit higher than those of Rural area

LME: Random effects and error variances

```
##  Groups    Name        Std.Dev.
##  unique_id (Intercept) 0.986
##  Residual              0.656
```

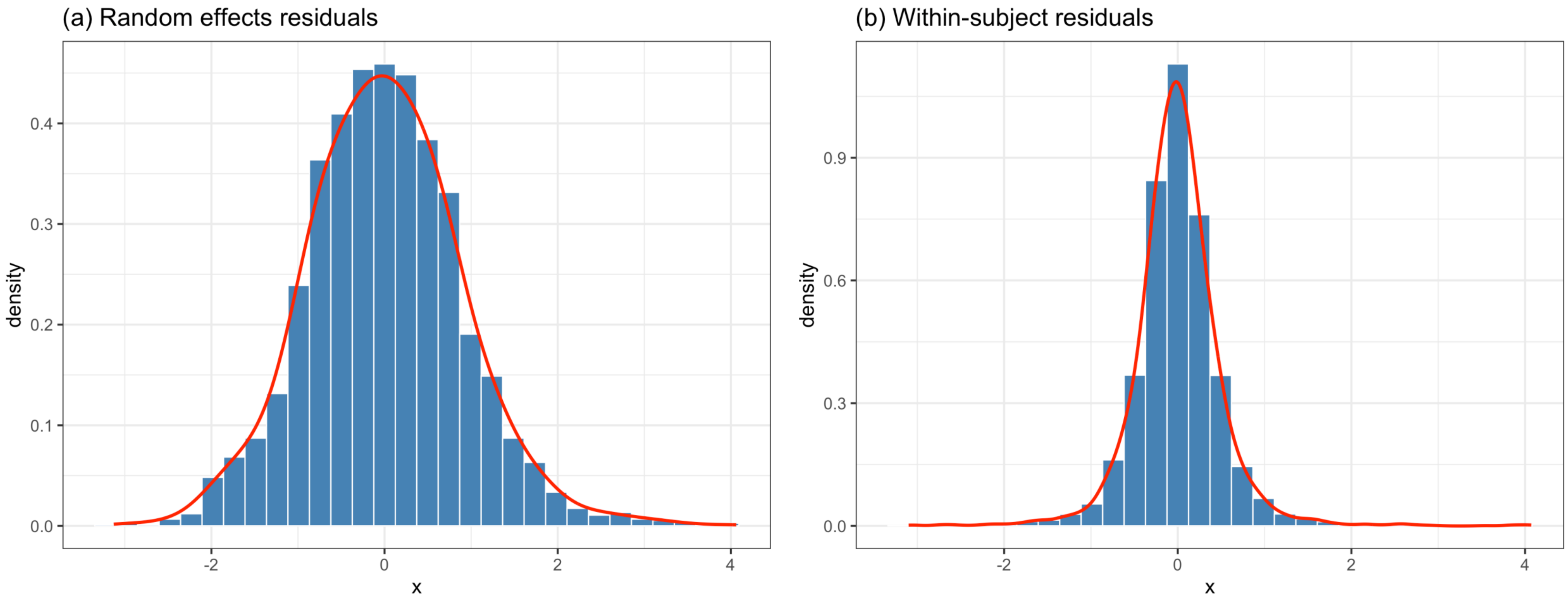
LME: Model selection criteria

```
## # A tibble: 1 × 7
##     nobs sigma logLik    AIC    BIC REMLcrit df.residual
##    <int> <dbl>  <dbl>  <dbl>  <dbl>    <dbl>       <int>
## 1   5959 0.656 -8509. 17034. 17087.   17018.        5951
```

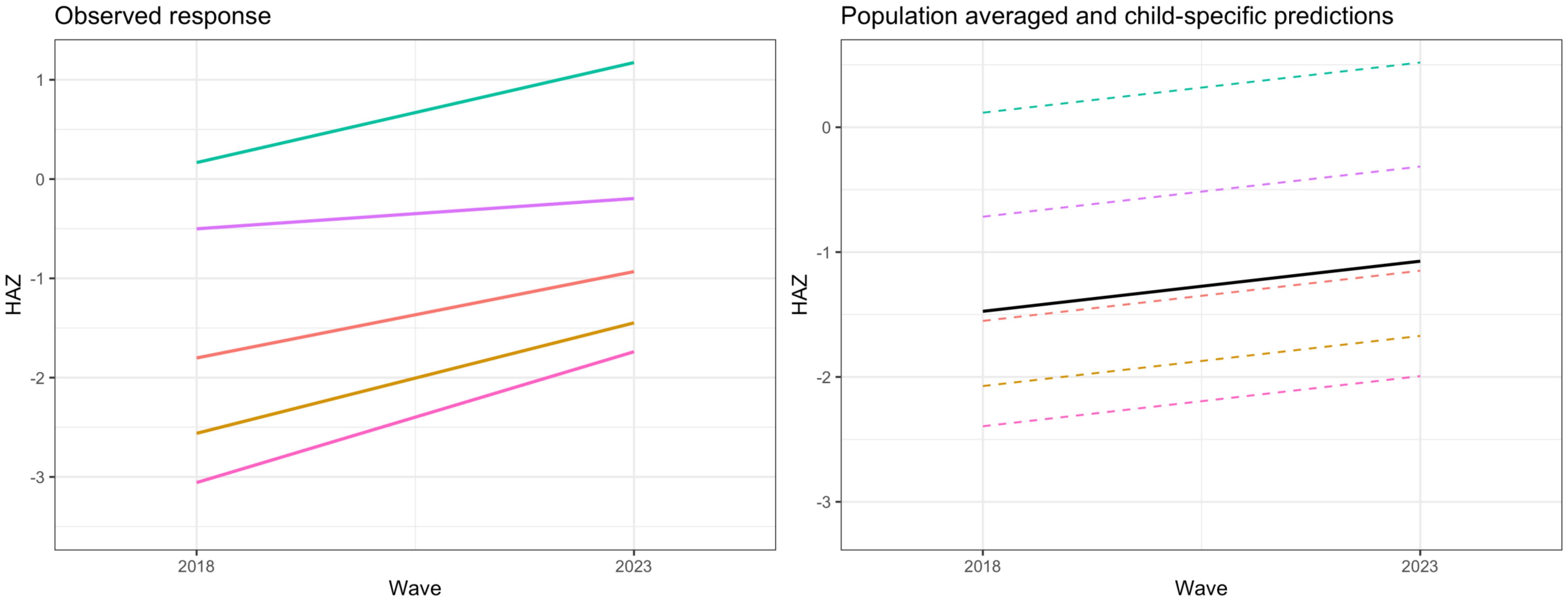
LME: ICC

```
## # Intraclass Correlation Coefficient
##
##                 ICC: 0.693
##
```

LME: Model Diagnostics



Observed response and predictions of 2-year-old girls from food-secured families in rural area



Generalized estimating equations (GEE)…1

- We consider modeling HAZ score with wave, age at 2018, sex of the child, and area as predictors

- Since there are only two responses for each child, there will be only one association parameter, i.e., exchangeable and unstructured correlation structures are the same

- R function geepack::geeglm is used for the fit, which requires specifying the id-variable and correlation structure

Generalized estimating equations (GEE)…2

Code for GEE

```
> library(geepack)
> mod2 <- geeglm(
  formula = HAZ ~ wave + age2018c + Girl + area,
  id = unique_id,
  corstr = "exchangeable",
  family = "gaussian",
  data = dat_f
  )
```

Estimates of model parameters

```
##                       Estimate Std.err Pr(>|W|)

## (Intercept)             -1.509   0.040    0.000
## wave2023                 0.402   0.017    0.000
## age2018c                 0.009   0.008    0.271
## Girl                     0.035   0.040    0.382
## areaUrban-non-           0.546   0.071    0.000
   slum
## areaUrban-slum          -0.098   0.063    0.122
```

Interpretations are similar to multiple linear regression models

Generalized estimating equations (GEE)…3

Correlation parameter estimate

```
## # A tibble: 1 × 2
##   alpha se_alpha
##   <dbl>  <dbl>
## 1 0.675 0.0169
```

Model selection criteria

```
##       QIC      QICu Quasi Lik       CIC    params      QICC
##  8294.836  8289.003 -4138.502     8.916     6.000  8294.873
```

Generalized estimating equations (GEE)…4

- The GEE method can be used to model time-dependent predictors, such as a household's food security status, wealth index, etc.

- To model a time-dependent predictor, the correlation structure must be specified as *independence* in the GEE routine

Time-dependent predictor food security status (Food_s)

```
## # A tibble: 10 × 4
##    unique_id  wave   Food_s         sex
##    <chr>      <fct> <fct>           <fct>
##  1 1179       2018   Mild-insecure  boy
##  2 1179       2023   Mild-insecure  boy
##  3 24212      2018   Mild-insecure  girl
##  4 24212      2023   Secure         girl
##  5 24645      2018   Mild-insecure  girl
##  6 24645      2023   Secure         girl
##  7 78187      2018   Secure         boy
##  8 78187      2023   Secure         boy
##  9 85268      2018   Secure         girl
## 10 85268      2023   Secure         girl
```

Code for GEE with time-dependent predictor Food_s

```
> mod2a <- geeglm(
  formula = HAZ ~ wave + age2018c + Girl + area + Food_s,
  id = unique_id, corstr = "independence", family = "gaussian",
  data = dat_f)
```

GEE: Estimates of model parameters

```
##                             Estimate  Std.err  Pr(>|W|)
## (Intercept)                   -1.412    0.042     0.000
## wave2023                       0.404    0.018     0.000
## age2018c                       0.009    0.008     0.230
## Girl                           0.034    0.039     0.386
## areaUrban-non-slum             0.522    0.070     0.000
## areaUrban-slum                -0.053    0.063     0.399
## Food_sMild-insecure           -0.217    0.040     0.000
## Food_sMod+sev-insecure        -0.306    0.045     0.000
```

Comparisons between two models

GEE: Model selection criteria

```
##          QIC  QICu   Quasi Lik     CIC  params  QICC
##    mod2 8295  8289       -4139   8.912       6  8295
##   mod2a 8201  8194       -4089  11.090       8  8201
```

The final model

- ❖ We considered the GEE method to fit the final model for HAZ
- ❖ The final model contains the following predictors
  - Wave, parents' education level, wealth index, food security, access to improved sanitation, area, division, household size, religion
  - Child age and sex

Estimates of the final model parameters

Term	Estimate	P-value
wave2023	0.393	**<0.001**
edu_m5-9	0.128	**0.003**
edu_m>=10	0.326	**<0.001**
edu_f5-9	0.089	**0.030**
edu_f>=10	0.286	**<0.001**
divisionChattogram	-0.269	**<0.001**
divisionDhaka	-0.170	0.053
divisionMymensingh	-0.267	**0.001**
divisionKhulna	-0.187	**0.023**
divisionRajshahi	-0.145	**0.068**
divisionRangpur	0.004	0.960
divisionSylhet	-0.652	**<0.001**

Term	Estimate	P-value
areaUrban-non-slum	0.298	<**0.001**
areaUrban-slum	-0.067	0.290
quint_comsecond	0.016	0.745
quint_commiddle	0.059	0.220
quint_comfourth	0.192	**<0.001**
quint_comhighest	0.232	**<0.001**
proc_foodtimes	-0.003	**0.020**
Food_sMild-insecure	-0.081	**0.037**
Food_sMod+sev-insecure	-0.117	**0.010**
sexgirl	0.037	0.322
age2018c	0.011	0.159
religionOther than Islam	-0.029	0.633
imp_saniNot improved	-0.045	0.175
insuf_ddFood groups >= 5	0.030	0.334

Summary

- Linear mixed effects (LME) model and generalized estimating equations (GEE) are considered for modeling correlated HAZ scores
- R codes for preparing the data for fitting LME and GEE are discussed
- Interpretations of estimates, model selection criteria, etc. are described for LME and GEE fits

Summary from the Final model

- Average HAZ score increased by 0.393 units over the years 2018 to 2023
- The average HAZ score improved for children whose parents had at least 5 years of education compared to those with less than 5 years
- The average HAZ score for children is higher in non-slum urban areas compared to rural areas, with no significant difference between children from rural and urban slum areas

Summary from the Final model

- Average HAZ score is relatively higher among children from wealthy and lower in food-insecure families and who eat processed food

- Sex and age of the child, household size, religion, dietary diversity, and access to improved sanitation were not significantly associated with the average HAZ score

Acknowledgement

- Participants of the study
- Investigators of the study
- IPHN, BBS, JPGSPH staff
- TAG Members
- District Commissioners
- Civil surgeons
- Upazilla Nirbahi Officers
- Upazilla Health & Family Planning Officers, Upazilla Family Planning Officers
- Mayors
- Councilors
- Union Parishad Chairmen
- Union Parishad Members
- Local elites

Thank You