Brief description of WASH Benefits endline survey data sets

# Counts in the WASH Benefits data sets:

We started with 5,551 complete household surveys in the baseline. In the midline, the survey team collected 4,716 complete surveys.

|  |  |
| --- | --- |
| Household status | Counts |
| Complete  Incomplete | 4,716  835 |

The team did not able to collect survey from the remaining 835 mothers due to following reasons:

|  |  |
| --- | --- |
| Reason | Counts |
| ABORTION | 5 |
| ABSENT | 186 |
| CHILD DEATH | 197 |
| FALSE PREGNANCY | 20 |
| MIGRATION OUT | 57 |
| MISCARRIAGE | 137 |
| REFUSE | 32 |
| STILL BIRTH | 201 |

In the endline, we decided to survey all the households with complete survey and the households who were either absent or refused to participate in the midline survey. During the endline, the team approached 4,937 households and able to collect complete survey from 4,639 mothers with at least one living target child.

|  |  |
| --- | --- |
| Household status | Counts |
| Complete  Incomplete | 4,639  298 |

The team failed to collect complete survey from the remaining households due to following reasons:

|  |  |
| --- | --- |
| Reason | Counts |
| ABSENT | 173 |
| CHILD DEATH | 23 |
| MIGRATION OUT | 55 |
| REFUSE | 47 |

In the endline data sets, we have 4611 mothers with one living child and 28 mothers with twin living children. So, the final household count is 4,639 and the total target children count is 4667. One exception is the anthropometry data set. It does not contain information from 35 households among all the complete survey households in the endline. In addition, they collect information from 6 households who refused to participate in the main survey.

|  |  |
| --- | --- |
| Data sets | Module |
| 01. WASHB\_Midline\_Endline\_data\_count\_cleaned.dta | Data counts of baseline, midline, and endline |
| 02. WASHB\_Endline\_Arm\_Identification.dta | Intervention arm |
| 03. WASHB\_intervention\_delivery\_date.dta | Intervention delivery date |
| 04. WASHB\_Endline\_main\_survey\_cleaned.dta | Module 40, 1, 0, 7, 8, 9, 10, 13, 41, 42, 61 |
| 05. WASHB\_Endline\_mother\_movement\_cleaned.dta | Module 1 (Mother's movement part) |
| 06. WASHB\_Endline\_census\_cleaned.dta | Module 0 (Census part) |
| 07. WASHB\_Endline\_childinformation\_cleaned.dta | Module 40 (Children information part) |
| 08. WASHB\_Endline\_childhealthinfo\_cleaned.dta | Module 2 |
| 09. WASHB\_Endline\_anthropometry\_cleaned.dta | Module 4 |
| 10. WASHB\_Endline\_vaccination\_cleaned.dta | Module 5 |
| 11. WASHB\_Endline\_LNS\_cleaned.dta | Module 14 |
| 12. WASHB\_Endline\_ffq\_cleaned.dta | Module 6 |
| 13. WASHB\_Endline\_STH\_mothers\_parity\_fathers\_age.dta | Mother's parity and father's age from STH survey |
| WASHB\_Endline\_Codebook.xlsx | Codebook of all the modules |
| WASHB\_Endline\_survey\_questionnaire.docx | Survey questionnaire except module 4 |
| WASHB\_Endline\_anthtropometry\_questionnaire.docx | Module 4 survey questionnaire |

# List of WASHB midline datasets and codebook:

# Description of each datasets:

01. WASHB\_Midline\_Endline\_data\_count\_cleaned.dta

This data set contains all the 5,551 WASH Benefits survey households. dataid variable is the unique id variable. It contains complete and incomplete status of all the households and their reason for incompleteness in the both midline and endline survey. Each row represents a single mother/household.

02. WASHB\_Endline\_Arm\_Identification.dta

This data set contains arm identification of all the households of WASH Benefits study. Here dataid variable is the unique id variable. Each row represents a single mother/household.

03. WASHB\_intervention\_delivery\_date.dta

This data set contains all the date of intervention delivery of all the clusters. Cluster\_ID is the unique id variable. Each row represents single cluster.

04. WASHB\_Endline\_main\_survey\_cleaned.dta

This dataset contains information from 4,639 households who have willingly participated in the endline survey and have at least one target child. Each row represents a household. Households are uniquely identified by the *dataid* key variable. First three digits of *dataid* represent cluster id and last two digits represent mother id; e.g. dataid "00102" means that this household belongs to a mother with id number "02" under cluster number "001". This *dataid* is consistent among all the datasets.

05. WASHB\_Endline\_mother\_movement\_cleaned.dta

This data set is in long format since mother moved multiple times. Here each row represents each mother's movement.

06. WASHB\_Endline\_census\_cleaned.dta

This dataset contains census related information from all the target households and all the households in a compound where a target household belongs. Each row represents single household and household may repeat if there are more than 1 target household in a compound. This dataset could be matched with "04. WASHB\_Endline\_main\_survey\_cleaned.dta" using *dataid* as key variable.

07. WASHB\_Endline\_childinformation\_cleaned.dta

This dataset contains identification of all the children who was <36 months old either during endline, midline or during baseline. It contains identification of 9964 children including 4667 target children. Each row represents a child and a child may repeat if there are more than 1 target household belong to same compound. This dataset could be matched with "04. WASHB\_Endline\_main\_survey\_cleaned.dta" using *dataid* as key variable.

Under *childid* variable, "T1" means first target child, "T2" means second target child, other value starting with "C" is a compound child.

08. WASHB\_Endline\_childhealthinfo\_cleaned.dta

This dataset contains health related information of the children who are listed in the "07. WASHB\_Endline\_childinformation\_cleaned.dta". Each row represents a child and this could be matched with "07. WASHB\_Endline\_childinformation\_cleaned.dta"dataset with a combination of *dataid childid* as key variables.

09. WASHB\_Endline\_anthropometry\_cleaned.dta

This dataset contains anthropometry related information of 4,637 children including 2,143 elder siblings and 4,637 target children. This dataset is in long format according to first target child, second target child, and elder siblings. childid starts with "T" is a target child and starts with "S" is an elder siblings.

It does not contain information from 35 households among all the complete survey households in endline. In addition, they collect information from 6 households who refused to participate in the main survey.

10. WASHB\_Endline\_vaccination\_cleaned.dta

This dataset contains vaccination related information of all the target children. This dataset could be matched with "07. WASHB\_Endline\_childinformation\_cleaned.dta" dataset using a combination of *dataid childid* as key variable. This dataset is in wide format according to first target child and second target child.

11. WASHB\_Endline\_LNS\_cleaned.dta

This dataset contains LNS related information of all the target children. This dataset could be matched with "07. WASHB\_Endline\_childinformation\_cleaned.dta" dataset using a combination of *dataid childid* as key variable.

12. WASHB\_Endline\_ffq\_cleaned.dta

This dataset contains child food frequency related information of all the target children. This dataset could be matched with "07. WASHB\_Endline\_childinformation\_cleaned.dta" dataset using a combination of *dataid childid* as key variable.

13. WASHB\_Endline\_STH\_mothers\_parity\_fathers\_age.dta

This data set contains mother's parity and father's age. This data set could be matched with "04. WASHB\_Endline\_main\_survey\_cleaned.dta" using dataid as the key variable.