Brief description of WASHB midline survey datasets

# List of WASHB midline datasets and codebook:

|  |  |
| --- | --- |
| Datasets | Module |
| 1. WASHB\_Midline\_main\_survey\_cleaned.dta | Module 40, 1, 0, 14, 7, 8, 9, 10, 13, 41, 42 |
| 2. WASHB\_Midline\_census\_cleaned.dta | Module 0 (Census part) |
| 3. WASHB\_Midline\_childinfo\_cleaned.dta | Module 40 (Children identification part) |
| 4. WASHB\_Midline\_childhealthinfo\_cleaned.dta | Module 2 |
| 5. WASHB\_Midline\_deworming\_cleaned.dta | Module 3 |
| 6. WASHB\_Midline\_vaccination\_cleaned.dta | Module 5 |
| 7. WASHB\_Midline\_ffq\_cleaned.dta | Module 6 |
| WASHB\_Midline\_codebook.xlsx | Module specific codebook of all the above module |
| 8. WASHB\_Midline\_anthropometry\_cleaned.dta | Module 4 |
| WASHB\_Midline\_anthropometry\_codebook.xlsx | Codebook of module 4 |

# Household count in midline datasets:

In the baseline survey there were 5551 households in total. During midline survey, field team collected complete information from 4716 households, partial information from 597 households, and no information from 238 households.

|  |  |
| --- | --- |
| Household status | Number of household |
| Full complete | 4716 |
| Partial complete | 597 |
| Incomplete | 238 |
| Total | 5551 |

In all the WASHB midline survey datasets, the total count of household is 4716. That's why if any dataset has household specific information there should be 4716 households in total. Incompleteness of information for 835 households occurred due to following reasons:

|  |  |
| --- | --- |
| Reason for incompleteness | Number of households |
| Absent | 186 |
| Abortion | 5 |
| Child death | 197 |
| False pregnancy | 20 |
| Migration out | 57 |
| Miscarriage | 137 |
| Refusal | 32 |
| Still birth | 201 |
| Total | 835 |

# Description of each datasets:

1. WASHB\_Midline\_main\_survey\_cleaned.dta

This dataset contains information from 4716 households who have willingly participated in the meadline 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 except the *8. WASHB\_Midline\_anthropometry\_cleaned.dta* dataset. In the anthropometry dataset, first three digit is the cluster number, the next two digits is the mother id and the last digit is the child id.

2. WASHB\_Midline\_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 *1. WASHB\_Midline\_main\_survey\_cleaned.dta* using *dataid* as key variable.

3. WASHB\_Midline\_childinfo\_cleaned.dta

This dataset contains identification of all the children who was <36 months old either during midline or during baseline. It contains identification of 8952 children including 4747 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 *1. WASHB\_Midline\_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.

4. WASHB\_Midline\_childhealthinfo\_cleaned.dta

This dataset contains health related information of the children who are listed in the *3. WASHB\_Midline\_childinfo\_cleaned.dta*  dataset. There are some mismatch which could be found in the *WASHB\_midline\_mismatch.xlsx* file. Each row represents a child and this could be matched with *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset with a combination of *dataid childid* as key variables.

5. WASHB\_Midline\_deworming\_cleaned.dta

This dataset contains deworming related information of each child listed in the *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset. There are some mismatch which could be found in the *WASHB\_midline\_mismatch.xlsx* file. Each row represents a child and this could be matched with *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset with a combination of *dataid childid* as key variables.

6. WASHB\_Midline\_vaccination\_cleaned.dta

This dataset contains vaccination related information of all the target children. This dataset could be matched with *1. WASHB\_Midline\_main\_survey\_cleaned.dta* using *dataid* as key variable. This dataset could also be matched with *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset using *dataid* as key variable. This dataset is in wide format according to first target child and second target child.

7. WASHB\_Midline\_ffq\_cleaned.dta

This dataset contains child food frequency related information of all target children. This dataset could be matched with *1. WASHB\_Midline\_main\_survey\_cleaned.dta* using *dataid* as key variable. This dataset could also be matched with *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset using *dataid* as key variable. This dataset is alos in wide format according to first target child and second target child.

8. WASHB\_Midline\_anthropometry\_cleaned.dta

This dataset contains anthropometry related information of 4711 target children listed in the *3. WASHB\_Midline\_childinfo\_cleaned.dta* dataset. In this dataset, first three digit of *dataid* is the cluster number, the next two digits is the mother id. This dataset is in long format according to first target child and second target child.