

# MCCT Dataset Development

Nicholus Tint Zaw

2022-06-30

## Purpose of the document

This document organized the record of the step-by-step process involved in the development of the MCCT baseline dataset for the respective module, which includes the interested outcome variables for analysis.

## Data Cleaning

Perform the data cleaning based on the `30-clean_hh.R` function package. As the data collection used two different surveys from HH and anthro, the data cleaning was performed separately for each respective raw data. Please note that the function did not work perfectly, and I am not sure why. Therefore, the individual syntax applied in the function package is used to perform data cleaning instead of the function in finally.

## HH data

```
# load datasets
load('hh.rda')
load('checks.rda')

# load function
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/30-clean_hh.R")

# data cleaning
# hh <- clean_hh(df = hh, checks = checks)
# function is not working perfectly. run individual commend instead of function

hh <- hh[row.names(hh) != 350, ]
hh <- hh[hh$geo_villward != "", ]
hh[hh$geo_vill == "168652" & !is.na(hh$geo_vill), "geo_rural"] <- "2"

for(i in checks$id) {
  hh[checks$index[checks$id == i],
    checks$variable[checks$id == i]] <- checks$newvalue[checks$id == i]
}
```

## Anthro data

```
load('anthroDF.rda')
load('anthroChecks.rda')

# data cleaning
# data cleaning - run by individual commend line from function package
anthroDF[anthroDF$geo_vill == "168652" & !is.na(anthroDF$geo_vill), "geo_rural"] <- "2"

for(i in anthroChecks$id) {
  anthroDF[anthroChecks$index[anthroChecks$id == i],
           anthroChecks$variable[anthroChecks$id == i]] <- anthroChecks$newvalue[anthroChecks$id == i]
}
```

## Outcome indicators Calculation

In this session, I calculated reported indicators from the MCCT baseline analysis based on our secondary data analysis's individual (interested) modules.

The process was simple. First, load the necessary raw `rda` files and then import the `r-function` script file to the `r` global environment. Then, apply the respective function to get the newly calculated outcome indicators variable. Please note that the function application for each module was based on the example session from the respective `r-function` file. For example, the calculation workflow for the child immunization and child health-seeking behavior were applied based on the example function usage session from the `06-recode_chealth.R` file. After calculating the respective module, each newly calculated dataset was saved as STATA `dta` file.

## Child Vaccinations & Child health

### Load the datasets

```
load('childHealth.rda')
load('hhMembers.rda')
```

### Load the functions

```
# run the 06 r script file
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/06-recode_chealth.R")
```

### Variables calculation

```
# creat the dataset with eligable children
chealth <- create_chealth(df = childHealth, x = hh, y = hhMembers)

# cimbined all child health module indicators
child_health <- recode_chealth(df = chealth)
```

```
child_health <- janitor::clean_names(child_health)

write_dta(child_health, file.path(getwd(), "stata_dta", "child_health_all.dta"))
```

## Child Anthro

### Load the datasets

```
load('childAnthro.rda')
```

### Load the functions

```
# run the 06 r script file
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/03-recode_anthro.R")
```

### Variables calculation

```
# created anthro indicator dataset
childanthro <- recode_anthro(df = create_canthro(df = childAnthro, x = anthroDF))

## =====
## =====
## =====

childanthro <- janitor::clean_names(childanthro)

write_dta(childanthro, file.path(getwd(), "stata_dta", "child_anthro_all.dta"))
```

## Maternal health (pregnancy)

### load the datasets

```
load('anc1.rda')
load('anc2.rda')
load('hhMembers.rda')
```

### load the functions

```
# run the 06 r script file
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/05-recode_anc.R")
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/14-recode_delivery.R")
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/17-recode_pnc.R")
```

## Variables calculation

```
# Recode anc indicators for currently pregnant women
x_current <- create_anc(df = anc1, x = hh, y = hhMembers, status = "current")
anc_current <- recode_anc(df = x_current, status = "current")

# Recode anc indicators for non-pregnant women
x_past <- create_anc(df = anc2, x = hh, y = hhMembers, status = "past")
anc_past <- recode_anc(df = x_past, status = "past")

# Recode birth/delivery indicators
delivery <- recode_birth(df = x_past)

# Recode postnatal care indicators
pnc <- recode_pnc(df = x_past)

anc_current <- janitor::clean_names(anc_current)
anc_past <- janitor::clean_names(anc_past)
delivery <- janitor::clean_names(delivery)
pnc <- janitor::clean_names(pnc)

write_dta(anc_current, file.path(getwd(), "stata_dta", "anc_current.dta"))
write_dta(anc_past, file.path(getwd(), "stata_dta", "anc_past.dta"))
write_dta(delivery, file.path(getwd(), "stata_dta", "delivery.dta"))
write_dta(pnc, file.path(getwd(), "stata_dta", "pnc.dta"))
```

## Coping Strategies

### load the datasets

This module will use the existing loaded `hh` module and not be required to load other datasets.

### Load the functions

```
# run the 06 r script file
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/10-recode_csi.R")
```

## Variables calculation

```
# consumption-based Coping Strategies Index
rcsi <- recode_csi_consumption(df = hh)

# livelihoods-based Coping Strategies Index
lcsi <- recode_csi_livelihoods(df = hh)

rcsi <- janitor::clean_names(rcsi)
lcsi <- janitor::clean_names(lcsi)
```

```
write_dta(rcsi, file.path(getwd(), "stata_dta", "rcsi.dta"))
write_dta(lcsi, file.path(getwd(), "stata_dta", "lcsi.dta"))
```

## PPI and Split into Quintiles

### Load the datasets

This module will use the existing loaded `hh` module and not be required to load other datasets.

### Load the functions

```
# run the 06 r script file
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/13-recode_ppi.R")
source("C:/Users/Nicholus Tint Zaw/Documents/GitHub/myanmarMCCTdata/R/25-split_to_quintiles.R")
```

### Variables calculation

```
# Recode poverty probability index indicators
ppiDF <- recode_ppi(df = hh)
ppiDF <- split_to_quintiles(df = ppiDF)

ppiDF <- janitor::clean_names(ppiDF)

write_dta(ppiDF, file.path(getwd(), "stata_dta", "ppiDF.dta"))
```

## Dataset Merging

We still need to perform dataset merging to create to combine the respective module **master data** and the **newly created dataset**. Here, I refer to **master data** as the original variables from the respective module dataset and the **newly created dataset** as the output data frame, which contains only the reported outcome variable. Therefore, we need to merge those two datasets to get one complete dataset for each module. Then, merge again with `hh` dataset to get the household information. However, those dataset merging will be performed with STATA dofile as it is easy to trace the result of merging.