

# Review notes of the Sudan Micronutrient Survey Data

Ernest Guevarra

3 May 2020

# 1 Context

This report documents our review of the national micronutrient survey data collected by the Federal Ministry of Health of Sudan supported by UNICEF in 2018-2019. This review is meant to assess the data in preparation for analysis and reporting.

## 2 Data structure

Data was provided as a Microsoft Excel spreadsheet in XLSX file format with a single worksheet containing all the data. The dataset has 19450 rows and 19 columns.

Given that no data dictionary or codebook has been provided with the dataset, we explored the various variables (columns) of the dataset and attempted to define and describe these variables. The variables are:

```
## [1] "state"      "bc"         "locality"   "psu"        "m.age"      "ch.age"
## [7] "sex"        "muac"       "ch.weight"  "ch.height"  "ch.oedema"  "hb"
## [13] "group"     "calcium"    "crp"        "ferritin"   "iodine"     "X18"
## [19] "X19"
```

Table 1 summarises and describes the variables and raises potential issues needing to be addressed prior to data analysis.

Table 1: Summary of Sudan micronutrient data variables

Variable	Expected data class/type	Actual data class/type
state	Either character giving state names or numeric giving state codes or identifiers	character
bc	This is most likely referring to barcodes which should be numeric	numeric
locality	Either character giving locality names or numeric giving locality codes or identifiers	numeric
psu	Numeric identifier for a PSU. Must match the numeric identifiers for PSUs in the population data	numeric
m.age	Numeric for age of mother in years	character
ch.age	Numeric for age of child in months	numeric
sex	Numeric; 1 for male and 2 for female	character
muac	Numeric for MUAC in millimetres	character
ch.weight	Numeric for child weight in kilograms	character
ch.height	Numeric for child height in centimetres	character
ch.oedema	Numeric indicating child's oedema status as being present or not present	numeric
hb	Numeric for haemoglobin in g/dL	character
group	Character for type of respondent	character
calcium	Numeric for serum calcium in mg/dL	numeric
crp	Numeric for c-reactive protein in mg/L	numeric
ferritin	Numeric for serum ferritin in mg/L	numeric
iodine	Numeric for urinary iodine in mg/L	numeric
X18	Unknown variable	character
X19	Unknown variable	character

### 2.1 state variable

The micronutrient data uses character values for the **state** variable specifying the name of the states. On review of the **state** names and comparing it with the **state** names in the locality list, we again note inconsistencies with spelling of **states** between the two datasets. Given

that no other state identifiers are used in the micronutrient data, the spelling of the states in the micronutrient data will need to be edited to match the spellings of the states in the locality list.

State names from the micronutrient data

```
## [1] "Northern"      "Central Darfur" "Gezira"         "Kassala"
## [5] "Red Sea"       "River Nile"     "Sennar"         "South Kordofan"
## [9] "West Kordofan" "White Nile"     "East Darfur"    "West Darfur"
## [13] "Khartoum"      "South Darfur"   "Blue Nile"      "Gedaref"
## [17] "North Kordofan" "North Darfur"   NA
```

State names from the full locality list

```
## [1] "Northern"      "River Nile"     "Khartoum"       "Al-Gazeera"
## [5] "Red Sea"       "Kassala"        "Al-Gadarif"     "Sinar"
## [9] "Blue Nile"     "White Nile"     "North Kordofan" "South Kordofan"
## [13] "West Kordofan" "North Darfur"   "South Darfur"   "Central Darfur"
## [17] "East Darfur"   "West Darfur"
```

## 2.2 locality variable

The micronutrient data identifies localities as numeric identifiers. These should in theory match with the locality list. When checked against the locality list, the micronutrient data has data points that have NA value or 0 value for locality ID.

```
unique(mnData$locality)[!unique(mnData$locality) %in% unique(locNames$localityID)]
```

```
## [1] NA 0
```

For the micronutrient data analysis, locality identification is not as important as state identification and PSU identification so this issue is less critical but it is noted here as this is a sign of a poorly organised dataset.

## 2.3 psu variable

The `psu` variable in the micronutrient data identifies the sampling point from which data has been collected. This is a critical variable as this will determine the population weighting to be used on the specific data when state level aggregate estimates are calculated. The `psu` variable in the micronutrient data should be found in the full list of PSUs in the PSU data. We check for this below:

```
unique(mnData$psu)[!unique(mnData$psu) %in% unique(psuData$psu)]
```

```
## [1] NA 2302610 6864 6630 2124111 2123110 188708 620420 3103505
## [10] 3103509 21231 20231
```

and we get 12 issues with PSU identifiers in the micronutrient data not matching any PSU identifier in the PSU data with population. This is a problem as the data associated with these 12 PSU identifiers cannot be properly weighted for micronutrient data analysis.

This requires FMoH and UNICEF Sudan intervention to check on why these PSUs are missing from the PSU dataset and rectify. If these cannot be rectified, the only possible approach will be to drop all data points from the micronutrient data that have these PSU identifiers. If this is done, then 1095 data points will have to be discarded.

## **2.4 m.age variable**

Data for age of mother is character/string values. These will need to be converted to numeric. Assumption is that these values are in years. Need confirmation from FMoH and UNICEF Sudan.

## **2.5 ch.age variable**

Data for age of child is numeric as expected. Assumption is that these values are in months. Need confirmation from FMoH and UNICEF Sudan.

## **2.6 sex variable**

Data for sex of child is in character/string format but with a numerical encoding of 1 and 2. These will be converted to numeric format for ease of coding. Assumption is that these values correspond to 1 = Male and 2 = Female. Need confirmation from FMoH and UNICEF Sudan.

## **2.7 muac variable**

Data for MUAC is in character/string format but with numerical encoding. These will be converted to numeric format. Assumption is that these values are in millimetres. Need confirmation from FMoH and UNICEF Sudan.

## **2.8 ch.weight and ch.height variable**

Data for weight and height is in character/string format but with numerical coding. These will be converted to numeric format. Assumption is that weight values are in kilograms and height values are in centimetres. Need confirmation from FMoH and UNICEF Sudan.

## **2.9 ch.oedema variable**

Data for oedema is in numeric format as below:

```
table(mnData$ch.oedema)
```



### **3 Additional information needed**

Given that no prior indicator definitions for micronutrient analysis has been provided by FMoH UNICEF Sudan despite earlier requests, we again ask that indicator definitions for the micronutrients to be assessed be provided as soon as possible in preparation for the analysis. These indicator definitions should reflect the various groupings to which these indicators pertain.