**My notes for the MERL R (Data exported on 23/12/2024)**

**Meeting agenda and plan**

1. Give general update on the data used documents used as guidelines.
2. Technoserve/ Melissa to give update on the indicator missing and its related data.
3. Go through the scripts and afterwards ask your questions.
4. Allow feedback from team
5. Ask the team for possible update on the survey itself. Eg: inclusion of respondent basic information.
6. Any other business
7. Next steps: incorporate feedback in the work and continue working on the other indicators.

**Questions for the MERL team during the meeting:**

1. Is there any test data in the dataset?
2. Are you generally in the baseline year. I am asking this because, so far 10 submissions are recorded under “follow up”.
3. \*\*\*Do you want any of the responses plotted on a graph?

**General**:

I used the [DQQ](https://drive.google.com/file/d/1eplRm9i5_109-a5Ac1Lqj-lUI3VgVIFx/view) indicator guide in making these calculations.

**Note to self**:

1. Create sub dataset using relevant variables.
2. Clean all relevant data by replacing the string to factors (yes = 1, no & --- = 0)
3. Change data type of numbers that appear as sting into numbers
4. First create the calculation for children and then adults.

Observations & Questions:

1. Are all the respondent’s female? It is not clear from the survey which genders were interviewed. I cannot assume it.
2. Questions did not specifically ask about quantity fruits and veg. It is rather assumed.
3. I did not see the adult-related submissions in the data downloaded/ ODATA Feed. Therefore questions under this section “ler\_\_agora\_gostaria\_de\_lhe\_fazer\_algumas\_perguntas\_de\_sim\_ou\_no\_sobre\_os\_al” do not appear in the survey. I confirmed it by filling a form myself. After the “KFC, Debonairs Pizza, Teka Famba, Burger House, Flaming Dogs or other places that serve pizza or

burgers?” question, the survey moves to Section III. ----- Solved

Child:

I have exempted general indicators that do not directly answers the DQQ questions. Examples are as follows:

1. alguma\_vez\_qual\_\_o\_nome\_da\_criana\_entre\_0\_e\_24\_meses\_foi\_amamentado
2. durante\_os\_primeiros\_2\_dias\_depois\_do\_parto\_deram\_para\_qual\_\_o\_nome\_da\_cria
3. durante\_os\_primeiros\_2\_dias\_depois\_do\_parto\_deram\_para\_qual\_\_o\_nome\_da\_cria
4. o\_leite\_era\_adoado\_ou\_com\_sabor

MDD-W calculation

1. In developing the Vitamin A fruits & vegetables, the survey grouped some vegetables and fruits. The assumption I made was; if one or more of the group contains Vitamin A, I categorized the indicator under Vitamin A.
2. \*According to the DQQ indictor guide, It is difficult to calculate the MDD-W for women between 15-49 years because the data does not provide ages of the respondent.
3. Calculations for **DQQ.**
4. First make calculations for **All-5**. Make calculation such that if there is a “Yes = 1” in any of the starchy foods, it will be a “Yes” for that category. Do same for the other categories. Then calculate the All-5.
5. For the MDD-W food groups, use the DQQ specified indicators to do the calculation.
6. Comment on NCD- Protect score: I have to research which of the vegetables in the questionnaire are Vitamin-A rich. Use specific foods in the DQQ to calculation.
7. For the NCD-Risk score (-), I used indicators from the following questions “amendoim\_manteiga\_de\_amendoim\_amendoim\_modo\_castanhas\_de\_caj\_ou\_sementes\_de” to “refrescos\_como\_coca-cola\_fanta\_frozy\_ou\_fizz\_ou\_bebidas\_energticas\_como\_o\_r”.

**Note**: I could not find the indicator “kfc\_debonairs\_pizza\_teka\_famba\_burger\_house\_flaming\_dogs\_ou\_outros\_lugares\_” in the data exported.

1. Calculations for **“(A2) # of people whose nutritional situation became more resilient to shocks”.** I will use indicators starting from **“missed\_food”.** How to calculate indicator **“Months\_missed\_food”** are as follows.

**Agriculture indicators:**

**General**:

**Questions and observations**:

1. For the **price** per unit sold, the beneficiaries only answered when they answer “Yes” to the previous question of “Didi you sell the product?”. This leads to missing data for the price stated. Is the data still reliable for crop productivity calculation? I calculated for each household level. This leads to missing data as the household which did not sell their crops will get zero for productivity.
2. Labour: There are no Indicator questions on permanent labour.
3. In calculating for total cost, do you expect the use of cost of inputs for labour \*\*\*“form.Section\_IV\_ProductionAreas.Section\_IV\_I\_DroughtSeason.crop\_drought\_season.expenditures\_on\_labor\_dry\_season” or the already calculated expenditure on labour to include family labour”

Missing data:

The price of fish indicator is not present.

**PPP conversion factor**:

The current PPP is 24.95. Source: [IMF Data mapper](https://www.imf.org/external/datamapper/PPPEX@WEO/MOZ?zoom=MOZ&highlight=MOZ)

**Next steps**:

1. Connect the scripts directly to CommCare data source
2. Calculate the Gross PPP per person and not the population average as currently done.

######------------------------------------------------------------------------------------------------------------########

General:

1. The indicator names shared with Nelson from the New App are different from the original names.
2. The team keeps mentioning that the New Long survey is under test. The team should let me know when actual data is collected so I filter the dummy irrelevant data.
3. I have a question, at this stage can the form be updated?

Calculating productivity for agroforestry.

1. The number of hours of work for family members and seasonal workers were not given. I therefore assumed and used 8 hours.

**Note to self**: Calculating the percentage of ALL-5, check how the outcome goes

The R codes debugging.

Calculating the percentage of people consuming a specific food group (meaning the people indicating a score of at least 1 for that food group.

# Select only the relevant columns

food\_data <- df %>%

select(starchy\_staple, vegetables, fruits, pulses)

# Calculate percentage for each category

food\_percentage <- food\_data %>%

summarise(across(everything(), ~ mean(. == 1) \* 100)) %>%

pivot\_longer(cols = everything(), names\_to = "Food Group", values\_to = "Percentage")

# Print the data

print(food\_percentage)

**Plot**

ggplot(food\_percentage, aes(x = Food Group, y = Percentage, fill = Food Group)) + geom\_bar(stat = "identity") + # Use precomputed values labs(title = "Food Consumption Distribution", x = "Food Groups", y = "Percentage of Population") + theme\_minimal() + # Clean style theme(axis.text.x = element\_text(angle = 20, hjust = 1))