RHoMIS Calculations

# 1 About

The Rural Household Multi-Indicator Survey (RHoMIS) is designed to rapidly collect household survey data on smallholder farmers. The survey is rich in information, covering a range of topics from dietary diversity to livelihoods and gendered control of resources. This public dataset has been processed using the RHoMIS R-package.

This document explains:

1. How the R-package was used to process this dataset
2. How it can be used to reprocess this dataset with different inputs
3. The logic of the calculations, and an overview of key indicators/outputs

For examples of how to use the R-package to explore the data, please see the quickstart code (‘analysis\_resources/quickstart.R’)

# 2 Data Processing

RHoMIS data processing takes place in three stages.

1. Unit Extraction: The R-package is used to search through surveys for all unit conversions, crops, and livestock which were encountered. The data processor can examine these units and provide numeric conversion factors so that key indicators can be calculated. For new species of crops and livestock, the user can check all of the entries and standardise (e.g. ensure that there are no duplicates with mispelling, such as “maize” and “maze”).
2. Secondary Cleaning: After the crop and livestock names are standardised (in step 1), the R-package can be used to identify all of the calorie conversions and prices needed for each survey (e.g. based on crops identified in step one we need calorie and prices for “maize”). The units converted in step 1 are used to calculate crop and livestock prices (see [Indicators Explained](#indicators-explained)).
3. Indicator Calculation: Finally, with all conversions and units identified, the R-package calculates a suite of indicators, and outputs, presented in this public dataset.

## 2.1 Installation

To re-calculate the indicators directly from the survey data, you will need to install the RHoMIS R-package. Currently RHoMIS is not available on CRAN, but we hope one day it will be. In the meantime, to install the RHoMIS R-package, you will need to install devtools, you can that by entering the following into the R console:

install.packages("devtools")

Then to install the most recent version of the R-package you can enter the following command:

library(devtools)  
devtools::install\_github("https://github.com/rhomis/rhomis-R-package")

## 2.2 Initial Setup

You will need to make sure you have the survey dataset in the directory you are working in. This is an example directory structure which could be used. If using RStudio interactively, set your working directory to this folder using the setwd() command. Your directory structure should look something like this:

📂 project  
┃  
┣📂 data  
┃ ┗📜 processed\_data.csv  
┃  
┣📂 conversions\_stage\_1  
┃ ┣📜 country.csv  
┃ ┗📜 ...  
┃  
┗📂 conversions\_stage\_2  
 ┣📜 crop\_calories.csv  
 ┣📜 mean\_crop\_pice\_lcu\_per\_kg  
 ┣📜 eggs\_calories.csv  
 ┣📜 mean\_eggs\_price\_per\_kg  
 ┗📜 ...

In this case, your working directory should be the rhomis\_dataverse\_download folder. Please note, if you intend to use the same conversion factors included in this data release, it is necessary to include ‘conversions\_stage\_1’ and ‘conversions\_stage\_2’ in this folder. If not, all unique units and conversions will be extracted from the dataset, and will need to be reconverted.

## 2.3 Stage 1: Unit Extraction and Conversion

### 2.3.1 Extracting Units

To extract units from the dataset, run the code below.

library(rhomis)  
  
extract\_units\_and\_conversions\_csv(  
 file\_path="data/processed\_data.csv",  
 id\_type = "column",  
 proj\_id = "id\_proj",  
 form\_id = "id\_form",

unique\_id\_col = "id\_unique",

hh\_id\_col = "id\_hh",

overwrite = F)

If you did not use conversion factors from the publication, you should be able to see that the conversions\_stage\_1 folder has been added. The conversions\_stage\_1 folder should will include the units which you will verify and change (see below).

### 2.3.2 Converting Units

**Please note, the units in this dataset have already been converted. These conversions were compiled in collaboration with project owners, and those involved in designing the survey. If you feel any of the unit conversions are incorrect, please do not hesitate to get in touch.**

In the project/conversions\_stage\_1 directory, you will find a series of csv files with conversion factors that need to be filled in. Each file will look like the table below. The survey\_value column indicates a value which was found in the survey. The conversion column shows the conversion factor which ought to be used. Where the conversion factor is unknown, the column will display NA. Where the conversion factor is known it will be provided. Known conversion factors are already embedded in the R-package.

| unit\_type | id\_rhomis\_dataset | survey\_value | conversion |
| --- | --- | --- | --- |
| crop\_unit | x | foo | NA |
| crop\_unit | x | bar | NA |

There are a range of conversions to enter and it is important to be aware which units are needed. Each file is described in more detail below:

|  |  |  |  |
| --- | --- | --- | --- |
| File | Description | Example Survey Value | Example Conversion |
| honey\_amount\_to\_l.csv | The conversion factor should be used to convert to kilograms/litres. | buckets\_12\_litre | 12 |
| country\_to\_iso2.csv | The conversion factor should be used to convert country names to [two-letter ISO country codes](https://www.iban.com/country-codes). This is important as it is used to find currency conversion factors | uganda | UG |
| crop\_name\_to\_std.csv | The name of crops entered in the survey. Often enumerators may specify “other” crops in a free-text-entry field. Sometimes these crops can be mispelled, or in a language other than English. Here we correct misspellings and translations into standard forms (**all lower case**) | maze | maize |
| crop\_price\_to\_lcu\_per\_kg.csv | The price units for crops which were sold. Please note that only one unit will be converted into a string, “total\_income\_per\_year” will be converted to “total\_income\_per\_year” as this is treated as a special case in the analysis scripts. | price\_per\_50kg\_sack | 0.02 |
| crop\_amount\_to\_kg.csv | The unit of crops which have been collected. This needs to be converted to kilograms | cart\_250kg | 250 |
| eggs\_price\_to\_lcu\_per\_year.csv | The amount of money made per unit time for selling eggs. This needs to be converted into an income per year | income for 3 months | 4 |
| eggs\_amount\_to\_pieces\_per\_year.csv | The number of eggs collected per year | pieces/day | 365 |
| fertiliser\_amount\_to\_kg.csv | The amount of fertiliser in kg | sacks\_25kg | 25 |
| livestock\_name\_to\_std.csv | The names of livestock entered in the survey. As with crop names, enumerators can also enter extra livestock names which are in a different language or mispelt. This conversion is used to standardise livestock names entered in the survey | catel | cattle |
| milk\_price\_to\_lcu\_per\_l.csv | The amount of money made per unit volume or per unit time. For unit time, the options are “day”, “week”, “month”, “year”. These time units must be entered as strings. For the unit volume, numeric conversions must be entered | month | month |
| 0.3litre | 0.3 |
| milk\_amount\_to\_l.csv | The amount of milk collected in litres per year. There are a number of exceptions which must be kept as text strings, and are dealt with in the analysis scripts (e.g. “per animal per day” and “l/animal/day” | l/day | 365 |
| land\_area\_to\_ha.csv | The amount of land in hectares | acres | 0.4 |

## 2.4 Stage 2: Calculating Prices, Preparing Calorie Conversions, and Extracting Secondary Units

As with unit extraction, ensure you are in the project directory. Then enter the commands below:

library(rhomis)  
  
calculate\_prices\_csv(  
 file\_path="data/processed\_data.csv",  
 id\_type = "column",  
 proj\_id = "id\_proj",  
 form\_id = "id\_form",

unique\_id\_col = "id\_unique",

hh\_id\_col = "id\_hh",

overwrite = F  
)

Again, if you did not already use conversions from this release, you will now see that the conversions\_stage\_2 folder has been added.

### 2.4.1 Converting Calories, Prices, and Other Units

#### 2.4.1.1 Calories and Prices

As with the units which were extracted, price and calorie conversions have been determined with the support of project owners, field staff, and those involved in implementing each survey (in the conversions\_stage\_2 folder). Where field staff were unable to verify prices, estimates were obtained from FAOstat. Prices are be in lcu/kg for crops, meat, and eggs (where lcu is local currency units). Prices are in lcu/l for milk and honey. Prices are in lcu/animal for whole livestock sales. For calorie values, conversions are in kcal/kg or kcal/l.

#### 2.4.1.2 Secondary Units

Aside from prices and calories, there are other conversion tables which depend on conversions from stage 1. For example livestock\_count\_tlu requires us to know the standard livestock names (identified in stage 1). The secondary units currently processed in the R-package are:

|  |  |  |  |
| --- | --- | --- | --- |
| File | Description | Example Survey Value | Example Conversion |
| livestock\_weight\_kg.csv | The conversion factor used to estimate the amount of meat collected from a whole animal. | sheep | 25 |
| livestock\_count\_tlu.csv | “Tropical Livestock Units” (TLU) for individual animals. | cattle | 0.7 |

## 2.5 Stage 3: Calculating Indicators

Using the calories and prices you have converted, you will then be able to produce the final indicators. Either run the code below, or run the 03-calculate-final-indicators.R file.

library(rhomis)  
calculate\_indicators\_local(  
 file\_path="data/processed\_data.csv",  
 id\_type="column",

proj\_id="id\_proj",  
 form\_id = "id\_form",

unique\_id\_col = "id\_unique",

hh\_id\_col = "id\_hh",

overwrite = F)

You will now see that the directory structure again looks quite different. Extra folders and files have been created. There is a lot of information here, for more explanation, please see [an explanation of indicators and outputs](#indicators-explained)

📂 project  
┃  
┣📂 data  
┃ ┗📜 processed\_data.csv  
┃  
┣📂 conversions\_stage\_1  
┃ ┣📜 country.csv  
┃ ┗📜 ...  
┃  
┣📂 conversions\_stage\_2  
┃ ┣📜 crop\_calories.csv  
┃ ┣📜 mean\_crop\_pice\_lcu\_per\_kg  
┃ ┗📜 ...  
┃  
┣📂 crop\_data  
┃ ┣📜 crop\_consumed\_kg\_per\_year.csv  
┃ ┗📜 ...  
┃  
┣📂 indicator\_data  
┃ ┗📜 indicators.csv  
┃  
┣📂 livestock\_data  
┃ ┣📜 crop\_consumed\_kg\_per\_year.csv  
┃ ┗📜 ...  
┃  
┣📂 off\_farm\_data  
┃ ┣📜 offfarm\_income\_name.csv  
┃ ┗📜 ...  
┃  
┗📂 processed\_data  
┃ ┗📜 processed\_data.csv  
┃  
┣📂 consumption\_calorie\_values  
┃ ┣📜 crop\_calories\_consumed\_kcal.csv  
┃ ┗📜 ...  
┃  
┗📂 extra\_outputs  
 ┣📂 consumption\_calorie\_values  
 ┣📂 consumption\_lcu\_values  
 ┗📂 gender\_control

# 3. Main Principle in Calculations

RHoMIS is designed to collect information on wide range of topics in a short amount of time. Questions in the survey are asked in a way that we can make estimates, whilst not forcing the respondent to provide a level of precision they are not able to.

To calculate indicators such as income, productivity, gendered control of resources, we rely on proportions and recall. This is best explained using worked examples (see below)

### 3.1.1 Worked Example

When calculating crop productivity, respondents are asked to give the amount harvested over the past 12 months, they are then asked to give the unit of this amount (e.g. ‘kg’, ‘tons’, ‘gorogoro’ etc.). By multiplying the two it is possible to calculate the number of kilograms harvested over the past 12 months (let’s label this ).

Respondents are also asked how much income they gain from the sale of this crop and the units of this income (e.g. ‘total income per year’, ‘price per kg’, ‘price per gorogoro’ etc.). This can either be used to calculate total income per year or the income per kg (let’s label these and respectively).

Using these variables it is necessary to calculate: number of kg sold (), number of kg consumed (), price per kilo (), income per year (), value controlled by the man (), value controlled by the woman () and the values of the crops which were consumed (). These calculations mainly rely on proportion responses. The respondent is asked what proportion of the harvest was consumed () and what proportion was sold (), where none=0, little=0.1, under half=0.2, half=0.5, most=0.7, all=0.9. Therefore:

Using the amount sold () it is possible to work out income per kg and price per kilo () and income per year (), depending on which sale units were given by the respondent:

The price per kg is then averaged across the whole project and verified against FAOstat prices, or verified by field staff. Please not this is step 2 of the indicator calculations. Price verification then makes it possible to calculate the value of crop consumed ():

Finally, the value controlled by the female () and the male () in the household can be calculated. The respondent is asked what proportion of the crop consumed is controlled by the male () or the woman (), where none=0, little=0.1, under half=0.2, half=0.5, most=0.7, all=0.9. The respondent is then asked what proportion of the income from the crop sale is controlled by the male () or the woman (). Which finally gives:

**This style of proportion calculation is used throughout the indicator calculations, for further details please see individual explanations of each indicator included in the ‘documentation/codebook.xls’ file.**

# 4 Outputs and Data

If you have used the R-package to re-process indicators, additional folders containing outputs on specific topics are produced (such as crop data, gender, or livestock, see section 2). The contents of each folder are explained in the table below. These outputs can also be found in the processed\_data.csv file (see documentation/codebook.xls). Please note each row in these extra outputs correspond to a single observation, as in the processed\_data.

For every file, each row represents one household (in the same row order as the original dataset). The variable measured and corresponding unit are contained within the file name. For example (crop\_data/crop\_harvest\_kg\_per\_year.csv corresponds to the amount of a crop harvested in kg/year). The name of the indicator (e.g. crop\_harvest\_kg\_per\_year) corresponds to the name of an indicator described in the documentation/codebook.csv file.

# 5 Reporting Issues

Please report any problems you encounter or feature requests in the issues section of [this repo](https://github.com/RHoMIS/rhomis-R-package/issues).