📙 Malaria case study - Part 1

Table of contents

## 1 Introduction

### 1.1 Overview

These pages will demonstrate how to use Quarto to data from Tanzania.

### 1.2 Learning objectives

* Test

## 2 Malaria research questions

Research questions in an enumerated list

## 3 Getting started

### 3.1 Access the Quarto template

Download the Quarto template used for this case study (add link) using GitHub.

#### 3.1.1 Quarto resources

Review the R Basics page “Reading a RMarkdown Script”, or reference these resources: Rmarkdown introduction and Rmarkdown authoring basics.

```{r}  
install.packages("ggplot2")  
install.packages("ggthemes")  
install.packages("networkD3")  
install.packages("apyramid")  
```

### 3.2 Dataset description

We will be using data and examples from a real consultation data which occurred in Tanzania between **2021-07-29** and **2021-12-17** within the Integrated Management of Childhood Illness (TIMCI) project.

|  |
| --- |
| Important |
| Data are made available by the Ifakara Health Institute (IHI) for training purposes only. Please note, that some data has been adapted in order to best achieve training objectives. No personally indentifiable information have been kept in this dataset. |

Information about the consultations of **10,308 children [1 day - 59 months]** from **18 facilities** (dispensaries and health centres) in **Kaliua** District, **Sengerema** District and **Tanga** District, Tanzania.

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | (a) | | |  | | --- | | (b) | |

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | (c) | | |  | | --- | | (d) | |

Figure 1: Study area

### 3.3 Data collection

Data were collected using *ODK* (ODK Collect, ODK Central) between **2021-07-29** and **2021-12-17**. Research assistants recorded the following information from different sources.

Types and sources of information

| Information | Prefix | Source |
| --- | --- | --- |
| Context | CTX | Metadata |
| Sociodemographics | SDC | Caregiver |
| Clinical presentation | CLIN | Caregiver |
| Laboratory investigations | TEST | Child booklet or facility MTUHA book |
| Diagnoses | DX | Child booklet or facility MTUHA book |
| Treatments | RX |  |
| * before consultation |  | Caregiver |
| * as a result of the consultation |  | Child booklet or facility MTUHA book |
| Referral advice | MGMT |  |
|  |  | Caregiver |
|  |  | Child booklet or facility MTUHA book |

### 3.4 Data preparation

Data cleaning and data de-identification

Personally identifiable information (PII) were removed.

### 3.5 Import the data

Data are stored in **dataset1.dta**.

#### ✏️ Exercise 1

1. Import the dataset and store it into a dataframe called **df**.
2. Select columns **child\_ID**, **CTX\_month**, **CTX\_district**, **SDC\_age\_in\_months** for the first **5** rows in data frame df.
3. Display in a table with the following characteristics:
   1. Align the content of the first column to the left and the content of the other 3 columns to the centre.
   2. Add a caption.

Extract of the database

| child\_ID | CTX\_month | CTX\_district | SDC\_age\_in\_months |
| --- | --- | --- | --- |
| 1 | 7 | Kaliua | 10 |
| 2 | 7 | Kaliua | 6 |
| 3 | 7 | Kaliua | 6 |
| 4 | 7 | Kaliua | 11 |
| 5 | 8 | Kaliua | 21 |

## 4 Population characteristics

### 4.1 Codebook

| Variable | Coding |  |
| --- | --- | --- |
| SDC\_age\_in\_months | | | |
| SDC\_sex | 1: male 2: female 98: unknown |  |
| CLIN\_fever | 0: no 1: yes 98: not sure |  |
| CLIN\_fever\_onset |  |  |
| CLIN\_cough | 0: no 1: yes 98: not sure |  |
| CLIN\_diarrhoea | 0: no 1: yes 98: not sure |  |
| RX\_preconsult\_antibiotics |  |  |
| RX\_preconsult\_antimalarials |  |  |
| CTX\_district | Kaliua  | | Sengerema  | | Tanga | | | |
| CTX\_area | urban  | | rural | | | |
| CTX\_facility\_type | dispensary  | | health centre | | | |

### 4.2 Structure of the data

#### ✏️ Exercise 2

Add the following two new variables to data frame df

| Variable | Coding |
| --- | --- |
| SDC\_age\_category | <2 months 2-11 months 12-23 months 24-35 months 36-47 months 48-59 months |
| CLIN\_fever\_onset\_category | <2 days 2-3 days 4-6 days ≥ 7 days |

|  |
| --- |
| Tip |
| * Stata: use the gen command * R: use the [mutate](https://dplyr.tidyverse.org/reference/mutate.html) and [case\_when](https://dplyr.tidyverse.org/reference/case_when.html) functions from the dplyr package |

#### ✏️ Exercise 3

Examine the structure of the data, including variable names, labels.

|  |
| --- |
| Tip |
| * Stata: use the [codebook](https://www.stata.com/manuals/dcodebook.pdf) command * R: use the [skim](https://docs.ropensci.org/skimr/reference/skim.html) function from the skimr package |

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 10308 |
| Number of columns | 53 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 10 |
| numeric | 1 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| SDC\_sex | 4 | 1.0 | FALSE | 2 | 1: 5229, 2: 5075 |
| CLIN\_fever | 0 | 1.0 | FALSE | 3 | 1: 7225, 0: 3068, 98: 15 |
| CLIN\_fever\_onset | 3083 | 0.7 | FALSE | 15 | 2: 2469, 1: 1997, 3: 1917, 4: 337 |
| CLIN\_diarrhoea | 0 | 1.0 | FALSE | 3 | 0: 7982, 1: 2306, 98: 20 |
| CLIN\_cough | 0 | 1.0 | FALSE | 3 | 1: 5635, 0: 4658, 98: 15 |
| RX\_preconsult\_antibiotics | 0 | 1.0 | FALSE | 2 | 0: 8573, 1: 1735 |
| RX\_preconsult\_antimalarials | 0 | 1.0 | FALSE | 2 | 0: 9866, 1: 442 |
| CTX\_district | 0 | 1.0 | FALSE | 3 | Tan: 5176, Sen: 2703, Kal: 2429 |
| CTX\_area | 0 | 1.0 | FALSE | 2 | urb: 6220, rur: 4088 |
| CTX\_facility\_type | 0 | 1.0 | FALSE | 2 | dis: 5599, hea: 4709 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SDC\_age\_in\_months | 0 | 1 | 18.75 | 14.9 | 0 | 7 | 15 | 27 | 59 | ▇▆▃▂▁ |

#### ✏️ Exercise 4

Display descriptive statistics for the population characteristics.

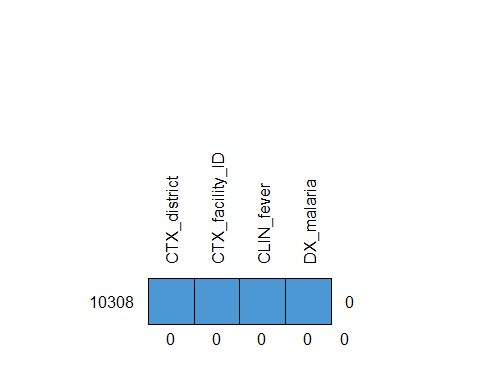
|  |
| --- |
| Tip |
| * Stata * R: use the [tbl\_summary](https://www.danieldsjoberg.com/gtsummary/reference/tbl_summary.html) function from the gtsummary package |

| **Characteristic** | **N = 10,308** |
| --- | --- |
| SDC\_age\_category |  |
| <2 months | 597 (5.8%) |
| 02-11 months | 3,576 (35%) |
| 12-23 months | 2,947 (29%) |
| 24-35 months | 1,529 (15%) |
| 36-47 months | 980 (9.5%) |
| 48-59 months | 679 (6.6%) |
| SDC\_sex |  |
| 1 | 5,229 (51%) |
| 2 | 5,075 (49%) |
| (Missing) | 4 |
| CLIN\_fever |  |
| 0 | 3,068 (30%) |
| 1 | 7,225 (70%) |
| 98 | 15 (0.1%) |
| CLIN\_fever\_onset\_category |  |
|  | 3,083 (30%) |
| <2 days | 1,998 (19%) |
| >= 7 days | 343 (3.3%) |
| 2-3 days | 4,386 (43%) |
| 4-6 days | 498 (4.8%) |
| CLIN\_diarrhoea |  |
| 0 | 7,982 (77%) |
| 1 | 2,306 (22%) |
| 98 | 20 (0.2%) |
| CLIN\_cough |  |
| 0 | 4,658 (45%) |
| 1 | 5,635 (55%) |
| 98 | 15 (0.1%) |
| RX\_preconsult\_antibiotics |  |
| 0 | 8,573 (83%) |
| 1 | 1,735 (17%) |
| RX\_preconsult\_antimalarials |  |
| 0 | 9,866 (96%) |
| 1 | 442 (4.3%) |
| CTX\_district |  |
| Kaliua | 2,429 (24%) |
| Sengerema | 2,703 (26%) |
| Tanga | 5,176 (50%) |
| CTX\_area |  |
| rural | 4,088 (40%) |
| urban | 6,220 (60%) |
| CTX\_facility\_type |  |
| dispensary | 5,599 (54%) |
| health center | 4,709 (46%) |

### 4.3 Missing values

Identify missing values in each variable: missing\_plot

/\ /\  
{ `---' }  
{ O O }  
==> V <== No need for mice. This data set is completely observed.  
 \ \|/ /  
 `-----'



CTX\_district CTX\_facility\_ID CLIN\_fever DX\_malaria   
10308 1 1 1 1 0  
 0 0 0 0 0

## 5 Healthcare provider actions

### 5.1 Codebook

* Temperature measured
  + Fever measured
* Fever (temp or history)
* Malaria test
* Any severe diagnosis
* Malaria diagnosis
* Malaria treatment
* Referral

| Variable | Coding |
| --- | --- |
| MEAS\_temperature |  |
|  |  |
| TEST\_malaria\_result | 0: negative 1: positive 2: indeterminate 95: unreadable result 98: not sure |
| DX\_malaria | 0: no 1: yes |
| RX\_antimalarials | 0: no 1: yes |
| MGMT\_referral\_src\_caregiver |  |
| MGMT\_referral\_src\_registry |  |

### 5.2 Structure of the data

#### 5.2.1 ✏️ Exercise 5

Examine the structure of the data, including variable names, labels.

|  |
| --- |
| Tip |
| * Stata: use the [codebook](https://www.stata.com/manuals/dcodebook.pdf) command * R: use the [skim](https://docs.ropensci.org/skimr/reference/skim.html) function from the skimr package |

#### 5.2.2 R

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 10308 |
| Number of columns | 53 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 1 |
| numeric | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| TEST\_malaria\_done | 0 | 1 | FALSE | 2 | 1: 5763, 0: 4545 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MEAS\_temperature | 9271 | 0.10 | 37.08 | 0.98 | 34.5 | 36.5 | 37 | 37.5 | 42.5 | ▃▇▃▁▁ |
| TEST\_malaria\_result | 4550 | 0.56 | 1.20 | 9.93 | 0.0 | 0.0 | 0 | 0.0 | 98.0 | ▇▁▁▁▁ |
| DX\_malaria | 0 | 1.00 | 0.17 | 0.38 | 0.0 | 0.0 | 0 | 0.0 | 1.0 | ▇▁▁▁▂ |
| RX\_antimalarials | 0 | 1.00 | 0.13 | 0.33 | 0.0 | 0.0 | 0 | 0.0 | 1.0 | ▇▁▁▁▁ |
| MGMT\_referral\_src\_caregiver | 0 | 1.00 | 0.22 | 4.50 | 0.0 | 0.0 | 0 | 0.0 | 98.0 | ▇▁▁▁▁ |
| MGMT\_referral\_src\_registry | 0 | 1.00 | 0.01 | 0.10 | 0.0 | 0.0 | 0 | 0.0 | 1.0 | ▇▁▁▁▁ |

#### 5.2.3 ✏️ Exercise 3

Add the following two new variables to data frame df

* MEAS\_fever
* Fever (temp or history)

|  |
| --- |
| Tip |
| * Stata: use the gen command * R: use the mutate function from the dplyr package |

#### 5.2.4 ✏️ Exercise 6

Display descriptive statistics for the following healthcare provider actions:

|  |
| --- |
| Tip |
| * R: use the [tbl\_summary](https://www.danieldsjoberg.com/gtsummary/reference/tbl_summary.html) function from the gtsummary package |

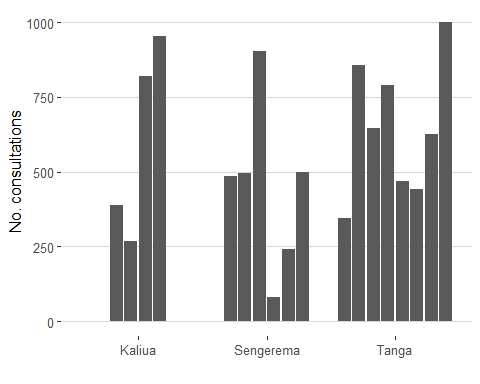
| **Characteristic** | **N = 10,308** |
| --- | --- |
| CALC\_temperature\_measured | 1,037 (10%) |
| CALC\_fever | 326 (31%) |
| (Missing) | 9,271 |
| CALC\_fever\_or\_temp | 7,252 (97%) |
| (Missing) | 2,842 |
| TEST\_malaria\_done |  |
| 0 | 4,545 (44%) |
| 1 | 5,763 (56%) |
| TEST\_malaria\_result |  |
| 0 | 4,665 (81%) |
| 1 | 1,032 (18%) |
| 2 | 1 (<0.1%) |
| 95 | 3 (<0.1%) |
| 98 | 57 (1.0%) |
| (Missing) | 4,550 |
| DX\_malaria | 1,800 (17%) |
| RX\_antimalarials | 1,290 (13%) |
| MGMT\_referral\_src\_caregiver |  |
| 0 | 10,122 (98%) |
| 1 | 164 (1.6%) |
| 97 | 9 (<0.1%) |
| 98 | 13 (0.1%) |
| MGMT\_referral\_src\_registry | 114 (1.1%) |

## 6 Number of consultations by facility

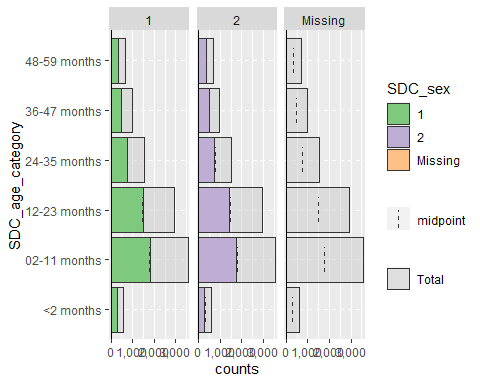
### 6.1 ✏️ Exercise 6

Plot the number of consultations by facility in bars, grouped by district.

|  |
| --- |
| Tip |
| * Stata: * R: |

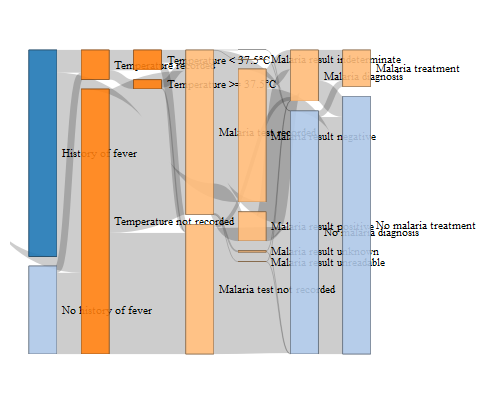


### 6.2 Age pyramid



## 7 Fever assessment

Comparison of systems effectiveness decay for malaria assessment and treatment in facilities. Steps reflecting adherence to guidelines are shown in blue, non-adherence shown in red, and not applicable in grey.



### 7.1 Facility F4153



### 7.2 Facility F9222



Sequence

* temp measurement by reported fever; by facility

Pearson's Chi-squared test  
  
data: df$CLIN\_fever and df$TEST\_malaria\_result  
X-squared = 61.449, df = 8, p-value = 2.42e-10

| **Characteristic** | **F5477**, N = 267 | **F9514**, N = 821 | **F9551**, N = 954 | **F7476**, N = 240 | **F4199**, N = 496 | **F0554**, N = 344 | **F4000**, N = 470 | **F1116**, N = 645 | **F1604**, N = 790 | **F5674**, N = 441 | **F0938**, N = 858 | **F7223**, N = 1,001 | **F3621**, N = 387 | **F5650**, N = 81 | **F3844**, N = 486 | **F9784**, N = 498 | **F4965**, N = 902 | **F6589**, N = 627 | **p-value** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MEAS\_temperature | 37.50 (37.00, 37.50) | NA (NA, NA) | 39.50 (39.25, 39.75) | 36.50 (36.50, 37.00) | 36.50 (36.00, 37.00) | 37.00 (37.00, 37.00) | 38.50 (37.00, 38.50) | 37.00 (36.50, 38.00) | 37.25 (37.00, 38.88) | 37.00 (37.00, 38.00) | 37.00 (37.00, 37.50) | 38.00 (37.62, 38.38) | 38.25 (38.12, 38.38) | 39.00 (38.25, 39.50) | 37.00 (36.50, 38.00) | 36.50 (36.50, 37.50) | 37.00 (36.50, 37.50) | 36.00 (36.00, 37.00) | <0.001 |
| Unknown | 210 | 821 | 952 | 86 | 467 | 338 | 445 | 586 | 776 | 202 | 837 | 995 | 385 | 78 | 322 | 297 | 880 | 594 |  |

* also showing ‘prevalence’ of fever when of whole clinic vs of those who measure to indicate bias

## 8 Malaria tests

* malaria tests of those with history or measured fever

0 1 2 95 98  
 FALSE 70 2 0 1 1  
 TRUE 3952 966 1 2 47

Pearson's Chi-squared test  
  
data: df$CALC\_fever\_or\_temp and df$TEST\_malaria\_result  
X-squared = 33.92, df = 4, p-value = 7.74e-07

| **Characteristic** | **F5477**, N = 190 | **F9514**, N = 581 | **F9551**, N = 780 | **F7476**, N = 197 | **F4199**, N = 406 | **F0554**, N = 209 | **F4000**, N = 358 | **F1116**, N = 467 | **F1604**, N = 491 | **F5674**, N = 301 | **F0938**, N = 550 | **F7223**, N = 603 | **F3621**, N = 270 | **F5650**, N = 62 | **F3844**, N = 358 | **F9784**, N = 330 | **F4965**, N = 683 | **F6589**, N = 416 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TEST\_malaria\_result |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0 | 83 (67%) | 151 (33%) | 319 (54%) | 125 (82%) | 285 (75%) | 135 (98%) | 159 (95%) | 371 (93%) | 226 (95%) | 213 (96%) | 221 (97%) | 346 (97%) | 21 (25%) | 32 (57%) | 242 (92%) | 225 (91%) | 511 (90%) | 287 (97%) |
| 1 | 38 (31%) | 300 (66%) | 270 (46%) | 28 (18%) | 95 (25%) | 0 (0%) | 2 (1.2%) | 25 (6.3%) | 7 (2.9%) | 4 (1.8%) | 6 (2.6%) | 6 (1.7%) | 59 (69%) | 20 (36%) | 20 (7.6%) | 21 (8.5%) | 55 (9.7%) | 10 (3.4%) |
| 2 | 0 (0%) | 0 (0%) | 1 (0.2%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| 95 | 1 (0.8%) | 0 (0%) | 1 (0.2%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| 98 | 2 (1.6%) | 5 (1.1%) | 1 (0.2%) | 0 (0%) | 0 (0%) | 3 (2.2%) | 6 (3.6%) | 2 (0.5%) | 6 (2.5%) | 6 (2.7%) | 0 (0%) | 3 (0.8%) | 5 (5.9%) | 4 (7.1%) | 0 (0%) | 1 (0.4%) | 3 (0.5%) | 0 (0%) |
| Unknown | 66 | 125 | 188 | 44 | 26 | 71 | 191 | 69 | 252 | 78 | 323 | 248 | 185 | 6 | 96 | 83 | 114 | 119 |

## 9 Malaria treatments

* malaria diagnoses vs. positive tests vs. treatment.

| Var1 | Var2 | Var3 | Freq |
| --- | --- | --- | --- |
| 0 | 0 | 0 | 3890 |
| 1 | 0 | 0 | 15 |
| 2 | 0 | 0 | 1 |
| 95 | 0 | 0 | 2 |
| 98 | 0 | 0 | 31 |
| 0 | 1 | 0 | 645 |
| 1 | 1 | 0 | 49 |
| 2 | 1 | 0 | 0 |
| 95 | 1 | 0 | 1 |
| 98 | 1 | 0 | 17 |
| 0 | 0 | 1 | 99 |
| 1 | 0 | 1 | 50 |
| 2 | 0 | 1 | 0 |
| 95 | 0 | 1 | 0 |
| 98 | 0 | 1 | 2 |
| 0 | 1 | 1 | 31 |
| 1 | 1 | 1 | 918 |
| 2 | 1 | 1 | 0 |
| 95 | 1 | 1 | 0 |
| 98 | 1 | 1 | 7 |

Pearson's Chi-squared test  
  
data: df$TEST\_malaria\_result and df$DX\_malaria  
X-squared = 2582, df = 4, p-value < 2.2e-16

Pearson's Chi-squared test  
  
data: df$TEST\_malaria\_result and df$RX\_antimalarials  
X-squared = 4508.8, df = 4, p-value < 2.2e-16

| **Characteristic** | **F5477**, N = 41 | **F9514**, N = 320 | **F9551**, N = 288 | **F7476**, N = 28 | **F4199**, N = 105 | **F0554**, N = 0 | **F4000**, N = 2 | **F1116**, N = 25 | **F1604**, N = 7 | **F5674**, N = 4 | **F0938**, N = 6 | **F7223**, N = 9 | **F3621**, N = 61 | **F5650**, N = 21 | **F3844**, N = 22 | **F9784**, N = 21 | **F4965**, N = 59 | **F6589**, N = 13 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DX\_malaria | 40 (98%) | 311 (97%) | 285 (99%) | 27 (96%) | 71 (68%) | 0 (NA%) | 2 (100%) | 17 (68%) | 5 (71%) | 4 (100%) | 6 (100%) | 6 (67%) | 59 (97%) | 21 (100%) | 22 (100%) | 21 (100%) | 57 (97%) | 13 (100%) |
| RX\_antimalarials | 40 (98%) | 288 (90%) | 283 (98%) | 26 (93%) | 104 (99%) | 0 (NA%) | 2 (100%) | 23 (92%) | 5 (71%) | 2 (50%) | 4 (67%) | 5 (56%) | 57 (93%) | 21 (100%) | 21 (95%) | 21 (100%) | 55 (93%) | 11 (85%) |