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

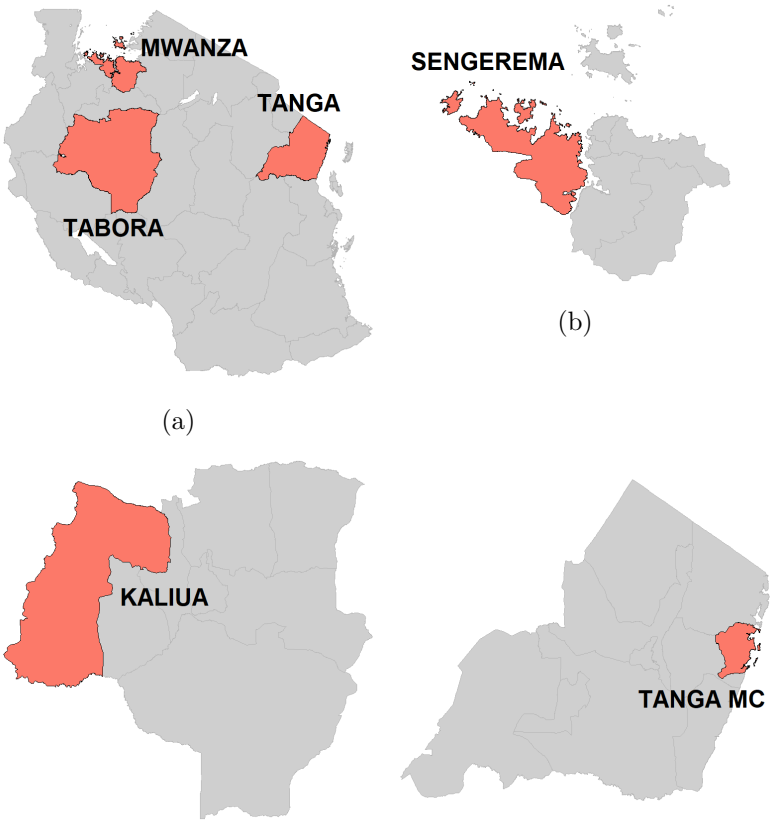
We will be using data and examples from a real consultation data which occurred in Tanzania between **2021-07-29** and **2021-12-17** within the Integrated Management of Childhood Illness (TIMCI) project.

> **!** 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.

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.



(a) (b)

(c) (d)

Figure 1: Study area

Table 1: 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:

a. Align the content of the first column to the left and the content of the other 3 columns to the centre.
b. Add a caption.

Table 2: 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 |	

Variable	Coding
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 and case_when functions from the `dplyr` package

Exercise 3

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

> **Tip**
>
> - Stata: use the codebook command
> - R: use the skim function from the `skimr` package

Table 5: 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_0in_months	1	1	18.75	14.90	7	15	27	59		

Exercise 4

Display descriptive statistics for the population characteristics.

> **Tip**
>
> - Stata
> - R: use the tbl_summary function from the `gtsummary` package

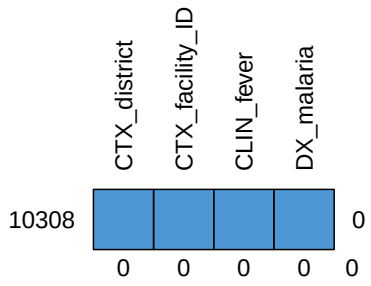
Characteristic	**N = 10,308**
SDC_age_category	
<2 months	597 (5.8%)
02-11 months	3,576 (35%)
12-23 months	2,947 (29%)
24-35 months	1,529 (15%)
36-47 months	980 (9.5%)
48-59 months	679 (6.6%)
SDC_sex	
1	5,229 (51%)
2	5,075 (49%)
(Missing)	4
CLIN_fever	
0	3,068 (30%)
1	7,225 (70%)
98	15 (0.1%)
CLIN_fever_onset_category	
	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%)

Characteristic	**N = 10,308**
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	

Variable	Coding
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 command
> - R: use the skim function from the `skimr` package

5.2.2 R

Table 10: Data summary

Name	Piped data
Number of rows	10308
Number of columns	53
Column type frequency:	
factor	1
numeric	6

Table 10: Data summary

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	927	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_doc_category	0	1.00	0.22	1.50	0.0	0.0	0	0.0	98.0	
MGMT_referral_doc_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 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%)

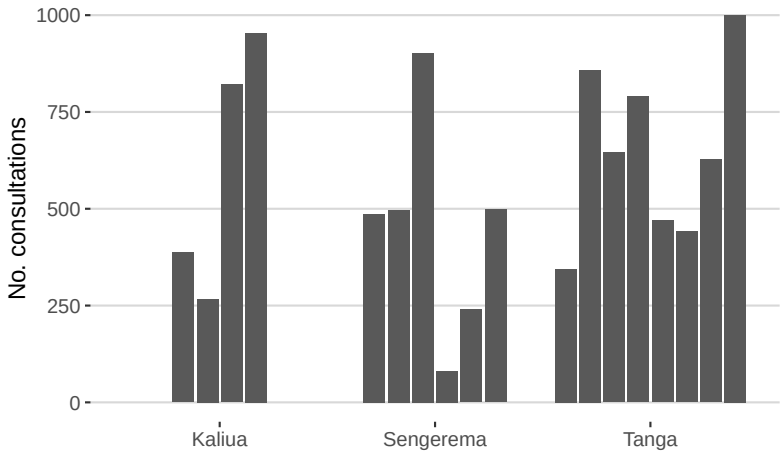
6 Number of consultations by facility

6.1 Exercise 6

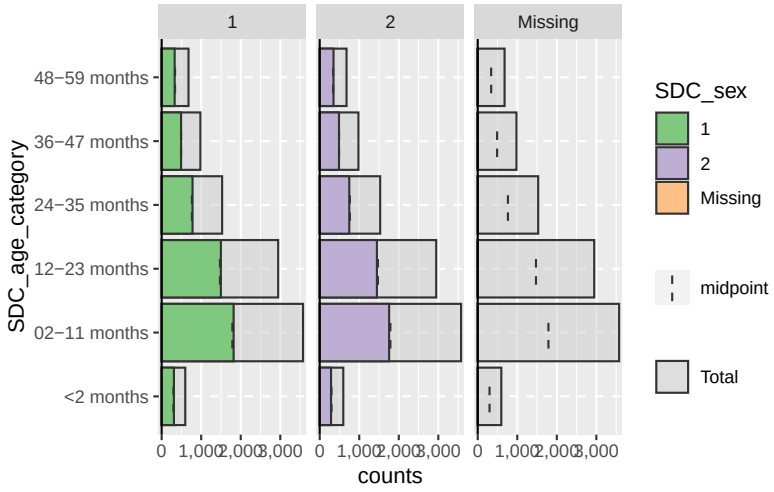
Plot the number of consultations by facility in bars, grouped by district.

Tip

- Stata:
- R:

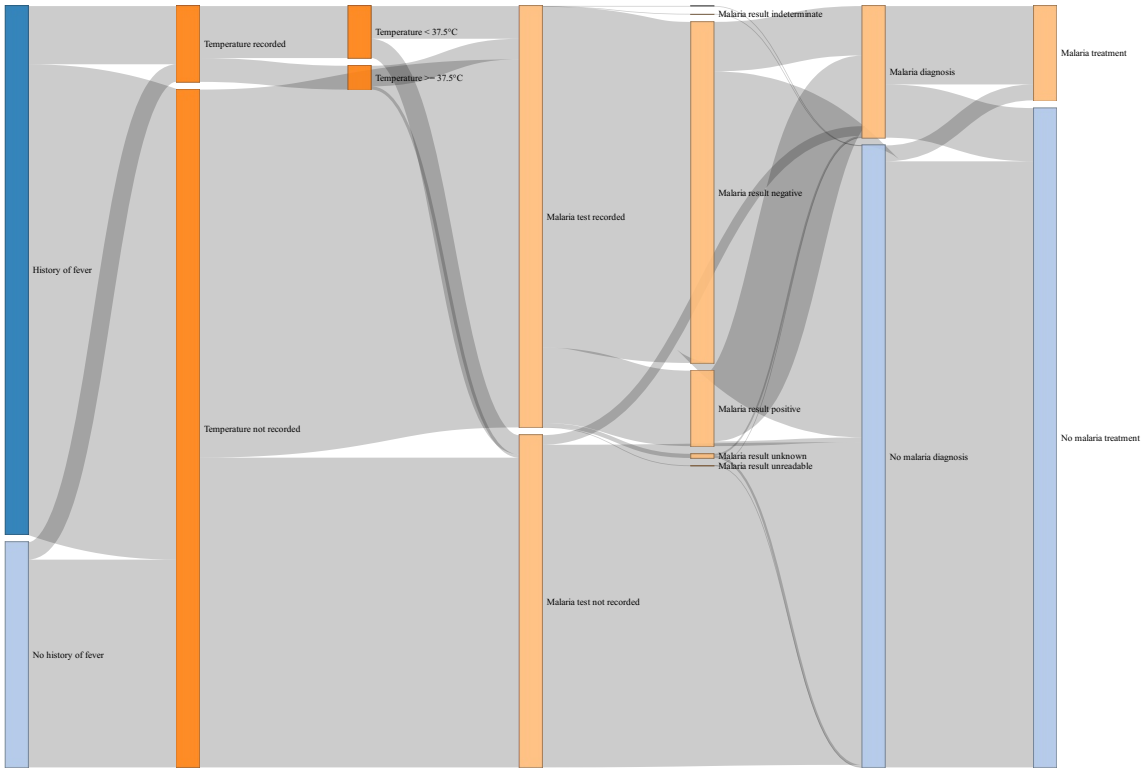


6.2 Age pyramid



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	[illegible]	p-value
MEAN_Temperature	[illegible]	<0.001
Unknown	[illegible]	

- 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	[illegible]	p-value
TEST_malaria_result		
0	83	

Characteristic	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
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

Var1	Var2	Var3	Freq
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	[illegible], N = 41	[illegible], N = [illegible]	[illegible], N = [illegible]	[illegible], N = [illegible]	[illegible], N = 105	[illegible], N = 0	[illegible], N = 2	[illegible], N = 25	[illegible], N = 7	[illegible], N = 4	[illegible], N = 6	[illegible], N = 9	[illegible], N = 61	[illegible], N = 21	[illegible], N = 22	[illegible], N = 21	[illegible], N = 59	[illegible], N = 13
DX_malaria	40 (98%)	[illegible] (97%)	[illegible] (99%)	[illegible] (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%)	[illegible] (90%)	[illegible] (98%)	[illegible] (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%)