

Plausibility bounds

In this document we describe the plausible values that are considered for each of the variables used in the imputation model. In addition we provide the rules that we applied on the creation of additional variables required for the imputation model.

Plausibility bounds

The following table describes the plausibility bounds for the continuous variables and the time-related variables used in the imputation process. There you can see for each variable, the classification, the units and the plausibility bounds (minimum, maximum). We plan to set all values outside the plausible bounds as NA.

| Classification | Variable name | Variable type | Units | Min | Max |
|----------------|-----------------------|---------------|-----------------|-----|------|
| Pwoman | age | Continuous | (years) | 10 | 55 |
| Pwoman | weight | Continuous | (Kg) | 30 | 200 |
| Pwoman | pre_pregweight | Continuous | (Kg) | 30 | 200 |
| Pwoman | height | Continuous | (cm) | 100 | 210 |
| Exposure | zikh_pcr_ga_1 | Time | (weeks) | 0 | 46 |
| Exposure | zikh_pcr_vl_1 | Continuous | (copies/microL) | 0 | Inf |
| Exposure | zikh_elisa_ga_1 | Time | (weeks) | 0 | 46 |
| Exposure | zikh_ga | Time | (weeks) | 0 | 46 |
| Exposure | symp_ga | Time | (weeks) | 0 | 46 |
| Exposure | arb_clindiag_ga | Time | (weeks) | 0 | 46 |
| Outcome | miscarriage_ga | Time | (weeks) | 0 | 20 |
| Outcome | loss_ga | Time | (weeks) | 20 | 46 |
| Outcome | endga | Time | (weeks) | 0 | 46 |
| Outcome | birth_ga | Time | (weeks) | 0 | 46 |
| Outcome | inf_weight | Continuous | (grams) | 100 | 6000 |
| Outcome | inf_length | Continuous | (cm) | 18 | 50 |
| Outcome | inf_head_circ_birth | Continuous | (cm) | 0 | Inf |
| Outcome | inf_head_circ_fu1 | Continuous | (cm) | 0 | Inf |
| Outcome | inf_head_circ_age_fu1 | Time | (months) | 0 | Inf |
| Outcome | inf_head_circ_fu2 | Continuous | (cm) | 0 | Inf |
| Outcome | inf_head_circ_age_fu2 | Time | (months) | 0 | Inf |
| Outcome | inf_head_circ_fu3 | Continuous | (cm) | 0 | Inf |
| Outcome | inf_head_circ_age_fu3 | Time | (months) | 0 | Inf |

By running this code, the “obs_to_check.RData” file is automatically generated with the observations outside the plausibility range. It can be found in the “3_Output_data” file. Inside the obs_to_check dataset, there is an additional column called “Var_to_check” that indicates the variable with possible implausible values for each observation.

Logic rules for exposures

Set to NA values

SAS and R manage missing values differently. To make the imputation process easier, we set to NA all values equal to 666,777,888,999. Except for the arb_clindiag variable, for which “777” values were assigned to a “6” level as it is used later on to define the CZS variable.

Zika test time of measurement (zikv_ga)

In case the zikv_ga value is not specified for an observation, but the time of measurement for the PCR “zikv_pcr_ga_1” or the elisa test “zikv_elisa_ga_1” was specified, we set the zikv_ga value with the date of measurement of the earliest test.

if is.na (zikv_ga) then zikv_ga = min (zikv_pcr_ga_1,zikv_elisa_ga_1)

Creation of new variables

For the following variables, we combine different variables in one general variable x_{com} in order to include the information in a simple way that helps us to avoid any imputation convergence problem. For instance if we want to create a new Xm variable that combines information of x_1, x_2, \dots, x_n . If for any of this x_i variables with any response, there is the presence of a characteristic, outcome or a positive test $x_i == 1$, then the new variable x_{com} is assigned with overall presence.

$$if(any(x_1 == 1, x_2 == 1, ..x_n = 1)), then, x_{com} = 1$$

In case that no presence is found for all the x_i variables with any response then the x_{com} is assigned to no presence.

$$if(all(x_1 == 0, x_2 == 0, ..x_n = 0)), then, x_{com} = 0$$

If for a certain patient, there was no information on any of these x_i variables then the observation is assigned to NA.

$$if(all(is.na(x_1), is.na(x_2), ..is.na(x_n = 1)), then, x_{com} = NA$$

storch__patho

Denotes the Intrauterine exposure to storch pathogens. It was created by combining the following variables with the previous explained rule.

- ifelse(storch==0,0,ifelse(!is.na(storch),1,0))]
- storch_bin
- toxo
- toxo_treat
- syphilis
- syphilis_treat
- varicella
- parvo
- rubella
- cmv
- herpes

- listeria
- chlamydia
- gonorrhea
- genitalwarts

arb__ever

It refers to the presence of a prior arb virus infection, and it combines the following variables:

- `ifelse(zikv_pcr_everpos==1,1,ifelse(zikv_pcr_everpos==0,0,NA))]`
- `zikv_elisa_everpos`
- `denv__ever`
- `chikv__ever`

flavi__alpha__virus

This variable denotes the presence of concurrent or prior flavi- or alpha virus infection, and we used the following variables to create it.

- `ifelse(arb_clindiag_plus==0,0,ifelse(!is.na(arb_clindiag_plus),1,0))`
- `ifelse(arb_clindiag!=0&arb_clindiag!=1,0,ifelse(!is.na(arb_clindiag),1,0))`
- `denv__ever`
- `chikv__ever`

arb__preg__nz

This variable is related to the presence of any arbo virus in the current pregnancy without considering zika virus.

- `ifelse(arb_clindiag==0|arb_clindiag==1,0,ifelse(!is.na(arb_clindiag),1,NA))`
- `denv__preg`
- `chikv__preg`