

Plausibility bounds

Under this document we describe the plausible values consider 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 set all values outside the plausible bounds as NA.

| Classification | Variable name | Variable type | Units | Min | Max |
|----------------|-----------------------|---------------|-----------------|-----|------|
| Pwoman | age | Continuous | (years) | 13 | Inf |
| Pwoman | weight | Continuous | (Kg) | 0 | Inf |
| Pwoman | pre_pregweight | Continuous | (Kg) | 0 | Inf |
| Pwoman | height | Continuous | (cm) | 0 | Inf |
| Exposure | zikhv_pcr_ga_1 | Time | (weeks) | 0 | Inf |
| Exposure | zikhv_pcr_vl_1 | Continuous | (copies/microL) | 0 | Inf |
| Exposure | zikhv_elisa_ga_1 | Time | (weeks) | 0 | Inf |
| Exposure | zikhv_ga | Time | (weeks) | 0 | Inf |
| Exposure | symp_ga | Time | (weeks) | 0 | Inf |
| Exposure | arb_clindiag_ga | Time | (weeks) | 0 | Inf |
| Outcome | miscarriage_ga | Time | (weeks) | 0 | Inf |
| Outcome | loss_ga | Time | (weeks) | 0 | Inf |
| Outcome | endga | Time | (weeks) | 0 | Inf |
| Outcome | birth_ga | Time | (weeks) | 0 | Inf |
| Outcome | inf_weight | Continuous | (grams) | 100 | 6000 |
| Outcome | inf_length | Continuous | (cm) | 18 | Inf |
| 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, it is automatically generated the “obs_to_check.RData” file with the observations outside the plausibility range, it can be found in the “3_Output_data” folder. Inside obs_to_check dataset, there is an additional column called “Var_to_check” that indicates the variable with implausible values for each observation.

Logic rules for exposures

Set to NA values

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

Zika test time of measurement (zikv_ga)

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

On the following variables, we combine different variables in one general variable x_{com} in order to include the information in a simple way that help us to avoid any imputation convergence problem. For instance if we want to create a new Xm variable that combine 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 or a positive test $x_i == 1$, then the new variable x_{com} is assigned with a 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 certain patient, it was not recordered any value for all the considered x_i variables then 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 consider zika virus.

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