COVID-19: impact of original, Gamma, Delta, and Omicron variants of SARS-CoV-2 in children up to two years old

2022

Contents

R packages

R packages (https://www.r-project.org) used for reading, processing and analyzing the data.

```r
# Load packages
loadlibrary <- function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = T)
  if (!require(x, character.only = TRUE))
  stop("Package not found")
  }
}

packages <- c(
  "readr",
  "readxl",
  "janitor",
  "dplyr",
  "forcats",
  "stringr",
  "lubridate",
  "summarytools",
  "magrittr",
  "questionr",
  "knitr",
  "data.table",
  "writexl",
  "modelsummary",
  "abjData",
  "aweek",
  "ggplot2",
  "viridis"
)
lapply(packages, loadlibrary)

# Global changes for the summarytools packages
st_options(freq.cumul = FALSE,
           display.labels = TRUE,
           lang = "pt",
           bootstrap.css = FALSE,
           plain.ascii = FALSE,
           dfSummary.silent = TRUE,
           footnote = NA,
           subtitle.emphasis = FALSE)
st_css()
```

Load, treatment and join of the datasets

The Influenza Epidemiological Surveillance Information System, SIVEP-Gripe (Sistema de Informação de Vigilância Epidemiológica da Gripe), is a nationwide surveillance database used to monitor severe acute respiratory infections in Brazil.

The Severe Acute Respiratory Syndrome databases were used with the last update on November 23, 2022. The databases are from the years 2020, 2021 and 2022 and only cases up to 2 years old were considered. The data was limited to the first notified case of COVID-19 in February 2020 until June 30, 2022.

The data was obtained from the address https://opendatasus.saude.gov.br/.

```
#################### Loading the datasets ######################
data_2020 <- fread("SRAG_2020.csv", sep = ";")
data_2021 <- fread("SRAG_2021.csv", sep = ";")
data_2022 <- fread("SRAG_2022.csv", sep = ";")
```

```
# Creating dt_sint and dt_nasc variables for age study
data_list <- list("data_2020" = data_2020,
                  "data_2021" = data_2021,
                  "data_2022" = data_2022)

for (i in seq_along(data_list)) {
  data_list[[i]] <- data_list[[i]] %>%
    dplyr::mutate(
      dt_sint = as.Date(DT_SIN_PRI, format = "%d/%m/%Y"),
      dt_nasc = as.Date(DT_NASC, format = "%d/%m/%Y")
    )
}
```

We will select cases up to 2 years old. To have a better understanding, we are going to check the behavior of this variable in the available databases. We are going to create an age variable based on the day of birth up to the date of the first symptom.

Study of age variables

```
# Creating our year variable as the difference between dt_sint and dt_nasc
for (i in seq_along(data_list)) {
    data_list[[i]]$aux <-
      as.numeric(data_list[[i]]$dt_sint - data_list[[i]]$dt_nasc)
}
```

```
# Checking missing data in the dt_nasc variable
for (i in seq_along(data_list)) {
  cat("Are there missing data for dt_nasc in", names(data_list)[i], "?")
  print(table(is.na(data_list[[i]]$dt_nasc)))
}
```

```
## Are there missing data for dt_nasc in data_2020 ?
##    FALSE     TRUE
## 1198872     2178
## Are there missing data for dt_nasc in data_2021 ?
##    FALSE     TRUE
## 1732733     1215
## Are there missing data for dt_nasc in data_2022 ?
##   FALSE    TRUE
## 490478      188
```

Note that there are missing data for the date of birth. Therefore, we will consider the NU_IDADE_N variable for cases without date of birth information. The NU_IDADE_N variable indicates how many units we have of TP_IDADE, with TP_IDADE representing which unit of measurement we are using (1 = days, 2 = months and 3 = years).

```
# Transforming NU_IDADE_N into days when considering TP_IDADE: 1 = days; 2 = months; 3 = years
for (i in seq_along(data_list)) {
  cat("Are there negative values of NU_IDADE_N in", names(data_list)[i], "?")
```

```
    print(table(data_list[[i]]$NU_IDADE_N < 0))
  }
```

```
## Are there negative values of NU_IDADE_N in data_2020 ?
##   FALSE      TRUE
## 1201045         5
## Are there negative values of NU_IDADE_N in data_2021 ?
##   FALSE      TRUE
## 1733942         6
## Are there negative values of NU_IDADE_N in data_2022 ?
##  FALSE     TRUE
## 490663        3
```

```
data_list[["data_2020"]] <-
  filter(data_list[["data_2020"]], NU_IDADE_N >= 0)

data_list[["data_2021"]] <-
  filter(data_list[["data_2021"]], NU_IDADE_N >= 0)

data_list[["data_2022"]] <-
  filter(data_list[["data_2022"]], NU_IDADE_N >= 0)

for (i in seq_along(data_list)) {
  data_list[[i]] <- data_list[[i]] %>%
    mutate(nu_idade_n = ifelse(TP_IDADE == 1,
                               NU_IDADE_N,
                               ifelse(TP_IDADE == 2,
                                      NU_IDADE_N*30,
                                      NU_IDADE_N*365)))
}
```

```
# New age variable(idade_dias) without missing date of birth data
for (i in seq_along(data_list)) {
  data_list[[i]] <- data_list[[i]] %>%
    mutate(idade_dias = ifelse(is.na(dt_nasc),
                               nu_idade_n,
                               aux))
}

for (i in seq_along(data_list)) {
  cat("Are there missing data for idade_dias in", names(data_list)[i], "?")
  print(table(is.na(data_list[[i]]$idade_dias)))
}
```

```
## Are there missing data for idade_dias in data_2020 ?
##   FALSE
## 1201045
## Are there missing data for idade_dias in data_2021 ?
##   FALSE
## 1733942
## Are there missing data for idade_dias in data_2022 ?
##  FALSE
## 490663
```

We have concluded the study of age-related variables.

Now, we have to filter the databases for cases with idade_dias <= 730. In other words, we are going to filter cases up to 2 years old.

```
# Filtering for idade_dias <= 730
data2_list <- list("data2_2020" = NULL,
                   "data2_2021" = NULL,
                   "data2_2022" = NULL)

for (i in seq_along(data_list)) {
  data2_list[[i]] <- data_list[[i]] %>%
  filter(idade_dias <= 730)
}
```

Merging datasets

```
# Changing FATOR_RISC from the 2021 and 2022 datasets to character, as it is in 2020 data
data2_list[["data2_2021"]] <- data2_list[["data2_2021"]] %>%
  mutate(FATOR_RISC = case_when(FATOR_RISC == 1 ~ "S", # S = yes
                                FATOR_RISC == 2 ~ "N")) # N = no

data2_list[["data2_2022"]] <- data2_list[["data2_2022"]] %>%
  mutate(FATOR_RISC = case_when(FATOR_RISC == 1 ~ "S", # S = yes
                                FATOR_RISC == 2 ~ "N")) # N = no

# Join datasets
data2 <- data2_list[["data2_2020"]] %>%
  full_join(data2_list[["data2_2021"]]) %>%
  full_join(data2_list[["data2_2022"]])
```

```
# Case year variable
data2 <- data2 %>%
  dplyr::mutate(
    dt_sint = as.Date(DT_SIN_PRI, format = "%d/%m/%Y"),
    dt_nasc = as.Date(DT_NASC, format = "%d/%m/%Y"),
    year_case = lubridate::year(dt_sint),
    ) %>%
  filter(dt_sint >= as.Date("16-02-2020", format = "%d-%m-%Y") &
         dt_sint <= as.Date("30-06-2022", format = "%d-%m-%Y")) # omicron < 30-06-2022
```

There are 158654 cases in the complete dataset.

Filtering and selection

We are going to select the hospitalized cases.

```
# Filtering only hospitalized cases
data2 <- data2 %>%
  filter(HOSPITAL == 1)
```

There are 153889 cases.

Now we are going to select only the cases confirmed with COVID-19 (CLASSI_FIN = 5).

```r
# Confirmed with COVID-19
data2 <- data2 %>%
  filter(CLASSI_FIN == 5)
```

There are 19265 cases.

Selecting cases with positive TR-PCR.

```r
# Case with PCR
data2 <- data2 %>%
  mutate(pcr_SN = case_when(
    (PCR_SARS2 == 1) |
    (str_detect(DS_PCR_OUT, "SARS|COVID|COV|CORONA|CIVID")) ~
    "yes",
    TRUE ~ "no")
    )

# Identify sorology
data2$res_igg <- ifelse(is.na(data2$RES_IGG) == TRUE, 0, data2$RES_IGG)
data2$res_igm <- ifelse(is.na(data2$RES_IGM) == TRUE, 0, data2$RES_IGM)
data2$res_iga <- ifelse(is.na(data2$RES_IGA) == TRUE, 0, data2$RES_IGA)
data2$sorologia_SN <- ifelse(data2$res_igg == 1 |
                                data2$res_igm == 1 |
                                data2$res_iga == 1,
                                "yes",
                                "no")

# Identify antigen
data2 <- data2 %>%
  mutate(antigeno_SN = case_when(
    (AN_SARS2 == 1) | # positive
    (str_detect(DS_AN_OUT, "SARS|COVID|COV|CORONA|CONA")) ~
    "yes",
    TRUE ~ "no")
    )

# Covid-19 classification variable
data2 <- data2 %>%
  mutate(classi_covid = case_when(
    CLASSI_FIN == 5 & pcr_SN == "yes" ~ "pcr",
    CLASSI_FIN == 5 & pcr_SN == "no" &
    antigeno_SN == "yes" ~ "antigenio",
    CLASSI_FIN == 5 & sorologia_SN == "yes" &
    antigeno_SN == "no" &
    pcr_SN == "no" ~ "sorologia",
    CLASSI_FIN != 5 ~ "no", # This is not "another etiologic agent or unspecified"
    TRUE ~ "outro")
    )

data3 <- data2 %>%
  filter(classi_covid == "pcr")
```

There are 11153 cases.

Now, we are going to create a variant identifier.

```r
# "variants" variable
in_gama <- as.Date("01-02-2021",format="%d-%m-%Y")
in_delta <- as.Date("01-08-2021",format="%d-%m-%Y")
in_ommi <- as.Date("01-01-2022",format="%d-%m-%Y")

data4 <- data3 %>%
  mutate(variants = case_when(
    dt_sint < in_gama ~ "original",
    dt_sint >= in_gama & dt_sint < in_delta ~ "gamma",
    dt_sint >= in_delta & dt_sint < in_ommi ~ "delta",
    dt_sint >= in_ommi ~ "omicron") # omicron < 30-06-2022
    )

knitr::kable(freq(data4$variants, total = TRUE))
```

	n	%	val%
delta	795	7.1	7.1
gamma	2941	26.4	26.4
omicron	3711	33.3	33.3
original	3706	33.2	33.2
Total	11153	100.0	100.0

Regions of Brazil

```r
# Creating the region variable
regions <- function(state){
  southeast <- c("SP", "RJ", "ES", "MG")
  south <- c("PR", "SC", "RS")
  central <- c("GO", "MT", "MS", "DF")
  northeast <- c("AL", "BA", "CE", "MA", "PB", "PE", "PI", "RN", "SE")
  north <- c("AC", "AP", "AM", "PA", "RO", "RR", "TO")
  out <- ifelse(any(state == southeast),
                "southeast",
                ifelse(any(state == south),
                       "south",
                       ifelse(any(state == central),
                              "central",
                              ifelse(any(state == northeast),
                                     "northeast",
                                     "north")))
                )
  return(out)
}

data4$region <- sapply(data4$SG_UF, regions)
data4$region <- ifelse(is.na(data4$region) == TRUE, 0, data4$region)

# Age in months variable
data4$age_month <- floor(data4$idade_dias/30)
```

Epidemiologic characteristics

```r
# Ethnicity
data4 <- data4 %>%
  mutate(ethnicity = case_when(
    CS_RACA == 1 ~ "white",
    CS_RACA == 2 ~ "black",
    CS_RACA == 3 ~ "yellow",
    CS_RACA == 4 ~ "brown",
    CS_RACA == 5 ~ "indigenous",
    TRUE ~ NA_character_)
    )

# Sex
data4 <- data4 %>%
  mutate(sex = case_when(
    CS_SEXO == "F" ~ "female",
    CS_SEXO == "M" ~ "male",
    TRUE ~ NA_character_)
    )

# Travel history
data4 <- data4 %>%
  mutate(travel_hist = case_when(
    HISTO_VGM == 1 ~ "yes",
    HISTO_VGM == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Flu syndrome evolved into SRAG
data4 <- data4 %>%
  mutate(flu_into_srag = case_when(
    SURTO_SG == 1 ~ "yes",
    SURTO_SG == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Contact with poultry or swine
data4 <- data4 %>%
  mutate(cont_pty_swine = case_when(
    AVE_SUINO == 1 ~ "yes",
    AVE_SUINO == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Flu vaccine
data4 <- data4 %>%
  mutate(vacine = case_when(
    VACINA == 1 ~ "yes",
    VACINA == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Antiviral
data4 <- data4 %>%
  mutate(antiviral = case_when(
    ANTIVIRAL == 1 ~ "Oseltamivir",
```

```r
    ANTIVIRAL == 2 ~ "Zanamivir",
    TRUE ~ NA_character_)
    )

# Area of residence
data4 <- data4 %>%
  mutate(zone = case_when(
    CS_ZONA == 1 ~ "urban",
    CS_ZONA == 2 ~ "rural",
    CS_ZONA == 3 ~ "periurban",
    TRUE ~ NA_character_)
    )

# If there was a change of municipality for service
data4 <- data4 %>%
  mutate(changed_muni = case_when(
    (CO_MUN_RES == CO_MU_INTE) &
      !is.na(CO_MU_INTE) &
      !is.na(CO_MUN_RES) ~ "no",
    (CO_MUN_RES != CO_MU_INTE) &
      !is.na(CO_MU_INTE) &
      !is.na(CO_MUN_RES) ~ "yes",
    TRUE ~ NA_character_)
    )

# Infection acquired in the hospital
data4 <- data4 %>%
  mutate(inf_inter = case_when(
    NOSOCOMIAL == 1 ~ "yes",
    NOSOCOMIAL == 2 ~ "no",
    TRUE ~ NA_character_)
    )
```

Symptoms

```r
# Fever
data4 <- data4 %>%
  mutate(fever = case_when(
    FEBRE == 1 ~ "yes",
    FEBRE == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Cough
data4 <- data4 %>%
  mutate(cough = case_when(
    TOSSE == 1 ~ "yes",
    TOSSE == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Sore throat
data4 <- data4 %>%
  mutate(sore_throat = case_when(
```

```r
    GARGANTA == 1 ~ "yes",
    GARGANTA == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Dyspnea
data4 <- data4 %>%
  mutate(dyspnea = case_when(
    DISPNEIA == 1 ~ "yes",
    DISPNEIA == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Respiratory discomfort
data4 <- data4 %>%
  mutate(resp_disc = case_when(
    DESC_RESP == 1 ~ "yes",
    DESC_RESP == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Desaturation SpO2 <95%
data4 <- data4 %>%
  mutate(desaturation = case_when(
    SATURACAO == 1 ~ "yes",
    SATURACAO == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Diarrhea
data4 <- data4 %>%
  mutate(diarrhea = case_when(
    DIARREIA == 1 ~ "yes",
    DIARREIA == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Vomit
data4 <- data4 %>%
  mutate(vomit = case_when(
    VOMITO == 1 ~ "yes",
    VOMITO == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Abdominal pain
data4 <- data4 %>%
  mutate(abd_pain = case_when(
    DOR_ABD == 1 ~ "yes",
    DOR_ABD == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Fatigue
data4 <- data4 %>%
```

```r
  mutate(fatigue = case_when(
    FADIGA == 1 ~ "yes",
    FADIGA == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Olfactory loss
data4 <- data4 %>%
  mutate(olfac_loss = case_when(
    PERD_OLFT == 1 ~ "yes",
    PERD_OLFT == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Loss of taste
data4 <- data4 %>%
  mutate(loss_taste = case_when(
    PERD_PALA == 1 ~ "yes",
    PERD_PALA == 2 ~ "no",
    TRUE ~ NA_character_)
    )
```

Comorbidities

For the comorbidity variables, the cases listed as "ignored" will be considered as "no".

```r
# Heart disease
data4 <- data4 %>%
  mutate(cardiac = case_when(
    CARDIOPATI == 1 ~ "yes",
    CARDIOPATI == 2 ~ "no",
    CARDIOPATI == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Hematologic
data4 <- data4 %>%
  mutate(hematologic = case_when(
    HEMATOLOGI == 1 ~ "yes",
    HEMATOLOGI == 2 ~ "no",
    HEMATOLOGI == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Hepatic
data4 <- data4 %>%
  mutate(hepatic = case_when(
    HEPATICA == 1 ~ "yes",
    HEPATICA == 2 ~ "no",
    HEPATICA == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Asthma
```

```r
data4 <- data4 %>%
  mutate(asthma = case_when(
    ASMA == 1 ~ "yes",
    ASMA == 2 ~ "no",
    ASMA == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Diabetes
data4 <- data4 %>%
  mutate(diabetes = case_when(
    DIABETES == 1 ~ "yes",
    DIABETES == 2 ~ "no",
    DIABETES == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Neurologic
data4 <- data4 %>%
  mutate(neurologic = case_when(
    NEUROLOGIC == 1 ~ "yes",
    NEUROLOGIC == 2 ~ "no",
    NEUROLOGIC == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Lung disease
data4 <- data4 %>%
  mutate(pneumologic = case_when(
    PNEUMOPATI == 1 ~ "yes",
    PNEUMOPATI == 2 ~ "no",
    PNEUMOPATI == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Imunossupression
data4 <- data4 %>%
  mutate(imuno = case_when(
    IMUNODEPRE == 1 ~ "yes",
    IMUNODEPRE == 2 ~ "no",
    IMUNODEPRE == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Renal
data4 <- data4 %>%
  mutate(renal = case_when(
    RENAL == 1 ~ "yes",
    RENAL == 2 ~ "no",
    RENAL == 9 ~ "no",
    TRUE ~ NA_character_)
    )

# Obesity
data4 <- data4 %>%
```

```
  mutate(obesity = case_when(
    OBESIDADE == 1 ~ "yes",
    OBESIDADE == 2 ~ "no",
    OBESIDADE == 9 ~ "no",
    TRUE ~ NA_character_)
    )
```

Outcomes

```
# ICU admission
data4 <- data4 %>%
  mutate(icu = case_when(
    UTI == 1 ~ "yes",
    UTI == 2 ~ "no",
    TRUE ~ NA_character_)
    )

# Invasive respiratory support
data4 <- data4 %>%
  mutate(intubation = case_when(
    SUPORT_VEN == 1 ~ "yes",
    SUPORT_VEN == 2 ~ "no",
    SUPORT_VEN == 3 ~ "no",
    TRUE ~ NA_character_)
    )

# Respiratory support
data4 <- data4 %>%
  mutate(suport_ven = case_when(
    SUPORT_VEN == 1 ~ "invasivo",
    SUPORT_VEN == 2 ~ "não invasivo",
    SUPORT_VEN == 3 ~ "não",
    TRUE ~ NA_character_)
)

data4$suport_ven <- factor(
  data4$suport_ven,
  levels = c("invasivo", "não invasivo", "não")
)

# Evolution
data4 <- data4 %>%
  mutate(death = case_when(
    EVOLUCAO == 1 ~ "cure",
    EVOLUCAO == 2 ~ "death",
    EVOLUCAO == 3 ~ "death",
    EVOLUCAO == 9 ~ NA_character_,
    TRUE ~ NA_character_)
    )
```

Final dataset

```
saveRDS(data4, file = "data_paper_1000d.rds")
write.csv(data4, "data_paper_1000d.csv", row.names = FALSE)
```

Flowchart

Case selection flowchart

Records in SIVEP-Gripe (epidemiological weeks 1-53/2020- 1-52/2021 1-47/2022): n = 3,425,650

→ Age > 2 years: n = 3,243,807

Children up to 2 years old: n= 158,654

→ Cases with date of first symptoms < 02/16/2020 or > 06/30/2022: n = 23,189

Hospital admissions: n= 153,889

→ COVID-19 unconfirmed: n= 134,624

→ COVID-19 unconfirmed by RT-PCR: n= 8,112

Children up to 2 years old hospitalized due to COVID-19 (RT-PCR): n= 11,153

| Children with original variant n= 3706 | Children with gamma variant n= 2941 | Children with delta variant n= 795 | Children with omicron variant n= 3711 |