COVID-19 outcomes in hospitalized puerperal, pregnant, and neither pregnant nor puerperal women: a population study

Codes and outputs of Propensity Score Matching

08/26/2021

1. Description

This file presents the documentation of the analysis of Propensity Scoring Method (PSM) of the article "COVID-19 outcomes in hospitalized puerperal, pregnant, and neither pregnant nor puerperal women: a population study" with authors Fabiano Elisei Serra, Rossana Pulcineli Vieira Francisco, Patricia de Rossi, Maria de Lourdes Brizot, and Agatha Sacramento Rodrigues.

2. R packages used, functions and dataset import

The data are analyzed using the free-software R (https://www.R-project.org) in version 4.0.3. Next, we present and load the libraries used in the data analysis process.

```
#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",
    "magrittr",
    "dplyr",
    "stringr",
    "questionr",
    "knitr",
    "forcats",
    "lubridate",
    "summarytools",
    "modelsummary",
    "kableExtra",
    "epitools",
    "WeightIt",
    "jtools",
```

```
    "survey",
    "cobalt",
    "nnet"
  )
lapply(packages, loadlibrary)
```

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.

Notification is mandatory for Influenza Syndrome (characterized by at least two of the following signs and symptoms: fever, even if referred, chills, sore throat, headache, cough, runny nose, olfactory or taste disorders) and who has dyspnea/respiratory discomfort or persistent pressure in the chest or O2 saturation less than 95% in room air or bluish color of the lips or face. Asymptomatic individuals with laboratory confirmation by molecular biology or immunological examination for COVID-19 infection are also reported.

For notifications in Sivep-Gripe, hospitalized cases in both public and private hospitals and all deaths due to severe acute respiratory infections regardless of hospitalization must be considered.

The analyzed period comprised data from epidemiological weeks 1 to 53 of 2020 (12/29/2019 - 01/02/2021) with the database downloaded on 01/11/2021 on the site https://opendatasus.saude.gov.br/dataset/bd-srag-2020. The data are loaded below:

```
#loading the dataset
data_all <- readr::read_delim(
  "INFLUD-11-01-2021.csv",
  ";",
  escape_double = FALSE,
  locale = locale(encoding = "ISO-8859-2"),
  trim_ws = TRUE
)
```

There are 1136681 cases in the complete dataset. The case selection is presented in the following according to the flowchart presented in the article.

3. Case selection and data treatment

The first filter consists of selecting the hospitalized cases. For that, the `HOSPITAL` variable is considered, in which 1-Yes, 2-No, and 9-Ignored.

```
#Selecting only hospitalization cases
data1 <- dplyr::filter(data_all, HOSPITAL == 1)
```

When considering only confirmed hospitalized cases, we get 1061254 observations.

The second filtering consists of the cases classified as COVID-19 in the database. The variable indicating the classification is `CLASSI_FIN`, with the following categories: 1-SRAG by influenza, 2-SRAG by another respiratory virus, 3-SRAG by another etiological agent, 4-SRAG not specified, and 5-SRAG by COVID-19.

```
questionr::freq(
  data1$CLASSI_FIN,
  cum = FALSE,
  total = TRUE,
```

```
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for case classification", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 1: Frequency table for case classification

	n	%
1	2507	0.2
2	4137	0.4
3	2929	0.3
4	365992	34.5
5	588711	55.5
NA	96978	9.1
Total	1061254	100.0

```
#Filtering COVID-19 cases
data2 <- dplyr::filter(data1, CLASSI_FIN == 5)
```

There are 588711 selected cases for now.

Only cases of COVID-19 confirmed by RT-PCR are selected. The selection is made as follows:

```
#Selecting COVID-19 confirmed by RT-PCR
data3 <- data2 %>%
  dplyr::filter((PCR_SARS2 == 1) |
                  (
                    stringr::str_detect(DS_PCR_OUT, "SARS|COVID|COV|CORONA|CIVID") &
                      !stringr::str_detect(DS_PCR_OUT, "63|43|229|HK|RINO|SINCI|PARE")
                  ) |
                  (
                    PCR_RESUL == 1 &
                      CRITERIO == 1 &
                      is.na(DS_PCR_OUT) &
                      (PCR_RINO  != 1 |
                         is.na(PCR_RINO)) &
                      (POS_PCRFLU != 1 | is.na(POS_PCRFLU)) &
                      (PCR_OUTRO != 1 | is.na(PCR_OUTRO)) &
                      (POS_PCROUT != 1 | is.na(POS_PCROUT)) &
                      (is.na(PCR_VSR)) &
                      (is.na(PCR_METAP)) &
                      (is.na(PCR_PARA1))
                  )
  )
```

After this selection, 454830 cases are selected.

The next step consists of selecting female cases. The sex variable is `CS_SEXO`, in which F-Female, M-Male and I-Ignored.

```
questionr::freq(
  data3$CS_SEXO,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for sex", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 2: Frequency table for sex

	n	%
F	199931	44
I	74	0
M	254825	56
Total	454830	100

```
#Filtering female cases
data4 <- dplyr::filter(data3, CS_SEXO == "F")
```

Now there are 199931 cases. The next selection is to consider female people over 9 years old and under 50 (not inclusive). The variable that indicates the cases age is `NU_IDADE_N`.

```
#Filtering female people over 9 years old and under 50
data5 <- dplyr::filter(data4, NU_IDADE_N > 9 & NU_IDADE_N < 50)
```

The number of cases results in 50845 cases.

Now we are going to identify pregnant people. For this, we will analyze the variable `CS_GESTANT`. This variable assumes the values: 1-1st gestational trimester; 2-2nd gestational trimester; 3-3rd gestational trimester; 4-Ignored gestational age; 5-No; 6-Does not apply; 9-Ignored.

```
questionr::freq(
  data5$CS_GESTANT,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for pregnancy variable", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

The next step is filtering cases we have information about pregnancy (yes - any gestational age - or not).

```
#Not considering do not apply and ignored
data6 <- dplyr::filter(data5, CS_GESTANT >= 1 & CS_GESTANT <= 5)
```

After the above filtering, we get 40640 observations.

The pregnancy indicator variable (independent of the gestational period) is created below.

Table 3: Frequency table for pregnancy variable

	n	%
0	1	0.0
1	295	0.6
2	829	1.6
3	2089	4.1
4	159	0.3
5	37268	73.3
6	2056	4.0
9	8148	16.0
Total	50845	100.0

```
#Creating pregnancy indicator variable
data6 <- data6 %>%
  dplyr::mutate(gestante_SN = ifelse(CS_GESTANT == 5, "no", "yes"))
```

```
questionr::freq(
  data6$gestante_SN,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for pregnancy indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 4: Frequency table for pregnancy indicator

	n	%
no	37268	91.7
yes	3372	8.3
Total	40640	100.0

The next step is considering the postpartum indicator variable. The `PUERPERA` variable has three categories: 1-yes, 2-no, and 9-Ignored.

```
questionr::freq(
  data6$PUERPERA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for postpartum indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Now we can create the group variable with the categories: preg - for pregnant women, puerp - for postpartum, and no - for woman of reprodutive age.

Table 5: Frequency table for postpartum indicator

	n	%
1	983	2.4
2	12183	30.0
9	230	0.6
NA	27244	67.0
Total	40640	100.0

```
#Creating postpartum indicator
data6 <- data6 %>%
  dplyr::mutate(puerpera = ifelse(is.na(PUERPERA) == TRUE, 0, PUERPERA))

#Creating group varible with three categories:
##preg - for pregnant women,
##puerp - for postpartum and
##no - woman of reprodutive age
data6 <- data6 %>%
  dplyr::mutate(gest_puerp = ifelse(
    gestante_SN == "yes",
    1,
    ifelse(gestante_SN == "no" & puerpera == 1, 2, 0))
    )

data6$gest_puerp <- factor(data6$gest_puerp,
                       levels = c(0,1,2),
                       labels = c("no", "preg", "puerp"))
```

```
questionr::freq(
  data6$gest_puerp,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for group variable", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 6: Frequency table for group variable

	n	%
no	36474	89.7
preg	3372	8.3
puerp	794	2.0
Total	40640	100.0

3.1 Characterization variables and comorbities

The age information is in `NU_IDADE_N`. We create the age group variable (`faixa_et`) with categories: "<20", "20-34" and ">34".

```r
#age group variable
data6 <-
  dplyr::mutate(data6, faixa_et = ifelse(
    NU_IDADE_N <= 19,
    "<20",
    ifelse(NU_IDADE_N >= 20 & NU_IDADE_N <= 34, "20-34", ">34")
  ))

data6$faixa_et <- factor(data6$faixa_et,
                         levels = c("<20", "20-34", ">34"))
```

```r
questionr::freq(
  data6$faixa_et,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for group age", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 7: Frequency table for group age

	n	%
<20	1320	3.2
20-34	12367	30.4
>34	26953	66.3
Total	40640	100.0

For ethnicity (`CS_RACA`), the categories are: 1-white; 2-black; 3-yellow; 4-brown; 5-Indigenous; 6-Ignored.

```r
questionr::freq(
  data6$CS_RACA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for ethnicity", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

We will now label this variable, creating the variable `raca`, considering only the valid categories.

```r
#ethnicity variable
data6$raca <- factor(
  data6$CS_RACA,
  levels = c("1", "2", "3", "4", "5"),
  labels = c("white", "black", "yellow", "brown", "indigenous")
)
```

For education (`CS_ESCOL_N`), the categories are: 0-no education/illiterate; 1-fundamental 1st cycle; 2-fundamental 2nd cycle; 3-high school; 4-superior; 5-not applicable, 9-ignored.

Table 8: Frequency table for ethnicity

	n	%
1	16758	41.2
2	1917	4.7
3	396	1.0
4	12706	31.3
5	114	0.3
9	6135	15.1
NA	2614	6.4
Total	40640	100.0

```
questionr::freq(
  data6$CS_ESCOL_N,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for school", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 9: Frequency table for school

	n	%
0	228	0.6
1	1546	3.8
2	2556	6.3
3	8501	20.9
4	4403	10.8
5	1	0.0
9	11495	28.3
NA	11910	29.3
Total	40640	100.0

We will now label this variable, creating the variable `escol`, considering only the valid categories and considering the following categories: no education/illiterate (`CS_ESCOL_N` = 0), up to high school (`CS_ESCOL_N` = 1 or 2), high school (`CS_ESCOL_N` = 3) and higher education (`CS_ESCOL_N` = 4).

```
#school variable
data6$escol <- factor(
  data6$CS_ESCOL_N,
  levels = c("0", "1", "2", "3", "4"),
  labels = c(
    "no education",
    "up to high school",
    "up to high school",
    "high school",
    "higher education"
  )
)
```

```
questionr::freq(
  data6$escol,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for school (new categories)", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 10: Frequency table for school (new categories)

	n	%
no education	228	0.6
up to high school	4102	10.1
high school	8501	20.9
higher education	4403	10.8
NA	23406	57.6
Total	40640	100.0

For comorbidities, the categories are: 1-yes, 2-no and 9-ignored. The comorbidities considered are: cardiopathy, hematology, liver disease, asthma, diabetes, neurological diseases, pneumopathy, immunosuppression, kidney disease, and obesity, and their frequency tables are presented below, respectively:

```
questionr::freq(
  data6$CARDIOPATI,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for cardiopathy", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 11: Frequency table for cardiopathy

	n	%
1	6071	14.9
2	8905	21.9
9	230	0.6
NA	25434	62.6
Total	40640	100.0

```
questionr::freq(
  data6$HEMATOLOGI,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
```

```
  knitr::kable(caption = "Frequency table for hematology", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 12: Frequency table for hematology

	n	%
1	383	0.9
2	12518	30.8
9	282	0.7
NA	27457	67.6
Total	40640	100.0

```
questionr::freq(
  data6$HEPATICA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for liver disease", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 13: Frequency table for liver disease

	n	%
1	201	0.5
2	12568	30.9
9	286	0.7
NA	27585	67.9
Total	40640	100.0

```
questionr::freq(
  data6$ASMA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for asthma", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

```
questionr::freq(
  data6$DIABETES,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for diabetes", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 14: Frequency table for asthma

	n	%
1	2055	5.1
2	11371	28.0
9	264	0.6
NA	26950	66.3
Total	40640	100.0

Table 15: Frequency table for diabetes

	n	%
1	5127	12.6
2	9448	23.2
9	217	0.5
NA	25848	63.6
Total	40640	100.0

```
questionr::freq(
  data6$NEUROLOGIC,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for neurological diseases", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 16: Frequency table for neurological diseases

	n	%
1	598	1.5
2	12335	30.4
9	273	0.7
NA	27434	67.5
Total	40640	100.0

```
questionr::freq(
  data6$PNEUMOPATI,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for pneumopathy", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 17: Frequency table for pneumopathy

	n	%
1	607	1.5
2	12338	30.4
9	281	0.7
NA	27414	67.5
Total	40640	100.0

```
questionr::freq(
  data6$IMUNODEPRE,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for immunosuppression", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 18: Frequency table for immunosuppression

	n	%
1	1347	3.3
2	11790	29.0
9	281	0.7
NA	27222	67.0
Total	40640	100.0

```
questionr::freq(
  data6$RENAL,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for kidney disease", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 19: Frequency table for kidney disease

	n	%
1	1116	2.7
2	11931	29.4
9	275	0.7
NA	27318	67.2
Total	40640	100.0

```r
questionr::freq(
  data6$OBESIDADE,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for obesity", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 20: Frequency table for obesity

	n	%
1	3937	9.7
2	9776	24.1
9	462	1.1
NA	26465	65.1
Total	40640	100.0

We will label in the following the comorbidities indicators, considering only the valid categories.

```r
#cardiopathy
data6$cardiopati <- factor(data6$CARDIOPATI,
                           levels = c("1", "2"),
                           labels = c("yes", "no"))
```

```r
#hematology
data6$hematologi <- factor(data6$HEMATOLOGI,
                           levels = c("1", "2"),
                           labels = c("yes", "no"))
```

```r
#liver disease
data6$hepatica <- factor(data6$HEPATICA,
                         levels = c("1", "2"),
                         labels = c("yes", "no"))
```

```r
#asthma
data6$asma <- factor(data6$ASMA,
                     levels = c("1", "2"),
                     labels = c("yes", "no"))
```

```r
#diabetes
data6$diabetes <- factor(data6$DIABETES,
                         levels = c("1", "2"),
                         labels = c("yes", "no"))
```

```r
#neurological diseases
data6$neuro <- factor(data6$NEUROLOGIC,
                      levels = c("1", "2"),
                      labels = c("yes", "no"))
```

```
#pneumopathy
data6$pneumopati <- factor(data6$PNEUMOPATI,
                           levels = c("1", "2"),
                           labels = c("yes", "no"))
```

```
#immunosuppression
data6$imunodepre <- factor(data6$IMUNODEPRE,
                           levels = c("1", "2"),
                           labels = c("yes", "no"))
```

```
#kidney disease
data6$renal <- factor(data6$RENAL,
                      levels = c("1", "2"),
                      labels = c("yes", "no"))
```

```
#obesity
data6$obesidade <- factor(data6$OBESIDADE,
                          levels = c("1", "2"),
                          labels = c("yes", "no"))
```

One variable we want to analyze is the comorbities group (`gr_comorb`) with the categories: "none", "1 or 2", ">2".

```
comorbidades <-
  c(
    "CARDIOPATI_aux",
    "HEMATOLOGI_aux",
    "HEPATICA_aux",
    "ASMA_aux",
    "DIABETES_aux",
    "NEUROLOGIC_aux",
    "PNEUMOPATI_aux",
    "IMUNODEPRE_aux",
    "RENAL_aux",
    "OBESIDADE_aux"
  )

comorbidades1 <-
  c(
    "CARDIOPATI_aux1",
    "HEMATOLOGI_aux1",
    "HEPATICA_aux1",
    "ASMA_aux1",
    "DIABETES_aux1",
    "NEUROLOGIC_aux1",
    "PNEUMOPATI_aux1",
    "IMUNODEPRE_aux1",
    "RENAL_aux1",
    "OBESIDADE_aux1"
  )

data6 <-
```

```
  mutate(
    data6,
    CARDIOPATI_aux = CARDIOPATI,
    HEMATOLOGI_aux = HEMATOLOGI,
    HEPATICA_aux = HEPATICA,
    ASMA_aux = ASMA,
    DIABETES_aux = DIABETES,
    NEUROLOGIC_aux = NEUROLOGIC,
    PNEUMOPATI_aux = PNEUMOPATI,
    IMUNODEPRE_aux = IMUNODEPRE,
    RENAL_aux = RENAL,
    OBESIDADE_aux = OBESIDADE
  )

data6 <-
  mutate(
    data6,
    CARDIOPATI_aux1 = CARDIOPATI,
    HEMATOLOGI_aux1 = HEMATOLOGI,
    HEPATICA_aux1 = HEPATICA,
    ASMA_aux1 = ASMA,
    DIABETES_aux1 = DIABETES,
    NEUROLOGIC_aux1 = NEUROLOGIC,
    PNEUMOPATI_aux1 = PNEUMOPATI,
    IMUNODEPRE_aux1 = IMUNODEPRE,
    RENAL_aux1 = RENAL,
    OBESIDADE_aux1 = OBESIDADE
  )

data6 <-  data6 %>%
  dplyr::mutate_at(dplyr::all_of(comorbidades), function(x) {
    dplyr::case_when(x == "1" ~ 1, TRUE ~ 0)
  }) %>%
  dplyr::mutate_at(dplyr::all_of(comorbidades1), function(x) {
    dplyr::case_when(x == "1" ~ 1, x == "2" ~ 0, TRUE ~ NA_real_)
  }) %>%
  dplyr::mutate(
    cont_comorb = CARDIOPATI_aux + HEMATOLOGI_aux +  HEPATICA_aux + ASMA_aux +
      DIABETES_aux + NEUROLOGIC_aux + PNEUMOPATI_aux + IMUNODEPRE_aux +
      RENAL_aux + OBESIDADE_aux
  ) %>%
  dplyr::mutate(
    num_comorb = dplyr::case_when(
      is.na(CARDIOPATI_aux1) |
        is.na(HEMATOLOGI_aux1) |
        is.na(HEPATICA_aux1) |
        is.na(ASMA_aux1) |
        is.na(DIABETES_aux1) |
        is.na(NEUROLOGIC_aux1) | is.na(PNEUMOPATI_aux1) |
        is.na(IMUNODEPRE_aux1) |
        is.na(RENAL_aux1) | is.na(OBESIDADE_aux1) ~ NA_real_,
      TRUE ~ cont_comorb
    ),
```

```
    gr_comorb = dplyr::case_when(
      num_comorb == 0 ~ 0,
      num_comorb == 1 ~ 1,
      num_comorb == 2 ~ 1,
      num_comorb >  2 ~ 2,
      TRUE ~ NA_real_
    )
  )

# Comorbities group
data6$gr_comorb <- factor(data6$gr_comorb,
                          levels = c(0, 1, 2),
                          labels = c("none", "1 or 2", ">2"))
```

```
questionr::freq(
  data6$gr_comorb,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for comorbities group", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 21: Frequency table for comorbities group

	n	%
none	2933	7.2
1 or 2	8195	20.2
>2	830	2.0
NA	28682	70.6
Total	40640	100.0

Another variable of interest is metabolic syndrome defined here if one has diabetes, heart disease and obesity. The variable name is `gr_sind_met` with the categories "yes" and "no".

```
sind_met <-  c("CARDIOPATI_aux",
               "DIABETES_aux", "OBESIDADE_aux")

sind_met1 <-  c("CARDIOPATI_aux1",
                "DIABETES_aux1",
                "OBESIDADE_aux1")

data6 <-
  mutate(
    data6,
    CARDIOPATI_aux = CARDIOPATI,
    DIABETES_aux = DIABETES,
    OBESIDADE_aux = OBESIDADE
  )

data6 <-
```

```
  mutate(
    data6,
    CARDIOPATI_aux1 = CARDIOPATI,
    DIABETES_aux1 = DIABETES,
    OBESIDADE_aux1 = OBESIDADE
  )

data6 <-  data6 %>%
  mutate_at(all_of(sind_met), function(x) {
    case_when(x == "1" ~ 1, TRUE ~ 0)
  }) %>%
  mutate_at(all_of(sind_met1), function(x) {
    case_when(x == "1" ~ 1, x == "2" ~ 0, TRUE ~ NA_real_)
  }) %>%
  mutate(cont_sind_met = CARDIOPATI_aux + DIABETES_aux + OBESIDADE_aux) %>%
  mutate(
    num_sind_met = case_when(
      is.na(CARDIOPATI_aux1) |
        is.na(DIABETES_aux1) | is.na(OBESIDADE_aux1) ~ NA_real_,
      TRUE ~ cont_sind_met
    ),
    gr_sind_met = case_when(
      num_sind_met == 0 ~ 0,
      num_sind_met == 1 ~ 0,
      num_sind_met == 2 ~ 0,
      num_sind_met == 3 ~ 1,
      TRUE ~ NA_real_
    )
  )

#metabolic syndrome indicator
data6$gr_sind_met <- factor(data6$gr_sind_met,
                            levels = c(1, 0),
                            labels = c("yes", "no"))
```

```
questionr::freq(
  data6$gr_sind_met,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for metabolic syndrome", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

3.2 Symptom variables and indicator of hospital-acquired infection

For the indicator of a case arising from an infection acquired in the hospital (`NOSOCOMIAL`), the categories are 1-yes, 2-no and 9-ignored.

Table 22: Frequency table for metabolic syndrome

	n	%
yes	436	1.1
no	12073	29.7
NA	28131	69.2
Total	40640	100.0

```
questionr::freq(
  data6$NOSOCOMIAL,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for hospital-acquired infection", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 23: Frequency table for hospital-acquired infection

	n	%
1	831	2.0
2	29891	73.6
9	2802	6.9
NA	7116	17.5
Total	40640	100.0

We will now label this variable, creating the variable `inf_inter`, considering only the valid categories.

```
data6$inf_inter <- factor(data6$NOSOCOMIAL,
                          levels = c("1", "2"),
                          labels = c("yes", "no"))
```

The symptoms are fever, cough, sore throat, dyspnoea, vomiting, abdominal pain, fatigue, respiratory distress, saturation, diarrhea, olfactory loss and loss of taste. In the original dataset they are `FEBRE`, `TOSSE`, `GARGANTA`, `DISPNEIA`, `VOMITO`, `DOR_ABD`, `FADIGA`, `DESC_RESP`, `SATURACAO`, `DIARREIA`, `PERD_OLFT`, `PERD_PALA`, respectively. The categories of these variables are 1-yes, 2-no and 9-ignored.

```
questionr::freq(
  data6$FEBRE,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for fever indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 24: Frequency table for fever indicator

	n	%
1	26759	65.8
2	9612	23.7
9	360	0.9
NA	3909	9.6
Total	40640	100.0

```
questionr::freq(
  data6$TOSSE,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for cough indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 25: Frequency table for cough indicator

	n	%
1	30188	74.3
2	6913	17.0
9	296	0.7
NA	3243	8.0
Total	40640	100.0

```
questionr::freq(
  data6$GARGANTA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for sore throat indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 26: Frequency table for sore throat indicator

	n	%
1	9734	24.0
2	21740	53.5
9	612	1.5
NA	8554	21.0
Total	40640	100.0

```
questionr::freq(
  data6$DISPNEIA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for dyspnea indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 27: Frequency table for dyspnea indicator

	n	%
1	27276	67.1
2	8996	22.1
9	295	0.7
NA	4073	10.0
Total	40640	100.0

```
questionr::freq(
  data6$VOMITO,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for vomiting", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 28: Frequency table for vomiting

	n	%
1	4535	11.2
2	25807	63.5
9	681	1.7
NA	9617	23.7
Total	40640	100.0

```
questionr::freq(
  data6$DOR_ABD,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for abdominal pain", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 29: Frequency table for abdominal pain

	n	%
1	1484	3.7
2	14291	35.2
9	493	1.2
NA	24372	60.0
Total	40640	100.0

```
questionr::freq(
  data6$FADIGA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for fatigue", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 30: Frequency table for fatigue

	n	%
1	4693	11.5
2	11523	28.4
9	497	1.2
NA	23927	58.9
Total	40640	100.0

```
questionr::freq(
  data6$DESC_RESP,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for respiratory discomfort", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 31: Frequency table for respiratory discomfort

	n	%
1	21873	53.8
2	12150	29.9
9	423	1.0
NA	6194	15.2
Total	40640	100.0

```
questionr::freq(
  data6$SATURACAO,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for saturation", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 32: Frequency table for saturation

	n	%
1	18260	44.9
2	15222	37.5
9	529	1.3
NA	6629	16.3
Total	40640	100.0

```
questionr::freq(
  data6$DIARREIA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for diarrhea", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 33: Frequency table for diarrhea

	n	%
1	6787	16.7
2	24120	59.4
9	621	1.5
NA	9112	22.4
Total	40640	100.0

```
questionr::freq(
  data6$PERD_OLFT,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for olfactory loss", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 34: Frequency table for olfactory loss

	n	%
1	3602	8.9
2	12600	31.0
9	542	1.3
NA	23896	58.8
Total	40640	100.0

```
questionr::freq(
  data6$PERD_PALA,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for loss of taste", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 35: Frequency table for loss of taste

	n	%
1	3463	8.5
2	12640	31.1
9	550	1.4
NA	23987	59.0
Total	40640	100.0

We will now label the sympthoms variables by considering only the valid categories and creating the variable `febre`, `tosse`, `garganta`, `dispneia`, `vomito`, `dor_abd`, `fadiga`, `desc_resp`, `saturacao`, `diarreia`, `perd_olft` and `perd_pala` that represent fever, cough, sore throat, dyspnoea, vomiting, abdominal pain, fatigue, respiratory distress, saturation, diarrhea, olfactory loss and loss of taste, respectively.

```
#fever
data6$febre <- factor(data6$FEBRE,
                      levels = c("1", "2"),
                      labels = c("yes", "no"))
```

```
#cough
data6$tosse <- factor(data6$TOSSE,
                      levels = c("1", "2"),
                      labels = c("yes", "no"))
```

```
#sore throat
data6$garganta <- factor(data6$GARGANTA,
                         levels = c("1", "2"),
                         labels = c("yes", "no"))
```

```r
#dyspnoea
data6$dispneia <- factor(data6$DISPNEIA,
                     levels = c("1", "2"),
                     labels = c("yes", "no"))
```

```r
#vomiting
data6$vomito <- factor(data6$VOMITO,
                       levels = c("1", "2"),
                       labels = c("yes", "no"))
```

```r
#abdominal pain
data6$dor_abd <- factor(data6$DOR_ABD,
                        levels = c("1", "2"),
                        labels = c("yes", "no"))
```

```r
#fatigue
data6$fadiga <- factor(data6$FADIGA,
                       levels = c("1", "2"),
                       labels = c("yes", "no"))
```

```r
#respiratory distress
data6$desc_resp <- factor(data6$DESC_RESP,
                          levels = c("1", "2"),
                          labels = c("yes", "no"))
```

```r
#saturation
data6$saturacao <- factor(data6$SATURACAO,
                     levels = c("1", "2"),
                     labels = c("yes", "no"))
```

```r
#diarrhea
data6$diarreia <- factor(data6$DIARREIA,
                     levels = c("1", "2"),
                     labels = c("yes", "no"))
```

```r
#olfactory loss
data6$perd_olft <- factor(data6$PERD_OLFT,
                          levels = c("1", "2"),
                          labels = c("yes", "no"))
```

```r
#loss of taste
data6$perd_pala <- factor(data6$PERD_PALA,
                          levels = c("1", "2"),
                          labels = c("yes", "no"))
```

Besides the indicator variable of each symptom, the variable group of symptoms has three categories: “none”, “1 or 2” and “>2” (`gr_sintomas`) and another indicator variable of at least one symptom (`sintomas_SN`), with categories “yes” or “no”. The symptoms are fever, cough, sore throat, dyspnoea, respiratory distress, saturation, diarrhea, vomiting, abdominal pain, fatigue, olfactory loss and loss of taste.

```
sintomas <-
  c(
    "FEBRE_aux",
    "TOSSE_aux",
    "GARGANTA_aux",
    "DISPNEIA_aux",
    "DESC_RESP_aux",
    "SATURACAO_aux",
    "DIARREIA_aux",
    "VOMITO_aux",
    "DOR_ABD_aux",
    "FADIGA_aux",
    "PERD_OLFT_aux",
    "PERD_PALA_aux"
  )

sintomas1 <-
  c(
    "FEBRE_aux1",
    "TOSSE_aux1",
    "GARGANTA_aux1",
    "DISPNEIA_aux1",
    "DESC_RESP_aux1",
    "SATURACAO_aux1",
    "DIARREIA_aux1",
    "VOMITO_aux1",
    "DOR_ABD_aux1",
    "FADIGA_aux1",
    "PERD_OLFT_aux1",
    "PERD_PALA_aux1"
  )

data6 <-
  mutate(
    data6,
    FEBRE_aux = FEBRE,
    TOSSE_aux = TOSSE,
    GARGANTA_aux = GARGANTA,
    DISPNEIA_aux = DISPNEIA,
    DESC_RESP_aux = DESC_RESP,
    SATURACAO_aux = SATURACAO,
    DIARREIA_aux = DIARREIA,
    VOMITO_aux = VOMITO,
    DOR_ABD_aux = DOR_ABD,
    FADIGA_aux = FADIGA,
    PERD_OLFT_aux = PERD_OLFT,
    PERD_PALA_aux = PERD_PALA
  )

data6 <-
  mutate(
    data6,
    FEBRE_aux1 = FEBRE,
```

```
    TOSSE_aux1 = TOSSE,
    GARGANTA_aux1 = GARGANTA,
    DISPNEIA_aux1 = DISPNEIA,
    DESC_RESP_aux1 = DESC_RESP,
    SATURACAO_aux1 = SATURACAO,
    DIARREIA_aux1 = DIARREIA,
    VOMITO_aux1 = VOMITO,
    DOR_ABD_aux1 = DOR_ABD,
    FADIGA_aux1 = FADIGA,
    PERD_OLFT_aux1 = PERD_OLFT,
    PERD_PALA_aux1 = PERD_PALA
  )

data6 <-  data6 %>%
  mutate_at(all_of(sintomas), function(x) {
    case_when(x == "1" ~ 1, TRUE ~ 0)
  }) %>%
  mutate_at(all_of(sintomas1), function(x) {
    case_when(x == "1" ~ 1, x == "2" ~ 0, TRUE ~ NA_real_)
  }) %>%
  mutate(
    cont_sintomas = FEBRE_aux + TOSSE_aux +  GARGANTA_aux + DISPNEIA_aux + DESC_RESP_aux +
      SATURACAO_aux + DIARREIA_aux + VOMITO_aux + DOR_ABD_aux + FADIGA_aux +
     PERD_OLFT_aux + PERD_PALA_aux
  ) %>%
  mutate(
    num_sintomas = case_when(
      is.na(FEBRE_aux1) |
        is.na(TOSSE_aux1) |
        is.na(GARGANTA_aux1) |
        is.na(DISPNEIA_aux1) |
        is.na(DESC_RESP_aux1) |
        is.na(SATURACAO_aux1) | is.na(DIARREIA_aux1) |
        is.na(VOMITO_aux1) |
        is.na(DOR_ABD_aux1) |
        is.na(FADIGA_aux1) |
        is.na(PERD_OLFT_aux1) | is.na(PERD_PALA_aux1) ~ NA_real_,
      TRUE ~ cont_sintomas
    ),
    gr_sintomas = case_when(
      num_sintomas == 0 ~ 0,
      num_sintomas == 1 ~ 1,
      num_sintomas == 2 ~ 1,
      num_sintomas >  2 ~ 2,
      TRUE ~ NA_real_
    ),
    sintomas_SN = case_when(
      gr_sintomas == 0 ~ 0,
      gr_sintomas == 1 ~ 1,
      gr_sintomas == 2 ~ 1,
      TRUE ~ NA_real_
    )
  )
```

```r
#Symptom group
data6$gr_sintomas <- factor(
  data6$gr_sintomas,
  levels = c(0, 1, 2),
  labels = c("none", "1 or 2", ">2")
)

#At least one symptom indicator
data6$sintomas_SN <- factor(data6$sintomas_SN,
                            levels = c(1, 0),
                            labels = c("yes", "no"))
```

```r
questionr::freq(
  data6$gr_sintomas,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for symptom group", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 36: Frequency table for symptom group

	n	%
none	338	0.8
1 or 2	2347	5.8
>2	12028	29.6
NA	25927	63.8
Total	40640	100.0

```r
questionr::freq(
  data6$sintomas_SN,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for at least one symptom indicator", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 37: Frequency table for at least one symptom indicator

	n	%
yes	14375	35.4
no	338	0.8
NA	25927	63.8
Total	40640	100.0

An indicator variable of at least one respiratory symptom (`sint_resp`) is created in the following.

```r
resp <-  c("DISPNEIA_aux", "DESC_RESP_aux", "SATURACAO_aux")

resp1 <-  c("DISPNEIA_aux1", "DESC_RESP_aux1", "SATURACAO_aux1")

data6 <-
  mutate(
    data6,
    DISPNEIA_aux = DISPNEIA,
    DESC_RESP_aux = DESC_RESP,
    SATURACAO_aux = SATURACAO
  )

data6 <-
  mutate(
    data6,
    DISPNEIA_aux1 = DISPNEIA,
    DESC_RESP_aux1 = DESC_RESP,
    SATURACAO_aux1 = SATURACAO
  )

data6 <-  data6 %>%
  mutate_at(all_of(resp), function(x) {
    case_when(x == "1" ~ 1, TRUE ~ 0)
  }) %>%
  mutate_at(all_of(resp1), function(x) {
    case_when(x == "1" ~ 1, x == "2" ~ 0, TRUE ~ NA_real_)
  }) %>%
  mutate(cont_resp = DISPNEIA_aux + DESC_RESP_aux + SATURACAO_aux) %>%
  mutate(
    num_resp = case_when(
      (cont_resp == 0) &
        (
          is.na(DISPNEIA_aux1) |
            is.na(DESC_RESP_aux1) | is.na(SATURACAO_aux1)
        ) ~ NA_real_,
      TRUE ~ cont_resp
    ),
    sint_resp = case_when(
      num_resp == 0 ~ 0,
      num_resp == 1 ~ 1,
      num_resp == 2 ~ 1,
      num_resp == 3 ~ 1,
      TRUE ~ NA_real_
    )
  )

# Any respiratory symptom indicator
data6$sint_resp <- factor(data6$sint_resp,
                          levels = c(1, 0),
                          labels = c("yes", "no"))
```

```r
questionr::freq(
  data6$sint_resp,
```

```
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for any respiratory symptom", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 38: Frequency table for any respiratory symptom

	n	%
yes	32875	80.9
no	4851	11.9
NA	2914	7.2
Total	40640	100.0

The SARI (severe acute respiratory infection) indicator (`sari`) is "yes" if one has fever and cough or sore throat and respiratory distress or dyspnoea or saturation. The SARI without fever indicator (`sari_sfebre`) is what the name says.

```
data6 <-  data6 %>%
  mutate(
    sari = case_when(
      FEBRE == "1" &
        (TOSSE == "1" | GARGANTA == "1") &
        (DESC_RESP == "1" |
           DISPNEIA == "1" | SATURACAO == "1") ~ 1,
      is.na(FEBRE_aux1) |
        (is.na(TOSSE_aux1)  &
           is.na(GARGANTA_aux1)) |
        (
          is.na(DESC_RESP_aux1) &
            is.na(DISPNEIA_aux1) &  is.na(SATURACAO_aux1)
        ) ~ NA_real_,
      TRUE ~ 0
    ),
    sari_sfebre = case_when(
      (TOSSE == "1" | GARGANTA == "1") &
        (DESC_RESP == "1" | DISPNEIA == "1" | SATURACAO == "1") ~ 1,
      (is.na(TOSSE_aux1)  &
         is.na(GARGANTA_aux1)) |
        (
          is.na(DESC_RESP_aux1) &
            is.na(DISPNEIA_aux1) &  is.na(SATURACAO_aux1)
        ) ~ NA_real_,
      TRUE ~ 0
    )
  )

#SARI
data6$sari <- factor(data6$sari,
                     levels = c(1, 0),
```

```
                        labels = c("yes", "no"))
#SARI without fever
data6$sari_sfebre <- factor(data6$sari_sfebre,
                            levels = c(1, 0),
                            labels = c("yes", "no"))
```

```
questionr::freq(
  data6$sari,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for SARI", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 39: Frequency table for SARI

	n	%
yes	19639	48.3
no	14479	35.6
NA	6522	16.0
Total	40640	100.0

```
questionr::freq(
  data6$sari_sfebre,
  cum = FALSE,
  total = TRUE,
  na.last = FALSE,
  valid = FALSE
) %>%
  knitr::kable(caption = "Frequency table for SARI without fever", digits = 2) %>%
  kable_styling(latex_options = "hold_position")
```

Table 40: Frequency table for SARI without fever

	n	%
yes	26671	65.6
no	9268	22.8
NA	4701	11.6
Total	40640	100.0

4. Propensity score matching (PSM)

4.1 PSM for symptoms variables

We considered as control variables: age, ethnicity, cardiopathy, asthma, diabetes, immunodepression and obesity.

First, we present the difference result among the groups regarding the control variables before the PSM. We consider as "balanced" the cases with mean difference greater than 0.05. As we can see, for most categories of control variables the groups are not balanced before the matching.

```
bal.tab(gest_puerp ~ NU_IDADE_N + raca + cardiopati + asma + diabetes + imunodepre + obesidade, data = d
```

```
## Balance summary across all treatment pairs
##                     Type Max.Diff.Un      M.Threshold.Un
## NU_IDADE_N_     Contin.      1.1385 Not Balanced, >0.05
## raca_white       Binary      0.1686 Not Balanced, >0.05
## raca_black       Binary      0.0130     Balanced, <0.05
## raca_yellow      Binary      0.0063     Balanced, <0.05
## raca_brown       Binary      0.1590 Not Balanced, >0.05
## raca_indigenous  Binary      0.0028     Balanced, <0.05
## raca:<NA>        Binary      0.0251     Balanced, <0.05
## cardiopati_no    Binary      0.2811 Not Balanced, >0.05
## cardiopati:<NA>  Binary      0.2976 Not Balanced, >0.05
## asma_no          Binary      0.0914 Not Balanced, >0.05
## asma:<NA>        Binary      0.2880 Not Balanced, >0.05
## diabetes_no      Binary      0.2486 Not Balanced, >0.05
## diabetes:<NA>    Binary      0.2758 Not Balanced, >0.05
## imunodepre_no    Binary      0.0742 Not Balanced, >0.05
## imunodepre:<NA>  Binary      0.2884 Not Balanced, >0.05
## obesidade_no     Binary      0.1981 Not Balanced, >0.05
## obesidade:<NA>   Binary      0.2805 Not Balanced, >0.05
##
## Balance tally for mean differences
##                       count
## Balanced, <0.05         4
## Not Balanced, >0.05    13
##
## Variable with the greatest mean difference
##     Variable Max.Diff.Un      M.Threshold.Un
##  NU_IDADE_N_      1.1385 Not Balanced, >0.05
##
## Sample sizes
##        no preg puerp
## All 36474 3372   794
```

After PSM, all categories of control variables are balanced, with the exception of age (table below). Although the mean difference is greater than 0.05, it substantially decreased after the PSM: from 1.1385 to 0.2684.

```
w.out <-
  weightit(
    gest_puerp ~ NU_IDADE_N + raca + cardiopati + asma + diabetes + imunodepre + obesidade,
    use.mlogit = FALSE,
    data = data6,
    focal = "puerp",
    method = "ps",
    estimand = "ATT"
  )
```

```r
cobalt::bal.tab(w.out, m.threshold = 0.05, disp.v.ratio = TRUE)
```

```
## Call
##  weightit(formula = gest_puerp ~ NU_IDADE_N + raca + cardiopati +
##     asma + diabetes + imunodepre + obesidade, data = data6, method = "ps",
##     estimand = "ATT", focal = "puerp", use.mlogit = FALSE)
##
## Balance summary across all treatment pairs
##                     Type Max.Diff.Adj         M.Threshold Max.V.Ratio.Adj
## NU_IDADE_N      Contin.       0.2684 Not Balanced, >0.05          2.2665
## raca_white       Binary       0.0080     Balanced, <0.05               .
## raca_black       Binary       0.0035     Balanced, <0.05               .
## raca_yellow      Binary       0.0011     Balanced, <0.05               .
## raca_brown       Binary       0.0105     Balanced, <0.05               .
## raca_indigenous  Binary       0.0008     Balanced, <0.05               .
## raca:<NA>        Binary       0.0062     Balanced, <0.05               .
## cardiopati_no    Binary       0.0104     Balanced, <0.05               .
## cardiopati:<NA>  Binary       0.0099     Balanced, <0.05               .
## asma_no          Binary       0.0045     Balanced, <0.05               .
## asma:<NA>        Binary       0.0122     Balanced, <0.05               .
## diabetes_no      Binary       0.0077     Balanced, <0.05               .
## diabetes:<NA>    Binary       0.0120     Balanced, <0.05               .
## imunodepre_no    Binary       0.0019     Balanced, <0.05               .
## imunodepre:<NA>  Binary       0.0120     Balanced, <0.05               .
## obesidade_no     Binary       0.0095     Balanced, <0.05               .
## obesidade:<NA>   Binary       0.0122     Balanced, <0.05               .
##
## Balance tally for mean differences
##                       count
## Balanced, <0.05          16
## Not Balanced, >0.05       1
##
## Variable with the greatest mean difference
##    Variable Max.Diff.Adj         M.Threshold
##  NU_IDADE_N       0.2684 Not Balanced, >0.05
##
## Effective sample sizes
##                   no    preg puerp
## Unadjusted 36474.   3372.     794
## Adjusted    3476.39 2296.34   794
```

4.2 PSM for outcomes

For the analysis of the outcomes, only the cases that we know whether it is a case of death or cure are selected.

```r
data6 <-
  data6 %>% mutate(
    evolucao = case_when(
      EVOLUCAO == 1 ~ "cure",
      EVOLUCAO == 2 ~ "death",
      EVOLUCAO == 3 ~ "death",
```

```
      TRUE ~ "in progress"
    )
  )
```

Now we exclude cases "in progress".

```
data7 <- filter(data6, evolucao != "in progress")

data7$evolucao <- factor(
  data7$evolucao,
  levels = c("death", "cure"),
  labels = c("death", "cure")
)
```

In this propensity score, we considered as control variables: age, ethnicity, school, Brazilian Federative Unit, cardiopathy, asthma, diabetes, immunodepression, obesity and respiratory symptoms.

We present below the difference result among the groups regarding the control variables before the PSM. We consider as "balanced" the cases with mean difference greater than 0.05. As we can see, the groups are not balanced before the matching for many categories of the control variables.

```
bal.tab(gest_puerp ~ NU_IDADE_N + raca + escol + SG_UF + cardiopati + asma +
                     diabetes + imunodepre + obesidade + sint_resp, data = data7, estimand = "ATT", m.th
```

```
## Balance summary across all treatment pairs
##                              Type Max.Diff.Un      M.Threshold.Un
## NU_IDADE_N_                Contin.      1.1475 Not Balanced, >0.05
## raca_white                  Binary      0.1723 Not Balanced, >0.05
## raca_black                  Binary      0.0126     Balanced, <0.05
## raca_yellow                 Binary      0.0070     Balanced, <0.05
## raca_brown                  Binary      0.1658 Not Balanced, >0.05
## raca_indigenous             Binary      0.0036     Balanced, <0.05
## raca:<NA>                   Binary      0.0330     Balanced, <0.05
## escol_no education          Binary      0.0098     Balanced, <0.05
## escol_up to high school     Binary      0.0089     Balanced, <0.05
## escol_high school           Binary      0.0720 Not Balanced, >0.05
## escol_higher education      Binary      0.0711 Not Balanced, >0.05
## escol:<NA>                  Binary      0.0077     Balanced, <0.05
## SG_UF_AC                    Binary      0.0037     Balanced, <0.05
## SG_UF_AL                    Binary      0.0025     Balanced, <0.05
## SG_UF_AM                    Binary      0.0222     Balanced, <0.05
## SG_UF_AP                    Binary      0.0007     Balanced, <0.05
## SG_UF_BA                    Binary      0.0293     Balanced, <0.05
## SG_UF_CE                    Binary      0.0349     Balanced, <0.05
## SG_UF_DF                    Binary      0.0315     Balanced, <0.05
## SG_UF_ES                    Binary      0.0036     Balanced, <0.05
## SG_UF_GO                    Binary      0.0161     Balanced, <0.05
## SG_UF_MA                    Binary      0.0130     Balanced, <0.05
## SG_UF_MG                    Binary      0.0237     Balanced, <0.05
## SG_UF_MS                    Binary      0.0132     Balanced, <0.05
## SG_UF_MT                    Binary      0.0086     Balanced, <0.05
## SG_UF_PA                    Binary      0.0128     Balanced, <0.05
## SG_UF_PB                    Binary      0.0105     Balanced, <0.05
```

```
## SG_UF_PE                 Binary      0.0631 Not Balanced, >0.05
## SG_UF_PI                 Binary      0.0137     Balanced, <0.05
## SG_UF_PR                 Binary      0.0198     Balanced, <0.05
## SG_UF_RJ                 Binary      0.0178     Balanced, <0.05
## SG_UF_RN                 Binary      0.0095     Balanced, <0.05
## SG_UF_RO                 Binary      0.0052     Balanced, <0.05
## SG_UF_RR                 Binary      0.0009     Balanced, <0.05
## SG_UF_RS                 Binary      0.0281     Balanced, <0.05
## SG_UF_SC                 Binary      0.0090     Balanced, <0.05
## SG_UF_SE                 Binary      0.0088     Balanced, <0.05
## SG_UF_SP                 Binary      0.1169 Not Balanced, >0.05
## SG_UF_TO                 Binary      0.0021     Balanced, <0.05
## SG_UF:<NA>               Binary      0.0014     Balanced, <0.05
## cardiopati_no            Binary      0.2775 Not Balanced, >0.05
## cardiopati:<NA>          Binary      0.3056 Not Balanced, >0.05
## asma_no                  Binary      0.0954 Not Balanced, >0.05
## asma:<NA>                Binary      0.2868 Not Balanced, >0.05
## diabetes_no              Binary      0.2381 Not Balanced, >0.05
## diabetes:<NA>            Binary      0.2843 Not Balanced, >0.05
## imunodepre_no            Binary      0.0728 Not Balanced, >0.05
## imunodepre:<NA>          Binary      0.2930 Not Balanced, >0.05
## obesidade_no             Binary      0.1970 Not Balanced, >0.05
## obesidade:<NA>           Binary      0.2858 Not Balanced, >0.05
## sint_resp_no             Binary      0.1953 Not Balanced, >0.05
## sint_resp:<NA>           Binary      0.0485     Balanced, <0.05
##
## Balance tally for mean differences
##                       count
## Balanced, <0.05          34
## Not Balanced, >0.05      18
##
## Variable with the greatest mean difference
##     Variable Max.Diff.Un      M.Threshold.Un
##  NU_IDADE_N_      1.1475 Not Balanced, >0.05
##
## Sample sizes
##        no preg puerp
## All 32081 2904   715
```

After PSM, all categories of control variables are balanced, with the exception of age and the "up to high school" category (table below). Despite this, we can see that the mean difference value for the "up to high school" category (value of 0.0502) is very close to the threshold and for age, the mean difference is substantially decreased after the PSM: from 1.1475 to 0.3693.

```
remove(w.out)

w.out <- weightit(gest_puerp ~ NU_IDADE_N + raca + escol + SG_UF + cardiopati + asma +
                    diabetes + imunodepre + obesidade + sint_resp, use.mlogit = FALSE,
                  data = data7, focal = "puerp", method = "ps", estimand = "ATT")
```

```
cobalt::bal.tab(w.out, m.threshold = 0.05, disp.v.ratio = TRUE)
```

```
## Call
```

```
##  weightit(formula = gest_puerp ~ NU_IDADE_N + raca + escol + SG_UF +
##     cardiopati + asma + diabetes + imunodepre + obesidade + sint_resp,
##     data = data7, method = "ps", estimand = "ATT", focal = "puerp",
##     use.mlogit = FALSE)
##
## Balance summary across all treatment pairs
##                            Type Max.Diff.Adj         M.Threshold
## NU_IDADE_N              Contin.       0.3693 Not Balanced, >0.05
## raca_white               Binary       0.0108     Balanced, <0.05
## raca_black               Binary       0.0077     Balanced, <0.05
## raca_yellow              Binary       0.0025     Balanced, <0.05
## raca_brown               Binary       0.0207     Balanced, <0.05
## raca_indigenous          Binary       0.0014     Balanced, <0.05
## raca:<NA>                Binary       0.0064     Balanced, <0.05
## escol_no education       Binary       0.0041     Balanced, <0.05
## escol_up to high school  Binary       0.0502 Not Balanced, >0.05
## escol_high school        Binary       0.0450     Balanced, <0.05
## escol_higher education   Binary       0.0066     Balanced, <0.05
## escol:<NA>               Binary       0.0342     Balanced, <0.05
## SG_UF_AC                 Binary       0.0002     Balanced, <0.05
## SG_UF_AL                 Binary       0.0011     Balanced, <0.05
## SG_UF_AM                 Binary       0.0054     Balanced, <0.05
## SG_UF_AP                 Binary       0.0003     Balanced, <0.05
## SG_UF_BA                 Binary       0.0007     Balanced, <0.05
## SG_UF_CE                 Binary       0.0010     Balanced, <0.05
## SG_UF_DF                 Binary       0.0119     Balanced, <0.05
## SG_UF_ES                 Binary       0.0032     Balanced, <0.05
## SG_UF_GO                 Binary       0.0043     Balanced, <0.05
## SG_UF_MA                 Binary       0.0010     Balanced, <0.05
## SG_UF_MG                 Binary       0.0082     Balanced, <0.05
## SG_UF_MS                 Binary       0.0020     Balanced, <0.05
## SG_UF_MT                 Binary       0.0037     Balanced, <0.05
## SG_UF_PA                 Binary       0.0008     Balanced, <0.05
## SG_UF_PB                 Binary       0.0048     Balanced, <0.05
## SG_UF_PE                 Binary       0.0328     Balanced, <0.05
## SG_UF_PI                 Binary       0.0002     Balanced, <0.05
## SG_UF_PR                 Binary       0.0045     Balanced, <0.05
## SG_UF_RJ                 Binary       0.0051     Balanced, <0.05
## SG_UF_RN                 Binary       0.0031     Balanced, <0.05
## SG_UF_RO                 Binary       0.0008     Balanced, <0.05
## SG_UF_RR                 Binary       0.0002     Balanced, <0.05
## SG_UF_RS                 Binary       0.0058     Balanced, <0.05
## SG_UF_SC                 Binary       0.0038     Balanced, <0.05
## SG_UF_SE                 Binary       0.0029     Balanced, <0.05
## SG_UF_SP                 Binary       0.0129     Balanced, <0.05
## SG_UF_TO                 Binary       0.0008     Balanced, <0.05
## SG_UF:<NA>               Binary       0.0014     Balanced, <0.05
## cardiopati_no            Binary       0.0149     Balanced, <0.05
## cardiopati:<NA>          Binary       0.0101     Balanced, <0.05
## asma_no                  Binary       0.0060     Balanced, <0.05
## asma:<NA>                Binary       0.0138     Balanced, <0.05
## diabetes_no              Binary       0.0084     Balanced, <0.05
## diabetes:<NA>            Binary       0.0128     Balanced, <0.05
## imunodepre_no            Binary       0.0026     Balanced, <0.05
```

```
## imunodepre:<NA>          Binary       0.0116     Balanced, <0.05
## obesidade_no             Binary       0.0095     Balanced, <0.05
## obesidade:<NA>           Binary       0.0138     Balanced, <0.05
## sint_resp_no             Binary       0.0061     Balanced, <0.05
## sint_resp:<NA>           Binary       0.0159     Balanced, <0.05
##                        Max.V.Ratio.Adj
## NU_IDADE_N                      2.4696
## raca_white                           .
## raca_black                           .
## raca_yellow                          .
## raca_brown                           .
## raca_indigenous                      .
## raca:<NA>                            .
## escol_no education                   .
## escol_up to high school              .
## escol_high school                    .
## escol_higher education               .
## escol:<NA>                           .
## SG_UF_AC                             .
## SG_UF_AL                             .
## SG_UF_AM                             .
## SG_UF_AP                             .
## SG_UF_BA                             .
## SG_UF_CE                             .
## SG_UF_DF                             .
## SG_UF_ES                             .
## SG_UF_GO                             .
## SG_UF_MA                             .
## SG_UF_MG                             .
## SG_UF_MS                             .
## SG_UF_MT                             .
## SG_UF_PA                             .
## SG_UF_PB                             .
## SG_UF_PE                             .
## SG_UF_PI                             .
## SG_UF_PR                             .
## SG_UF_RJ                             .
## SG_UF_RN                             .
## SG_UF_RO                             .
## SG_UF_RR                             .
## SG_UF_RS                             .
## SG_UF_SC                             .
## SG_UF_SE                             .
## SG_UF_SP                             .
## SG_UF_TO                             .
## SG_UF:<NA>                           .
## cardiopati_no                        .
## cardiopati:<NA>                      .
## asma_no                              .
## asma:<NA>                            .
## diabetes_no                          .
## diabetes:<NA>                        .
## imunodepre_no                        .
## imunodepre:<NA>                      .
```

```
## obesidade_no                          .
## obesidade:<NA>                        .
## sint_resp_no                          .
## sint_resp:<NA>                        .
##
## Balance tally for mean differences
##                     count
## Balanced, <0.05        50
## Not Balanced, >0.05     2
##
## Variable with the greatest mean difference
##    Variable Max.Diff.Adj         M.Threshold
##  NU_IDADE_N       0.3693 Not Balanced, >0.05
##
## Effective sample sizes
##                   no    preg puerp
## Unadjusted 32081.   2904.     715
## Adjusted    1062.01 1739.87   715
```