Early-life socioeconomic inequality and biological aging in Colombia 2\_ Preparing data for analysis

Juan Rivillas

28/02/2022

## Contents Code 2: Data processing and preparation for analysis.

Two main steps are described in this markdown.

**Step 1. Preparing work space (import imputed data)**. This step includes: 1.1. Working directory and loading of packages required for the tidying data. 1.2. Importing of imputed data base. 1.3. Verifying upload and exploring imputed data base. 1.4. Overview of the entire dataframe (missing values, complete rates, histograms).

**Step 2. Tidying data: organize data in a consistent data structure**. This includes: 2.1. Vector transformation (re-coding, standardizing values and creating new functions) and computing frequency tables. 2.2. Create new combination of vectors not present in the data (simultaneously with 2.2). 2.3. Cleaning and selecting final data by eliminating columns no needed to calculate BHS. 2.4. Baseline characteristics of population study and applying preliminary statistical tests.

## Step 1. Preparing work space (import imputed data).

**1.1. Working directory and loading of packages required for the tidying data**.

Assign file pathway to save automatically plots and outcomes in pdf format.

setwd("/Users/macbookpro/Documents/PhD Project/Data/SABE/Markdowns")

Install CRAN packages and install the relevant libraries. This code chunk shows the loading of packages required for the analyses. install.packages: dplyr, ggplot2, vcdExtra,epiDisplay, data.table, formattable, table1, haven, devtools, DescTools.

Loading the relevant libraries into the work space to tidying and transformation data. library(tidyverse) #Tidy data and transformation library(dplyr)  
library(epiDisplay) #to make frequency table in r (nicer version) library(DescTools) #single and cumulative frequencies values are reported. library(table1) #to Create HTML Tables of Descriptive Statistics in epi library(“ggplot2”) #improved plots and bar charts library(haven) library(devtools)

library(tidyverse) #Tidy data and transformation

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.5 ✓ dplyr 1.0.8  
## ✓ tidyr 1.1.4 ✓ stringr 1.4.0  
## ✓ readr 2.1.0 ✓ forcats 0.5.1

## Warning: package 'dplyr' was built under R version 4.1.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)   
library(epiDisplay) #to make frequency table in r (nicer version)

## Loading required package: foreign

## Loading required package: survival

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

## Loading required package: nnet

##   
## Attaching package: 'epiDisplay'

## The following object is masked from 'package:ggplot2':  
##   
## alpha

library(DescTools) #single and cumulative frequencies values are reported.  
library(table1) #to Create HTML Tables of Descriptive Statistics in epi

##   
## Attaching package: 'table1'

## The following objects are masked from 'package:base':  
##   
## units, units<-

library("ggplot2") #improved plots and bar charts  
library(haven)  
library(devtools)

## Loading required package: usethis

**1.2. Importing of imputed SABE data base (Loading data EXCEL files)**.

library(readxl)  
biomarkers.imp <- read\_excel("~/Documents/PhD Project/Data/SABE/Dataframes/biomarkers.imp.xlsx")

Another option for CSV files: biomarkers.imp <- read\_csv(“biomarkers.imp.csv”)

**1.3. Verifying upload and exploring imputed data base**. Get and overview of the data (types and dimensions of observations and columns). We already did this in Code 1 at the end, but for quality control: inspect the variables, the total number of observations, a full list of the variables names, the data type of each variable, the first few observation for each variable, and dimension of data.

class(biomarkers.imp) names(biomarkers.imp) glimpse(biomarkers.imp) dim(biomarkers.imp) head(biomarkers.imp, n=20) #Print out first observations in data frame tail(biomarkers.imp, n=20) #Print out last observations in data frame

Data summary: names columns, number of observations, n missing values, completed rates, and type of data (caracter or numeric)

skimr::skim(biomarkers.imp)

Data summary

|  |  |
| --- | --- |
| Name | biomarkers.imp |
| Number of rows | 4092 |
| Number of columns | 52 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 9 |
| numeric | 43 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| area\_residence | 0 | 1.00 | 1 | 1 | 0 | 3 | 0 |
| P835 | 1080 | 0.74 | 3 | 61 | 0 | 1218 | 0 |
| P835\_1 | 1614 | 0.61 | 3 | 128 | 0 | 1087 | 0 |
| P835\_2 | 2181 | 0.47 | 2 | 65 | 0 | 899 | 0 |
| P835\_3 | 2746 | 0.33 | 3 | 43 | 0 | 650 | 0 |
| P835\_4 | 3217 | 0.21 | 3 | 44 | 0 | 508 | 0 |
| P835\_5 | 3527 | 0.14 | 3 | 38 | 0 | 347 | 0 |
| P835\_6 | 3752 | 0.08 | 3 | 34 | 0 | 242 | 0 |
| P835\_7 | 3882 | 0.05 | 5 | 42 | 0 | 167 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | 0 | 1 | 11844.32 | 7213.21 | 2.0 | 5063.25 | 11773.0 | 18281.50 | 23646.0 | ▇▅▇▆▇ |
| current\_sep | 0 | 1 | 1.98 | 0.84 | 1.0 | 1.00 | 2.0 | 3.00 | 6.0 | ▇▂▁▁▁ |
| sep\_occupation | 0 | 1 | 5.43 | 11.32 | 1.0 | 1.00 | 3.0 | 5.00 | 99.0 | ▇▁▁▁▁ |
| family\_sep | 0 | 1 | 1.83 | 0.91 | 1.0 | 1.00 | 2.0 | 2.00 | 9.0 | ▇▂▁▁▁ |
| educational\_level | 0 | 1 | 3.20 | 6.44 | 1.0 | 2.00 | 2.0 | 3.00 | 99.0 | ▇▁▁▁▁ |
| health\_insurance | 0 | 1 | 1.69 | 0.77 | 1.0 | 1.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| neglected\_food | 0 | 1 | 1.78 | 0.73 | 1.0 | 1.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| household\_violence | 0 | 1 | 1.85 | 0.54 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| migration | 0 | 1 | 1.84 | 0.37 | 1.0 | 2.00 | 2.0 | 2.00 | 2.0 | ▂▁▁▁▇ |
| migration\_yo | 0 | 1 | 42.28 | 22.08 | 1.0 | 21.00 | 48.0 | 58.00 | 99.0 | ▆▃▇▅▁ |
| emotional\_abuse | 0 | 1 | 4.00 | 0.78 | 1.0 | 4.00 | 4.0 | 4.00 | 9.0 | ▁▇▁▁▁ |
| poor\_health2 | 0 | 1 | 1.90 | 0.55 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection1 | 0 | 1 | 2.02 | 0.46 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection2 | 0 | 1 | 2.00 | 0.47 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection3 | 0 | 1 | 1.65 | 1.08 | 1.0 | 1.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection4 | 0 | 1 | 2.01 | 0.34 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection5 | 0 | 1 | 2.00 | 0.53 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection6 | 0 | 1 | 1.97 | 0.39 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| early\_infection7 | 0 | 1 | 2.00 | 0.66 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| hdl | 0 | 1 | 45.56 | 13.54 | 5.0 | 36.00 | 44.0 | 53.00 | 120.0 | ▁▇▃▁▁ |
| ldl | 0 | 1 | 126.20 | 35.97 | 30.0 | 101.00 | 125.0 | 149.00 | 338.0 | ▂▇▂▁▁ |
| total\_cholesterol | 0 | 1 | 194.57 | 41.79 | 60.0 | 166.00 | 193.0 | 221.00 | 439.0 | ▁▇▃▁▁ |
| triglycerides | 0 | 1 | 162.99 | 96.85 | 42.0 | 105.00 | 141.0 | 192.25 | 2062.0 | ▇▁▁▁▁ |
| glucose | 0 | 1 | 100.76 | 36.79 | 31.0 | 86.00 | 93.0 | 103.00 | 543.0 | ▇▁▁▁▁ |
| glycated\_haemoglobin | 0 | 1 | 13.80 | 1.73 | 0.2 | 12.70 | 13.8 | 14.90 | 24.1 | ▁▁▇▃▁ |
| weight | 0 | 1 | 73.45 | 76.97 | 30.0 | 56.00 | 65.0 | 73.25 | 777.0 | ▇▁▁▁▁ |
| height | 0 | 1 | 146.59 | 125.20 | 1.0 | 144.00 | 153.0 | 162.00 | 777.0 | ▇▅▁▁▁ |
| wc | 0 | 1 | 105.47 | 93.05 | 66.0 | 85.00 | 93.0 | 101.00 | 777.0 | ▇▁▁▁▁ |
| systolic\_bprh | 0 | 1 | 221.01 | 606.94 | 84.0 | 122.00 | 137.0 | 156.00 | 8888.0 | ▇▁▁▁▁ |
| dyastolic\_bprh | 0 | 1 | 140.14 | 610.19 | 43.0 | 68.00 | 76.0 | 90.00 | 8888.0 | ▇▁▁▁▁ |
| systolic\_bplh | 0 | 1 | 283.15 | 1027.34 | 60.0 | 116.00 | 132.0 | 151.00 | 8888.0 | ▇▁▁▁▁ |
| dyastolic\_bplh | 0 | 1 | 94.97 | 227.39 | 42.0 | 68.00 | 75.0 | 83.00 | 8888.0 | ▇▁▁▁▁ |
| hta | 0 | 1 | 1.45 | 0.58 | 1.0 | 1.00 | 1.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| diabetes | 0 | 1 | 1.85 | 0.52 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| cvd | 0 | 1 | 1.86 | 0.40 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| medication | 0 | 1 | 1.26 | 0.44 | 1.0 | 1.00 | 1.0 | 2.00 | 2.0 | ▇▁▁▁▃ |
| age | 0 | 1 | 70.49 | 8.03 | 60.0 | 64.00 | 69.0 | 76.00 | 105.0 | ▇▅▂▁▁ |
| sex | 0 | 1 | 1.60 | 0.49 | 1.0 | 1.00 | 2.0 | 2.00 | 2.0 | ▅▁▁▁▇ |
| ethnic\_group | 0 | 1 | 4.43 | 1.61 | 1.0 | 4.00 | 5.0 | 5.00 | 9.0 | ▂▆▇▁▁ |
| alcohol | 0 | 1 | 4.58 | 1.21 | 1.0 | 5.00 | 5.0 | 5.00 | 9.0 | ▁▁▇▁▁ |
| smoking | 0 | 1 | 2.93 | 1.14 | 1.0 | 2.00 | 4.0 | 4.00 | 9.0 | ▇▇▁▁▁ |
| cash\_transfer1 | 0 | 1 | 1.98 | 0.39 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |
| cash\_transfer2 | 0 | 1 | 1.79 | 0.59 | 1.0 | 2.00 | 2.0 | 2.00 | 9.0 | ▇▁▁▁▁ |

**Step 2. Tidying data: organize data in a consistent data structure**.

Note: original SABE database is a “tidy” dataset: each column is a variable, each row is an observation, and each cell is a single value. However, there are variables not present in the data we need create (for instance early infection, adverse childhood experiences, BMI, and age groups). Additionally, levels of the variables need to be changed and SEP variables were regrouped accordingly to synthetise number of levels.

**2.1. Vector transformation and computing frequency tables**. Start by converting into a tibble data frame (biomarkers), assigning column labels in the data, and computing frequency tables (mainly demographics, SEP variables, and health outcomes).

**Sociodemographic data** There are four demographics characteristics variables: sex, age, ethnic groups, and area of residence.

Sex

biomarkers.imp$sex <- factor(biomarkers.imp$sex,  
 levels = c("1","2"),  
 labels = c("male", "female"))  
  
#Desc(biomarkers.imp$sex)  
#Freq(biomarkers.imp$sex)

Age: to create vector age groups.

biomarkers.imp$age\_group <- cut( x = biomarkers.imp$age, breaks = c(60, 70, 80, Inf))  
  
table(biomarkers.imp$age\_group)

##   
## (60,70] (70,80] (80,Inf]   
## 2068 1235 545

#Desc(biomarkers.imp$age\_group)  
#Freq(biomarkers.imp$age\_group, useNA = "always")

Verify that each age value was assigned to the correct category by cross-tabulating the numeric and category columns. Cross tabulation of the numeric and category columns.

table("Numeric Values" = biomarkers.imp$age, # names specified in table for clarity.  
 "Categories" = biomarkers.imp$age\_group,  
 useNA = "always") # don't forget to examine NA values

## Categories  
## Numeric Values (60,70] (70,80] (80,Inf] <NA>  
## 60 0 0 0 244  
## 61 230 0 0 0  
## 62 246 0 0 0  
## 63 245 0 0 0  
## 64 221 0 0 0  
## 65 213 0 0 0  
## 66 191 0 0 0  
## 67 180 0 0 0  
## 68 195 0 0 0  
## 69 186 0 0 0  
## 70 161 0 0 0  
## 71 0 164 0 0  
## 72 0 142 0 0  
## 73 0 131 0 0  
## 74 0 145 0 0  
## 75 0 139 0 0  
## 76 0 128 0 0  
## 77 0 102 0 0  
## 78 0 100 0 0  
## 79 0 91 0 0  
## 80 0 93 0 0  
## 81 0 0 75 0  
## 82 0 0 79 0  
## 83 0 0 59 0  
## 84 0 0 70 0  
## 85 0 0 56 0  
## 86 0 0 39 0  
## 87 0 0 40 0  
## 88 0 0 23 0  
## 89 0 0 29 0  
## 90 0 0 26 0  
## 91 0 0 13 0  
## 92 0 0 11 0  
## 93 0 0 7 0  
## 94 0 0 5 0  
## 95 0 0 3 0  
## 96 0 0 1 0  
## 97 0 0 3 0  
## 98 0 0 2 0  
## 99 0 0 2 0  
## 101 0 0 1 0  
## 105 0 0 1 0  
## <NA> 0 0 0 0

Ethnic group: Seven ethnic groups regrouped into five main groups according to race and ethnic classification in Colombia.

biomarkers.imp$ethnic\_group <- factor(biomarkers.imp$ethnic\_group,  
 levels = c("1","2","3","4","5","7","8","9"),  
 labels = c("indigenous", "afrocolombian","afrocolombian","white","mestize","mestize","Not reply/Don't know", "Not reply/Don't know"))  
  
#Desc(biomarkers.imp$ethnic\_group)  
#Freq(biomarkers.imp$ethnic\_group, useNA = "always")

Area of residence regrouped into two main groups.

biomarkers.imp$area\_residence <- factor(biomarkers.imp$area\_residence,  
 levels = c("1","2","3"),  
 labels = c("urban", "rural","rural"))  
  
#Desc(biomarkers.imp$area\_residence)  
#Freq(biomarkers.imp$area\_residence)

Summary demographic characteristics by sex.

table\_1<- table1(~age\_group + ethnic\_group + area\_residence| sex, data=biomarkers.imp)  
  
table\_1

|  | male (N=1635) | female (N=2457) | Overall (N=4092) |
| --- | --- | --- | --- |
| **age\_group** |  |  |  |
| (60,70] | 795 (48.6%) | 1273 (51.8%) | 2068 (50.5%) |
| (70,80] | 518 (31.7%) | 717 (29.2%) | 1235 (30.2%) |
| (80,Inf] | 241 (14.7%) | 304 (12.4%) | 545 (13.3%) |
| Missing | 81 (5.0%) | 163 (6.6%) | 244 (6.0%) |
| **ethnic\_group** |  |  |  |
| indigenous | 159 (9.7%) | 116 (4.7%) | 275 (6.7%) |
| afrocolombian | 182 (11.1%) | 231 (9.4%) | 413 (10.1%) |
| white | 462 (28.3%) | 758 (30.9%) | 1220 (29.8%) |
| mestize | 776 (47.5%) | 1223 (49.8%) | 1999 (48.9%) |
| Not reply/Don't know | 56 (3.4%) | 129 (5.3%) | 185 (4.5%) |
| **area\_residence** |  |  |  |
| urban | 1277 (78.1%) | 2074 (84.4%) | 3351 (81.9%) |
| rural | 358 (21.9%) | 383 (15.6%) | 741 (18.1%) |

**Socioeconomic position data**.

There are five socioeconomic position variables:

Two of early-life SEP variables: i) family SEP and ii) Adverse Childhood Experiences ACEs. Three of current SEP variables: i) current SEP, ii) occupation status, and iii) educational level.

An ordered factor with the categories of early-life and current socioeconomic position was created to deal with the natural ordering (Except for ACEs, which are unordered and it is described below).

levels: a) Low b) Intermediate c) High

Family socioeconomic position (reported before 15 year old).

biomarkers.imp$family\_sep <- factor(biomarkers.imp$family\_sep,  
 levels = c("1","2","3","8","9"),  
 labels = c("high", "intermediate","low","Not reply/Don't know","Not reply/Don't know"))  
  
#compute frequency tables  
#Desc(biomarkers.imp$family\_sep)  
#Freq(biomarkers.imp$family\_sep, useNA = "always")

Current socioeconomic position regrouped from five into three main categories.

biomarkers.imp$current\_sep <- factor(biomarkers.imp$current\_sep,  
 levels = c("1","2","3","4","5","6"),  
 labels = c("low", "low","intermediate","intermediate","high", "high"))  
  
#Desc(biomarkers.imp$current\_sep)

Occupation status was recoded according to the International Standard Classification of Occupations. Ten major occupations were classified into three groups:

1. Non manual b) Skilled manual c) Semi-skilled manual

biomarkers.imp$sep\_occupation <- factor(biomarkers.imp$sep\_occupation,  
 levels = c("1","2","3","4","5","6","7","8","9","77","98","99"),  
 labels = c("skilled manual","skilled manual","semi-skilled","non-manual","semi-skilled",  
 "non-manual","non-manual","skilled manual","semi-skilled","non-manual","Not reply/Don't know","Not reply/Don't know"))  
  
#Desc(biomarkers.imp$sep\_occupation)

**Adverse Childhood Experiences ACEs data**

Early-life infection. Column creation, transformation data and compute frequency tables of early-life infection variables: summarize six variables reporting early infection before 15 years old. Question 803: do you remember having any of these diseases? measles, viral hepatitis, CKD, TB, rheumatic fever, asthma, and bronchitis.

biomarkers.imp$CKD <- factor(biomarkers.imp$early\_infection1,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))  
  
biomarkers.imp$viral.hepatitis <- factor(biomarkers.imp$early\_infection2,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))   
  
biomarkers.imp$measles <- factor(biomarkers.imp$early\_infection3,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))  
  
biomarkers.imp$rheumatic.fever <- factor(biomarkers.imp$early\_infection4,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))  
  
biomarkers.imp$TB <- factor(biomarkers.imp$early\_infection5,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))   
  
biomarkers.imp$asthma <- factor(biomarkers.imp$early\_infection6,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))   
  
biomarkers.imp$bronchitis <- factor(biomarkers.imp$early\_infection7,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "Not reply/Don't know","Not reply/Don't know"))  
  
#Desc(biomarkers.imp$bronchitis)  
#Freq(biomarkers.imp$bronchitis, useNA = "always")

Create a new combination of vectors not present in the data (early-life infection).

#To explicitly coerce objects from one class to another  
biomarkers.imp$CKD <- as.numeric(as.character(biomarkers.imp$CKD))

## Warning: NAs introduced by coercion

biomarkers.imp$viral.hepatitis <- as.numeric(as.character(biomarkers.imp$viral.hepatitis))

## Warning: NAs introduced by coercion

biomarkers.imp$measles <- as.numeric(as.character(biomarkers.imp$measles))

## Warning: NAs introduced by coercion

biomarkers.imp$TB <- as.numeric(as.character(biomarkers.imp$TB))

## Warning: NAs introduced by coercion

biomarkers.imp$rheumatic.fever <- as.numeric(as.character(biomarkers.imp$rheumatic.fever))

## Warning: NAs introduced by coercion

biomarkers.imp$asthma <- as.numeric(as.character(biomarkers.imp$asthma))

## Warning: NAs introduced by coercion

biomarkers.imp$bronchitis <- as.numeric(as.character(biomarkers.imp$bronchitis))

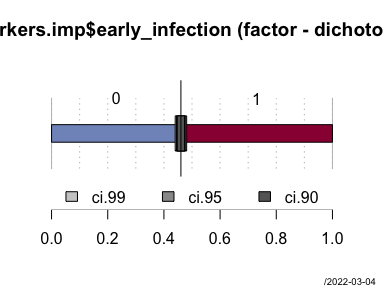
## Warning: NAs introduced by coercion

biomarkers.imp$early\_infection <- biomarkers.imp$CKD + biomarkers.imp$viral.hepatitis + biomarkers.imp$measles +   
biomarkers.imp$TB + biomarkers.imp$rheumatic.fever + biomarkers.imp$asthma + biomarkers.imp$bronchitis   
  
Freq(biomarkers.imp$early\_infection, useNA = "always")

## level freq perc cumfreq cumperc  
## 1 [0,0.5] 1'819 44.5% 1'819 44.5%  
## 2 (0.5,1] 1'710 41.8% 3'529 86.2%  
## 3 (1,1.5] 0 0.0% 3'529 86.2%  
## 4 (1.5,2] 320 7.8% 3'849 94.1%  
## 5 (2,2.5] 0 0.0% 3'849 94.1%  
## 6 (2.5,3] 81 2.0% 3'930 96.0%  
## 7 (3,3.5] 0 0.0% 3'930 96.0%  
## 8 (3.5,4] 18 0.4% 3'948 96.5%  
## 9 (4,4.5] 0 0.0% 3'948 96.5%  
## 10 (4.5,5] 0 0.0% 3'948 96.5%  
## 11 (5,5.5] 0 0.0% 3'948 96.5%  
## 12 (5.5,6] 0 0.0% 3'948 96.5%  
## 13 (6,6.5] 0 0.0% 3'948 96.5%  
## 14 (6.5,7] 2 0.0% 3'950 96.5%  
## 15 <NA> 142 3.5% 4'092 100.0%

biomarkers.imp$early\_infection <- factor(biomarkers.imp$early\_infection,  
 levels = c("0","1","2","3","4","5","6","7"),  
 labels = c("0","1","1","1","1","1","1","1"))  
  
Desc(biomarkers.imp$early\_infection)

## ------------------------------------------------------------------------------   
## biomarkers.imp$early\_infection (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 3'950 142 2  
## 96.5% 3.5%   
##   
## freq perc lci.95 uci.95'  
## 0 1'819 46.1% 44.5% 47.6%  
## 1 2'131 53.9% 52.4% 55.5%  
##   
## ' 95%-CI (Wilson)



class(biomarkers.imp$early\_infection)

## [1] "factor"

class(biomarkers.imp$TB)

## [1] "numeric"

class(biomarkers.imp$measles)

## [1] "numeric"

Early-life infection of study population by sex and age groups with CI 95%.

biomarkers.imp$CKD <- as.factor(as.numeric(biomarkers.imp$CKD))  
biomarkers.imp$viral.hepatitis <- as.factor(as.numeric(biomarkers.imp$viral.hepatitis))  
biomarkers.imp$measles <- as.factor(as.numeric(biomarkers.imp$measles))  
biomarkers.imp$TB <- as.factor(as.numeric(biomarkers.imp$TB))  
biomarkers.imp$rheumatic.fever <- as.factor(as.numeric(biomarkers.imp$rheumatic.fever))  
biomarkers.imp$asthma <- as.factor(as.numeric(biomarkers.imp$asthma))  
biomarkers.imp$bronchitis <- as.factor(as.numeric(biomarkers.imp$bronchitis))  
biomarkers.imp$early\_infection <- as.factor(as.numeric(biomarkers.imp$early\_infection))  
  
table\_2 <-table1 (~CKD + viral.hepatitis + measles + TB + rheumatic.fever + asthma + bronchitis + early\_infection | sex + age\_group, data=biomarkers.imp)  
  
table\_2

|  | male | | | female | | | Overall | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (60,70] (N=795) | (70,80] (N=518) | (80,Inf] (N=241) | (60,70] (N=1273) | (70,80] (N=717) | (80,Inf] (N=304) | (60,70] (N=2068) | (70,80] (N=1235) | (80,Inf] (N=545) |
| **CKD** |  |  |  |  |  |  |  |  |  |
| 0 | 767 (96.5%) | 510 (98.5%) | 235 (97.5%) | 1253 (98.4%) | 708 (98.7%) | 299 (98.4%) | 2020 (97.7%) | 1218 (98.6%) | 534 (98.0%) |
| 1 | 25 (3.1%) | 5 (1.0%) | 3 (1.2%) | 15 (1.2%) | 5 (0.7%) | 3 (1.0%) | 40 (1.9%) | 10 (0.8%) | 6 (1.1%) |
| Missing | 3 (0.4%) | 3 (0.6%) | 3 (1.2%) | 5 (0.4%) | 4 (0.6%) | 2 (0.7%) | 8 (0.4%) | 7 (0.6%) | 5 (0.9%) |
| **viral.hepatitis** |  |  |  |  |  |  |  |  |  |
| 0 | 760 (95.6%) | 496 (95.8%) | 230 (95.4%) | 1219 (95.8%) | 695 (96.9%) | 297 (97.7%) | 1979 (95.7%) | 1191 (96.4%) | 527 (96.7%) |
| 1 | 31 (3.9%) | 19 (3.7%) | 9 (3.7%) | 50 (3.9%) | 19 (2.7%) | 6 (2.0%) | 81 (3.9%) | 38 (3.1%) | 15 (2.8%) |
| Missing | 4 (0.5%) | 3 (0.6%) | 2 (0.8%) | 4 (0.3%) | 3 (0.4%) | 1 (0.3%) | 8 (0.4%) | 6 (0.5%) | 3 (0.6%) |
| **measles** |  |  |  |  |  |  |  |  |  |
| 0 | 401 (50.4%) | 279 (53.9%) | 127 (52.7%) | 641 (50.4%) | 340 (47.4%) | 155 (51.0%) | 1042 (50.4%) | 619 (50.1%) | 282 (51.7%) |
| 1 | 387 (48.7%) | 234 (45.2%) | 107 (44.4%) | 604 (47.4%) | 356 (49.7%) | 142 (46.7%) | 991 (47.9%) | 590 (47.8%) | 249 (45.7%) |
| Missing | 7 (0.9%) | 5 (1.0%) | 7 (2.9%) | 28 (2.2%) | 21 (2.9%) | 7 (2.3%) | 35 (1.7%) | 26 (2.1%) | 14 (2.6%) |
| **TB** |  |  |  |  |  |  |  |  |  |
| 0 | 763 (96.0%) | 490 (94.6%) | 218 (90.5%) | 1232 (96.8%) | 690 (96.2%) | 283 (93.1%) | 1995 (96.5%) | 1180 (95.5%) | 501 (91.9%) |
| 1 | 25 (3.1%) | 27 (5.2%) | 21 (8.7%) | 34 (2.7%) | 23 (3.2%) | 20 (6.6%) | 59 (2.9%) | 50 (4.0%) | 41 (7.5%) |
| Missing | 7 (0.9%) | 1 (0.2%) | 2 (0.8%) | 7 (0.5%) | 4 (0.6%) | 1 (0.3%) | 14 (0.7%) | 5 (0.4%) | 3 (0.6%) |
| **rheumatic.fever** |  |  |  |  |  |  |  |  |  |
| 0 | 786 (98.9%) | 512 (98.8%) | 240 (99.6%) | 1265 (99.4%) | 712 (99.3%) | 300 (98.7%) | 2051 (99.2%) | 1224 (99.1%) | 540 (99.1%) |
| 1 | 8 (1.0%) | 5 (1.0%) | 1 (0.4%) | 3 (0.2%) | 2 (0.3%) | 3 (1.0%) | 11 (0.5%) | 7 (0.6%) | 4 (0.7%) |
| Missing | 1 (0.1%) | 1 (0.2%) | 0 (0%) | 5 (0.4%) | 3 (0.4%) | 1 (0.3%) | 6 (0.3%) | 4 (0.3%) | 1 (0.2%) |
| **asthma** |  |  |  |  |  |  |  |  |  |
| 0 | 762 (95.8%) | 497 (95.9%) | 228 (94.6%) | 1202 (94.4%) | 675 (94.1%) | 287 (94.4%) | 1964 (95.0%) | 1172 (94.9%) | 515 (94.5%) |
| 1 | 33 (4.2%) | 20 (3.9%) | 13 (5.4%) | 67 (5.3%) | 39 (5.4%) | 16 (5.3%) | 100 (4.8%) | 59 (4.8%) | 29 (5.3%) |
| Missing | 0 (0%) | 1 (0.2%) | 0 (0%) | 4 (0.3%) | 3 (0.4%) | 1 (0.3%) | 4 (0.2%) | 4 (0.3%) | 1 (0.2%) |
| **bronchitis** |  |  |  |  |  |  |  |  |  |
| 0 | 744 (93.6%) | 482 (93.1%) | 231 (95.9%) | 1186 (93.2%) | 653 (91.1%) | 284 (93.4%) | 1930 (93.3%) | 1135 (91.9%) | 515 (94.5%) |
| 1 | 44 (5.5%) | 31 (6.0%) | 9 (3.7%) | 78 (6.1%) | 56 (7.8%) | 17 (5.6%) | 122 (5.9%) | 87 (7.0%) | 26 (4.8%) |
| Missing | 7 (0.9%) | 5 (1.0%) | 1 (0.4%) | 9 (0.7%) | 8 (1.1%) | 3 (1.0%) | 16 (0.8%) | 13 (1.1%) | 4 (0.7%) |
| **early\_infection** |  |  |  |  |  |  |  |  |  |
| 1 | 357 (44.9%) | 247 (47.7%) | 105 (43.6%) | 564 (44.3%) | 299 (41.7%) | 128 (42.1%) | 921 (44.5%) | 546 (44.2%) | 233 (42.8%) |
| 2 | 415 (52.2%) | 259 (50.0%) | 124 (51.5%) | 662 (52.0%) | 387 (54.0%) | 166 (54.6%) | 1077 (52.1%) | 646 (52.3%) | 290 (53.2%) |
| Missing | 23 (2.9%) | 12 (2.3%) | 12 (5.0%) | 47 (3.7%) | 31 (4.3%) | 10 (3.3%) | 70 (3.4%) | 43 (3.5%) | 22 (4.0%) |

library(compareGroups)  
  
table\_2\_cg <- compareGroups( data = biomarkers.imp,   
 formula = sex ~ CKD + viral.hepatitis + measles + TB + rheumatic.fever + asthma + bronchitis + early\_infection)  
  
table\_2\_ci <- createTable(x = table\_2\_cg, show.ci = T, show.n = T, show.all = T )  
  
export2word( x = table\_2\_ci, file = '/Users/macbookpro/Documents/PhD Project/Data/SABE/Outputs/table\_2\_ci.docx')

Description of early-life infection data by family socioeconomic position data.

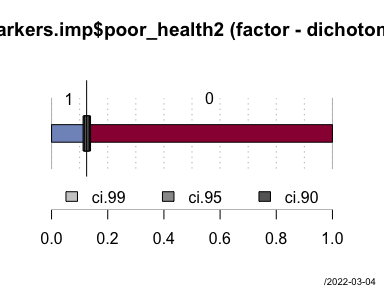
table\_3 <- table1(~sex + age\_group + CKD + viral.hepatitis + measles + TB + rheumatic.fever + asthma + bronchitis + early\_infection| family\_sep, data=biomarkers.imp)  
  
table\_3

|  | high (N=1617) | intermediate (N=1711) | low (N=738) | Not reply/Don't know (N=26) | Overall (N=4092) |
| --- | --- | --- | --- | --- | --- |
| **sex** |  |  |  |  |  |
| male | 635 (39.3%) | 695 (40.6%) | 298 (40.4%) | 7 (26.9%) | 1635 (40.0%) |
| female | 982 (60.7%) | 1016 (59.4%) | 440 (59.6%) | 19 (73.1%) | 2457 (60.0%) |
| **age\_group** |  |  |  |  |  |
| (60,70] | 752 (46.5%) | 902 (52.7%) | 406 (55.0%) | 8 (30.8%) | 2068 (50.5%) |
| (70,80] | 519 (32.1%) | 493 (28.8%) | 213 (28.9%) | 10 (38.5%) | 1235 (30.2%) |
| (80,Inf] | 258 (16.0%) | 215 (12.6%) | 65 (8.8%) | 7 (26.9%) | 545 (13.3%) |
| Missing | 88 (5.4%) | 101 (5.9%) | 54 (7.3%) | 1 (3.8%) | 244 (6.0%) |
| **CKD** |  |  |  |  |  |
| 0 | 1588 (98.2%) | 1678 (98.1%) | 719 (97.4%) | 22 (84.6%) | 4007 (97.9%) |
| 1 | 19 (1.2%) | 28 (1.6%) | 17 (2.3%) | 0 (0%) | 64 (1.6%) |
| Missing | 10 (0.6%) | 5 (0.3%) | 2 (0.3%) | 4 (15.4%) | 21 (0.5%) |
| **viral.hepatitis** |  |  |  |  |  |
| 0 | 1553 (96.0%) | 1654 (96.7%) | 708 (95.9%) | 19 (73.1%) | 3934 (96.1%) |
| 1 | 57 (3.5%) | 52 (3.0%) | 27 (3.7%) | 3 (11.5%) | 139 (3.4%) |
| Missing | 7 (0.4%) | 5 (0.3%) | 3 (0.4%) | 4 (15.4%) | 19 (0.5%) |
| **measles** |  |  |  |  |  |
| 0 | 850 (52.6%) | 861 (50.3%) | 349 (47.3%) | 14 (53.8%) | 2074 (50.7%) |
| 1 | 740 (45.8%) | 823 (48.1%) | 369 (50.0%) | 7 (26.9%) | 1939 (47.4%) |
| Missing | 27 (1.7%) | 27 (1.6%) | 20 (2.7%) | 5 (19.2%) | 79 (1.9%) |
| **TB** |  |  |  |  |  |
| 0 | 1553 (96.0%) | 1634 (95.5%) | 704 (95.4%) | 21 (80.8%) | 3912 (95.6%) |
| 1 | 57 (3.5%) | 70 (4.1%) | 28 (3.8%) | 1 (3.8%) | 156 (3.8%) |
| Missing | 7 (0.4%) | 7 (0.4%) | 6 (0.8%) | 4 (15.4%) | 24 (0.6%) |
| **rheumatic.fever** |  |  |  |  |  |
| 0 | 1604 (99.2%) | 1699 (99.3%) | 734 (99.5%) | 21 (80.8%) | 4058 (99.2%) |
| 1 | 10 (0.6%) | 8 (0.5%) | 4 (0.5%) | 0 (0%) | 22 (0.5%) |
| Missing | 3 (0.2%) | 4 (0.2%) | 0 (0%) | 5 (19.2%) | 12 (0.3%) |
| **asthma** |  |  |  |  |  |
| 0 | 1551 (95.9%) | 1618 (94.6%) | 693 (93.9%) | 20 (76.9%) | 3882 (94.9%) |
| 1 | 64 (4.0%) | 91 (5.3%) | 43 (5.8%) | 1 (3.8%) | 199 (4.9%) |
| Missing | 2 (0.1%) | 2 (0.1%) | 2 (0.3%) | 5 (19.2%) | 11 (0.3%) |
| **bronchitis** |  |  |  |  |  |
| 0 | 1522 (94.1%) | 1597 (93.3%) | 675 (91.5%) | 19 (73.1%) | 3813 (93.2%) |
| 1 | 81 (5.0%) | 104 (6.1%) | 58 (7.9%) | 0 (0%) | 243 (5.9%) |
| Missing | 14 (0.9%) | 10 (0.6%) | 5 (0.7%) | 7 (26.9%) | 36 (0.9%) |
| **early\_infection** |  |  |  |  |  |
| 1 | 746 (46.1%) | 757 (44.2%) | 305 (41.3%) | 11 (42.3%) | 1819 (44.5%) |
| 2 | 815 (50.4%) | 910 (53.2%) | 399 (54.1%) | 7 (26.9%) | 2131 (52.1%) |
| Missing | 56 (3.5%) | 44 (2.6%) | 34 (4.6%) | 8 (30.8%) | 142 (3.5%) |

Poor health. Vector transformation of “poor health” reported during the first 15 years of life. Question 804: were you in bed for a month or more due to any health problem?

biomarkers.imp$poor\_health2 <- factor(biomarkers.imp$poor\_health2,  
 levels = c("1","2","8","9"),  
 labels = c("1","0","0","0"))  
  
Desc (biomarkers.imp$poor\_health2)

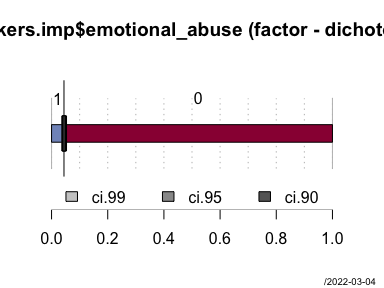
## ------------------------------------------------------------------------------   
## biomarkers.imp$poor\_health2 (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## 1 511 12.5% 11.5% 13.5%  
## 0 3'581 87.5% 86.5% 88.5%  
##   
## ' 95%-CI (Wilson)



Now, there are five childhood experiences variables (emotional abuse, migration, household violence, neglect food, poor health) and new variable created from present data (new variable). First step, vector transformation of emotional abuse YES/NOT as numerical values and convert them into characters. Second step rename levels in the aces columns as factors.

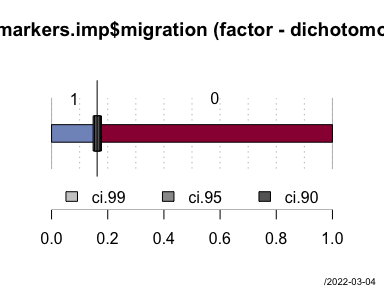
biomarkers.imp$emotional\_abuse <- factor(biomarkers.imp$emotional\_abuse,  
 levels = c("1","2","3","4","8","9"),  
 labels = c("1","1","1","0","0","0"))  
  
Desc(biomarkers.imp$emotional\_abuse)

## ------------------------------------------------------------------------------   
## biomarkers.imp$emotional\_abuse (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## 1 181 4.4% 3.8% 5.1%  
## 0 3'911 95.6% 94.9% 96.2%  
##   
## ' 95%-CI (Wilson)



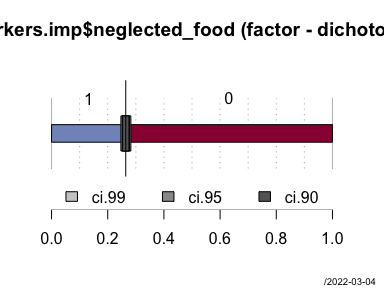
biomarkers.imp$migration <- factor(biomarkers.imp$migration,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))  
  
Desc(biomarkers.imp$migration)

## ------------------------------------------------------------------------------   
## biomarkers.imp$migration (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## 1 664 16.2% 15.1% 17.4%  
## 0 3'428 83.8% 82.6% 84.9%  
##   
## ' 95%-CI (Wilson)



biomarkers.imp$neglected\_food <- factor(biomarkers.imp$neglected\_food,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))  
  
Desc(biomarkers.imp$neglected\_food)

## ------------------------------------------------------------------------------   
## biomarkers.imp$neglected\_food (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## 1 1'081 26.4% 25.1% 27.8%  
## 0 3'011 73.6% 72.2% 74.9%  
##   
## ' 95%-CI (Wilson)

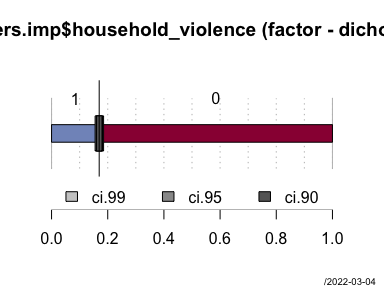


Freq(biomarkers.imp$neglected\_food, useNA = "always")

## level freq perc cumfreq cumperc  
## 1 1 1'081 26.4% 1'081 26.4%  
## 2 0 3'011 73.6% 4'092 100.0%  
## 3 <NA> 0 0.0% 4'092 100.0%

biomarkers.imp$household\_violence <- factor(biomarkers.imp$household\_violence,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))  
  
#compute frequency tables  
Desc(biomarkers.imp$household\_violence)

## ------------------------------------------------------------------------------   
## biomarkers.imp$household\_violence (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## 1 694 17.0% 15.8% 18.1%  
## 0 3'398 83.0% 81.9% 84.2%  
##   
## ' 95%-CI (Wilson)

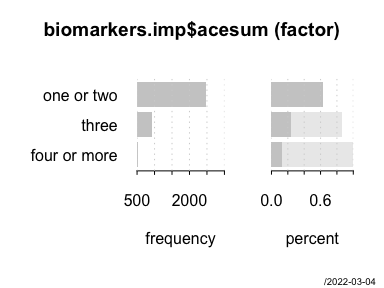


Then, we create new combination of vectors not present in the data (adverse childhood experiences). First step, it is calculate a new variable called aces (sum of six exposures reported). Second step, it was recoded into three main categories.

Levels: a) none b) one or two c) three d) four or more adverse childhood exposures

#class(biomarkers.imp$early\_infection)  
#class(biomarkers.imp$migration)  
#class(biomarkers.imp$emotional\_abuse)  
#class(biomarkers.imp$early\_infection)  
  
#To explicitly coerce objects from one class to another  
biomarkers.imp$neglected\_food <- as.numeric(as.character(biomarkers.imp$neglected\_food))  
biomarkers.imp$household\_violence <- as.numeric(as.character(biomarkers.imp$household\_violence))  
biomarkers.imp$migration <- as.numeric(as.character(biomarkers.imp$migration))  
biomarkers.imp$emotional\_abuse <- as.numeric(as.character(biomarkers.imp$emotional\_abuse))  
biomarkers.imp$poor\_health2 <- as.numeric(as.character(biomarkers.imp$poor\_health2))  
biomarkers.imp$early\_infection <- as.numeric(as.character(biomarkers.imp$early\_infection))  
  
biomarkers.imp$acesum <- biomarkers.imp$neglected\_food + biomarkers.imp$household\_violence + biomarkers.imp$migration + biomarkers.imp$emotional\_abuse + biomarkers.imp$poor\_health2 + biomarkers.imp$early\_infection   
  
biomarkers.imp$acesum <- factor(biomarkers.imp$acesum,  
 levels = c("1","2","3","4","5","6"),  
 labels = c("one or two", "one or two", "three","four or more", "four or more", "four or more"))  
  
#compute frequency tables  
Desc(biomarkers.imp$acesum)

## ------------------------------------------------------------------------------   
## biomarkers.imp$acesum (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 3'947 145 3 3 y  
## 96.5% 3.5%   
##   
## level freq perc cumfreq cumperc  
## 1 one or two 2'478 62.8% 2'478 62.8%  
## 2 three 938 23.8% 3'416 86.5%  
## 3 four or more 531 13.5% 3'947 100.0%



Summary childhood experiences of study population by sex and age groups.

biomarkers.imp$neglected\_food <- as.factor(as.numeric(biomarkers.imp$neglected\_food))  
biomarkers.imp$household\_violence <- as.factor(as.numeric(biomarkers.imp$household\_violence))  
biomarkers.imp$migration <- as.factor(as.numeric(biomarkers.imp$migration))  
biomarkers.imp$emotional\_abuse <- as.factor(as.numeric(biomarkers.imp$emotional\_abuse))  
biomarkers.imp$poor\_health2 <- as.factor(as.numeric(biomarkers.imp$poor\_health2))  
biomarkers.imp$early\_infection <- as.factor(as.numeric(biomarkers.imp$early\_infection))  
  
table\_4 <-table1(~neglected\_food + household\_violence + migration + emotional\_abuse + poor\_health2 + early\_infection + acesum| sex + age\_group, data=biomarkers.imp)  
  
table\_4

|  | male | | | female | | | Overall | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (60,70] (N=795) | (70,80] (N=518) | (80,Inf] (N=241) | (60,70] (N=1273) | (70,80] (N=717) | (80,Inf] (N=304) | (60,70] (N=2068) | (70,80] (N=1235) | (80,Inf] (N=545) |
| **neglected\_food** |  |  |  |  |  |  |  |  |  |
| 0 | 548 (68.9%) | 384 (74.1%) | 174 (72.2%) | 957 (75.2%) | 543 (75.7%) | 241 (79.3%) | 1505 (72.8%) | 927 (75.1%) | 415 (76.1%) |
| 1 | 247 (31.1%) | 134 (25.9%) | 67 (27.8%) | 316 (24.8%) | 174 (24.3%) | 63 (20.7%) | 563 (27.2%) | 308 (24.9%) | 130 (23.9%) |
| **household\_violence** |  |  |  |  |  |  |  |  |  |
| 0 | 654 (82.3%) | 448 (86.5%) | 216 (89.6%) | 1009 (79.3%) | 611 (85.2%) | 269 (88.5%) | 1663 (80.4%) | 1059 (85.7%) | 485 (89.0%) |
| 1 | 141 (17.7%) | 70 (13.5%) | 25 (10.4%) | 264 (20.7%) | 106 (14.8%) | 35 (11.5%) | 405 (19.6%) | 176 (14.3%) | 60 (11.0%) |
| **migration** |  |  |  |  |  |  |  |  |  |
| 0 | 645 (81.1%) | 429 (82.8%) | 195 (80.9%) | 1080 (84.8%) | 610 (85.1%) | 268 (88.2%) | 1725 (83.4%) | 1039 (84.1%) | 463 (85.0%) |
| 1 | 150 (18.9%) | 89 (17.2%) | 46 (19.1%) | 193 (15.2%) | 107 (14.9%) | 36 (11.8%) | 343 (16.6%) | 196 (15.9%) | 82 (15.0%) |
| **emotional\_abuse** |  |  |  |  |  |  |  |  |  |
| 0 | 769 (96.7%) | 495 (95.6%) | 234 (97.1%) | 1203 (94.5%) | 691 (96.4%) | 292 (96.1%) | 1972 (95.4%) | 1186 (96.0%) | 526 (96.5%) |
| 1 | 26 (3.3%) | 23 (4.4%) | 7 (2.9%) | 70 (5.5%) | 26 (3.6%) | 12 (3.9%) | 96 (4.6%) | 49 (4.0%) | 19 (3.5%) |
| **poor\_health2** |  |  |  |  |  |  |  |  |  |
| 0 | 695 (87.4%) | 458 (88.4%) | 221 (91.7%) | 1105 (86.8%) | 624 (87.0%) | 269 (88.5%) | 1800 (87.0%) | 1082 (87.6%) | 490 (89.9%) |
| 1 | 100 (12.6%) | 60 (11.6%) | 20 (8.3%) | 168 (13.2%) | 93 (13.0%) | 35 (11.5%) | 268 (13.0%) | 153 (12.4%) | 55 (10.1%) |
| **early\_infection** |  |  |  |  |  |  |  |  |  |
| 1 | 357 (44.9%) | 247 (47.7%) | 105 (43.6%) | 564 (44.3%) | 299 (41.7%) | 128 (42.1%) | 921 (44.5%) | 546 (44.2%) | 233 (42.8%) |
| 2 | 415 (52.2%) | 259 (50.0%) | 124 (51.5%) | 662 (52.0%) | 387 (54.0%) | 166 (54.6%) | 1077 (52.1%) | 646 (52.3%) | 290 (53.2%) |
| Missing | 23 (2.9%) | 12 (2.3%) | 12 (5.0%) | 47 (3.7%) | 31 (4.3%) | 10 (3.3%) | 70 (3.4%) | 43 (3.5%) | 22 (4.0%) |
| **acesum** |  |  |  |  |  |  |  |  |  |
| one or two | 465 (58.5%) | 337 (65.1%) | 151 (62.7%) | 759 (59.6%) | 431 (60.1%) | 201 (66.1%) | 1224 (59.2%) | 768 (62.2%) | 352 (64.6%) |
| three | 187 (23.5%) | 106 (20.5%) | 54 (22.4%) | 286 (22.5%) | 180 (25.1%) | 71 (23.4%) | 473 (22.9%) | 286 (23.2%) | 125 (22.9%) |
| four or more | 120 (15.1%) | 62 (12.0%) | 24 (10.0%) | 180 (14.1%) | 74 (10.3%) | 22 (7.2%) | 300 (14.5%) | 136 (11.0%) | 46 (8.4%) |
| Missing | 23 (2.9%) | 13 (2.5%) | 12 (5.0%) | 48 (3.8%) | 32 (4.5%) | 10 (3.3%) | 71 (3.4%) | 45 (3.6%) | 22 (4.0%) |

Adverse childhood experiences by early-life SEP.

table\_5 <- table1(~sex + age\_group + neglected\_food + household\_violence + migration + emotional\_abuse + poor\_health2 + early\_infection + acesum| family\_sep, data=biomarkers.imp)  
  
table\_5

|  | high (N=1617) | intermediate (N=1711) | low (N=738) | Not reply/Don't know (N=26) | Overall (N=4092) |
| --- | --- | --- | --- | --- | --- |
| **sex** |  |  |  |  |  |
| male | 635 (39.3%) | 695 (40.6%) | 298 (40.4%) | 7 (26.9%) | 1635 (40.0%) |
| female | 982 (60.7%) | 1016 (59.4%) | 440 (59.6%) | 19 (73.1%) | 2457 (60.0%) |
| **age\_group** |  |  |  |  |  |
| (60,70] | 752 (46.5%) | 902 (52.7%) | 406 (55.0%) | 8 (30.8%) | 2068 (50.5%) |
| (70,80] | 519 (32.1%) | 493 (28.8%) | 213 (28.9%) | 10 (38.5%) | 1235 (30.2%) |
| (80,Inf] | 258 (16.0%) | 215 (12.6%) | 65 (8.8%) | 7 (26.9%) | 545 (13.3%) |
| Missing | 88 (5.4%) | 101 (5.9%) | 54 (7.3%) | 1 (3.8%) | 244 (6.0%) |
| **neglected\_food** |  |  |  |  |  |
| 0 | 1410 (87.2%) | 1259 (73.6%) | 324 (43.9%) | 18 (69.2%) | 3011 (73.6%) |
| 1 | 207 (12.8%) | 452 (26.4%) | 414 (56.1%) | 8 (30.8%) | 1081 (26.4%) |
| **household\_violence** |  |  |  |  |  |
| 0 | 1426 (88.2%) | 1418 (82.9%) | 532 (72.1%) | 22 (84.6%) | 3398 (83.0%) |
| 1 | 191 (11.8%) | 293 (17.1%) | 206 (27.9%) | 4 (15.4%) | 694 (17.0%) |
| **migration** |  |  |  |  |  |
| 0 | 1365 (84.4%) | 1431 (83.6%) | 608 (82.4%) | 24 (92.3%) | 3428 (83.8%) |
| 1 | 252 (15.6%) | 280 (16.4%) | 130 (17.6%) | 2 (7.7%) | 664 (16.2%) |
| **emotional\_abuse** |  |  |  |  |  |
| 0 | 1577 (97.5%) | 1620 (94.7%) | 690 (93.5%) | 24 (92.3%) | 3911 (95.6%) |
| 1 | 40 (2.5%) | 91 (5.3%) | 48 (6.5%) | 2 (7.7%) | 181 (4.4%) |
| **poor\_health2** |  |  |  |  |  |
| 0 | 1422 (87.9%) | 1498 (87.6%) | 638 (86.4%) | 23 (88.5%) | 3581 (87.5%) |
| 1 | 195 (12.1%) | 213 (12.4%) | 100 (13.6%) | 3 (11.5%) | 511 (12.5%) |
| **early\_infection** |  |  |  |  |  |
| 1 | 746 (46.1%) | 757 (44.2%) | 305 (41.3%) | 11 (42.3%) | 1819 (44.5%) |
| 2 | 815 (50.4%) | 910 (53.2%) | 399 (54.1%) | 7 (26.9%) | 2131 (52.1%) |
| Missing | 56 (3.5%) | 44 (2.6%) | 34 (4.6%) | 8 (30.8%) | 142 (3.5%) |
| **acesum** |  |  |  |  |  |
| one or two | 1109 (68.6%) | 1039 (60.7%) | 320 (43.4%) | 10 (38.5%) | 2478 (60.6%) |
| three | 326 (20.2%) | 404 (23.6%) | 202 (27.4%) | 6 (23.1%) | 938 (22.9%) |
| four or more | 124 (7.7%) | 224 (13.1%) | 181 (24.5%) | 2 (7.7%) | 531 (13.0%) |
| Missing | 58 (3.6%) | 44 (2.6%) | 35 (4.7%) | 8 (30.8%) | 145 (3.5%) |

Adverse childhood experiences by demographics characteristics (ethnicity).

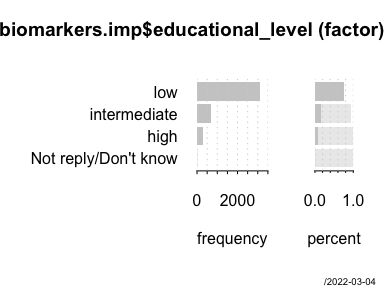
table\_6 <-table1(~neglected\_food + household\_violence + migration + emotional\_abuse + poor\_health2 + early\_infection + acesum| ethnic\_group, data=biomarkers.imp)  
  
table\_6

|  | indigenous (N=275) | afrocolombian (N=413) | white (N=1220) | mestize (N=1999) | Not reply/Don't know (N=185) | Overall (N=4092) |
| --- | --- | --- | --- | --- | --- | --- |
| **neglected\_food** |  |  |  |  |  |  |
| 0 | 177 (64.4%) | 291 (70.5%) | 926 (75.9%) | 1480 (74.0%) | 137 (74.1%) | 3011 (73.6%) |
| 1 | 98 (35.6%) | 122 (29.5%) | 294 (24.1%) | 519 (26.0%) | 48 (25.9%) | 1081 (26.4%) |
| **household\_violence** |  |  |  |  |  |  |
| 0 | 228 (82.9%) | 353 (85.5%) | 1005 (82.4%) | 1654 (82.7%) | 158 (85.4%) | 3398 (83.0%) |
| 1 | 47 (17.1%) | 60 (14.5%) | 215 (17.6%) | 345 (17.3%) | 27 (14.6%) | 694 (17.0%) |
| **migration** |  |  |  |  |  |  |
| 0 | 227 (82.5%) | 353 (85.5%) | 1041 (85.3%) | 1658 (82.9%) | 149 (80.5%) | 3428 (83.8%) |
| 1 | 48 (17.5%) | 60 (14.5%) | 179 (14.7%) | 341 (17.1%) | 36 (19.5%) | 664 (16.2%) |
| **emotional\_abuse** |  |  |  |  |  |  |
| 0 | 254 (92.4%) | 383 (92.7%) | 1172 (96.1%) | 1922 (96.1%) | 180 (97.3%) | 3911 (95.6%) |
| 1 | 21 (7.6%) | 30 (7.3%) | 48 (3.9%) | 77 (3.9%) | 5 (2.7%) | 181 (4.4%) |
| **poor\_health2** |  |  |  |  |  |  |
| 0 | 236 (85.8%) | 365 (88.4%) | 1067 (87.5%) | 1746 (87.3%) | 167 (90.3%) | 3581 (87.5%) |
| 1 | 39 (14.2%) | 48 (11.6%) | 153 (12.5%) | 253 (12.7%) | 18 (9.7%) | 511 (12.5%) |
| **early\_infection** |  |  |  |  |  |  |
| 1 | 122 (44.4%) | 203 (49.2%) | 544 (44.6%) | 867 (43.4%) | 83 (44.9%) | 1819 (44.5%) |
| 2 | 143 (52.0%) | 201 (48.7%) | 631 (51.7%) | 1058 (52.9%) | 98 (53.0%) | 2131 (52.1%) |
| Missing | 10 (3.6%) | 9 (2.2%) | 45 (3.7%) | 74 (3.7%) | 4 (2.2%) | 142 (3.5%) |
| **acesum** |  |  |  |  |  |  |
| one or two | 152 (55.3%) | 261 (63.2%) | 758 (62.1%) | 1191 (59.6%) | 116 (62.7%) | 2478 (60.6%) |
| three | 68 (24.7%) | 95 (23.0%) | 258 (21.1%) | 470 (23.5%) | 47 (25.4%) | 938 (22.9%) |
| four or more | 44 (16.0%) | 48 (11.6%) | 158 (13.0%) | 263 (13.2%) | 18 (9.7%) | 531 (13.0%) |
| Missing | 11 (4.0%) | 9 (2.2%) | 46 (3.8%) | 75 (3.8%) | 4 (2.2%) | 145 (3.5%) |

Education. Ten categories were regrouped into three main categories as follows: a) Low b) Intermediate c) High

biomarkers.imp$educational\_level <- factor(biomarkers.imp$educational\_level,  
 levels = c("1","2","3","4","5","6","7","8","9","10","11","98","99"),  
 labels = c("low", "low","low","intermediate","intermediate",  
 "high","high","high","high","high","high","Not reply/Don't know","Not reply/Don't know"))  
  
Desc(biomarkers.imp$educational\_level)

## ------------------------------------------------------------------------------   
## biomarkers.imp$educational\_level (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 4 4 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 low 3'115 76.1% 3'115 76.1%  
## 2 intermediate 684 16.7% 3'799 92.8%  
## 3 high 276 6.7% 4'075 99.6%  
## 4 Not reply/Don't know 17 0.4% 4'092 100.0%

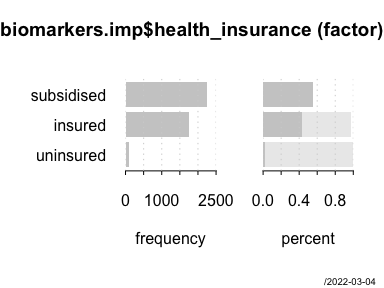


**Social protection data**

Health insurance regrouped into three main groups.

biomarkers.imp$health\_insurance <- factor(biomarkers.imp$health\_insurance,  
 levels = c("1","2","3","4","5","8","9"),  
 labels = c("insured", "subsidised","insured","insured","uninsured","uninsured","uninsured"))  
  
Desc(biomarkers.imp$health\_insurance)

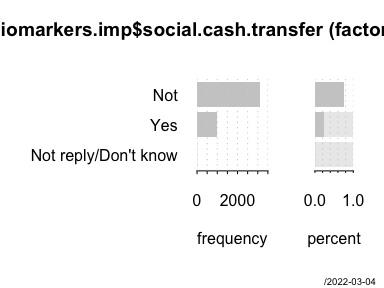
## ------------------------------------------------------------------------------   
## biomarkers.imp$health\_insurance (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 subsidised 2'246 54.9% 2'246 54.9%  
## 2 insured 1'753 42.8% 3'999 97.7%  
## 3 uninsured 93 2.3% 4'092 100.0%



Social cash transfers

biomarkers.imp$social.cash.transfer <- factor(biomarkers.imp$cash\_transfer2,  
 levels = c("1","2","8","9"),  
 labels = c("Yes","Not","Not reply/Don't know", "Not reply/Don't know"))  
  
#compute frequency tables  
Desc(biomarkers.imp$social.cash.transfer)

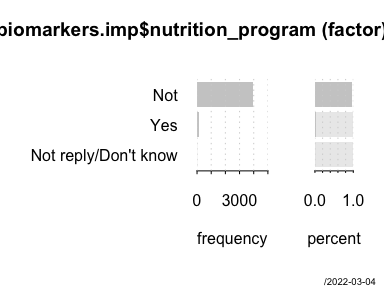
## ------------------------------------------------------------------------------   
## biomarkers.imp$social.cash.transfer (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Not 3'114 76.1% 3'114 76.1%  
## 2 Yes 964 23.6% 4'078 99.7%  
## 3 Not reply/Don't know 14 0.3% 4'092 100.0%



Elderly nutrition program.

biomarkers.imp$nutrition\_program <- factor(biomarkers.imp$cash\_transfer1,  
 levels = c("1","2","8","9"),  
 labels = c("Yes","Not","Not reply/Don't know", "Not reply/Don't know"))  
  
#compute frequency tables  
Desc(biomarkers.imp$nutrition\_program)

## ------------------------------------------------------------------------------   
## biomarkers.imp$nutrition\_program (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Not 3'936 96.2% 3'936 96.2%  
## 2 Yes 145 3.5% 4'081 99.7%  
## 3 Not reply/Don't know 11 0.3% 4'092 100.0%



Baseline characteristics by sex.

Baseline characteristics by age.

Baseline characteristics by ethnicity.

Baseline characteristics by social policies.

**Health outcomes data** There are three variables reporting chronic diseases (cardiovascular disease, HTN, and diabetes), and 13 variables reporting medication. Here, we create a new combination of vectors (comorbidities and medication sum).

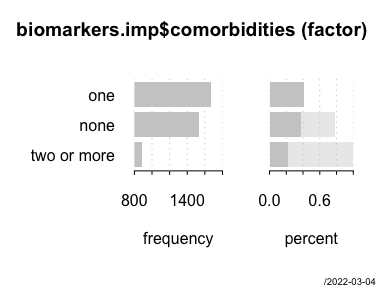
Chronic conditions and medication by demographic characteristics (sex and age groups).

#Rename values in the health outcomes columns and create a vectors of diseases  
biomarkers.imp$cvd <- factor(biomarkers.imp$cvd,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))  
   
biomarkers.imp$hta <- factor(biomarkers.imp$hta,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))  
   
biomarkers.imp$diabetes <- factor(biomarkers.imp$diabetes,  
 levels = c("1","2","8","9"),  
 labels = c("1","0", "0","0"))

Create new combination of vectors not present in the data (comorbidity or sum of three chronic conditions). formula: comorbidity = cvd + hta + diabetes. This was recoded into three main levels: a) none b) one c) two or more conditions

# Multiple conditions to adding new column to data frame:  
#To explicitly coerce objects from one class to another  
  
biomarkers.imp$hta <- as.numeric(as.character(biomarkers.imp$hta))  
biomarkers.imp$diabetes <- as.numeric(as.character(biomarkers.imp$diabetes))  
biomarkers.imp$cvd <- as.numeric(as.character(biomarkers.imp$cvd))  
  
biomarkers.imp$comorbidities <- biomarkers.imp$hta + biomarkers.imp$cvd + biomarkers.imp$diabetes  
  
#create a vector co-morbidities  
biomarkers.imp$comorbidities <- factor(biomarkers.imp$comorbidities,  
 levels = c("0","1","2","3"),  
 labels = c("none","one", "two or more","two or more"))  
  
#Desc(biomarkers.imp$hta)  
#Desc(biomarkers.imp$cvd)  
#Desc(biomarkers.imp$diabetes)  
Desc(biomarkers.imp$comorbidities)

## ------------------------------------------------------------------------------   
## biomarkers.imp$comorbidities (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 one 1'666 40.7% 1'666 40.7%  
## 2 none 1'536 37.5% 3'202 78.3%  
## 3 two or more 890 21.7% 4'092 100.0%



Freq(biomarkers.imp$comorbidities)

## level freq perc cumfreq cumperc  
## 1 none 1'536 37.5% 1'536 37.5%  
## 2 one 1'666 40.7% 3'202 78.3%  
## 3 two or more 890 21.7% 4'092 100.0%

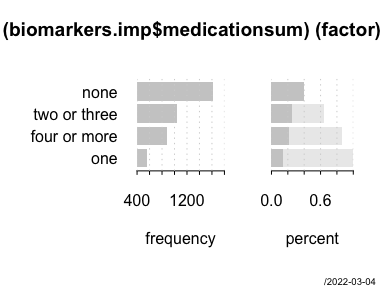
Vector transformation of medication. There are 13 medication variables in the dataframe. Vector transformation only constrained seven first variables. medication\_1

#To explicitly coerce objects from one class to another and vector transformation  
biomarkers.imp$medication\_1 <- ifelse(test = is.na(biomarkers.imp$P835\_1 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_2 <- ifelse(test = is.na(biomarkers.imp$P835\_2 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_3 <- ifelse(test = is.na(biomarkers.imp$P835\_3 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_4 <- ifelse(test = is.na(biomarkers.imp$P835\_4 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_5 <- ifelse(test = is.na(biomarkers.imp$P835\_5 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_6 <- ifelse(test = is.na(biomarkers.imp$P835\_6 ) == FALSE, yes = 1, no = 0 )  
biomarkers.imp$medication\_7 <- ifelse(test = is.na(biomarkers.imp$P835\_7 ) == FALSE, yes = 1, no = 0 )  
#biomarkers.imp$medication\_8 <- ifelse(test = is.na(biomarkers.imp$P835\_8 ) == FALSE, yes = 1, no = 0 )  
#biomarkers$medication\_9 <- ifelse(test = is.na(biomarkers$P835\_9 ) == FALSE, yes = 1, no = 0 )  
#biomarkers$medication\_10 <- ifelse(test = is.na(biomarkers$P835\_10 ) == FALSE, yes = 1, no = 0 )  
#biomarkers$medication\_11 <- ifelse(test = is.na(biomarkers$P835\_11 ) == FALSE, yes = 1, no = 0 )  
#biomarkers$medication\_12 <- ifelse(test = is.na(biomarkers$P835\_12 ) == FALSE, yes = 1, no = 0 )  
#biomarkers$medication\_0 <- ifelse(test = is.na(biomarkers$P835 ) == FALSE, yes = 1, no = 0 )  
  
#compute frequency tables  
#table(biomarkers.imp$medication\_1, useNA = "always")  
#table(biomarkers.imp$medication\_2, useNA = "always")  
#table(biomarkers.imp$medication\_3, useNA = "always")  
#table(biomarkers.imp$medication\_4, useNA = "always")  
#table(biomarkers.imp$medication\_5, useNA = "always")  
#table(biomarkers.imp$medication\_6, useNA = "always")  
#table(biomarkers.imp$medication\_7, useNA = "always")  
#table(biomarkers$medication\_8, useNA = "always")  
#table(biomarkers$medication\_9, useNA = "always")  
#table(biomarkers$medication\_10, useNA = "always")  
#table(biomarkers$medication\_11, useNA = "always")  
#table(biomarkers$medication\_12, useNA = "always")  
#table(biomarkers$medication\_0, useNA = "always")

Create new combination of vectors not present in the data (medication). The new variable “number of treatments” was recoded into three levels: a) none b) one c) two or three d) four or more

biomarkers.imp$medicationsum <- biomarkers.imp$medication\_1 + biomarkers.imp$medication\_2 + biomarkers.imp$medication\_3 + biomarkers.imp$medication\_4 + biomarkers.imp$medication\_5 + biomarkers.imp$medication\_6 + + biomarkers.imp$medication\_7  
  
biomarkers.imp$medicationsum <- factor(biomarkers.imp$medicationsum,  
 levels = c("0","1","2","3","4","5","6","7","8","9","10","11","12","13"),  
 labels = c("none","one","two or three", "two or three", "four or more", "four or more", "four or more","four or more","four or more","four or more","four or more","four or more","four or more", "four or more"))  
  
Desc((biomarkers.imp$medicationsum))

## ------------------------------------------------------------------------------   
## (biomarkers.imp$medicationsum) (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 4 4 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 none 1'614 39.4% 1'614 39.4%  
## 2 two or three 1'036 25.3% 2'650 64.8%  
## 3 four or more 875 21.4% 3'525 86.1%  
## 4 one 567 13.9% 4'092 100.0%



Freq(biomarkers.imp$medicationsum)

## level freq perc cumfreq cumperc  
## 1 none 1'614 39.4% 1'614 39.4%  
## 2 one 567 13.9% 2'181 53.3%  
## 3 two or three 1'036 25.3% 3'217 78.6%  
## 4 four or more 875 21.4% 4'092 100.0%

Chronic conditions and medication by demographics characteristics (sex and age groups).

Chronic conditions and medication by sex.

Chronic conditions and medication by age groups.

Chronic conditions and medication by ethnicity

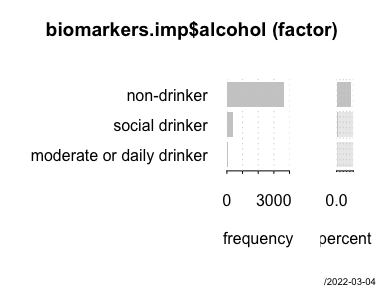
Description SEP by comorbidities.

**Health risk and behaviours data** There are three variables present in the data (smoking, alcohol consumption, waist circumference) and two new variables were created from data (BMI and blood pressure).

Alcohol. levels transformation of vector were recoded from five into three main groups/levels: a) social drinker b) moderate or daily drinker c) non-drinker

biomarkers.imp$alcohol <- factor(biomarkers.imp$alcohol,  
 levels = c("1","2","3","4","5","8","9"),  
 labels = c("social drinker","moderate or daily drinker", "moderate or daily drinker","moderate or daily drinker", "non-drinker","non-drinker","non-drinker"))  
  
Desc(biomarkers.imp$alcohol)

## ------------------------------------------------------------------------------   
## biomarkers.imp$alcohol (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 non-drinker 3'624 88.6% 3'624 88.6%  
## 2 social drinker 382 9.3% 4'006 97.9%  
## 3 moderate or daily drinker 86 2.1% 4'092 100.0%

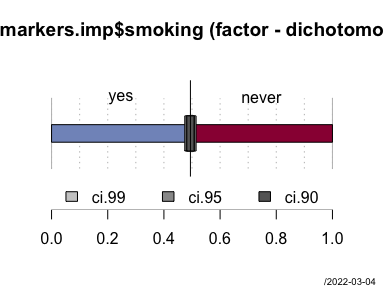


#Freq(biomarkers.imp$alcohol)

Smoking. Levels of smoking were recoded into binary variable indicating if the participants had ever smoked: a) never (4) and b) smoked (1-3)

biomarkers.imp$smoking <- factor(biomarkers.imp$smoking,  
 levels = c("1","2","3","4","8","9"),  
 labels = c("yes","yes", "yes","never", "never","never"))  
  
Desc(biomarkers.imp$smoking)

## ------------------------------------------------------------------------------   
## biomarkers.imp$smoking (factor - dichotomous)  
##   
## length n NAs unique  
## 4'092 4'092 0 2  
## 100.0% 0.0%   
##   
## freq perc lci.95 uci.95'  
## yes 2'023 49.4% 47.9% 51.0%  
## never 2'069 50.6% 49.0% 52.1%  
##   
## ' 95%-CI (Wilson)



Freq(biomarkers.imp$smoking)

## level freq perc cumfreq cumperc  
## 1 yes 2'023 49.4% 2'023 49.4%  
## 2 never 2'069 50.6% 4'092 100.0%

Before estimate BMI, higher waist circumference, and blood pressure we need to replace 777,999, and 8888 values (Do not apply) to NA values in seven columns.

#load library to replace the value 777 (Not apply) in the weight and height columns with a missing value (NA)  
library(naniar)

Replace 777 and 999 values (Do not apply) to NA values in seven columns: weight, height, waist circumference, systolic blood pressure right hand, systolic blood pressure left hand, diastolic blood pressure right hand, and diastolic blood pressure left hand.

biomarkers.imp$weight\_2 <- ifelse(test = biomarkers.imp$weight %in% c( 777 ), yes = NA, no = biomarkers.imp$weight )  
biomarkers.imp$height\_2 <- ifelse(test = biomarkers.imp$height %in% c( 777 ), yes = NA, no = biomarkers.imp$height )  
biomarkers.imp$wc\_2 <- ifelse(test = biomarkers.imp$wc %in% c( 777 ), yes = NA, no = biomarkers.imp$wc )  
  
biomarkers.imp$dyastolic\_bprh\_2 <- ifelse(test = biomarkers.imp$dyastolic\_bprh %in% c( 999, 8888 ), yes = NA, no = biomarkers.imp$dyastolic\_bprh )  
biomarkers.imp$dyastolic\_bplh\_2 <- ifelse(test = biomarkers.imp$dyastolic\_bplh %in% c( 999, 8888 ), yes = NA, no = biomarkers.imp$dyastolic\_bplh )  
biomarkers.imp$systolic\_bprh\_2 <- ifelse(test = biomarkers.imp$systolic\_bprh %in% c( 999, 8888 ), yes = NA, no = biomarkers.imp$systolic\_bprh )  
biomarkers.imp$systolic\_bplh\_2 <- ifelse(test = biomarkers.imp$systolic\_bplh %in% c( 999, 8888 ), yes = NA, no = biomarkers.imp$systolic\_bplh )  
  
#antropometric measures  
table\_confer <- biomarkers.imp %>%  
 filter( is.na( weight\_2 ) )   
 table(table\_confer$weight, table\_confer$weight\_2, useNA = "always" )

##   
## <NA>  
## 777 47  
## <NA> 0

#table\_confer\_1 <- biomarkers.imp %>%  
 #filter( is.na( height\_2 ) )   
 #table(table\_confer\_1$weight, table\_confer\_1$height\_2, useNA = "always" )  
   
table\_confer\_2 <- biomarkers.imp %>%  
 filter( is.na( wc\_2 ) )   
 table(table\_confer\_2$wc, table\_confer\_2$wc\_2, useNA = "always" )

##   
## <NA>  
## 777 76  
## <NA> 0

#blood pressure   
table\_confer\_3 <- biomarkers.imp %>%  
 filter( is.na( dyastolic\_bprh\_2 ) )   
 table(table\_confer\_3$dyastolic\_bprh, table\_confer\_3$dyastolic\_bprh\_2, useNA = "always" )

##   
## <NA>  
## 999 73  
## 8888 19  
## <NA> 0

table\_confer\_4 <- biomarkers.imp %>%  
 filter( is.na( dyastolic\_bplh\_2 ) )   
 table(table\_confer\_4$dyastolic\_bplh, table\_confer\_4$dyastolic\_bplh\_2, useNA = "always" )

##   
## <NA>  
## 999 67  
## 8888 2  
## <NA> 0

table\_confer\_5 <- biomarkers.imp %>%  
 filter( is.na( systolic\_bprh\_2 ) )   
 table(table\_confer\_5$systolic\_bprh, table\_confer\_5$systolic\_bprh\_2, useNA = "always" )

##   
## <NA>  
## 999 209  
## 8888 18  
## <NA> 0

table\_confer\_6 <- biomarkers.imp %>%  
 filter( is.na( systolic\_bplh\_2 ) )   
 table(table\_confer\_6$systolic\_bplh, table\_confer\_6$systolic\_bplh\_2, useNA = "always" )

##   
## <NA>  
## 999 153  
## 8888 56  
## <NA> 0

#table( biomarkers.imp[]$weight, biomarkers.imp$weight\_2,useNA = 'always' )  
#table( biomarkers.imp[]$height, biomarkers.imp$height\_2,useNA = 'always' )  
#table( biomarkers.imp[]$wc, biomarkers.imp$wc\_2,useNA = 'always' )  
#table( biomarkers.imp[]$dyastolic\_bprh, biomarkers.imp$dyastolic\_bprh\_2,useNA = 'always' )  
#table( biomarkers.imp[]$dyastolic\_bplh, biomarkers.imp$dyastolic\_bplh\_2, useNA = 'always' )  
#table( biomarkers.imp[]$systolic\_bprh, biomarkers.imp$systolic\_bprh\_2,useNA = 'always' )  
#table( biomarkers.imp[]$systolic\_bplh, biomarkers.imp$systolic\_bplh\_2,useNA = 'always' )

Body Mass Index (BMI): Create new combination of vectors (Body Mass Index BMI) not present in the data calculate Body Mass Index (BMI) through a function bmi that takes a person’s weight in lbs and height in inches and returns their [Body Mass Index](https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/).

formula: bmi =weight (kg) / height (m) x height (m)

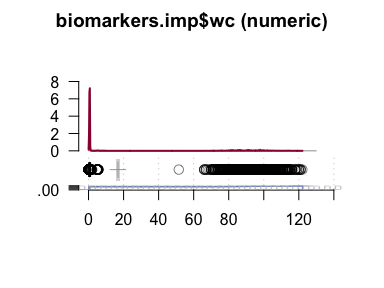
Levels: a) below 25 (normal) b) between 25 - 30 (Overweight) c) above 30 (Obese)

biomarkers.imp$height\_m <- biomarkers.imp$height\_2 / 100  
  
biomarkers.imp$height\_m\_squared <- (biomarkers.imp$height\_m)^2  
  
biomarkers.imp$bmi\_2 <- round(biomarkers.imp$weight\_2 / (biomarkers.imp$height\_m)^2, digits = 2 )  
  
biomarkers.imp$BMI <- cut( x = biomarkers.imp$bmi\_2, breaks = c(0, 25, 30, Inf),  
 levels = c("(0,25]", "(25.1,29.9]", "(30,Inf]"),   
 labels = c("below 25 normal", "25 and below 29.9 overweight","above 30 obese"))  
  
#Desc(biomarkers.imp$BMI)  
#Freq(biomarkers.imp$BMI, useNA = "always")

Waist circumference. Create new combination of vectors not present in the data (waist-hip ratio). vector transformation (WHR OMS Recommendations thresholds) <https://doi.org/10.1016/j.jshs.2017.03.011>

biomarkers.imp$wc <- (biomarkers.imp$wc) / (biomarkers.imp$height)  
  
Desc(biomarkers.imp$wc)

## ------------------------------------------------------------------------------   
## biomarkers.imp$wc (numeric)  
##   
## length n NAs unique 0s mean'  
## 4'092 4'092 0 1'414 0 16.78548496  
## 100.0% 0.0% 0.0%   
##   
## .05 .10 .25 median .75 .90  
## 0.46794872 0.50589653 0.55696203 0.61842105 0.71140940 90.00000000  
##   
## range sd vcoef mad IQR skew  
## 121.91377091 35.23410448 2.09908171 0.10644444 0.15444737 1.76893032  
##   
## meanCI  
## 15.70561359  
## 17.86535632  
##   
## .95  
## 99.00000000  
##   
## kurt  
## 1.25634072  
##   
## lowest : 0.08622909 (2), 0.08880309, 0.09266409, 0.09395109 (2), 0.0978121  
## highest: 118.0 (5), 119.0 (2), 120.0 (3), 121.0 (3), 122.0  
##   
## ' 95%-CI (classic)



Freq(biomarkers.imp$wc, useNA = "always")

## level freq perc cumfreq cumperc  
## 1 [0,10] 3'370 82.4% 3'370 82.4%  
## 2 (10,20] 0 0.0% 3'370 82.4%  
## 3 (20,30] 0 0.0% 3'370 82.4%  
## 4 (30,40] 0 0.0% 3'370 82.4%  
## 5 (40,50] 0 0.0% 3'370 82.4%  
## 6 (50,60] 1 0.0% 3'371 82.4%  
## 7 (60,70] 12 0.3% 3'383 82.7%  
## 8 (70,80] 91 2.2% 3'474 84.9%  
## 9 (80,90] 222 5.4% 3'696 90.3%  
## 10 (90,100] 240 5.9% 3'936 96.2%  
## 11 (100,110] 104 2.5% 4'040 98.7%  
## 12 (110,120] 48 1.2% 4'088 99.9%  
## 13 (120,130] 4 0.1% 4'092 100.0%  
## 14 <NA> 0 0.0% 4'092 100.0%

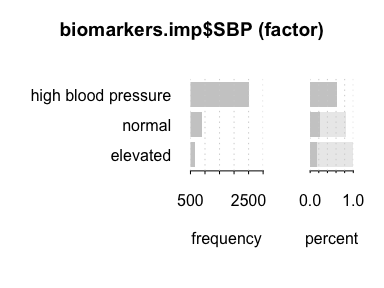
Blood pressure. Create new combination of vectors not present in the data (blood pressure). According to the WHO’s guidelines: Normal adult blood pressure is defined as a blood pressure of 120 mm Hg1 when the heart beats (systolic) and a blood pressure of 80 mm Hg when the heart relaxes (diastolic). When systolic blood pressure is equal to or above 140 mm Hg and/or a diastolic blood pressure equal to or above 90 mm Hg the blood pressure is considered to be raised or high.

Blood Pressure Category SBP DBP

Normal < 120 mm Hg And < 80 mm Hg Elevated 120-129 mm Hg And < 80-89 mm Hg High Blood Pressure ≥ 130-mm Hg ≥90 mm Hg

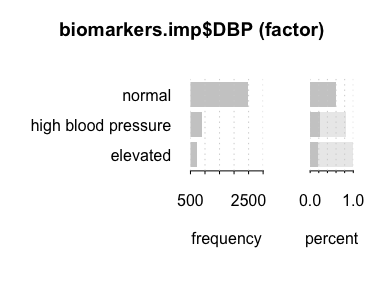
biomarkers.imp$SBP <- (biomarkers.imp$systolic\_bplh + biomarkers.imp$systolic\_bprh)/2  
biomarkers.imp$DBP <- (biomarkers.imp$dyastolic\_bplh + biomarkers.imp$dyastolic\_bprh)/2  
  
biomarkers.imp$SBP <- cut( x = biomarkers.imp$SBP, breaks = c(0, 120, 129, Inf),  
 levels = c("(0,120]","(120,129]","(130,Inf]"),  
 labels = c("normal","elevated","high blood pressure"))  
  
biomarkers.imp$DBP <- cut( x = biomarkers.imp$DBP, breaks = c(0, 80, 89,Inf),  
 levels = c("(0,80]","(80,89]","(90,Inf]"),  
 labels = c("normal","elevated", "high blood pressure"))  
  
Desc(biomarkers.imp$SBP)

## ------------------------------------------------------------------------------   
## biomarkers.imp$SBP (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 high blood pressure 2'522 61.6% 2'522 61.6%  
## 2 normal 908 22.2% 3'430 83.8%  
## 3 elevated 662 16.2% 4'092 100.0%



Desc(biomarkers.imp$DBP)

## ------------------------------------------------------------------------------   
## biomarkers.imp$DBP (factor)  
##   
## length n NAs unique levels dupes  
## 4'092 4'092 0 3 3 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 normal 2'484 60.7% 2'484 60.7%  
## 2 high blood pressure 894 21.8% 3'378 82.6%  
## 3 elevated 714 17.4% 4'092 100.0%



#Freq(biomarkers.imp$SBP, useNA = "always")  
#Freq(biomarkers.imp$DBP, useNA = "always")

Health risk by sex and age groups.

Chronic diseases and health risk by SEP.

chronic diseases and health risk by ACEs

chronic diseases and health risk by education

chronic diseases and health risk by cash transfers.

**3.4. Baseline characteristics of study population study and applying preliminary statistical tests**.

Description of population study by sex and age groups.

Description of population study by family socioeconomic position.

Table description of population study by ACEs.

Table description of population study by education.

Table description of population study by cash transfers.

**3.5. Cleaning and selecting final data by eliminating columns no needed to calculate BHS**. After transformation data we have 91 columns in the dataframe. We delete 66 columns which are no required to the calculation of BHS and regression tables by mapping Null value to them.

biomarkers.imp$P835\_6 <- biomarkers.imp$P835\_7 <- biomarkers.imp$P835\_1 <- biomarkers.imp$P835\_2 <- biomarkers.imp$P835\_3 <- biomarkers.imp$P835\_4 <- biomarkers.imp$P835\_5 <- biomarkers.imp$P835 <- biomarkers.imp$early\_infection1 <- biomarkers.imp$early\_infection2 <- biomarkers.imp$early\_infection3 <- biomarkers.imp$early\_infection4 <- biomarkers.imp$early\_infection5 <- biomarkers.imp$early\_infection7 <- biomarkers.imp$systolic\_bprh <-biomarkers.imp$early\_infection <- biomarkers.imp$early\_infection6 <- biomarkers.imp$systolic\_bplh <- biomarkers.imp$dyastolic\_bplh <- biomarkers.imp$cash\_transfer1 <- biomarkers.imp$cash\_transfer2 <- biomarkers.imp$medication\_1 <- biomarkers.imp$medication\_2 <- biomarkers.imp$medication\_3 <- biomarkers.imp$medication\_4 <- biomarkers.imp$medication\_5 <- biomarkers.imp$medication\_6 <- biomarkers.imp$medication\_7 <- biomarkers.imp$medication <-   
biomarkers.imp$height\_m <- biomarkers.imp$weight <- biomarkers.imp$height <- biomarkers.imp$height\_m\_squared <- biomarkers.imp$neglected\_food <- biomarkers.imp$emotional\_abuse <- biomarkers.imp$household\_violence <- biomarkers.imp$migration <- biomarkers.imp$migration\_yo <- biomarkers.imp$poor\_health2 <- biomarkers.imp$CKD <- biomarkers.imp$viral.hepatitis <-   
biomarkers.imp$measles <- biomarkers.imp$TB <- biomarkers.imp$rheumatic.fever <- biomarkers.imp$asthma <- biomarkers.imp$bronchitis <- biomarkers.imp$nutrition\_program <- biomarkers.imp$weight\_2 <- biomarkers.imp$height\_2 <- biomarkers.imp$wc\_2 <- NULL  
  
view(biomarkers.imp)

Save new dataframe for bivariate analysis and hypothesis testing

writexl::write\_xlsx( x = biomarkers.imp, path = '/Users/macbookpro/Documents/PhD Project/Data/SABE/dataframes/biomarkers.imp2.xlsx' )

Save biomarkers and covariates dataframe (selection of 14 variables).

myvars1 <- c("ID", "hdl", "ldl", "total\_cholesterol", "triglycerides", "glucose", "glycated\_haemoglobin", "wc", "bmi\_2",  
"systolic\_bprh\_2","dyastolic\_bprh\_2", "age\_group", "sex", "ethnic\_group")  
  
covbiomarkers <- biomarkers.imp[myvars1]

Save covbiomarkers in a separated dataframe for BHS calculation.

writexl::write\_xlsx( x = covbiomarkers, path = '/Users/macbookpro/Documents/PhD Project/Data/SABE/dataframes/covbiomarkers.xlsx' )

####END CODE 2####