

Preoperative Atelectasis

Part 1: Overview, selection criteria, and missing data

Javier Mancilla Galindo

2023-12-01

Setup

Packages used

```
if (!require("pacman", quietly = TRUE)) {  
  install.packages("pacman")  
}  
  
pacman::p_load(  
  tidyverse, # Used for basic data handling and visualization.  
  dataverse, # Retrieve dataset from the Harvard dataverse.  
  overviewR, # Used to assess missing data.  
  table1, #Used to add labels to variables.  
  report #Used to cite packages used in this session.  
)
```

Session and package dependencies

R version 4.3.2 (2023-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 22621)

Matrix products: default

locale:

[1] LC_COLLATE=Spanish_Mexico.utf8 LC_CTYPE=Spanish_Mexico.utf8
[3] LC_MONETARY=Spanish_Mexico.utf8 LC_NUMERIC=C

```
[5] LC_TIME=Spanish_Mexico.utf8
```

```
time zone: Europe/Berlin
```

```
tzcode source: internal
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices datasets  utils      methods    base
```

```
other attached packages:
```

```
[1] report_0.5.7      table1_1.4.3      overviewR_0.0.13  dataverse_0.3.13  
[5] lubridate_1.9.3   forcats_1.0.0     stringr_1.5.1     dplyr_1.1.4  
[9] purrr_1.0.2       readr_2.1.4       tidyr_1.3.0       tibble_3.2.1  
[13] ggplot2_3.4.4     tidyverse_2.0.0   pacman_0.5.1
```

General overview

```
summary(data)
```

ID	age	sex	weight	height
Length:243	Min. :20.00	Woman:221	Min. : 73.8	Min. :1.480
Class :character	1st Qu.:33.00	Man : 22	1st Qu.: 97.6	1st Qu.:1.620
Mode :character	Median :40.00		Median :111.2	Median :1.660
	Mean :40.37		Mean :115.9	Mean :1.670
	3rd Qu.:49.00		3rd Qu.:129.7	3rd Qu.:1.715
	Max. :65.00		Max. :264.6	Max. :1.910

BMI	type_obesity	ARISCAT	ARISCAT_group
Min. :30.00	Class 1 Obesity: 62	Min. :23.0	Low Risk :178
1st Qu.:34.81	Class 2 Obesity: 55	1st Qu.:23.0	Intermediate Risk: 65
Median :40.31	Class 3 Obesity:126	Median :23.0	High Risk : 0
Mean :41.51		Mean :24.6	
3rd Qu.:46.15		3rd Qu.:26.0	
Max. :77.31		Max. :42.0	

ASA	spo2_VPO	surgical_procedure	CORADS	atelectasis
ASA 1: 52	Min. :88.00	SG :192	CO-RADS 1:233	Yes: 81
ASA 2:149	1st Qu.:93.00	RYGB: 7	CO-RADS 2: 6	No :162
ASA 3: 33	Median :96.00	OAGB: 5	CO-RADS 3: 2	
NA's : 9	Mean :94.93	LBGS: 31	CO-RADS 4: 2	
	3rd Qu.:97.00	NA's: 8		

Max. :99.00

atelectasis_location	atelectasis_percent	hb	hct
Unilateral: 55	Min. : 0.000	Min. : 9.90	Min. :30.30
Bilateral : 26	1st Qu.: 0.000	1st Qu.:13.90	1st Qu.:40.52
NA's :162	Median : 0.000	Median :14.45	Median :42.60
	Mean : 2.778	Mean :14.52	Mean :42.68
	3rd Qu.: 5.000	3rd Qu.:15.20	3rd Qu.:44.67
	Max. :27.500	Max. :18.50	Max. :52.90
		NA's :5	NA's :5

leu	neu_percent	neu_absolute	linf_percent
Min. : 3.100	Min. :32.00	Min. : 1.600	Min. :14.00
1st Qu.: 6.600	1st Qu.:59.00	1st Qu.: 3.969	1st Qu.:30.00
Median : 7.700	Median :63.00	Median : 4.883	Median :35.00
Mean : 7.826	Mean :62.94	Mean : 4.956	Mean :34.87
3rd Qu.: 8.900	3rd Qu.:68.00	3rd Qu.: 5.894	3rd Qu.:39.00
Max. :13.300	Max. :84.00	Max. :10.773	Max. :66.00
NA's :5	NA's :5	NA's :5	NA's :5

linf_absolute	mon_percent	mon_absolute	platelets
Min. :0.616	Min. :0.000	Min. :0.616	Min. :154.0
1st Qu.:2.244	1st Qu.:1.000	1st Qu.:2.244	1st Qu.:272.2
Median :2.587	Median :2.000	Median :2.587	Median :316.0
Mean :2.703	Mean :1.626	Mean :2.703	Mean :317.2
3rd Qu.:3.095	3rd Qu.:2.000	3rd Qu.:3.095	3rd Qu.:354.8
Max. :5.676	Max. :4.000	Max. :5.676	Max. :495.0
NA's :5	NA's :5	NA's :5	NA's :5

glucose	urea	creatinine	rapid_covid_test
Min. : 59.00	Min. :10.0	Min. :0.5000	negative:243
1st Qu.: 74.00	1st Qu.:17.0	1st Qu.:0.6600	
Median : 83.00	Median :20.5	Median :0.7500	
Mean : 85.61	Mean :21.4	Mean :0.7577	
3rd Qu.: 92.00	3rd Qu.:26.0	3rd Qu.:0.8400	
Max. :200.00	Max. :49.0	Max. :1.5900	
NA's :5	NA's :5	NA's :5	

PCR_covid	surgery_performed	state_residence	altitude
negative: 3	No : 10	Texas :87	Min. : 31.0
NA's :240	Yes:233	Washington:39	1st Qu.: 519.0
		Utah :27	Median : 519.0
		Alberta :22	Mean : 650.1
		Florida :20	3rd Qu.: 806.0
		California:16	Max. :1861.0
		(Other) :32	

hypertension diabetes sleep_apnea hypothyroidism dyslipidemia

No :179	No :218	No :224	No :219	No :224
Yes: 64	Yes: 25	Yes: 19	Yes: 24	Yes: 19

antidepressant_use	prior_covid19	other_comorb
No :146	No :239	No :143
Yes: 97	Yes: 4	Yes:100

Exclude participants with CO-RADS 3:

Participants with higher probability of having a current diagnosis of COVID-19 are expected to have chest CT alterations due to COVID-19 pneumonia. Thus, will be excluded.

Number of patients according to CO-RADS:

CO-RADS 1	CO-RADS 2	CO-RADS 3	CO-RADS 4
233	6	2	2

Number of participants in the dataset (to keep track of no. of exclusions).

```
count(data)
```

```
# A tibble: 1 x 1
      n
<int>
1    243
```

```
data <- data %>%
  filter(as.numeric(CORADS) < 3) %>%
  droplevels(data$CORADS)
```

```
count(data)
```

```
# A tibble: 1 x 1
      n
<int>
1    239
```

Exclude participants with prior COVID-19:

Since prior COVID-19 is considered a confounder and since there are only 3 participants with prior COVID-19 which would provide difficult to assess the role of prior COVID-19 in analyses, participants with prior COVID-19 were excluded from the analysis.

```
count(data)
```

```
# A tibble: 1 x 1
      n
<int>
1    239
```

```
data <- data %>%
  filter(prior_covid19 == "No") %>%
  droplevels(data$prior_covid19)
```

```
count(data)
```

```
# A tibble: 1 x 1
      n
<int>
1    236
```

Will remove prior_covid19 column as it no longer provides information of a varying characteristic. Similarly, rapid_covid_test does not provide additional information.

```
length(data)
```

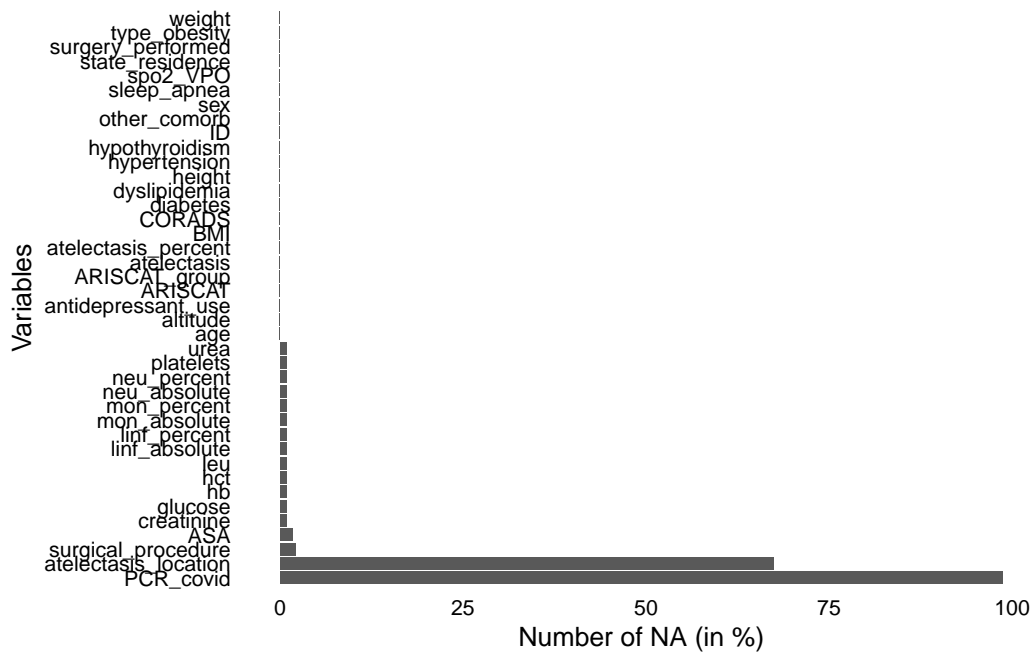
```
[1] 42
```

```
data <- data %>% select(-c(prior_covid19, rapid_covid_test))
```

```
length(data)
```

```
[1] 40
```

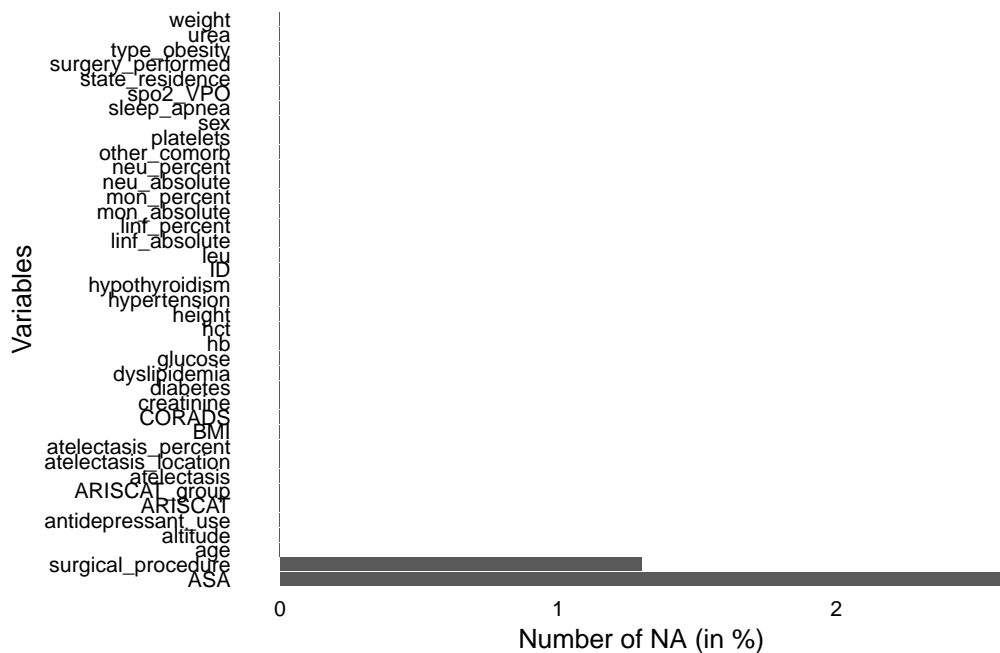
Missing data per variable



Missing data for PCR_covid and is explained since only patients who decided to have a test performed on their own will reported the result. The medical center did not require a negative PCR test at that time during the pandemic, reason why PCR tests were not systematically performed. As shown in earlier summary of variables, all available tests (n=3) were negative. This variable will not be analyzed further downstream:

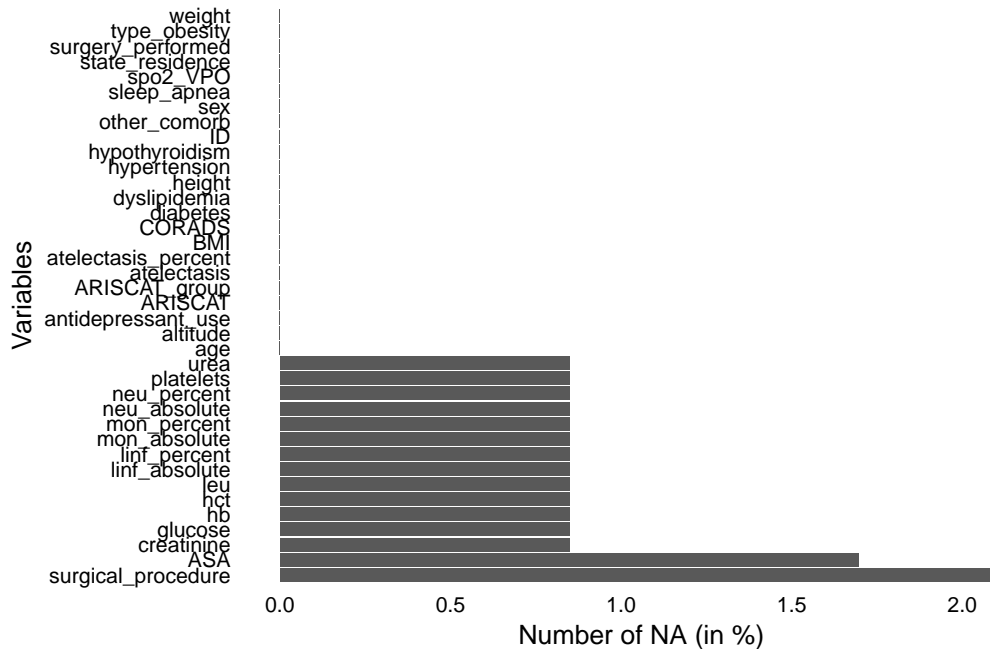
```
data <- data %>% select(-PCR_covid)
```

The variable atelectasis_location has missing data since those patients who did not have atelectasis naturally do not have a location recorded. Will assess if data are missing for those who had atelectasis:



There is no missing data for *atelectasis_location* after sub-setting only those who had atelectasis.

Lastly, I will subset all participants without the prior variable to further assess the extent of missing data for other variables:



Missing data constitutes **less than 5%** for all remaining variables. Thus, will proceed with **complete-case analysis** without performing any data imputation procedure.

Package references

- Grolemund G, Wickham H (2011). “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software*, 40(3), 1-25. <https://www.jstatsoft.org/v40/i03/>.
- Kuriwaki S, Beasley W, Leeper TJ (2023). *dataverse: R Client for Dataverse 4+ Repositories*. R package version 0.3.13.
- Meyer C, Hammerschmidt D (2023). *overviewR: Easily Extracting Information About Your Data*. R package version 0.0.13, <https://CRAN.R-project.org/package=overviewR>.
- Müller K, Wickham H (2023). *tibble: Simple Data Frames*. R package version 3.2.1, <https://CRAN.R-project.org/package=tibble>.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rich B (2023). *table1: Tables of Descriptive Statistics in HTML*. R package version 1.4.3, <https://CRAN.R-project.org/package=table1>.
- Rinker TW, Kurkiewicz D (2018). *pacman: Package Management for R*. version 0.5.0, <http://github.com/trinker/pacman>.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.

- Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0, <https://CRAN.R-project.org/package=forcats>.
- Wickham H (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1, <https://CRAN.R-project.org/package=stringr>.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.
- Wickham H, Henry L (2023). *purrr: Functional Programming Tools*. R package version 1.0.2, <https://CRAN.R-project.org/package=purrr>.
- Wickham H, Hester J, Bryan J (2023). *readr: Read Rectangular Text Data*. R package version 2.1.4, <https://CRAN.R-project.org/package=readr>.
- Wickham H, Vaughan D, Girlich M (2023). *tidyr: Tidy Messy Data*. R package version 1.3.0, <https://CRAN.R-project.org/package=tidyr>.