# Preoperative Atelectasis

## Part 4: Outcomes

Javier Mancilla Galindo

2023-12-01

## Table of contents



## Setup

### Packages used

```
if (!require("pacman", quietly = TRUE)) {
  install.packages("pacman")
}
```

```r
pacman::p_load(
  tidyverse, # Used for basic data handling and visualization.
  table1, #Used to add lables to variables.
  RColorBrewer, #Color palettes for data visualization.
  gridExtra, #Used to arrange multiple ggplots in a grid.
  grid, #Used to arrange multiple ggplots in a grid.
  mgcv, #Used to model non-linear relationships with a general additive model.
  ggmosaic, #Used to create mosaic plots.
  car, #Used assess distribution of continuous variables (stacked Q-Q plots).
  simpleboot, boot, # Used to calculate mean atelectasis coverage and
                  # 95%CI through bootstrapping.
  gt, #Used to present tables in html format.
  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] grid      stats     graphics  grDevices datasets  utils     methods
[8] base

other attached packages:
 [1] report_0.5.7       gt_0.10.0          boot_1.3-28.1      simpleboot_1.1-7
 [5] car_3.1-2          carData_3.0-5      ggmosaic_0.3.3     mgcv_1.9-0
 [9] nlme_3.1-164       gridExtra_2.3      RColorBrewer_1.1-3 table1_1.4.3
[13] lubridate_1.9.3    forcats_1.0.0      stringr_1.5.1      dplyr_1.1.4
```

```
[17] purrr_1.0.2        readr_2.1.4        tidyr_1.3.0        tibble_3.2.1
[21] ggplot2_3.4.4      tidyverse_2.0.0    pacman_0.5.1
```

Set seed (for reproducibility of bootstrapping) as the current year 2023:

```
seed <- 2023
```

## Outcome variable

Corroborate that atelectasis(Yes/No) matches atelectasis percent equal or different to 0%:

```
           atelectasis_percent
atelectasis   0 2.5   5 7.5  10 12.5  15 17.5 27.5
        Yes   0  11  14  33   6    1   4    7    1
        No  159   0   0   0   0    0   0    0    0
```

Yes, these do match.

### Prevalence of atelectasis

```
             Yes    No
frequencies 77.0 159.0
percent     32.6  67.4
```

Prevalence of atelectasis with 95% confidence interval

```
	1-sample proportions test without continuity correction

data:  frequencies, null probability 0.5
X-squared = 28.492, df = 1, p-value = 9.411e-08
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2696526 0.3884549
sample estimates:
        p 
0.3262712 
```

> The prevalence of atelectasis was **32.6 (95%CI: 26.97, 38.85)**.

### Atelectasis - obesity class

Mean expected frequency:

```
  mean_expected_freq
1           39.33333
```

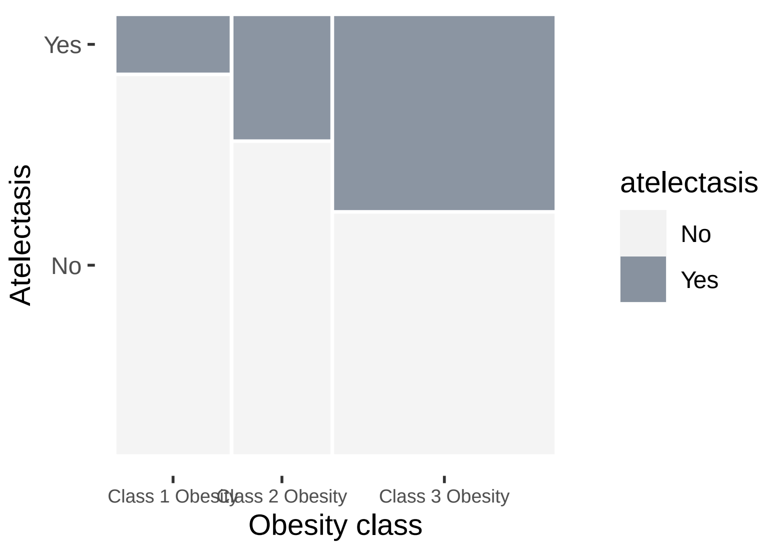
Frequencies:

```
                 atelectasis
type_obesity      Yes No
  Class 1 Obesity   8 54
  Class 2 Obesity  15 38
  Class 3 Obesity  54 67
```

Percentage:

```
                 atelectasis
type_obesity        Yes    No
  Class 1 Obesity 12.90 87.10
  Class 2 Obesity 28.30 71.70
  Class 3 Obesity 44.63 55.37
```

Mosaic Plot



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 19.352, df = 2, p-value = 6.279e-05
```

#### Atelectasis location by obesity class

Mean expected frequency:

```
  mean_expected_freq
1           12.83333
```

Mean expected frequency is greater than 5.0, so chi-squared without continuity correction is adequate.

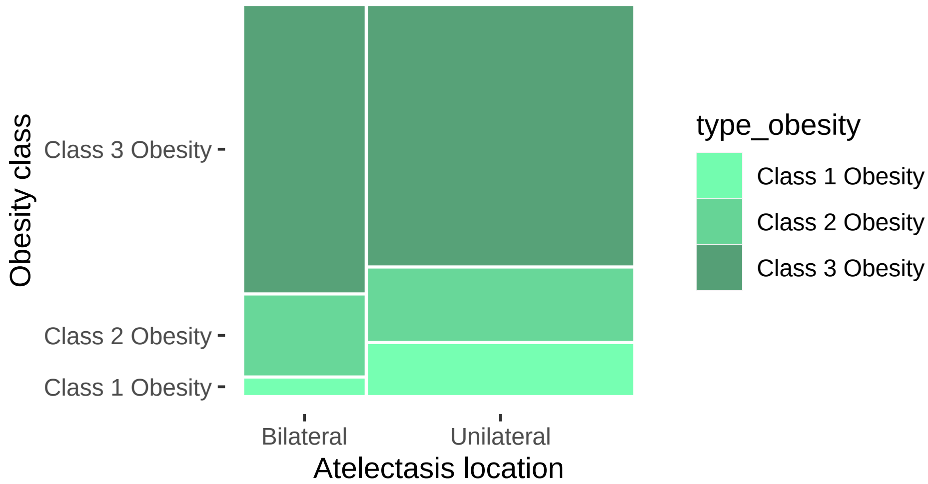
Frequencies:

```
                 atelectasis_location
type_obesity      Bilateral Unilateral
  Class 1 Obesity         1          7
  Class 2 Obesity         5         10
  Class 3 Obesity        18         36
```

Percentage:

```
                 atelectasis_location
type_obesity      Bilateral Unilateral
  Class 1 Obesity     12.50      87.50
  Class 2 Obesity     33.33      66.67
  Class 3 Obesity     33.33      66.67
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 1.4503, df = 2, p-value = 0.4843
```

Prevalence of atelectasis with 95% confidence intervals calculated with sourced script ***Prevalence_atelectasis.R***

> The prevalence of atelectasis was greater in higher obesity classes: class 1, n=8 (12.9%, 95%CI:6.13 - 24.4); class 2, n=15 (28.3%, 95%CI:17.2 - 42.56); and class 3, n=54 (44.63%, 95%CI:35.68 - 53.92) (p<0.001).

> Of those who had atelectasis, the most frequent presentation was unilateral n=53 (68.83%), compared to bilateral n=24 (31.17%). When examining this by obesity class, the observed distribution was not significantly different for those with class 1, 2, and 3 obesity categories (n=7 (87.5%), n=10 (66.67%), and n=36 (66.67%), respectively) (p=0.484).

### Atelectasis Percent

#### Mean atelectasis percentage

The following would be the mean atelectasis percentage coverage if a normal distribution were assumed, which is what has been done in some prior studies:

```
      mean       sd
1 2.658898 4.687145
```

And by obesity class:

```
# A tibble: 3 x 3
  type_obesity     mean    sd
  <fct>           <dbl> <dbl>
1 Class 1 Obesity 0.927  2.91
2 Class 2 Obesity 1.56   3.15
3 Class 3 Obesity 4.03   5.52
```

As is evident from these numbers, assuming normality causes standard deviation to capture negative values, which is impossible in reality for this variable.

Thus, bootstrapping the mean and 95%CI is expected to lead to more appropriate estimates.

Mean by bootstrapping for the total sample:

```
[1] 2.664441
```

Bootstrap 95% confidence intervals:

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_atel)

Intervals :
Level      Normal              Basic
95%   ( 2.058,  3.248 )   ( 2.044,  3.220 )

Level     Percentile            BCa
95%   ( 2.097,  3.273 )   ( 2.076,  3.263 )
Calculations and Intervals on Original Scale
```

The bias-corrected and accelerated (BCa) bootstrap interval is known to lead to more stable intervals with better coverage. Will report this. However, it is a good thing that here 95%CI through different methods do not lead to widely different results.

Now, I will calculate this for different BMI categories:

Class 1

Mean:

```
[1] 0.9268065
```

95% CI:

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_class1)

Intervals :
Level      Normal              Basic
95%   ( 0.1968,  1.6593 )   ( 0.1210,  1.5726 )

Level     Percentile            BCa
95%   ( 0.2823,  1.7339 )   ( 0.3226,  1.8145 )
Calculations and Intervals on Original Scale
```

Class 2

Mean:

```
[1] 1.552382
```

95% CI:

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_class2)

Intervals :
Level      Normal              Basic
95%   ( 0.723,  2.399 )   ( 0.660,  2.311 )

Level     Percentile            BCa
95%   ( 0.802,  2.453 )   ( 0.755,  2.453 )
Calculations and Intervals on Original Scale
```

Class 3

Mean:

```
[1] 4.036455
```

95% CI:

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_class3)

Intervals :
Level      Normal              Basic
95%   ( 3.043,  5.000 )   ( 2.996,  4.979 )

Level     Percentile            BCa
95%   ( 3.079,  5.062 )   ( 3.058,  5.021 )
Calculations and Intervals on Original Scale
```

> The mean atelectasis percentage coverage in the sample was 2.66% (95%CI:2.08-3.26) and according to obesity categories: class 1 (0.93%, 95%CI:0.32-1.81), class 2 (1.55%, 95%CI:0.75-2.45), and class 3 (4.04%, 95%CI:3.06-5.02).

What could happen if I took random subsamples between n=20 and n=30 similar to what has been done in other studies and calculate mean BMI and mean atelectasis percentage assuming normal distributions?

Random sample

```
  mean_BMI mean_atelectasis
1   42.024            2.375

  mean_BMI mean_atelectasis
1  40.9836              2.1

  mean_BMI mean_atelectasis
1   40.712             1.75
```

### Atelectasis percentage by obesity class

Now, I will continue assessing atelectasis percentage if assumed to be categorical ordinal:

Mean expected frequency:

```
  mean_expected_freq
1           8.740741
```

Mean expected frequency is greater than 5.0, so chi-squared without continuity correction is adequate.

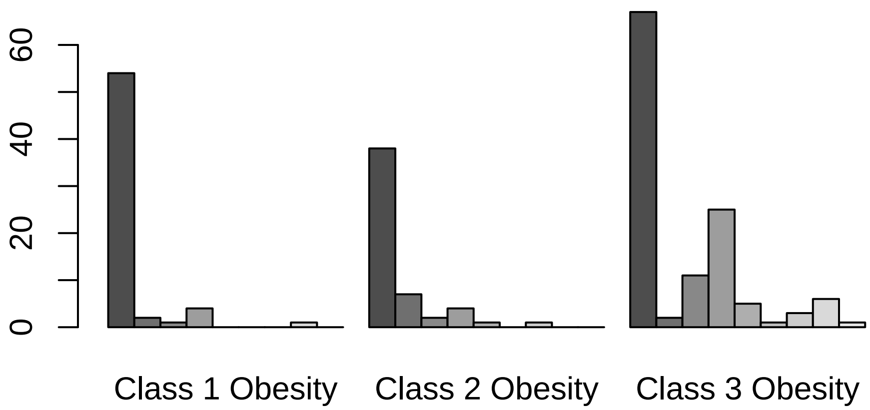
Frequencies:

```
                   type_obesity
atelectasis_percent Class 1 Obesity Class 2 Obesity Class 3 Obesity
               0                 54              38              67
               2.5                2               7               2
               5                  1               2              11
               7.5                4               4              25
               10                 0               1               5
               12.5               0               0               1
               15                 0               1               3
               17.5               1               0               6
               27.5               0               0               1
```

Percentage by obesity class

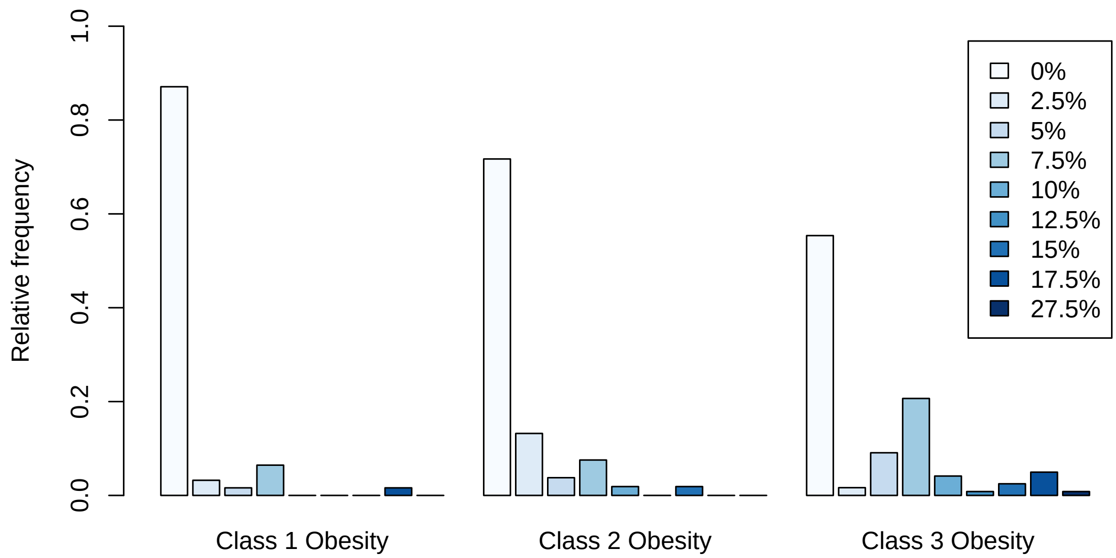
```
                   type_obesity
atelectasis_percent Class 1 Obesity Class 2 Obesity Class 3 Obesity
               0              87.10           71.70           55.37
               2.5             3.23           13.21            1.65
               5               1.61            3.77            9.09
               7.5             6.45            7.55           20.66
               10              0.00            1.89            4.13
               12.5            0.00            0.00            0.83
               15              0.00            1.89            2.48
               17.5            1.61            0.00            4.96
               27.5            0.00            0.00            0.83
```

Barplot of absolute frequencies:

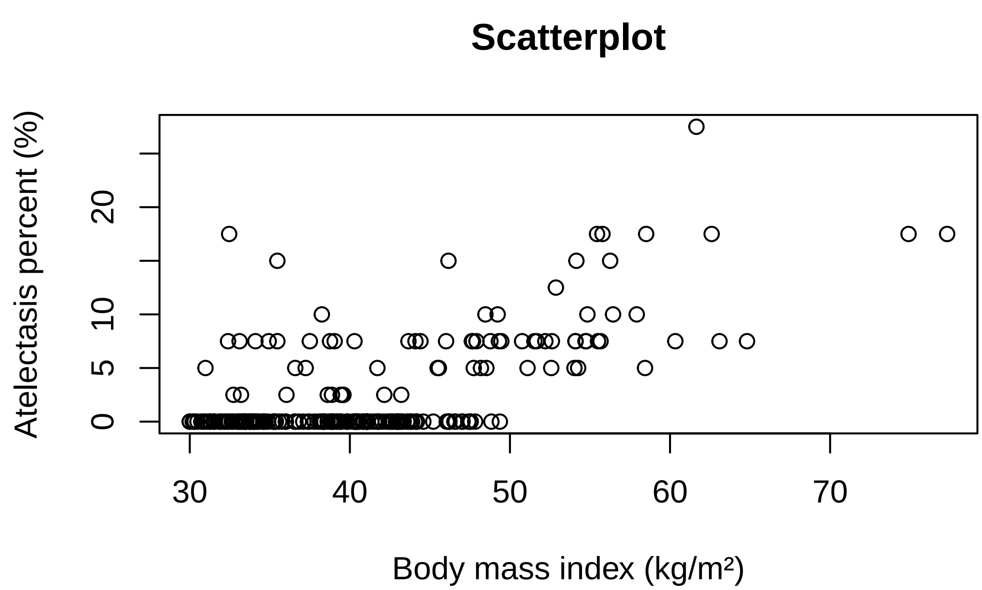



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 39.434, df = 16, p-value = 0.0009412
```

**Barplot of atelectasis percentage by obesity class category**


Relative frequency
0.0
0.2
0.4
0.6
0.8
1.0
Class 1 Obesity
Class 2 Obesity
Class 3 Obesity
0%
2.5%
5%
7.5%
10%
12.5%
15%
17.5%
27.5%

**Smooth term?**


Scatterplot
Atelectasis percent (%)
0
5
10
20
30
40
50
60
70
Body mass index (kg/m²)

Atelectasis percent seems to increase as BMI increases. However, relationship is not linear.

Models evaluated with the accompanying sourced script ***nonlinear__BMI__Atelectasis.R***

All models are significantly better than linear. Thus, using a smooth term for BMI to predict atelectasis percent is better than modelling a linear relationship.

Best AIC:

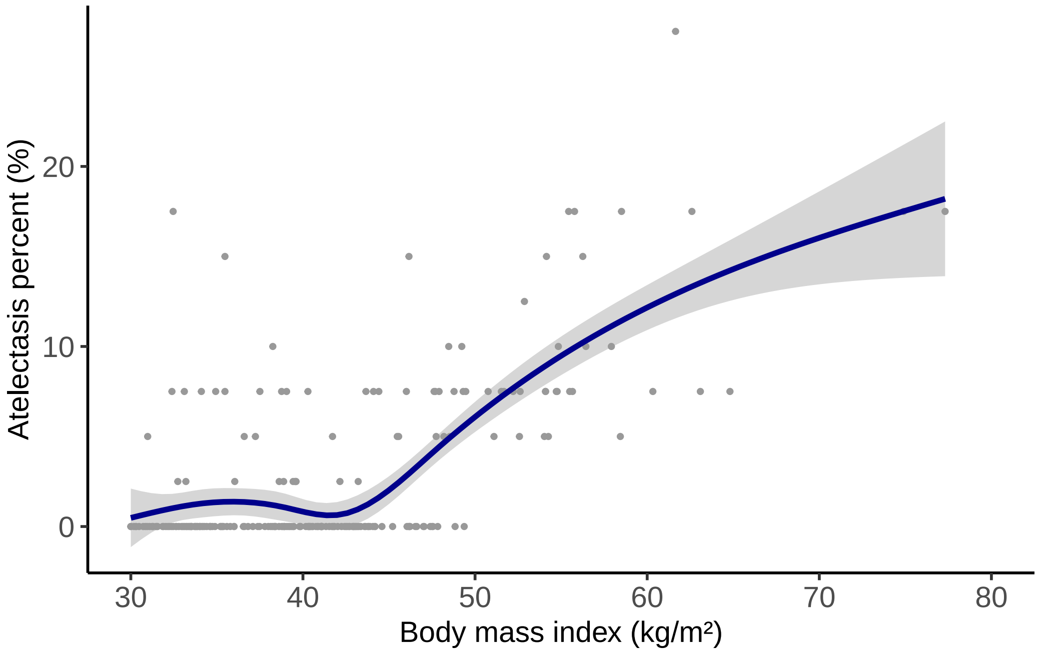
| model | AIC |
|---|---|
| k=2 | 1259.7 |
| k=4 | 1237.6 |
| k=6 | 1232.3 |
| k=8 | 1233.6 |

Regarding AIC, greatest improvement in AIC is k=6. Will model with k=5 and k=7 to compare

Best AIC:

| model | AIC |
|---|---|
| k=5 | 1235.7 |
| k=6 | 1232.3 |
| k=7 | 1233.2 |

k=6 offers the lowest AIC. Will keep k=6 to model.




Positive non-monotonic relationship since atelectasis increases as BMI increases only after ~BMI equal to 42.

Will assess Spearman's correlation again only to have a rough idea (will not report this in the paper since the relationship is not monotonic):

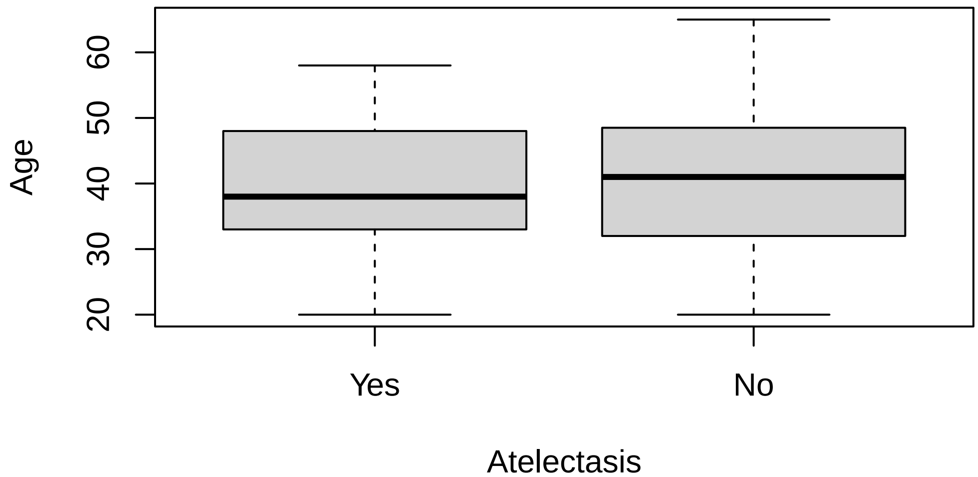
```
    Spearman's rank correlation rho

data:  spo2_VPO and atelectasis_percent
S = 3883525, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.7727568
```

> Atelectasis percent exhibited a negative non-linear non-monotonic relationship with SpO2 (**Figure 1A**, rho= -0.773, p<0.001).

Note that this p-value refers to the smooth term vs linear as assessed in GAM models.

Interestingly, this figure is almost a mirror image of the priorly created plot for SpO2 ~ BMI.

### Atelectasis - age




Assess distribution of age by atelectasis (yes/no):

Distribution near-normal, will assess mean and variance for further testing:

| Atelectasis | n | age_mean | sd | variance |
|---|---|---|---|---|
| Yes | 77 | 39.65 | 9.30 | 86.5728 |
| No | 159 | 40.55 | 10.14 | 102.8816 |

Variances near-similar, but group sizes differ. Welch's t-test more suitable:

```
    Welch Two Sample t-test

data:  age by atelectasis
t = -0.67931, df = 162.72, p-value = 0.4979
alternative hypothesis: true difference in means between group Yes and group No is not equal
95 percent confidence interval:
 -3.532230  1.724013
sample estimates:
mean in group Yes  mean in group No
         39.64935          40.55346
```

Age was similarly distributed among patients without atelectasis (40.5, sd:10.1) and those with atelectasis (39.6, sd:9.3) (p=0.498).

## Atelectasis - sex

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Mean expected frequency is greater than 5.0, so chi-squared without continuity correction is adequate.

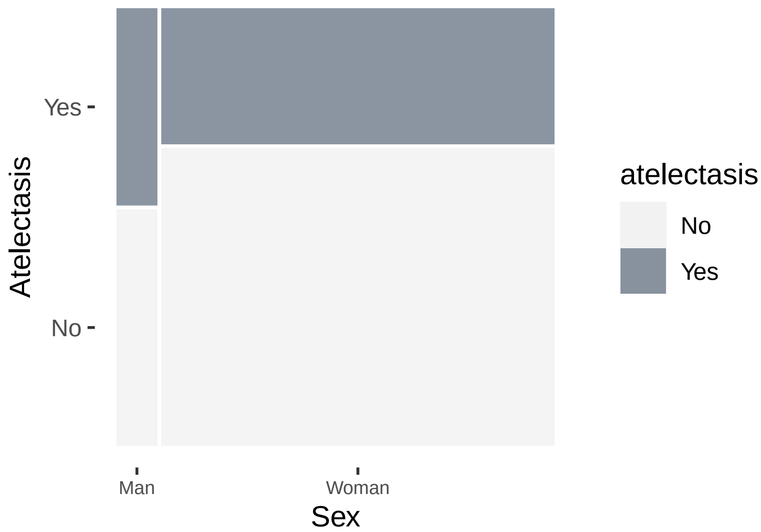
Frequencies:

```
       atelectasis
sex      Yes  No
  Man     10  12
  Woman   67 147
```

Percentage:

```
       atelectasis
sex        Yes    No
  Man    45.45 54.55
  Woman 31.31 68.69
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 1.8161, df = 1, p-value = 0.1778
```

> There were no significant differences in atelectasis ocurrence between men (31.3%) and women (45.5%) (p=0.178).

### Atelectasis - OSA

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Mean expected frequency is greater than 5.0, so chi-squared without continuity correction is adequate.

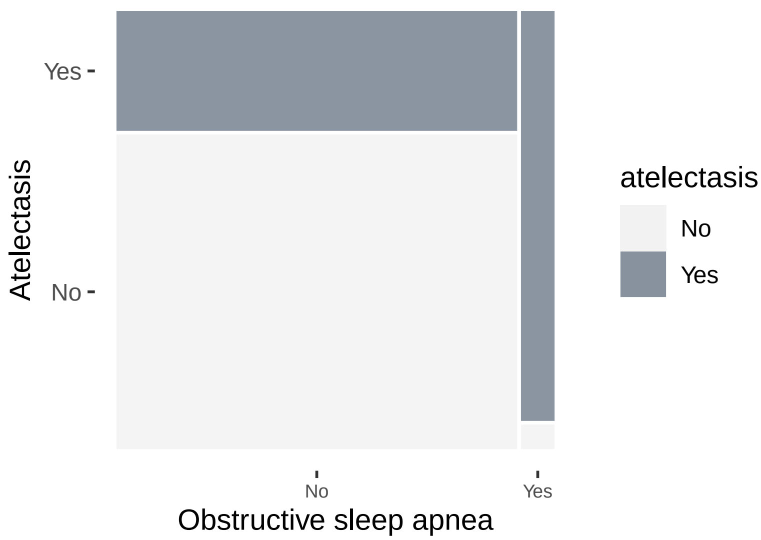
Frequencies:

```
           atelectasis
sleep_apnea Yes  No
        No   60 158
        Yes  17   1
```

Percentage:

```
           atelectasis
sleep_apnea   Yes    No
        No  27.52 72.48
        Yes 94.44  5.56
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 33.875, df = 1, p-value = 5.876e-09
```

Patients with a diagnosis of obstructive sleep apnea had atelectasis more frequently (94.4%) than those without the diagnosis (27.5%) (p<0.001).

### Atelectasis location by OSA

Mean expected frequency:

```
  mean_expected_freq
1              19.25
```

Mean expected frequency is greater than 5.0, so chi-squared without continuity correction is adequate.

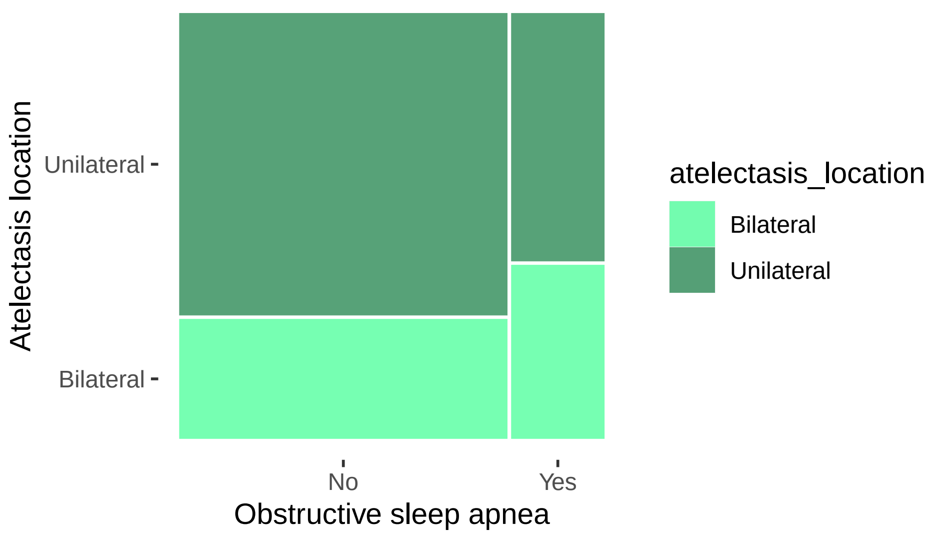
Frequencies:

```
           atelectasis_location
sleep_apnea Bilateral Unilateral
        No         17         43
        Yes         7         10
```

Percentage:

```
           atelectasis_location
sleep_apnea Bilateral Unilateral
        No      28.33      71.67
        Yes     41.18      58.82
```

Mosaic Plot

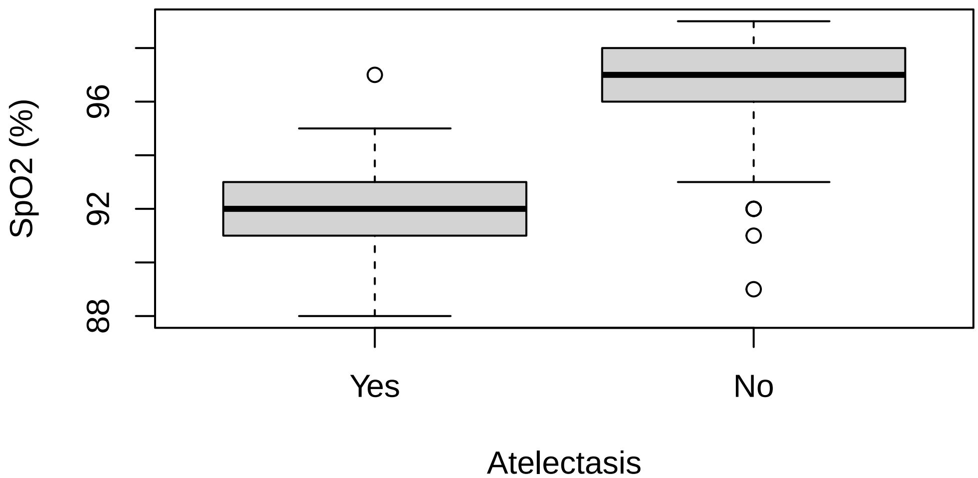



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 1.0185, df = 1, p-value = 0.3129
```

The location of atelectasis was not different among patients with and without OSA (p=0.313).

### Atelectasis - SpO2

| Atelectasis | n | spo2_median | Q1 | Q3 | min | max |
|---|---|---|---|---|---|---|
| Yes | 77 | 92 | 91 | 93 | 88 | 97 |
| No | 159 | 97 | 96 | 98 | 89 | 99 |




Distribution not normal and influential outliers. Will assess non-parametrically.

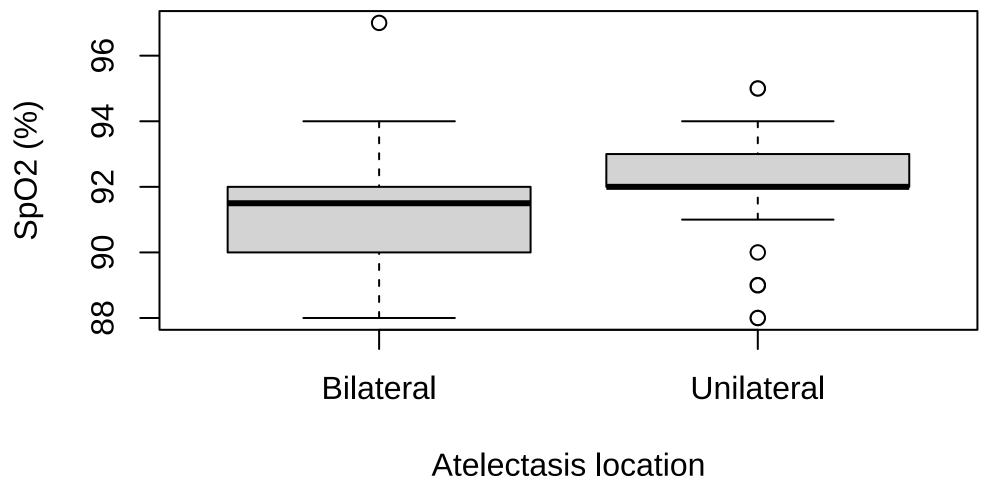
```
    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by atelectasis
W = 465.5, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

> The median SpO2 was significantly lower in patients with atelectasis (92, IQR: 91-93) compared to those without (97, IQR: 96-98) (p<0.001).

### Atelectasis location - SpO2

| Atelectasis location | n | spo2_median | Q1 | Q3 | min | max |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Bilateral | 24 | 91.5 | 90 | 92 | 88 | 97 |
| Unilateral | 53 | 92.0 | 92 | 93 | 88 | 95 |




Distribution not normal and likely influential outliers. Will assess non-parametrically.

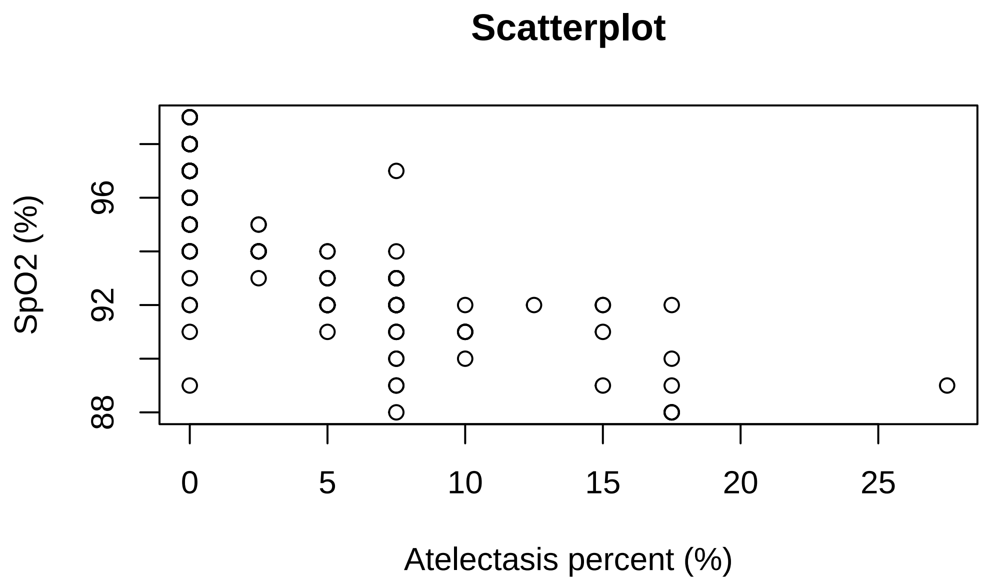
```
	Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by atelectasis_location
W = 393, p-value = 0.006227
alternative hypothesis: true location shift is not equal to 0
```

> The median SpO2 was significantly lower in patients with bilateral atelectasis (92, IQR: 92-93) compared to those with unilateral atelectasis (91.5, IQR: 90-92) (p=0.006).

### Atelectasis percent - SpO2

#### Smooth term?

Scatterplot




Decreasing SpO2 as atelectasis percent increases.

Would a smooth term be more useful to model SpO2?

Models evaluated with the accompanying sourced script ***nonlinear_Atelectasis_SpO2.R***

All models are significantly better than linear. Thus, using a smooth term for atelectasis percent is better than modelling a linear relationship.

Best AIC:

| model | AIC |
|---|---|
| k=2 | 892.1 |
| k=4 | 888.6 |
| k=6 | 885.8 |
| k=8 | 885.8 |

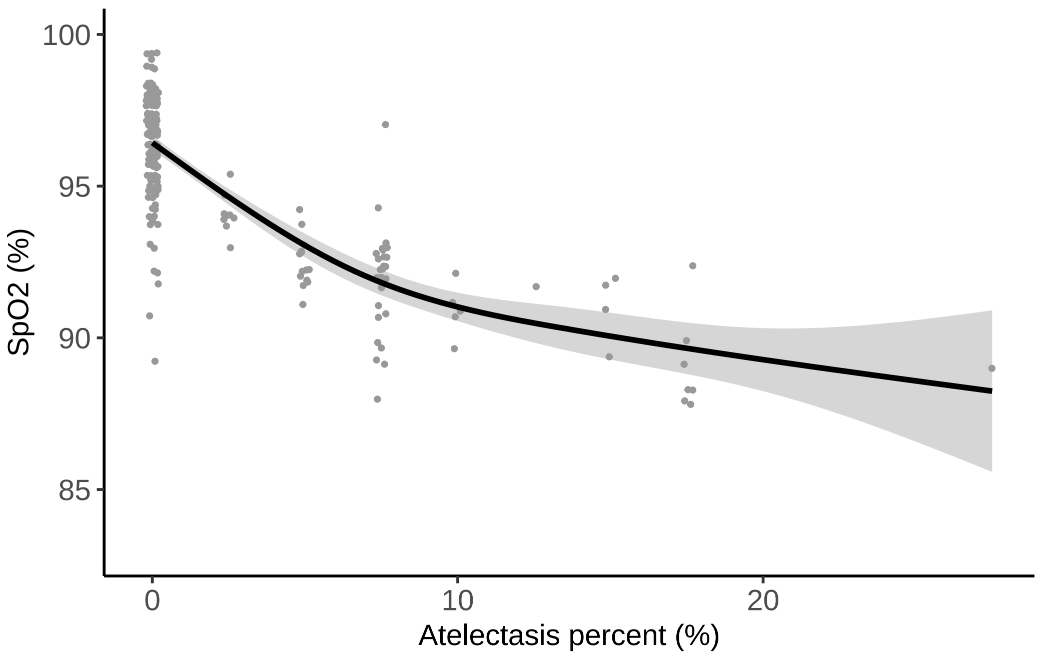
Regarding AIC, no model offers greater improvement in AIC than k=6. Will try a model with k=5.

Best AIC:

| model | AIC |
|---|---|
| k=4 | 888.6 |

| | |
|---|---|
| k=5 | 885.1 |
| k=6 | 885.8 |

There is a drop in AIC for k=5, which also offers the best k-index. Nonetheless, one problem with this is that the extra knot is explaining a clump around 12.%, for which there was only one single observation. Thus, it is likely that this clump and additional knot is only explaining noise in the data, and would thus not be a good representation of the trend in the variable. Thus, will keep k=4 to model as this model offers the best visual representation of the trend in all categories.



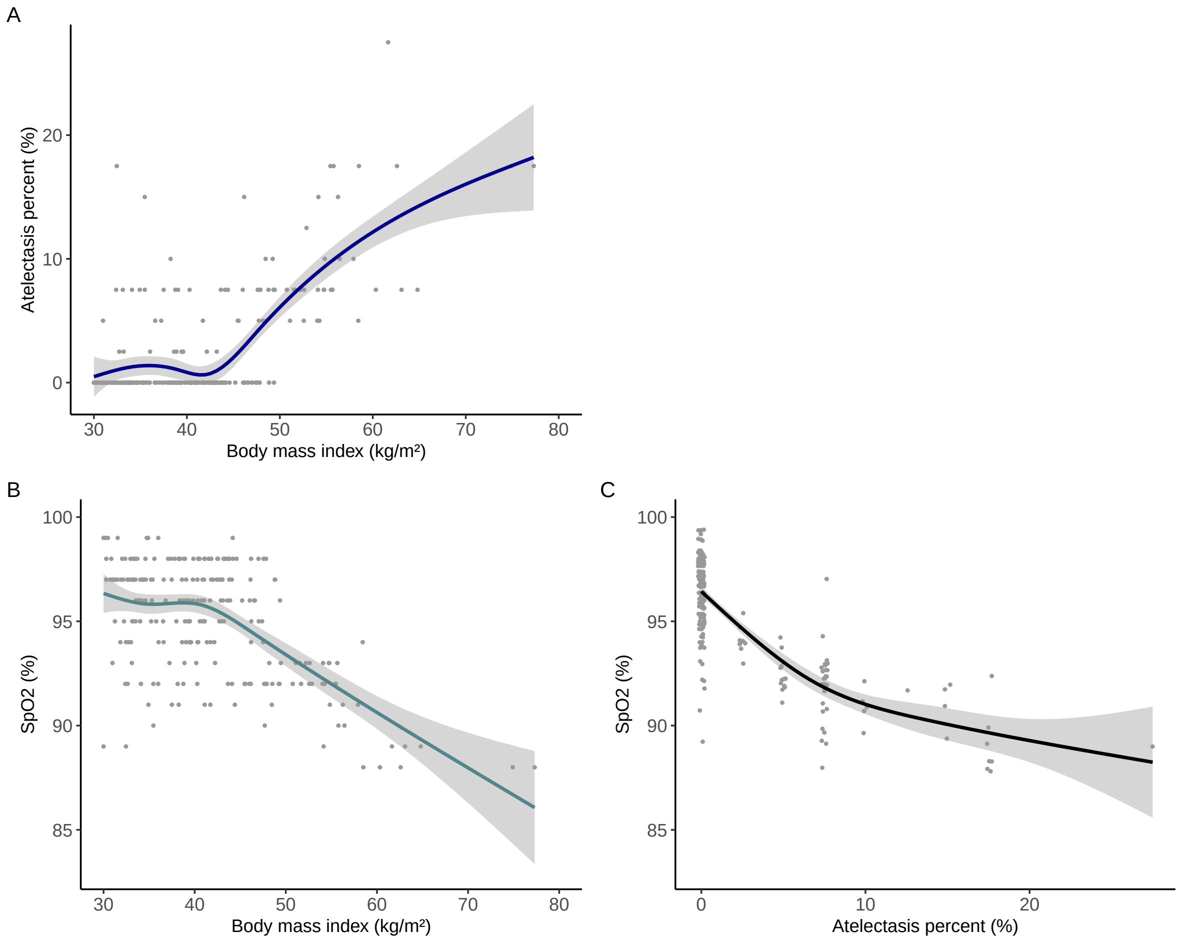

Negative monotonic relationship since SpO2 decreases as BMI increases. Will assess Spearman's correlation coefficient to report in paper:

```
    Spearman's rank correlation rho

data:  spo2_VPO and atelectasis_percent
S = 3883525, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.7727568
```

Atelectasis percent exhibited a negative non-linear monotonic relationship with SpO2 (**Figure 1C**, rho= -0.773, p<0.001).

## Figure 1

Created with the accompanying sourced script ***Figure1.R***




## Ordinal variable

Since there is only one participant in the 30% category, will collapse with the 20 category for further analyses:

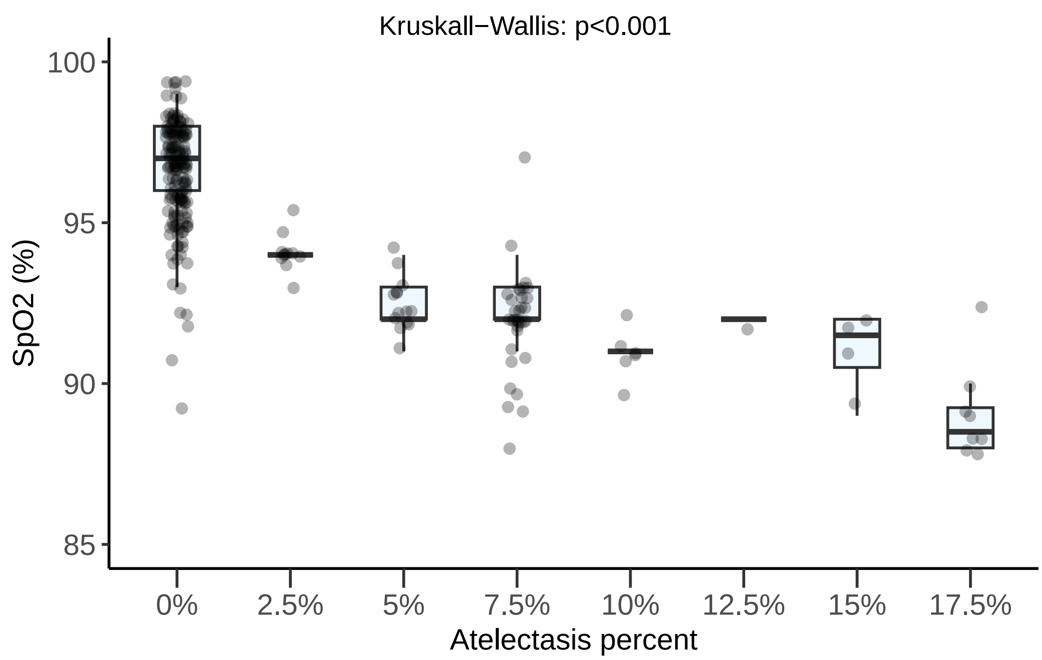
```
  0%  2.5%    5%  7.5%   10% 12.5%   15% 17.5%
 159    11    14    33     6     1     4     8
```

Distribution not normal, group sizes are different and there are outliers in both directions, depending where you are located. Thus, will proceed with non-parametric assessment:

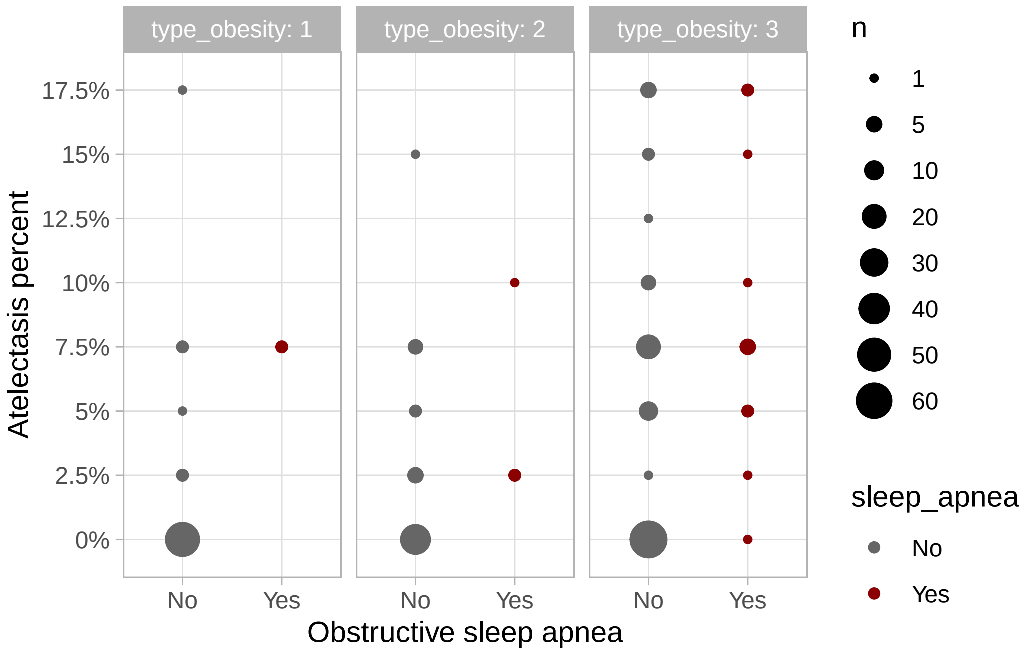
| atelectasis_percent | n | spo2_median | Q1 | Q3 | min | max |
|---|---|---|---|---|---|---|
| 0% | 159 | 97.0 | 96.0 | 98.00 | 89 | 99 |
| 2.5% | 11 | 94.0 | 94.0 | 94.00 | 93 | 95 |
| 5% | 14 | 92.0 | 92.0 | 93.00 | 91 | 94 |
| 7.5% | 33 | 92.0 | 92.0 | 93.00 | 88 | 97 |
| 10% | 6 | 91.0 | 91.0 | 91.00 | 90 | 92 |
| 12.5% | 1 | 92.0 | 92.0 | 92.00 | 92 | 92 |
| 15% | 4 | 91.5 | 90.5 | 92.00 | 89 | 92 |
| 17.5% | 8 | 88.5 | 88.0 | 89.25 | 88 | 92 |

```
    Kruskal-Wallis rank sum test

data:  spo2_VPO by atelectasis_percent
Kruskal-Wallis chi-squared = 141.19, df = 7, p-value < 2.2e-16
```

There was a decreasing trend in median SpO2 with higher atelectasis percentage extension ($p<0.001$).



**Relationship between OSA, obesity type and atelectasis percent:**




Sleep apnea was more common with higher BMI categories and also with higher atelectasis percentage. Atelectasis percent increases at higher obesity classes.

## Package References


- Auguie B (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3, https://CRAN.R-project.org/package=gridExtra.
- Canty A, Ripley BD (2022). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-28.1. Davison AC, Hinkley DV (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2, http://statwww.epfl.ch/davison/BMA/.
- Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
- Fox J, Weisberg S, Price B (2022). *carData: Companion to Applied Regression Data Sets*. R package version 3.0-5, https://CRAN.R-project.org/package=carData.
- 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/.
- Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J (2023). *gt: Easily Create Presentation-Ready Display Tables*. R package version 0.10.0, https://CRAN.R-project.org/package=gt.

- Jeppson H, Hofmann H, Cook D (2021). *ggmosaic: Mosaic Plots in the 'ggplot2' Framework*. R package version 0.3.3, https://CRAN.R-project.org/package=ggmosaic.
- Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023). "Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption." *CRAN*. https://easystats.github.io/report/.
- Müller K, Wickham H (2023). *tibble: Simple Data Frames*. R package version 3.2.1, https://CRAN.R-project.org/package=tibble.
- Neuwirth E (2022). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-3, https://CRAN.R-project.org/package=RColorBrewer.
- Peng RD (2019). *simpleboot: Simple Bootstrap Routines*. R package version 1.1-7, https://CRAN.R-project.org/package=simpleboot.
- Pinheiro J, Bates D, R Core Team (2023). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-164, https://CRAN.R-project.org/package=nlme. Pinheiro JC, Bates DM (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York. doi:10.1007/b98882 https://doi.org/10.1007/b98882.
- 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, Grolemund 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.
- Wood SN (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models." *Journal of the Royal Statistical*

*Society (B)*, *73*(1), 3-36. Wood S, N., Pya, S"afken B (2016). "Smoothing parameter and model selection for general smooth models (with discussion)." *Journal of the American Statistical Association*, *111*, 1548-1575. Wood SN (2004). "Stable and efficient multiple smoothing parameter estimation for generalized additive models." *Journal of the American Statistical Association*, *99*(467), 673-686. Wood S (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC. Wood SN (2003). "Thin-plate regression splines." *Journal of the Royal Statistical Society (B)*, *65*(1), 95-114.