Preoperative Atelectasis

Part 3: Assessment of Independent Variables

Javier Mancilla Galindo

2024-04-09

Table of contents

Setup

Packages used

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


pacman::p_load(
  tidyverse, # Used for basic data handling and visualization.
  table1, #Used to add lables to variables.
  mgcv, #Used to model non-linear relationships with a general additive model.
  ggmosaic, #Used to create mosaic plots.
  car, #Used to visualize distribution of continuous variables (stacked Q-Q plots).
  dagitty, #Used in conjunction with https://www.dagitty.net/ to create
          #directed acyclic graph to inform statistical modelling.
  report #Used to cite packages used in this session.
)
```

Session and package dependencies

```
R version 4.3.3 (2024-02-29 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 22631)

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:
```

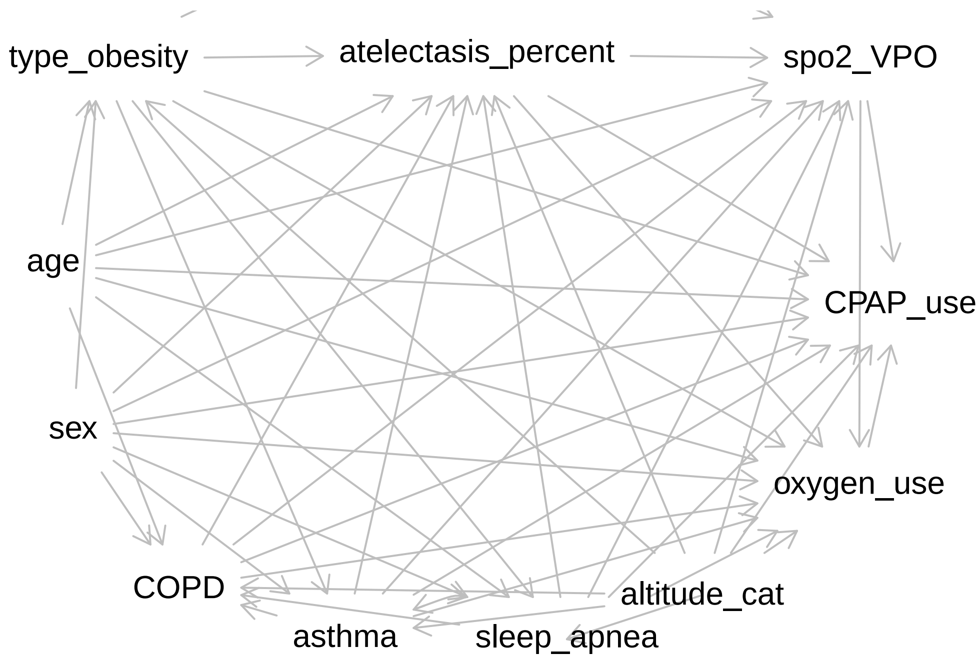
```
 [1] report_0.5.8    dagitty_0.3-4   car_3.1-2       carData_3.0-5  
 [5] ggmosaic_0.3.3  mgcv_1.9-1      nlme_3.1-164    table1_1.4.3   
 [9] lubridate_1.9.3 forcats_1.0.0   stringr_1.5.1   dplyr_1.1.4    
[13] purrr_1.0.2     readr_2.1.5     tidyr_1.3.1     tibble_3.2.1   
[17] ggplot2_3.5.0   tidyverse_2.0.0 pacman_0.5.1   
```

Assessment of independent variables

The selection of variables that will be assessed is according to the following directed acyclic graph which will be used again before statistical modelling, to assess conditional independencies.

DAG

DAG generated in the DAGitty website and sourced from the accompanying script ***DAG_atelectasis.R***



The rationale for variables in this DAG are as follows:

Exposure

The increasing degree of obesity, according to the WHO obesity class categories or BMI, is the exposure of interest.

Primary outcome

Having atelectasis (Yes or No) and an increasing degree of atelectasis (atelectasis_percent) are the main outcomes of interest. An arrow from type_obesity to atelectasis represents the exposure-outcome relationship of interest.

Secondary outcome

Decreasing preoperative SpO2 is hypothesized to be related to an increasing degree of obesity. An arrow from type_obesity to spo2_VPO represents this. Atelectasis_percentage is thought to be the main mediator of the effect of BMI on preoperative SpO2. An arrow from type_obesity to atelectasis_percent, followed by an arrow from atelectasis_percent to spo2_VPO.

Covariates

Sex and Age

These two variables are known to be associated with a higher risk of developing postoperative atelectasis in patients with obesity undergoing bariatric surgery. Baltieri, et al.. Arrows originating from these variables and going to type_obesity, atelectasis_percent, and spo2_VPO represent these relationships.
The implications for analyses is that ***sex*** and ***age*** are both **confonders** to be accounted for in both the models with atelectasis and SpO2 as outcomes.

Obstructive sleep apnea

Increasing BMI is a strong risk factor for OSA and OSA severity. Baltieri, et al. Therefore, an arrow originating in type_obesity, pointing towards OSA represents this relationship. OSA is hypothesized to lead to the degree of atelectasis and preoperative SpO2. Therefore, an arrow from OSA to atelectasis_percent and spo2_VPO represents these relationships. The implications for the analysis are the following:

1. OSA is a potential mediator of the effect of BMI on atelectasis percentage. Therefore, this variable should **not** be adjusted for in the models with ***atelectasis*** as the outcome.

2. OSA is a **confounder** of the mediator-outcome relationship in the models with ***SpO2*** as the outcome.

Asthma

It has been shown that obesity leads to asthma, whereas the inverse relationship is very unlikely to be possible. Thus, an arrow from type obesity to asthma was drawn. Yang-Ching, et al..

It has been reported that obesity-assiciated late onset non-allergic asthma is negatively related to atelectasis due to a tendency to develop more air trapping than atelectasis in these patients, compared to patients with obesity and no diagnosis of asthma in whom the airways collapse slowly and air is expelled, leading to atelectasis. Bhatawadekar , et al.. During sleep, asthma affects SpO2 independently of BMI and OSA. Sundbom, et al. For these reasons, an arrow from asthma to atelectasis, and an arrow from asthma to SpO2 was drawn.

The implications for the analysis are the following:

1. Asthma is a potential mediator of the effect of BMI on atelectasis percentage. Therefore, this variable should **not** be adjusted for in the models with ***atelectasis*** as the outcome.

2. Asthma is a potential **confounder** of the mediator-outcome relationship in the models with ***SpO2*** as the outcome.

COPD

Although there is a strong relationship between undernutrition and COPD, the releationship between obesity and COPD has been inconsistent among studies. Since this study only included patients with obesity, the potential relationship between underweight and COPD is likely not relevant for this particular study. Furthermore, there is still doubt regarding any potential role of obesity-related pathophysiological mechanisms which could potentially lead to COPD. Hanson, et al. For these reasons, an arrow between COPD and obesity (or the inverse) was not drawn. This assumption was checked through the conditional independencies check (see Part 4), and this assumption is consistent with the data.

Regarding a relationship between COPD and SpO2, there is a clear relationship between these variables, reason why an arrow going from COPD to SpO2 was drawn. As for atelectasis, studies have found atelectasis, especially in patients with wood smoke-realted COPD. González-García, et al and Carmo Moreira, et al Thus, an arrow from COPD to atelectasis was drawn.

Altitude

Although not directly linked to obesity, participants with OSA at an an altitude above 1600 meters can develop hypobaric hypoxia, which "promotes frequent central apneas in addition to obstructive events, resulting in combined intermittent and sustained hypoxia". Bloch, et al

For the atelectasis outcome, I could not find evidence either supporting or rejecting an association between altitude and prevalence of atelectasis. However, during the conditional independencies assumptions testing procedure, the data suggested a correlation, reason why an arrow from altitude to atelectasis was drawn as the reverse is less likely to be true (i.e., obesity would hardly determine the altitude of the place of residence).

The implications for analyses is that ***altitude_cat*** is a potential **confounder** to be accounted for in both the models with atelectasis and SpO2 as outcomes.

Oxygen use at home and CPAP use at home

These variables are descendants of the exposure, mediator, outcomes, and covariates of interest. The implications for analyses is that these 2 variables should **not** be adjusted for in any of the models.

Hemoglobin

There is no strong evidence supporting a link between BMI and hemoglobin. In any case, hemoglobin would be a descendant of all main variables of interest (exposure, mediator, and outcomes). Thus, hemoglobin was excluded from this DAG for simplification.

Other variables

Other variables that are potential confounders are not shown in this DAG since they were addressed by design in this study as follows:

- Current COVID-19: Exclusion criteria were applied to **n=2** patients with CO-RADS 3 and **n=2** with CO-RADS 4. Only participants with low probability of COVID-19 (CO-RADS 1 and 2) were included in this study.
- Prior COVID-19: This was an exclusion criterion (**n=3**).
- Bronchiectasis in chest CT: This was an exclusion criterion (n=0).
- Neuromuscular diseases: This was an exclusion criterion (n=0).
- Prior of current tuberculosis: This was an exclusion criterion (n=0).

Unmeasured variables

Due to the possibility of unmeasured confounders, E-values will be calculated and presented when possible as sensitivity analyses.

Description of independent variables

Age

Summary:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  20.00   32.75   40.00   40.26   48.25   65.00 
```

The mean age was 40.3 (SD: 9.87).

Sex

Frequencies:

```
sex
Woman   Man 
  214    22 
```

Percentage:

```
sex
Woman   Man 
 90.7   9.3 
```

Most patients in the sample were woman (n=214, 90.7%).

Body mass index (BMI)

Summary:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  30.00   34.63   40.30   41.37   46.02   77.31 
```

Frequencies:

```
type_obesity
Class 1 Obesity Class 2 Obesity Class 3 Obesity 
             63              53             120 
```

Percentage:

```
type_obesity
Class 1 Obesity Class 2 Obesity Class 3 Obesity
           26.7            22.5            50.8
```

Distribution of BMI was assessed earlier. It is right-skewed due to extreme values (verified outliers). The WHO classification of BMI for obesity class will be used to complement descriptions and for potential use later during statistical modelling.

> The median BMI was 40.295 (IQR: 34.63- 46.02). The distribution of BMI was right-skewed due to extreme BMI values (range: 30- 77.31). Most patients were in the class 3 obesity category (n=120, 50.8%), followed by class 1 (n=63, 26.7%) and 2 (n=53, 22.5%). a

Obstructive sleep apnea

Frequencies:

```
sleep_apnea
 No Yes
203  33
```

Percentage:

```
sleep_apnea
 No Yes
 86  14
```

> Patients with a diagnosis of OSA were 14% (n=33) of the sample.

Asthma

Frequencies:

```
asthma
 No Yes
216  20
```

Percentage:

```
asthma
  No  Yes
91.5  8.5
```

> Patients with a diagnosis of asthma were 8.5% (n=20) of the sample.

COPD

Frequencies:

```
COPD
 No Yes
228   8
```

Percentage:

```
COPD
  No  Yes
96.6  3.4
```

Patients with COPD were 3.4% (n=8) of the sample.

Altitude

Summary:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   31.0   519.0   519.0   652.7   806.0  1861.0
```

Distribution of altitude was assessed earlier. Distribution is very unclear due to very widespread datapoints. Thus, I will create a new variable categorizing values according to the study by Crocker ME, et al.

Frequencies:

```
altitude_cat
     Low altitude Moderate altitude
              205                31
```

Percentage:

```
altitude_cat
     Low altitude Moderate altitude
             86.9              13.1
```

SpO2

Summary:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
     88      93      96      95      97      99 
```

Distribution of SpO2 during the pre-anesthetic is left-skewed due to some participants exhibiting decreased SpO2. I will categorize according to clinical categories to assess the proportion of patients with decreased SpO2:

Proportion of patients with decreased SpO2

Frequencies:

```
spo2_cat
     90 90 to 94      >94 
      15       75      146 
```

Percentage:

```
spo2_cat
     90 90 to 94      >94 
     6.4     31.8     61.9 
```

> The median SpO2 during the pre-anethetic assessment was 96 (IQR: 93-97) %, with a minimum value of 88%. Of these, n=146 (61.9%) had normal SpO2 (above 94%), whereas n=75 (31.8%) had a value in the 90-94% range, and n=15 (6.4%) had 90%.

Oxygen use

Frequencies:

```
oxygen_use
 No Yes 
206  30 
```

Percentage:

```
oxygen_use
  No  Yes 
87.3 12.7 
```

> A total 30 (12.7%) patients used oxygen at home.

CPAP use

Frequencies:

```
CPAP_use
 No Yes
203  33
```

Percentage:

```
CPAP_use
 No Yes
 86  14
```

whereas 14% (n=33) reported using CPAP.

Hemoglobin

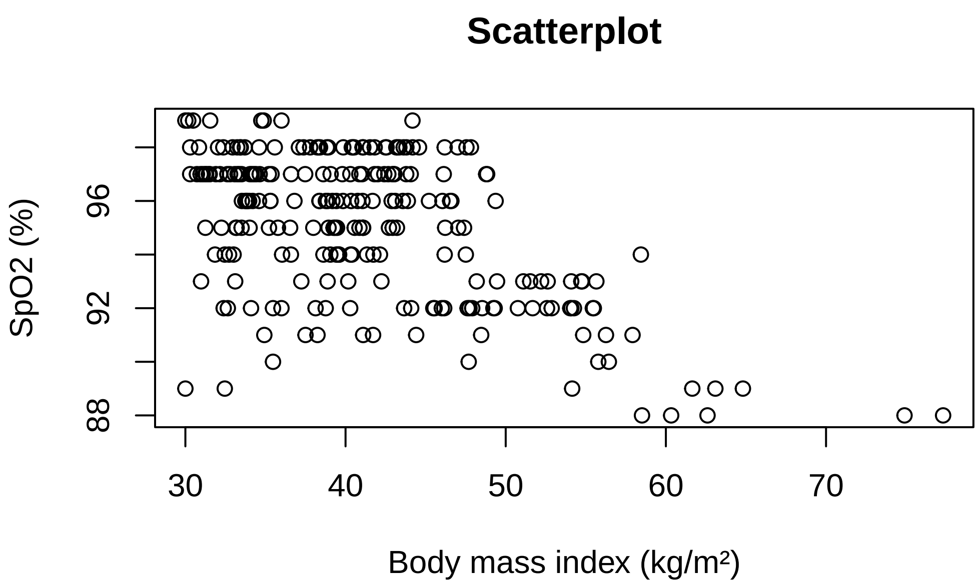
Summary:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   9.90   13.90   14.50   14.54   15.20   18.50       2
```

Distribution of hemoglobin was assessed and follows a normal distribution. Two participants don't have a hemoglobin value.

Relationships between independent variables

BMI and SpO2



Relationship does not seem to be linear (also, variables were not normally distributed, with outliers), but suggests a negative correlation. Will assess if a smooth BMI term explains SpO2 better, and if so, what is the best number of knots to model this relationship:

Models evaluated with the accompanying sourced script ***nonlinear_BMI_SpO2.R***

All non-linear models are significantly better than linear. Thus, using a smooth term for BMI is better than modelling a linear relationship.

Best AIC:

```
list(AIC_k2,AIC_k4,AIC_k6,AIC_k8,AIC_k12)
```

```
[[1]]
[1] 1048.14

[[2]]
[1] 1040.448

[[3]]
```

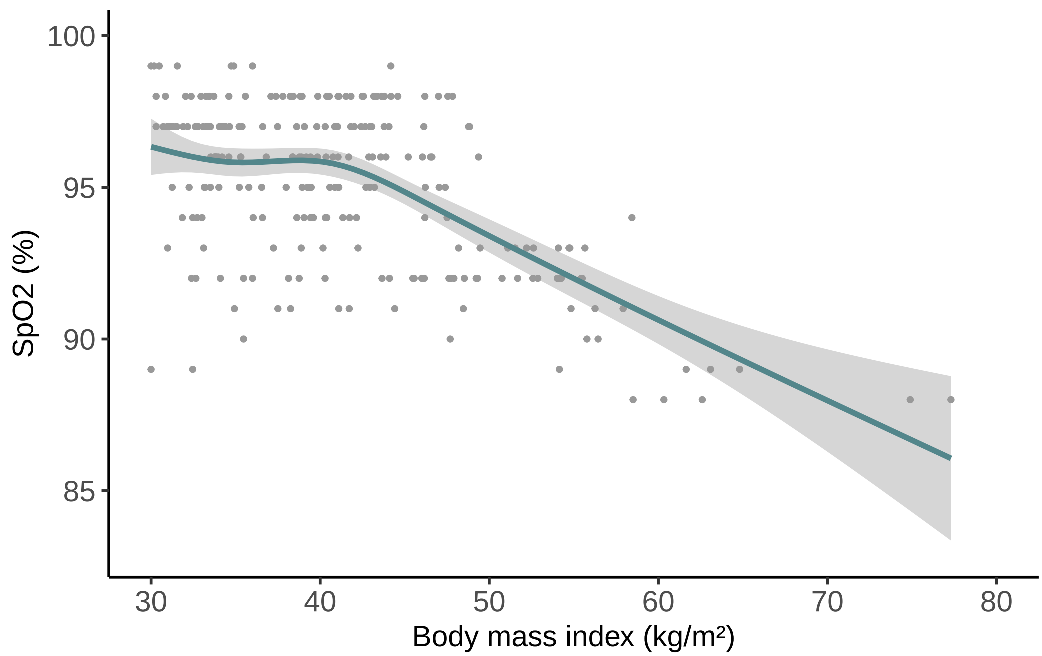
```
[1] 1036.959

[[4]]
[1] 1036.83

[[5]]
[1] 1037.165
```

Regarding AIC, the models with k>6 are not better at explaining the variance. Thus, I will with k=5 since the best model is expected to be anywhere between k=4 and k=6:

```r
list(AIC_k4,AIC_k5,AIC_k6)
```

```
[[1]]
[1] 1040.448

[[2]]
[1] 1037.475

[[3]]
[1] 1036.959
```

Model with k=5 still offers and advantage compared to k=4 (drop in AIC). No other improvements in k-index or visual representation are achieved with higher k. Thus, will use k=5 to model.



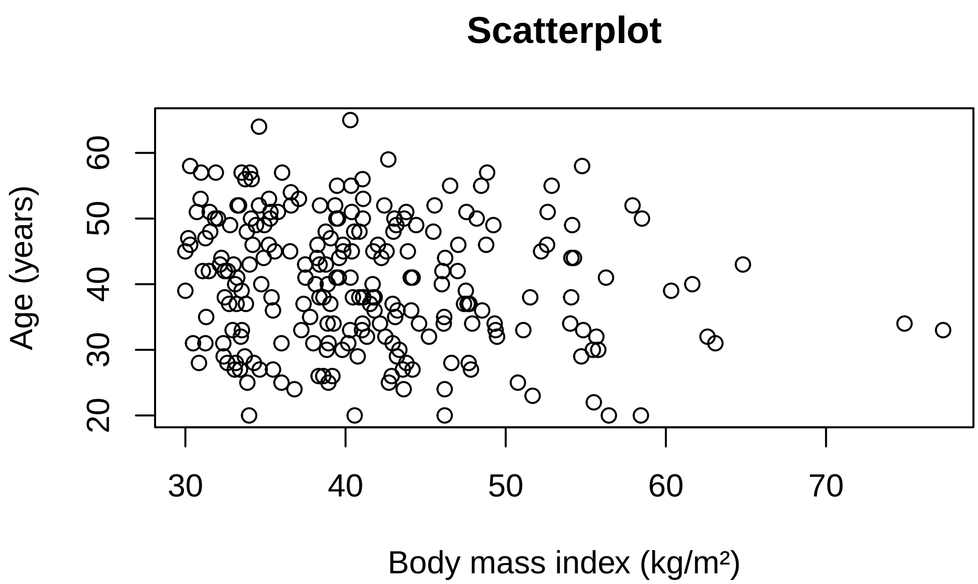
Negative non-monotonic relationship since SpO2 decreases, but then seems to increase slightly again at BMI 40, followed by a marked decrease as BMI decreases at values higher than ~42.

Spearman's correlation coefficient shouldn't be used due to relationship not being monotonically decreasing. However, I will calculate it just to have a rough idea (but will not report this in the paper).

```
    Spearman's rank correlation rho

data:  spo2_VPO and BMI
S = 3105183, p-value = 2.28e-11
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.417458
```

> BMI exhibited a negative non-linear monotonic relationship with SpO2 (**Figure 1B**, rho= -0.417, p<0.001).

BMI and age



Datapoints scattered. Relationship monotonic and probably linear, but there are influential true outliers with extreme BMI. Will assess with Spearman correlation analysis due to extreme BMI values.

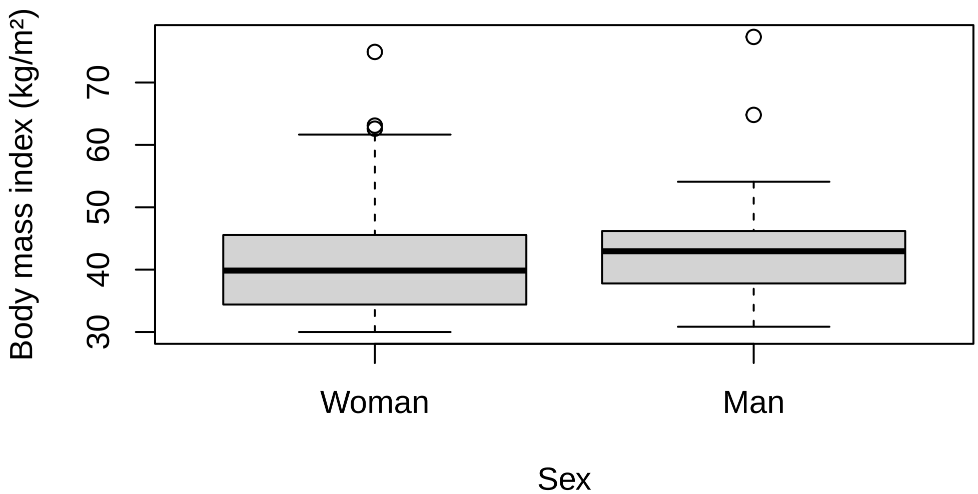
```
    Spearman's rank correlation rho

data:  age and BMI
S = 2530759, p-value = 0.017
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.1552445
```

Age had a weak negative correlation with BMI (rho= -0.155, p=0.017).

BMI and sex

Median BMI:

```
# A tibble: 2 x 7
  sex         n median    Q1    Q3   min   max
  <fct> <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 Woman   214   39.8  34.5  45.5  30    74.9
2 Man      22   43.0  37.9  46.2  30.8  77.3
```

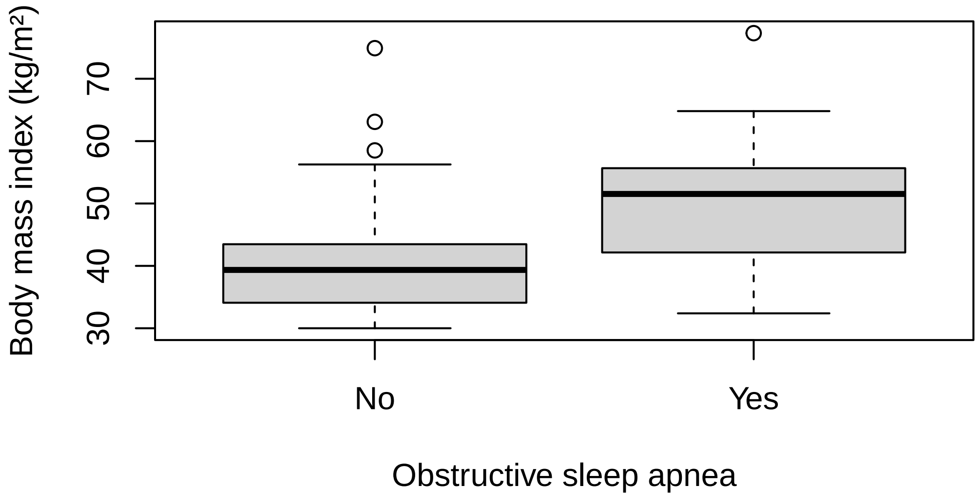


Distribution not normal and influential outliers. Will assess non-parametrically.

```
    Wilcoxon rank sum test with continuity correction

data:  BMI by sex
W = 1918.5, p-value = 0.1537
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was not different between men (43, IQR: 37.9-46.2) and women (39.9, IQR: 34.5-45.5) (p=0.154).

BMI and sleep apnea



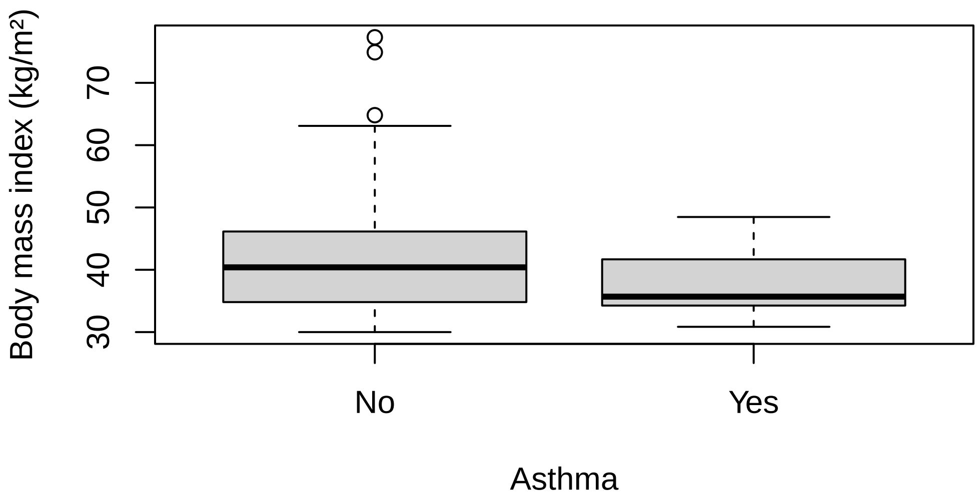
Distribution not normal and influential outliers. Will assess non-parametrically.

```
# A tibble: 2 x 7
  sleep_apnea     n median    Q1    Q3   min   max
  <fct>       <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 No            203   39.4  34.1  43.5  30    74.9
2 Yes            33   51.5  42.2  55.7  32.4  77.3
```

```
	Wilcoxon rank sum test with continuity correction

data:  BMI by sleep_apnea
W = 1301.5, p-value = 1.812e-08
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was significantly higher in participants with sleep apnea (51.5, IQR: 42.1-55.7) compared to those without OSA (39.4, IQR: 34.1-43.5) (p=0).

BMI and asthma

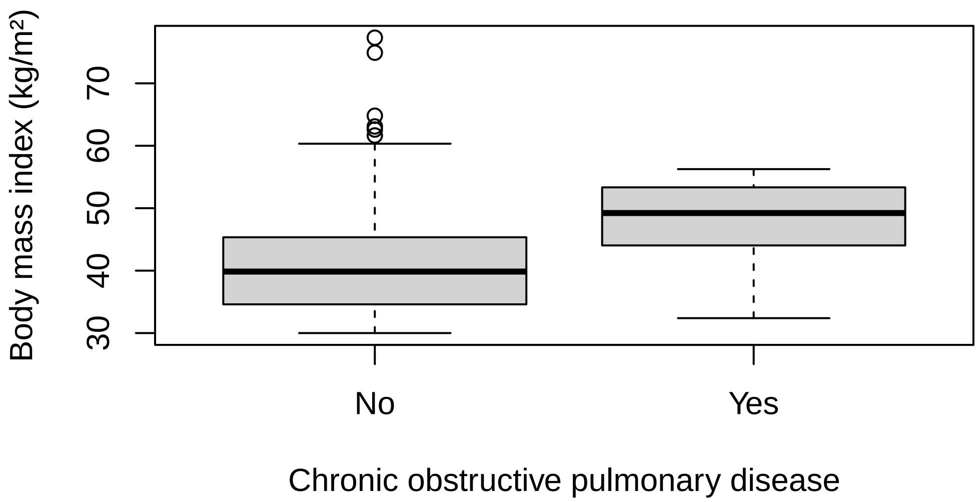


Distribution not normal and influential outliers. Will assess non-parametrically.

```
# A tibble: 2 x 7
  asthma      n median    Q1    Q3   min   max
  <fct>   <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 No        216   40.4  34.9  46.1  30    77.3
2 Yes        20   35.7  34.3  41.3  30.8  48.5

    Wilcoxon rank sum test with continuity correction

data:  BMI by asthma
W = 2695.5, p-value = 0.06701
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was not significantly different in patients with asthma (35.7, IQR: 34.3-41.3) compared to those without (40.4, IQR: 34.9-46.1) (p=0.067).

BMI and COPD



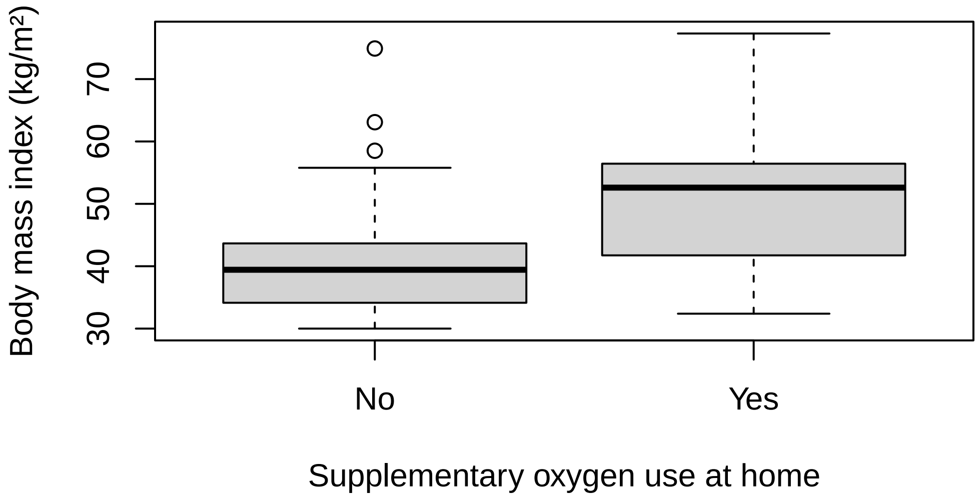
Distribution not normal and influential outliers. Will assess non-parametrically.

```
# A tibble: 2 x 7
  COPD      n median    Q1    Q3   min   max
  <fct> <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 No      228   39.8  34.6  45.3  30    77.3
2 Yes       8   49.2  44.2  53.0  32.4  56.3
```

```
    Wilcoxon rank sum test with continuity correction

data:  BMI by COPD
W = 453, p-value = 0.0157
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was significantly higher in participants with COPD (49.2, IQR: 44.2-53) than those without COPD (39.9, IQR: 34.6-45.3) (p=0.016).

BMI and oxygen use



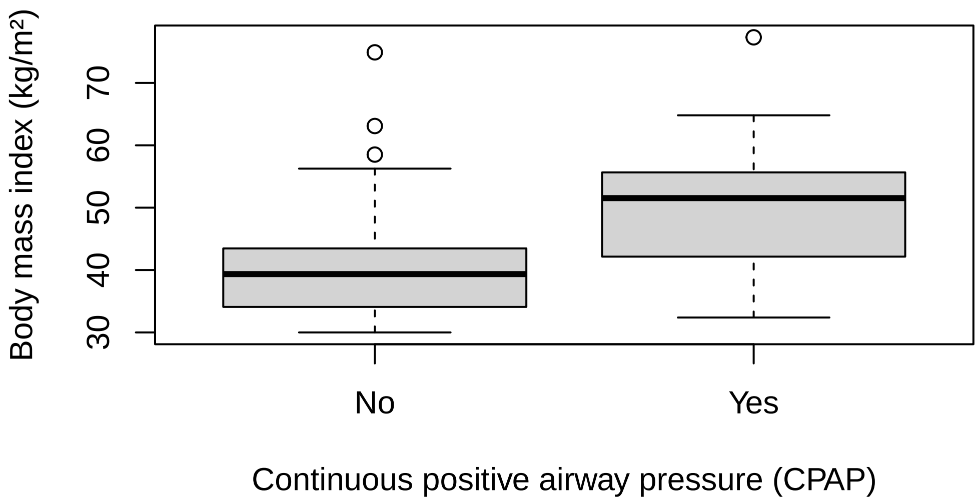
Distribution not normal and influential outliers. Will assess non-parametrically.

```
# A tibble: 2 x 7
  oxygen_use     n median    Q1    Q3   min   max
  <fct>      <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 No           206   39.4  34.2  43.7  30    74.9
2 Yes           30   52.6  41.8  56.4  32.4  77.3
```

```
    Wilcoxon rank sum test with continuity correction

data:  BMI by oxygen_use
W = 1213.5, p-value = 7.883e-08
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was significantly higher in patients who reported oxygen use at home (52.6, IQR: 41.8-56.4) compared to those with no supplementary oxygen use (39.4, IQR: 34.2-43.7) ($p<0.001$).

BMI and CPAP use



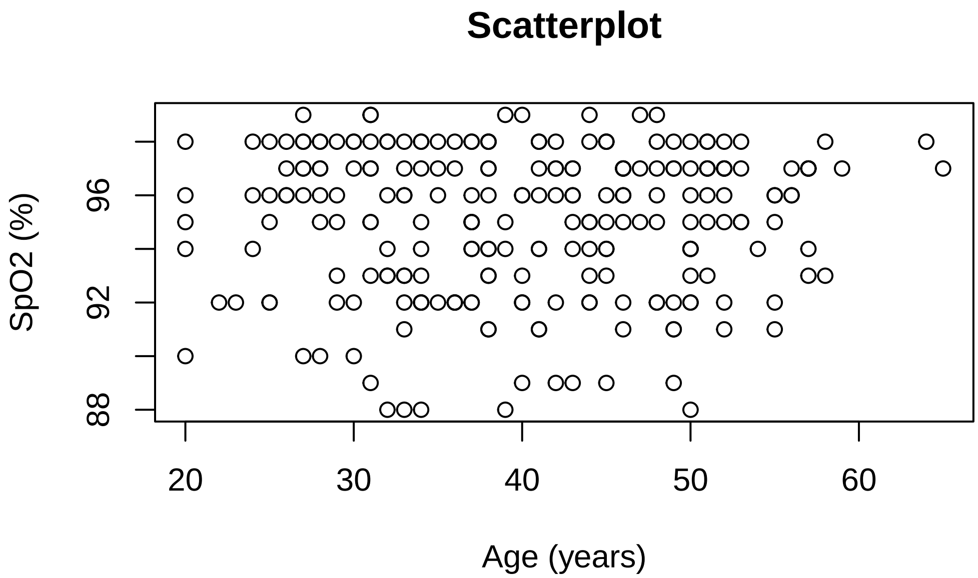
Distribution not normal and influential outliers. Will assess non-parametrically.

```
# A tibble: 2 x 7
  CPAP_use      n median    Q1    Q3   min   max
  <fct>     <int>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 No          203   39.4  34.1  43.5  30    74.9
2 Yes          33   51.5  42.2  55.7  32.4  77.3
```

```
	Wilcoxon rank sum test with continuity correction

data:  BMI by CPAP_use
W = 1301.5, p-value = 1.812e-08
alternative hypothesis: true location shift is not equal to 0
```

> The median BMI was significantly higher in participants with CPAP use at home (51.5, IQR: 42.1-55.7) compared to those who did not report CPAP use (39.4, IQR: 34.1-43.5) ($p<0.001$).

Age and SpO2

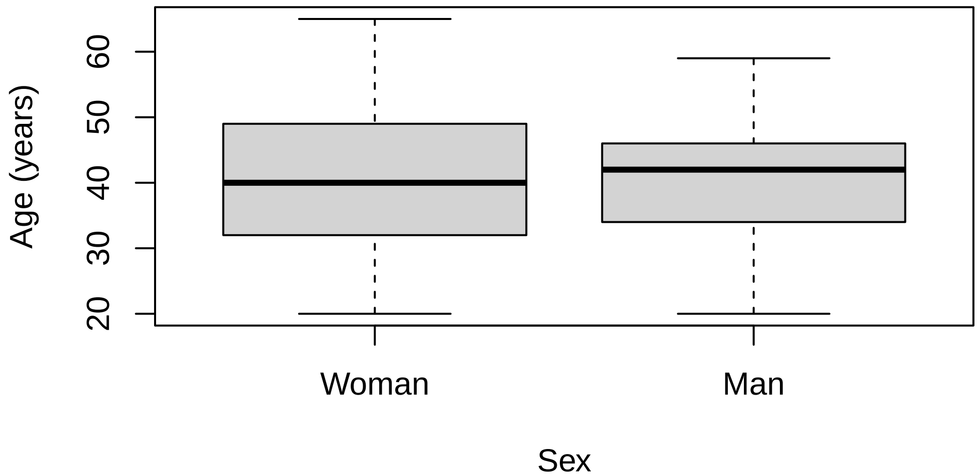


Do not seem to be correlated. Will apply Spearman's correlation test:

```
    Spearman's rank correlation rho

data:  spo2_VPO and age
S = 2143192, p-value = 0.7405
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
0.02167287
```

Age and SpO2 were not correlated (rho= 0.022, p=0.74).

Age and sex



Distribution near-normal, but light tails for women. However, t-test could be robust to deviations from normality and differences in group size. Will assess mean and variance for further testing:

```
# A tibble: 2 x 5
  sex        n age_mean    sd variance
  <fct> <int>    <dbl> <dbl>    <dbl>
1 Woman   214     40.2  9.94     98.9
2 Man      22     40.6  9.28     86.1
```

Variances are similar. However, group sizes differ my 10x. Welch's t-test more suitable:

```
    Welch Two Sample t-test

data:  age by sex
t = -0.19917, df = 26.213, p-value = 0.8437
alternative hypothesis: true difference in means between group Woman and group Man is not equ
95 percent confidence interval:
 -4.715913  3.882438
sample estimates:
```

```
mean in group Woman   mean in group Man
           40.21963            40.63636
```

> Mean age was similar bethween men (40.6, sd:9.3) and women (40.2, sd:9.9) (p=0.844).

Age and sleep apnea

Distribution near-normal. Will assess mean and variance for further testing.

```
# A tibble: 2 x 5
  sleep_apnea     n age_mean    sd variance
  <fct>       <int>    <dbl> <dbl>    <dbl>
1 No            203     40.1  9.88     97.7
2 Yes            33     41.2  9.87     97.5
```

Size per group very different, variances do not look similar. Welch's t-test more suitable:

```
    Welch Two Sample t-test

data:  age by sleep_apnea
t = -0.57922, df = 43.092, p-value = 0.5655
alternative hypothesis: true difference in means between group No and group Yes is not equal
95 percent confidence interval:
 -4.810659  2.663772
sample estimates:
 mean in group No mean in group Yes
         40.10837          41.18182
```

> Age was not significantly different between participants with OSA (41.2, sd:9.9) and those without (40.1, sd:9.9) (p=0.565).

Age and asthma

Distribution normal. Will assess mean and variance for further testing.

```
# A tibble: 2 x 5
  asthma     n age_mean    sd variance
  <fct>  <int>    <dbl> <dbl>    <dbl>
1 No       216     40.3  9.82     96.5
2 Yes       20     40.1 10.6     112.
```

Size per group very different, variances look similar. Welch's t-test more suitable due to differring group size:

```
    Welch Two Sample t-test

data:  age by asthma
t = 0.070295, df = 22.127, p-value = 0.9446
alternative hypothesis: true difference in means between group No and group Yes is not equal
95 percent confidence interval:
 -4.933425  5.279722
sample estimates:
 mean in group No mean in group Yes
         40.27315          40.10000
```

> Age was not significantly different between participants with asthma (40.1, sd:10.6) and those without (40.3, sd:9.8) (p=0.945).

Age and COPD

Group size low to conclude distribution for COPD positive patients. Will assess mean and variance for further testing.

```
# A tibble: 2 x 5
  COPD      n age_mean    sd variance
  <fct> <int>    <dbl> <dbl>    <dbl>
1 No      228     40.3  9.88     97.6
2 Yes       8     39   10.1     102.
```

Size per group very different. Welch's t-test more suitable:

```
    Welch Two Sample t-test

data:  age by COPD
t = 0.35835, df = 7.4762, p-value = 0.73
alternative hypothesis: true difference in means between group No and group Yes is not equal
95 percent confidence interval:
 -7.183172  9.788435
sample estimates:
 mean in group No mean in group Yes
         40.30263          39.00000
```

> Age was not significantly different between participants with COPD (39, sd:10.1) and those without (40.3, sd:9.9) (p=0.73).

Age and oxygen use

Distribution near-normal. Will assess mean and variance for further testing.

```
# A tibble: 2 x 5
  oxygen_use     n age_mean    sd variance
  <fct>      <int>    <dbl> <dbl>    <dbl>
1 No           206     40.2  9.85     97.0
2 Yes           30     41.0 10.1     103.
```

Size per group very different. Welch's t-test more suitable:

```
    Welch Two Sample t-test

data:  age by oxygen_use
t = -0.41098, df = 37.416, p-value = 0.6834
alternative hypothesis: true difference in means between group No and group Yes is not equal
95 percent confidence interval:
 -4.809765  3.187111
sample estimates:
 mean in group No mean in group Yes
         40.15534          40.96667
```

> Age was not significantly different between participants with self-reported oxygen use (41, sd:10.1) and those without (40.2, sd:9.8) (p=0.683).

Age and CPAP use

Distribution near-normal. Will assess mean and variance for further testing.

```
# A tibble: 2 x 5
  CPAP_use     n age_mean    sd variance
  <fct>    <int>    <dbl> <dbl>    <dbl>
1 No         203     40.1  9.88     97.7
2 Yes         33     41.2  9.87     97.5
```

Size per group very different, but equal variances. Conventional t-test expected to be robust:

```
	Two Sample t-test

data:  age by CPAP_use
t = -0.57882, df = 234, p-value = 0.5633
alternative hypothesis: true difference in means between group No and group Yes is not equal
95 percent confidence interval:
 -4.727155  2.580268
sample estimates:
 mean in group No mean in group Yes
         40.10837          41.18182
```

Age was not significantly different between participants with CPAP use (41.2, sd:9.9) and those without (40.1, sd:9.9) (p=0.563).

SpO2 and sex

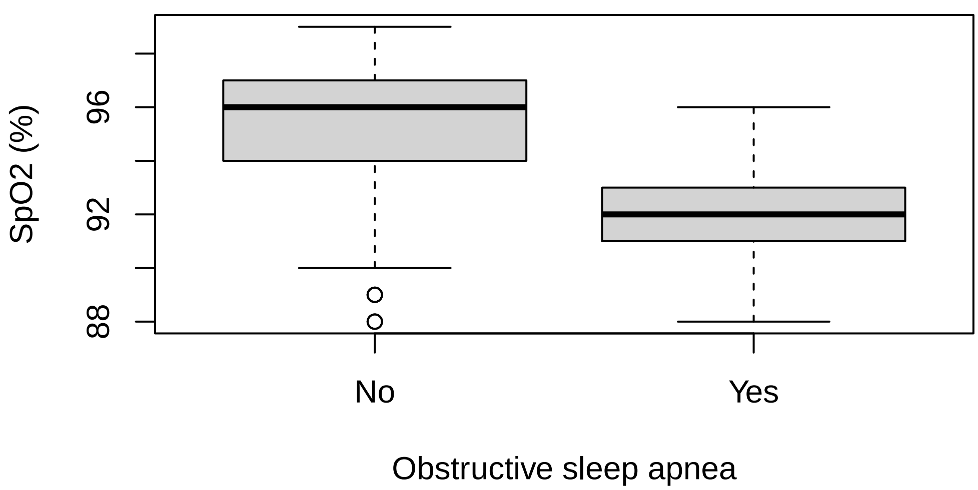
Distribution deviates from normal and small group size for men. Will assess non-parametrically.

```
# A tibble: 2 x 7
  sex       n spo2_median    Q1    Q3   min   max
  <fct> <int>       <dbl> <dbl> <dbl> <int> <int>
1 Woman   214          96    93  97      88    99
```

```
2 Man      22        94.5    92  97.8    88    98

    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by sex
W = 2602, p-value = 0.413
alternative hypothesis: true location shift is not equal to 0
```

The median SpO2 was not different between men (94.5, IQR: 92-97.8) and women (96, IQR: 93-97) (p=0.413).

SpO2 and sleep apnea



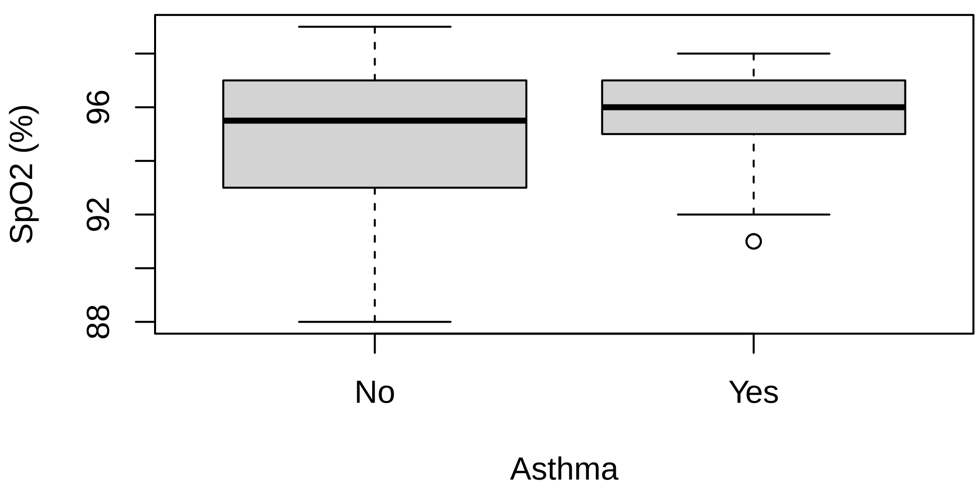
Distribution not normal, and smaller group size for those with sleep apnea. Will assess non-parametrically.

```
# A tibble: 2 x 7
  sleep_apnea     n spo2_median    Q1    Q3   min   max
  <fct>       <int>       <int> <dbl> <dbl> <int> <int>
1 No            203          96    94    97    88    99
2 Yes            33          92    91    93    88    96
```

```
    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by sleep_apnea
W = 5801, p-value = 1.069e-11
alternative hypothesis: true location shift is not equal to 0
```

> Patients with sleep apnea had a lower median SpO2 (92, IQR: 91-93) than those without OSA (96, IQR: 94-97) (p<0.001).

SpO2 and asthma



Distribution not normal, and smaller group size for those with the comorbidity. Will assess non-parametrically.

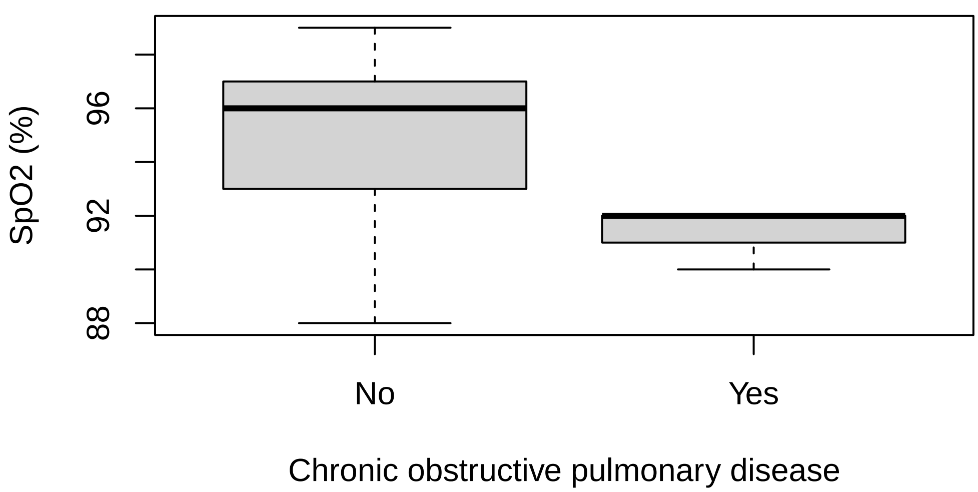
```
# A tibble: 2 x 7
  asthma     n spo2_median    Q1    Q3   min   max
  <fct>  <int>       <dbl> <dbl> <dbl> <int> <int>
1 No       216        95.5    93    97    88    99
2 Yes       20        96      95    97    91    98
```

```
    Wilcoxon rank sum test with continuity correction
```

```
data:  spo2_VPO by asthma
W = 1959, p-value = 0.4887
alternative hypothesis: true location shift is not equal to 0
```

> The median SpO2 was not significantly different among those with asthma (96, IQR: 95-97) compared to those without (95.5, IQR: 93-97) (p=0.489).

SpO2 and COPD

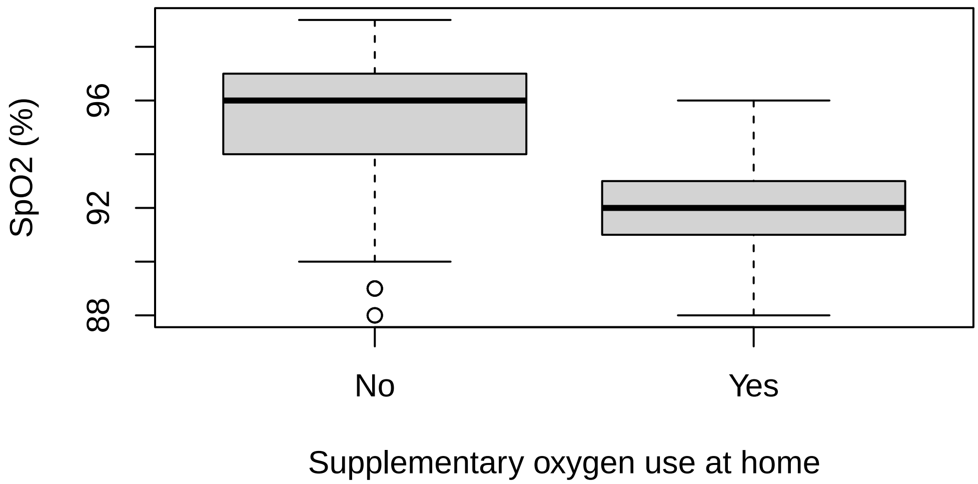


Distribution not normal, and smaller group size for those with the comorbidity. Will assess non-parametrically.

```
# A tibble: 2 x 7
  COPD      n spo2_median    Q1    Q3   min   max
  <fct> <int>       <dbl> <dbl> <dbl> <int> <int>
1 No      228          96    93    97    88    99
2 Yes       8          92    91    92    90    92

    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by COPD
W = 1605.5, p-value = 0.0002306
alternative hypothesis: true location shift is not equal to 0
```

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

SpO2 and oxygen use at home

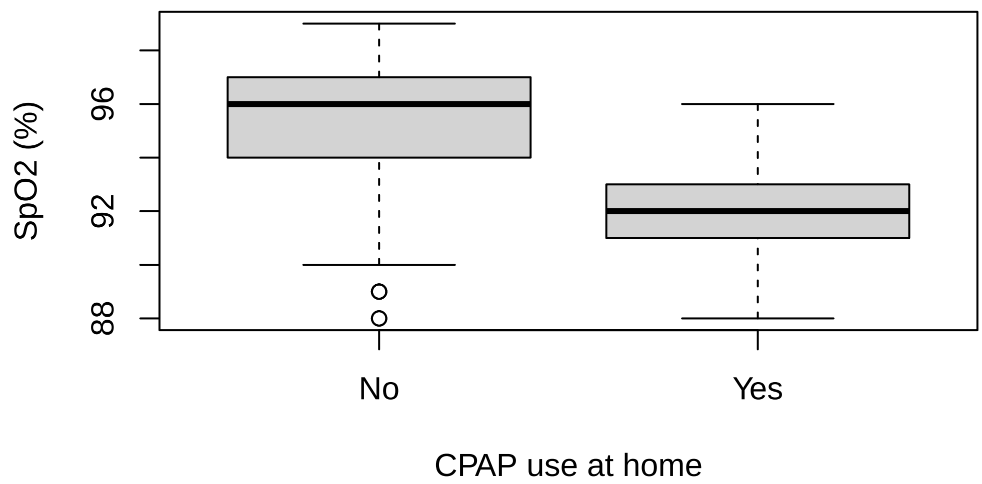


Distribution not normal, and smaller group size for those with the comorbidity. Will assess non-parametrically.

```
# A tibble: 2 x 7
  oxygen_use     n spo2_median    Q1    Q3   min   max
  <fct>      <int>       <dbl> <dbl> <dbl> <int> <int>
1 No           206          96    94    97    88    99
2 Yes           30          92    91    93    88    96

    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by oxygen_use
W = 5347, p-value = 7.271e-11
alternative hypothesis: true location shift is not equal to 0
```

> The median SpO2 was significantly lower among those with supplementary oxygen use at home (92, IQR: 91-93) compared to those without (96, IQR: 94-97) (p<0.001).

SpO2 and CPAP use



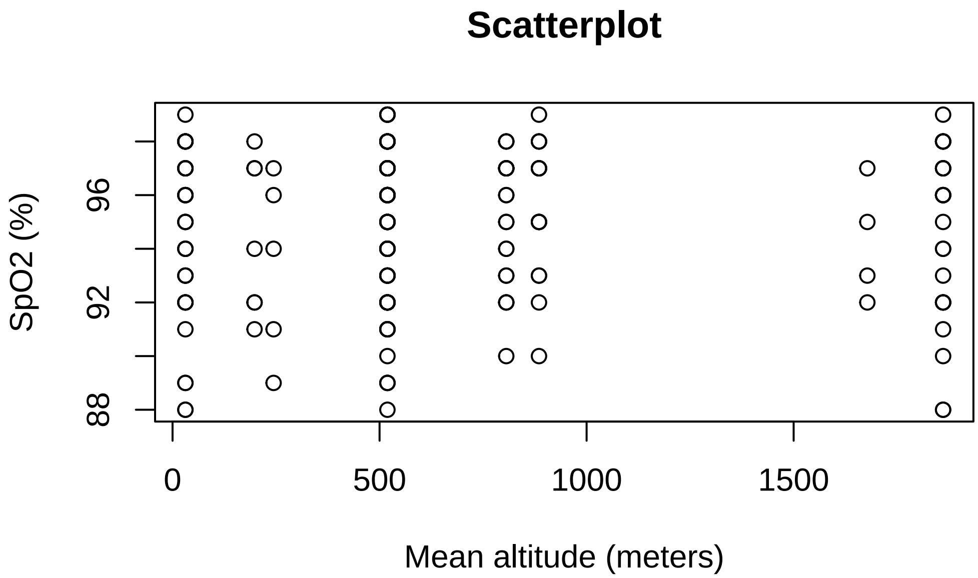
Distribution not normal, and smaller group size for those with the comorbidity. Will assess non-parametrically.

```
# A tibble: 2 x 7
  CPAP_use      n spo2_median    Q1    Q3   min   max
  <fct>     <int>       <int> <dbl> <dbl> <int> <int>
1 No          203          96    94    97    88    99
2 Yes          33          92    91    93    88    96
```

```
    Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by CPAP_use
W = 5801, p-value = 1.069e-11
alternative hypothesis: true location shift is not equal to 0
```

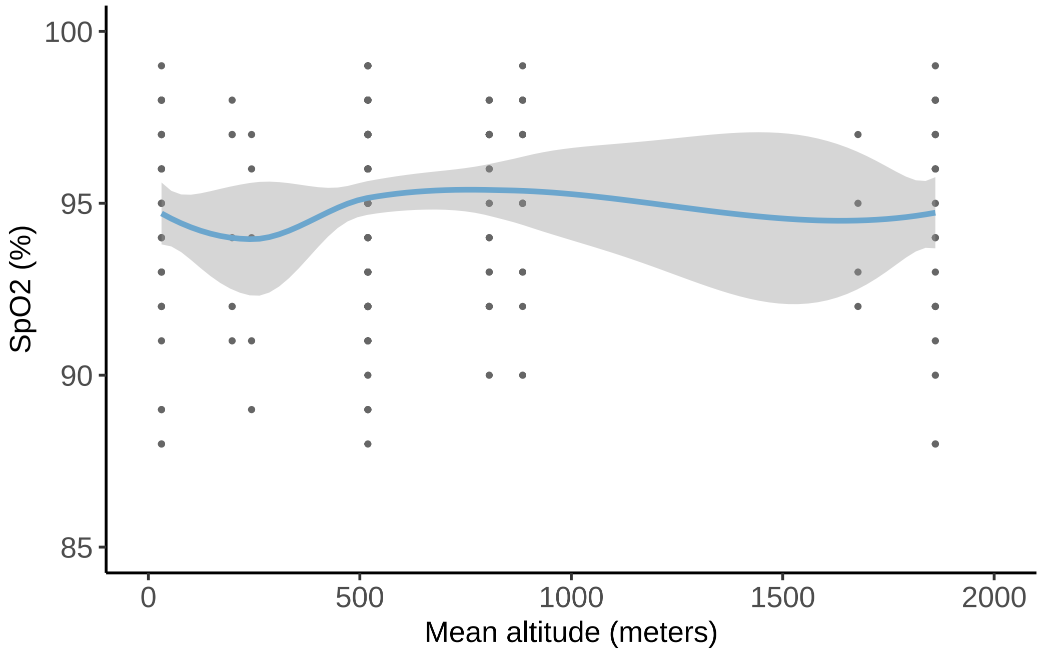
> The median SpO2 was significantly lower among those with CPAP use at home (92, IQR: 91-93) compared to those without (96, IQR: 94-97) ($p<0.001$).

SpO2 and altitude



There does not seem to be a pattern.

Would a smooth term be useful to model altitude?



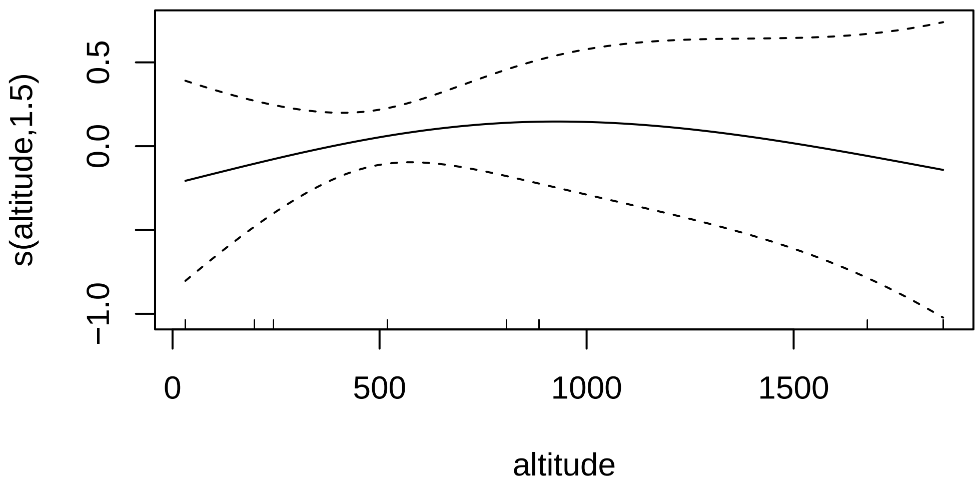
It is likely that a smooth term for SpO2 would be non-informative since there is no clear reasonable pattern in this smooth plot. Additionally, it is well known that any impacts in SpO2 due to altitudes up to 2000 are very limited (i.e 1 to 2 units). go to reference.

I will still check if a smooth term may be better than linear in case that adjustment for this variable is needed.

GAM model with k=4 (this was also checked with varying k from 2 to 10):

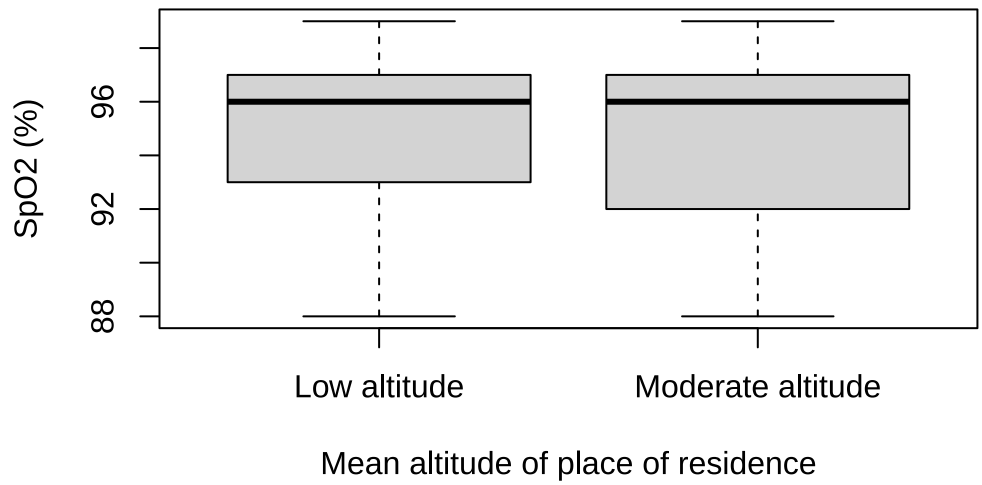
```
Family: gaussian
Link function: identity

Formula:
spo2_VPO ~ s(altitude, k = 4)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   94.996      0.178   533.7   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df     F p-value
s(altitude) 1.505  1.798 0.437   0.631
```

```
R-sq.(adj) =  0.000124   Deviance explained = 0.653%
GCV = 7.5559  Scale est. = 7.4757    n = 236
```



Smooth term is not significantly better than one assuming linearity. Furthermore, the relationship with SpO2 in smooth term does not make any sense (i.e., according to prior reference, SpO2 should decrease at higher altitudes). Thus, it would be very likely that including this term would only explain noise in any case, not the true known causal relationship between SpO2 and altitude.

Lastly, will check the pattern according to altitude categories, which may be a better term to use in models in any case.



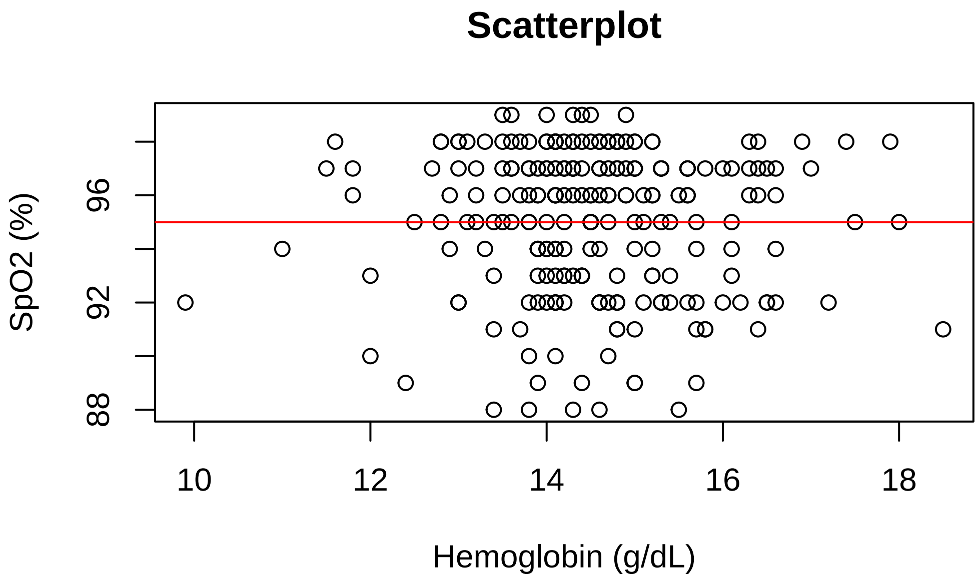
Distribution deviates from normal and small group size for the moderate altitude group. Will assess non-parametrically.

```
# A tibble: 2 x 7
  altitude_cat          n spo2_median    Q1    Q3   min   max
  <fct>             <int>       <int> <dbl> <dbl> <int> <int>
1 Low altitude        205          96    93    97    88    99
2 Moderate altitude    31          96    92    97    88    99
```

```
	Wilcoxon rank sum test with continuity correction

data:  spo2_VPO by altitude_cat
W = 3360, p-value = 0.6043
alternative hypothesis: true location shift is not equal to 0
```

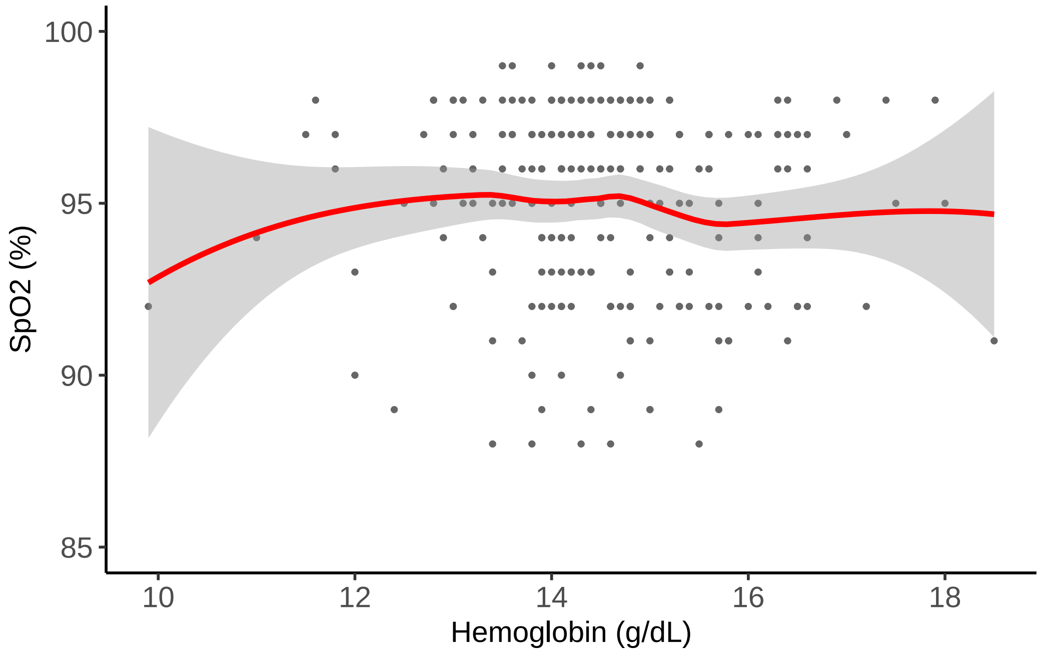
> The median SpO2 was not different between low and moderate altitude categories (p=0.604).

SpO2 and hemoglobin



There does not seem to be a clear pattern.

Would a smooth term be useful to model SpO2?

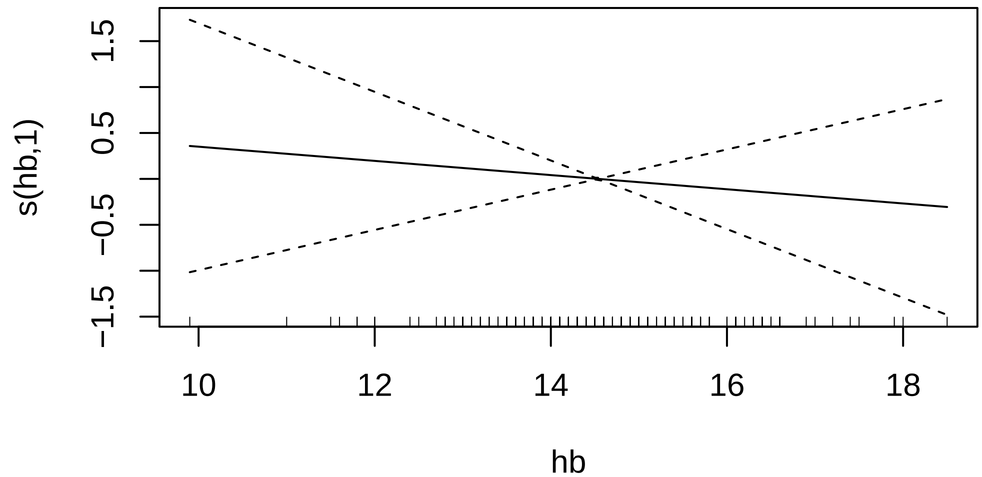


Hemoglobin likely has an effect on SpO2 at lower hemoglobin values, which makes sense with what is observed in the graph. Assuming a linear relationship could lead to incorrect conclusions according to this. Nonetheless, it looks like the apparent non-linear relationship at low Hb values is due to only 2 observations with wide confidence intervals showing that the true slope could go either up, straight or down, so it may also be incorrect to assume a non-linear relationship based only on this plot. I will model to see if there is an optimal smooth term for hemoglobin or if a linear term best fits the data:

GAM model with k=4 (this was also checked with varying k from 2 to 10):

```
Family: gaussian
Link function: identity

Formula:
spo2_VPO ~ s(hb, k = 4)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  94.9829     0.1789   530.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df     F p-value
```

```
s(hb)   1      1 0.272   0.603

R-sq.(adj) =  -0.00314   Deviance explained = 0.117%
GCV = 7.5555  Scale est. = 7.4909    n = 234
```



The estimated degrees of freedom (edf) in both cases were 1, plus p=0.6, meaning that a linear term is better fitted to this data than a non-linear term.

```
    Spearman's rank correlation rho

data:  spo2_VPO and hb
S = 2274841, p-value = 0.3201
alternative hypothesis: true rho is not equal to 0
sample estimates:
        rho
-0.06527711
```

SpO2 and hemoglobin were not correlated (rho= -0.065, p=0.32).

Sex and sleep apnea

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

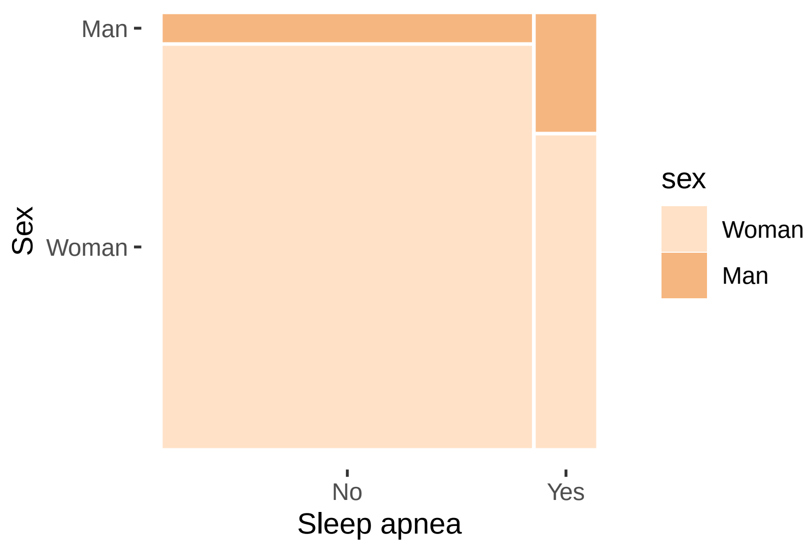
Frequencies:

```
       sleep_apnea
sex       No Yes
  Woman 190  24
  Man    13   9
```

Percentage:

```
       sleep_apnea
sex        No  Yes
  Woman 88.8 11.2
  Man   59.1 40.9
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 14.624, df = 1, p-value = 0.0001312
```

Sex was associated with OSA (p<0.001) as men had the diagnosis more frequently compared to women.

Sex and asthma

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

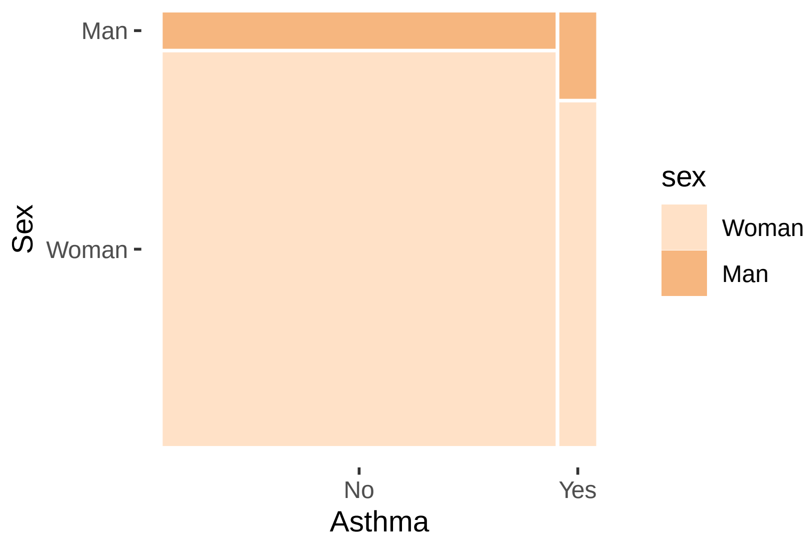
Frequencies:

```
       asthma
sex      No Yes
  Woman 198  16
  Man    18   4
```

Percentage:

```
       asthma
sex       No  Yes
  Woman 92.5  7.5
  Man   81.8 18.2
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 2.9475, df = 1, p-value = 0.08601
```

Sex was not associated with asthma (p=0.086).

Sex and COPD

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

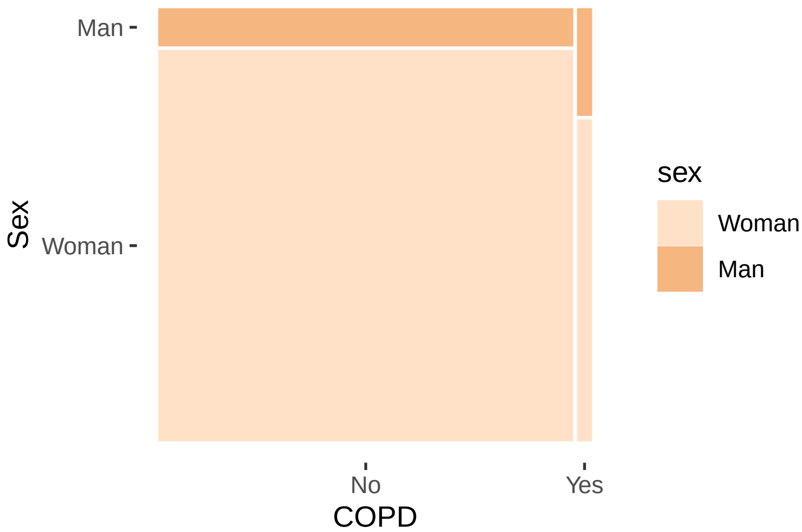
Frequencies:

```
       COPD
sex      No Yes
  Woman 208   6
  Man    20   2
```

Percentage:

```
       COPD
sex        No  Yes
  Woman 97.2  2.8
  Man   90.9  9.1
```

Mosaic Plot




```
	Pearson's Chi-squared test

data:  frequencies
X-squared = 2.4079, df = 1, p-value = 0.1207
```

Sex was not associated with COPD (p=0.121).

Sex and oxygen use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

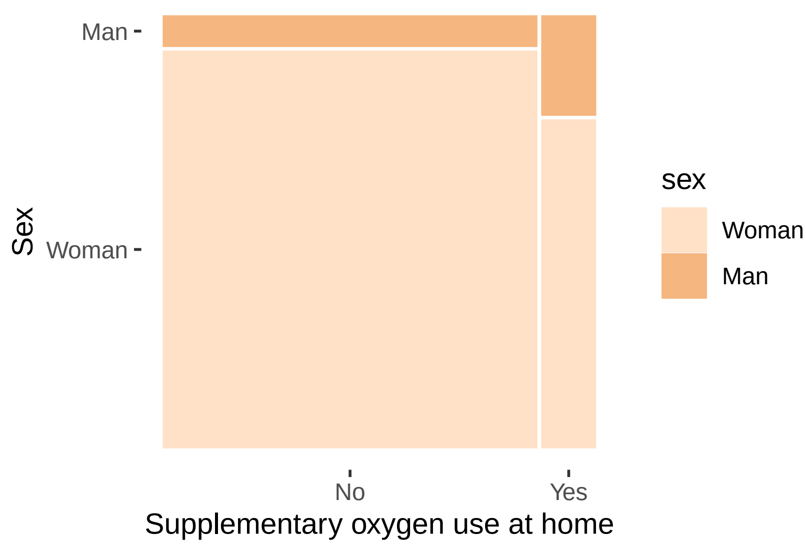
Frequencies:

```
       oxygen_use
sex       No Yes
  Woman 191  23
  Man    15   7
```

Percentage:

```
       oxygen_use
sex        No  Yes
  Woman 89.3 10.7
  Man   68.2 31.8
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 7.982, df = 1, p-value = 0.004725
```

Sex was associated with oxygen use at home (p=0.005), oxygen use was more frequent among men than women.

Sex and CPAP use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

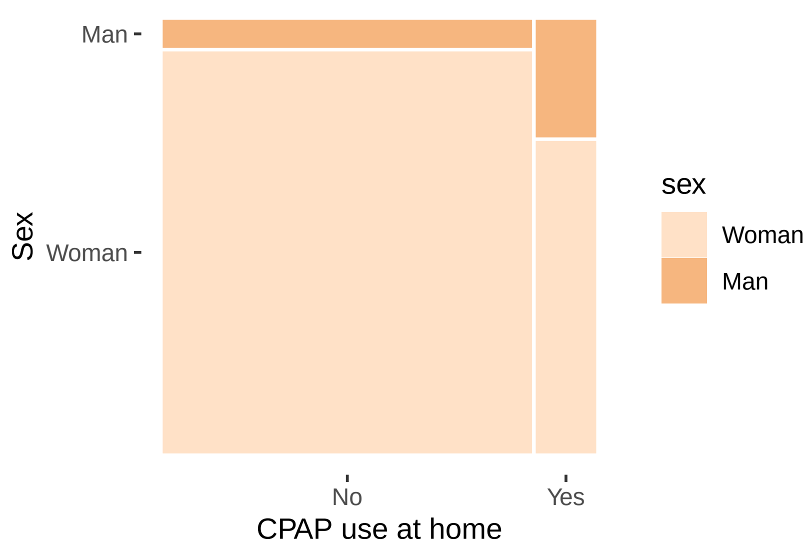
Frequencies:

```
       CPAP_use
sex       No Yes
  Woman 190  24
  Man    13   9
```

Percentage:

```
       CPAP_use
sex        No  Yes
  Woman 88.8 11.2
  Man   59.1 40.9
```

Mosaic Plot



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 14.624, df = 1, p-value = 0.0001312
```

> Sex was associated with CPAP use at home (p<0.001). CPAP use was more frequent among men than women.

Sex and altitude

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

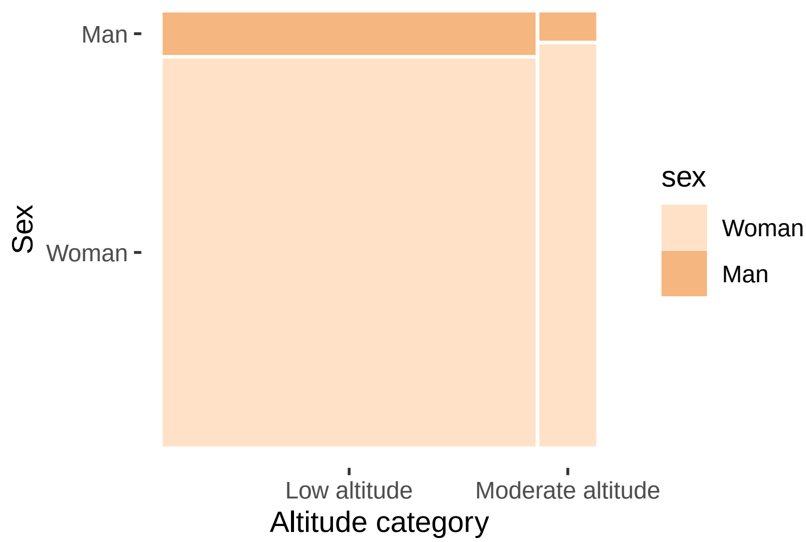
Frequencies:

```
       altitude_cat
sex     Low altitude Moderate altitude
  Woman          185                29
  Man             20                 2
```

Percentage:

```
       altitude_cat
sex     Low altitude Moderate altitude
  Woman         86.4              13.6
  Man           90.9               9.1
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 0.34786, df = 1, p-value = 0.5553
```

Sex was not associated with altitude category (p=0.555).

Sleep apnea and asthma

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

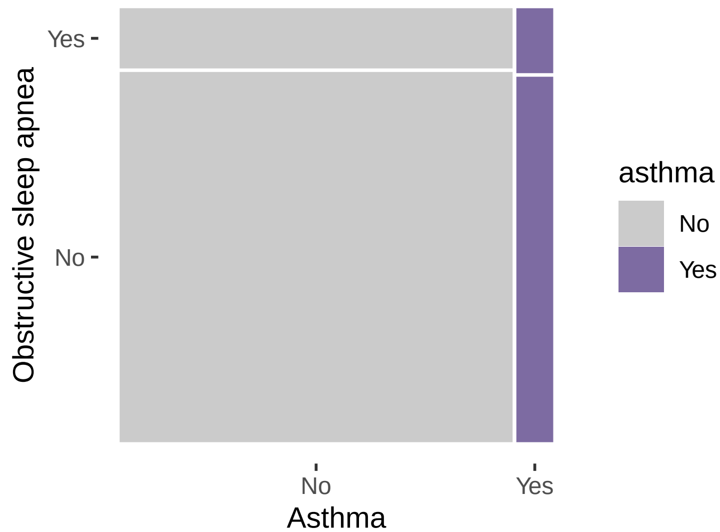
Frequencies:

```
           asthma
sleep_apnea  No Yes
        No  186  17
        Yes  30   3
```

Percentage:

```
           asthma
sleep_apnea   No  Yes
         No  91.6  8.4
         Yes 90.9  9.1
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 0.018789, df = 1, p-value = 0.891
```

Sleep apnea was not associated with asthma (p=0.891).

Sleep apnea and COPD

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

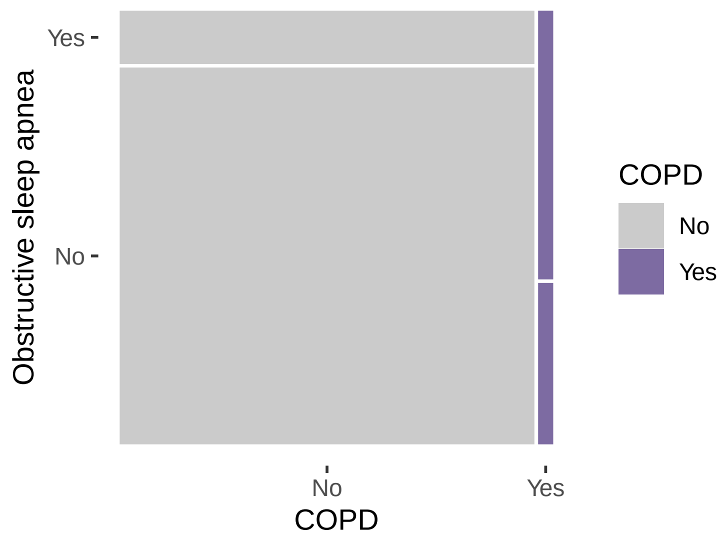
Frequencies:

```
            COPD
sleep_apnea  No Yes
        No  200   3
        Yes  28   5
```

Percentage:

```
            COPD
sleep_apnea   No  Yes
        No  98.5  1.5
        Yes 84.8 15.2
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 16.206, df = 1, p-value = 5.682e-05
```

Sleep apnea was associated with COPD (p<0.001).

Sleep apnea and oxygen use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

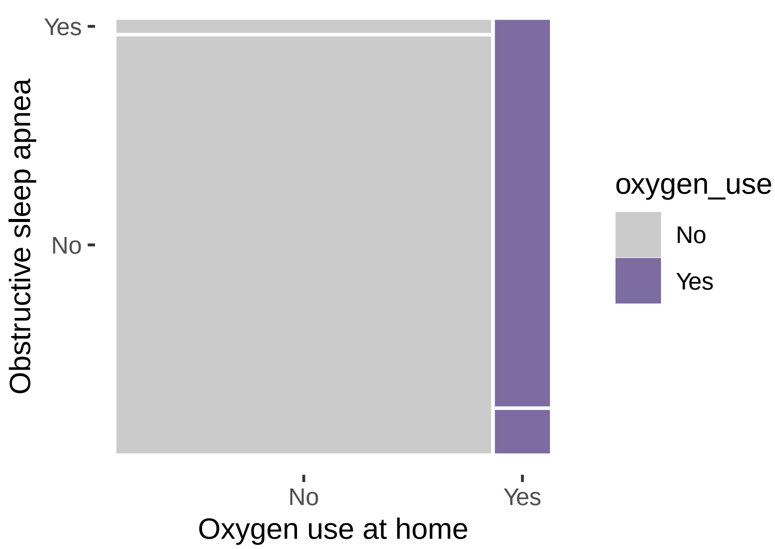
Frequencies:

```
           oxygen_use
sleep_apnea  No Yes
        No  200   3
        Yes   6  27
```

Percentage:

```
           oxygen_use
sleep_apnea   No  Yes
        No  98.5  1.5
        Yes 18.2 81.8
```

Mosaic Plot



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 165.12, df = 1, p-value < 2.2e-16
```

Sleep apnea was associated with oxygen use at home (p<0.001).

Sleep apnea and CPAP use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

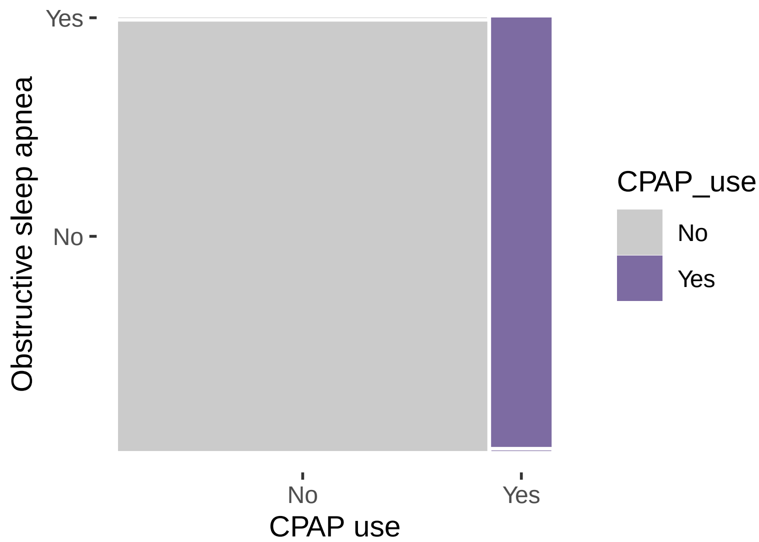
Frequencies:

```
           CPAP_use
sleep_apnea  No Yes
        No  203   0
        Yes   0  33
```

Percentage:

```
           CPAP_use
sleep_apnea  No Yes
        No  100   0
        Yes   0 100
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 236, df = 1, p-value < 2.2e-16
```

Sleep apnea was associated with CPAP use at home (p<0.001). All participants reporting a diagnosis of obstructive sleep apnea reported using CPAP at home.

COPD and asthma

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

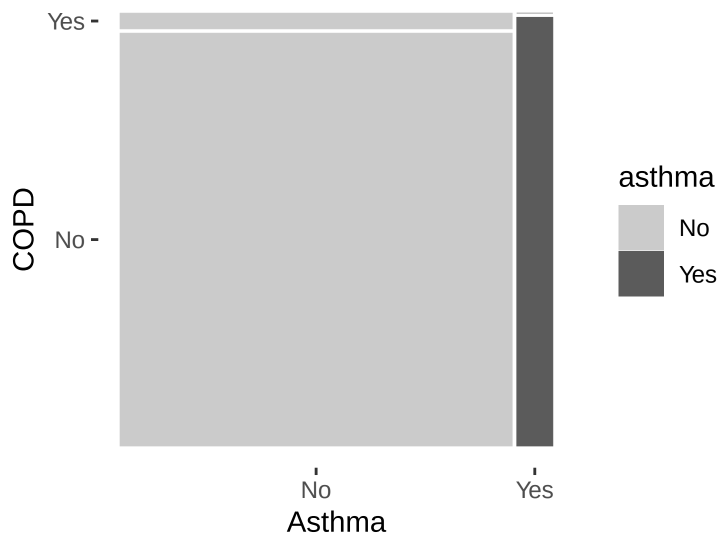
Frequencies:

```
     asthma
COPD   No Yes
  No  208  20
  Yes   8   0
```

Percentage:

```
     asthma
COPD     No   Yes
  No   91.2   8.8
  Yes 100.0   0.0
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 0.76673, df = 1, p-value = 0.3812
```

COPD was not associated with asthma (p=0.381).

COPD and oxygen use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

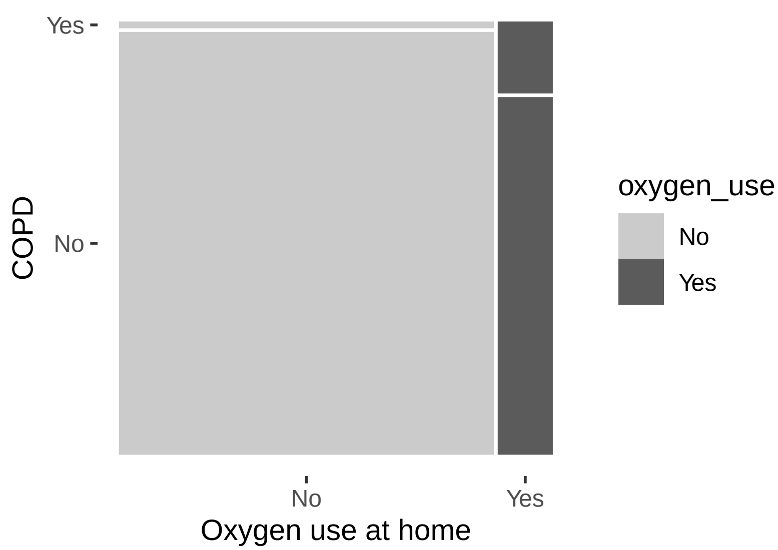
Frequencies:

```
     oxygen_use
COPD   No Yes
  No  203  25
  Yes   3   5
```

Percentage:

```
     oxygen_use
COPD    No  Yes
  No  89.0 11.0
  Yes 37.5 62.5
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 18.499, df = 1, p-value = 1.7e-05
```

COPD was associated with oxygen use at home ($p<0.001$).

COPD and CPAP use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

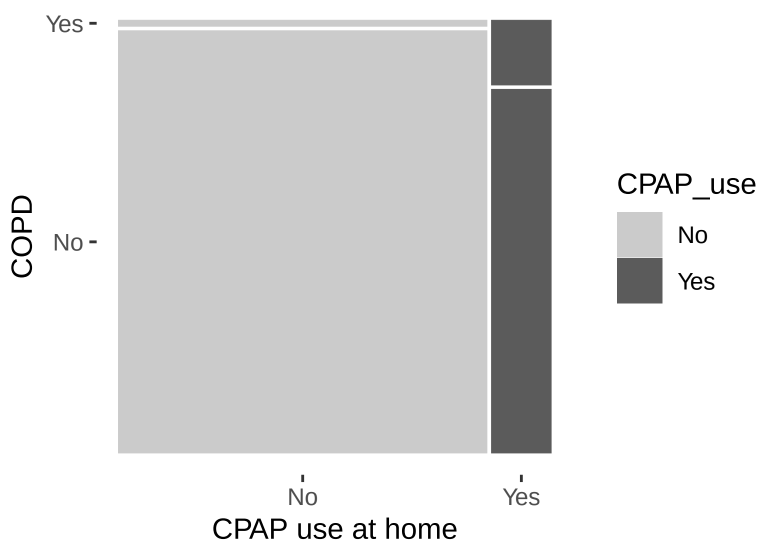
Frequencies:

```
     CPAP_use
COPD   No Yes
  No  200  28
  Yes   3   5
```

Percentage:

```
     CPAP_use
COPD    No  Yes
  No  87.7 12.3
  Yes 37.5 62.5
```

Mosaic Plot



```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 16.206, df = 1, p-value = 5.682e-05
```

COPD was associated with CPAP use at home ($p<0.001$).

Asthma and oxygen use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

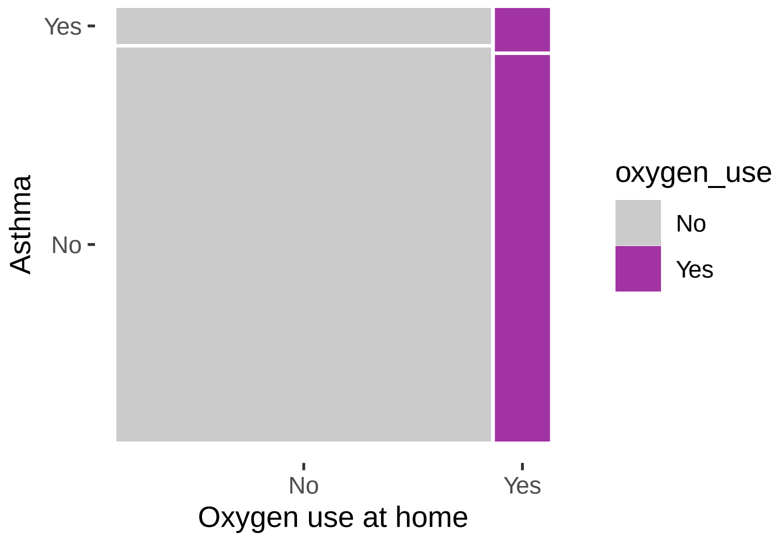
Frequencies:

```
      oxygen_use
asthma  No Yes
   No  189  27
   Yes  17   3
```

Percentage:

```
      oxygen_use
asthma   No  Yes
   No  87.5 12.5
   Yes 85.0 15.0
```

Mosaic Plot




```
    Pearson's Chi-squared test

data:  frequencies
X-squared = 0.10311, df = 1, p-value = 0.7481
```

Asthma was not associated with oxygen use at home (p=0.748).

Asthma and CPAP use

Mean expected frequency:

```
  mean_expected_freq
1                 59
```

Since value is grater than 5.0, chi-squared without continuity correction is appropriate.

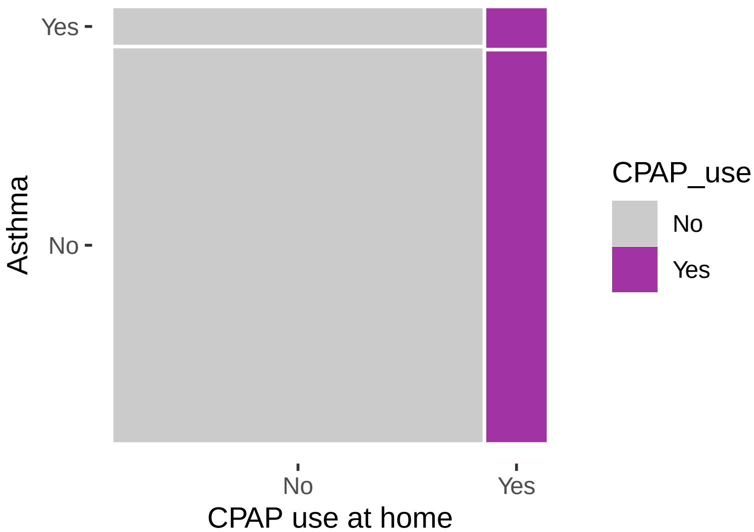
Frequencies:

```
      CPAP_use
asthma  No Yes
   No  186  30
   Yes  17   3
```

Percentage:

```
      CPAP_use
asthma   No  Yes
   No  86.1 13.9
   Yes 85.0 15.0
```

Mosaic Plot




```
	Pearson's Chi-squared test

data:  frequencies
X-squared = 0.018789, df = 1, p-value = 0.891
```

Asthma was not associated with CPAP use at home (p=0.891).

Package References

- 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/.
- 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.
- 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 (2024). *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.
- Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT (2016). "Robust causal inference using directed acyclic graphs: the R package 'dagitty'." *International Journal of Epidemiology*, *45*(6), 1887-1894. doi:10.1093/ije/dyw341 https://doi.org/10.1093/ije/dyw341.
- 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 (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5, https://CRAN.R-project.org/package=readr.
- Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, 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.