Preoperative Atelectasis

Part 8: Statistical Modelling of SpO2 (without removing outliers)

Javier Mancilla Galindo

2023-11-30

Table of contents

Setup

Packages used

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

```r
pacman::p_load(
  tidyverse, # Used for basic data handling and visualization.
  RColorBrewer, #Color palettes for data visualization.
  table1, #Used to add lables to variables.
  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.
    gt, #Used to present a summary of the results of regression models.
    gtsummary, #Used to create table to summarize regression models.
    gratia #Used together with gglopt2 to create smooth partial effects plot
          # from gam models.
  )
```

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] gratia_0.8.1       gtsummary_1.7.2    gt_0.10.0          mgcv_1.9-0
 [5] nlme_3.1-164       gridExtra_2.3      table1_1.4.3       RColorBrewer_1.1-3
 [9] lubridate_1.9.3    forcats_1.0.0      stringr_1.5.1      dplyr_1.1.4
[13] purrr_1.0.2        readr_2.1.4        tidyr_1.3.0        tibble_3.2.1
[17] ggplot2_3.4.4      tidyverse_2.0.0    pacman_0.5.1
```

Model SpO2

The SpO2 variable does not have a normal distribution. Furthermore, the distance between
1% increases in SpO2 cannot be considered equidistant increases since values are determined

from the S-shaped curve of hemoglobin saturation. This is the reason why the distribution of SpO2 is negatively skewed, with upper values reaching the saturation point of the hemoglobin curve.

Therefore, modelling SpO2 as a linear term could be potentially misleading. Nonetheless, a model assuming a gaussian distribution for SpO2 may potentially be easier to understand and communicate.

Thus, I first created model SpO2 assuming a gaussian distribution and then applied a fractional regression model which is more appropriate for the distribution of this variable. The rationale for this was that if conclusions were not different with both models, presenting a model assuming a gaussian distribution would have been easier to understand and communicate. However, since conclusions were indeed different, I will present the results for the more appropriate fractional regression model.

Fractional regression model

Convert SpO2 to fractional values between 0 and 1 to model.

```
data <- data %>% mutate(spo2_fraction = spo2_VPO/100)
```

Empty model

First, I will fit an empty model

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ 1

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.94355    0.03744   78.62   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =      0   Deviance explained = -3.04e-13%
GCV = 0.015586  Scale est. = 0.015728  n = 236
```

Smooth BMI

Model with a smooth BMI term as the only explanatory variable.

Since we are now using a different family function (quasibinomial with logit link) and we are no longer assuming a Gaussian distribution, it is important to determine the k value that offers the best representation of the change in the outcome variable with this function. I checked this by varying the value of k in the following code and **k=8***** offered the best visual representation with the largest increase in deviance explained and optimal k.value. This can be replicated by varying the value of k in the code:

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(BMI, k = 8)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.98458    0.03212   92.92   <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(BMI) 4.968  5.829 21.79  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.386   Deviance explained = 34.1%
GCV = 0.010718  Scale est. = 0.010826  n = 236
```
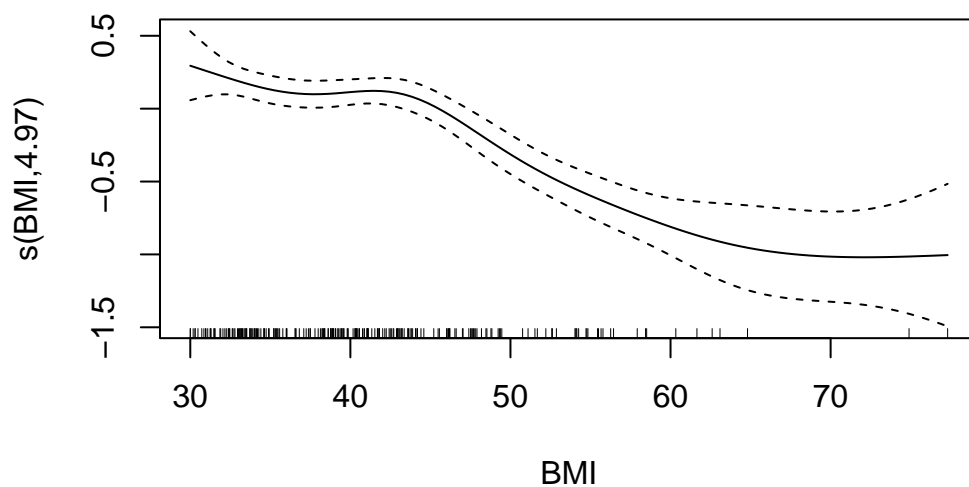
OSA only

Fit a model that only contains OSA:

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ sleep_apnea

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.00191    0.03833  78.326  < 2e-16 ***
sleep_apneaYes -0.59672    0.10971  -5.439 1.34e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.115   Deviance explained = 10.3%
GCV = 0.014099  Scale est. = 0.014441  n = 236
```

Atelectasis percent

Fit a model that only contains atelectasis percent:

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ atelectasis_percent_factor

Parametric coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    3.31028    0.03416  96.903  < 2e-16 ***
atelectasis_percent_factor2.5%  -0.54251    0.10713  -5.064 8.45e-07 ***
atelectasis_percent_factor5%    -0.79798    0.08751  -9.118  < 2e-16 ***
atelectasis_percent_factor7.5%  -0.87205    0.06127 -14.233  < 2e-16 ***
atelectasis_percent_factor10%   -0.99665    0.11831  -8.424 4.10e-15 ***
atelectasis_percent_factor12.5% -0.86794    0.29467  -2.945  0.00356 **
atelectasis_percent_factor15%   -0.99665    0.14287  -6.976 3.26e-11 ***
atelectasis_percent_factor17.5% -1.21954    0.09601 -12.703  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.676   Deviance explained = 62.6%
GCV = 0.006189  Scale est. = 0.0063048  n = 236
```

Atelectasis percent smooth term

Fit a model that contains atelectasis percent under a smooth term.

Note that this is a different function than models created in Part 4 since a gaussian function is no longer used. Thus, I determined the optimal k value which is $k=5$. This is comparable to having atelectasis percentage as categorical as will be shown later. However, the smooth term will allow to have a visual representation of the partial effect of atelectasis percent on SpO2. Thus, I will keep the smooth term to model the effect of atelectasis.
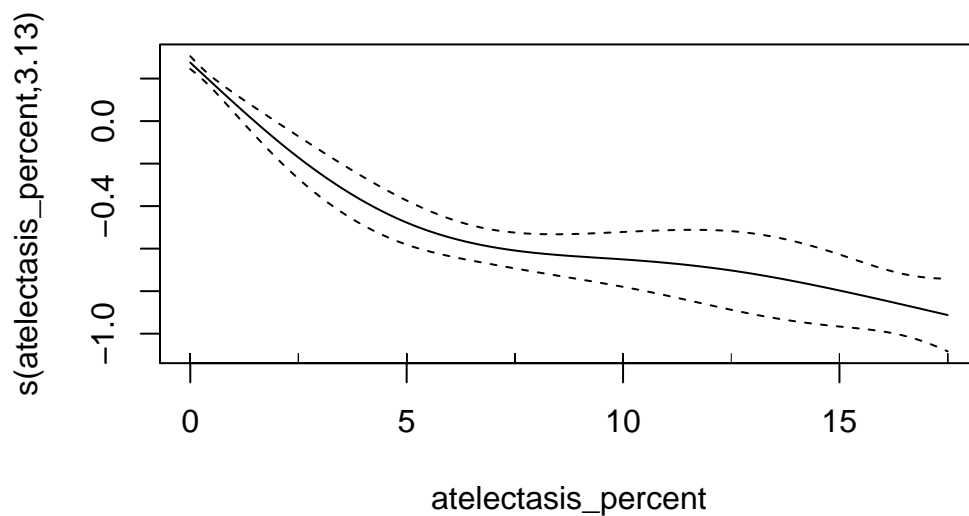
```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(atelectasis_percent, k = 5)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.02654    0.02526   119.8   <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(atelectasis_percent) 3.131  3.592 105.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.676   Deviance explained = 62.3%
GCV = 0.0060413  Scale est. = 0.0062171  n = 236
```

s(BMI) + s(atelectasis percentage)

Fit model sBMI plus atelectasis percentage:


```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(BMI, k = 8) + s(atelectasis_percent, k = 5)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.02704    0.02516   120.3   <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(BMI)                 3.461  4.248  1.12   0.314
s(atelectasis_percent) 3.059  3.524 54.23  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.688   Deviance explained = 63.3%
GCV = 0.0060456  Scale est. = 0.0061669  n = 236
```

OSA + BMI

Model sleep apnea plus sBMI:


```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(BMI, k = 8) + sleep_apnea

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.02564    0.03246  93.208  < 2e-16 ***
sleep_apneaYes -0.45017    0.09492  -4.743 3.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
         edf Ref.df     F p-value
s(BMI) 5.055   5.91 19.95  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.43   Deviance explained = 39.8%
GCV = 0.0098803  Scale est. = 0.0099659  n = 236
```

sBMI + OSA + atelectasis percent

Model sBMI + sleep apnea + atelectasis percent:

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(BMI, k = 8) + sleep_apnea + s(atelectasis_percent,
    k = 5)

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.03409    0.02603 116.576   <2e-16 ***
sleep_apneaYes -0.09196    0.07673  -1.199    0.232
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                      edf Ref.df      F p-value
s(BMI)                1.083  1.160  1.794   0.192
s(atelectasis_percent) 3.091  3.554 49.259  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.681   Deviance explained = 62.8%
GCV = 0.0060627  Scale est. = 0.0062147  n = 236
```

Adjusted model BMI

Fit model for BMI adjusted for confounders:

```
Family: quasibinomial
Link function: logit


Formula:
spo2_fraction ~ s(BMI, k = 8) + sex + age + sleep_apnea + hb +
    altitude_cat


Parametric coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 3.528269   0.428386   8.236 1.50e-14 ***
sexWoman                   -0.117513   0.112300  -1.046    0.296
age                        -0.001045   0.003139  -0.333    0.740
sleep_apneaYes             -0.474332   0.102125  -4.645 5.82e-06 ***
hb                         -0.024104   0.026525  -0.909    0.364
altitude_catModerate altitude -0.028400 0.088623  -0.320    0.749
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:
         edf Ref.df     F p-value
s(BMI) 5.032  5.886 19.54  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.424   Deviance explained = 40.5%
GCV = 0.010074  Scale est. = 0.0099703  n = 234
```

Adjusted model Atelectasis Percent

Fit model for atelectasis percent adjusted for confounders:

```
Family: quasibinomial
Link function: logit


Formula:
spo2_fraction ~ s(atelectasis_percent, k = 5) + sex + age + sleep_apnea +
    hb + altitude_cat


Parametric coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                2.9670765  0.3292438   9.012   <2e-16 ***
sexWoman                  -0.0317677  0.0875122  -0.363    0.717
```

```
age                            -0.0003715  0.0024969  -0.149    0.882
sleep_apneaYes                 -0.0973732  0.0836521  -1.164    0.246
hb                              0.0079574  0.0203634   0.391    0.696
altitude_catModerate altitude  -0.0525988  0.0688772  -0.764    0.446
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                        edf Ref.df    F p-value
s(atelectasis_percent) 3.099  3.563 89.51  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.678   Deviance explained = 63.2%
GCV = 0.0061173  Scale est. = 0.0061931  n = 234
```

Fully adjusted model

Fit fully adjusted model including both BMI and atelectasis percentage:

```
Family: quasibinomial
Link function: logit

Formula:
spo2_fraction ~ s(BMI, k = 8) + s(atelectasis_percent, k = 5) +
    sex + age + sleep_apnea + hb + altitude_cat

Parametric coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     3.0862682  0.3351823   9.208   <2e-16 ***
sexWoman                       -0.0680870  0.0900036  -0.756    0.450
age                            -0.0008286  0.0025147  -0.329    0.742
sleep_apneaYes                 -0.1097329  0.0844916  -1.299    0.195
hb                              0.0032139  0.0205653   0.156    0.876
altitude_catModerate altitude  -0.0350273  0.0699253  -0.501    0.617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                        edf Ref.df    F p-value
s(BMI)                 3.244  4.015  1.063   0.366
s(atelectasis_percent) 2.998  3.469 45.713  <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.687   Deviance explained = 64.2%
GCV = 0.0061332  Scale est. = 0.0061365  n = 234
```

Build a dataframe to compare models:

Models sorted by explained deviance (from higher to lower):

Model	aR2	dev
Fully_adjusted	68.7	64.2
sBMI_atel	68.8	63.3
adjusted_atelectasis	67.8	63.2
sBMI_atel_OSA	68.1	62.8
atel_only	67.6	62.6
atel_smooth	67.6	62.3
adjusted_BMI	42.4	40.5
sBMI_OSA	43.0	39.8
sBMI	38.6	34.1
OSA_only	11.5	10.3
empty	0.0	0.0

Create table for models:

Group	Characteristic	OR	95% CI	p-value
BMI only	s(BMI)			<0.001
OSA only	Obstructive sleep apnea			
	No	—	—	
	Yes	0.55	0.44, 0.68	<0.001
Atelectasis percent only	s(atelectasis_percent)			<0.001
BMI + OSA	Obstructive sleep apnea			
	No	—	—	
	Yes	0.64	0.53, 0.77	<0.001
	s(BMI)			<0.001
BMI + Atelectasis percent	s(BMI)			0.3
	s(atelectasis_percent)			<0.001

Group	Characteristic	OR	95% CI	p-value
BMI + OSA + Atelectasis percent	Obstructive sleep apnea			
	No	—	—	
	Yes	0.91	0.78, 1.06	0.2
	s(BMI)			0.2
	s(atelectasis_percent)			<0.001
BMI + OSA + age + sex + hb + altitude	Sex			
	Man	—	—	
	Woman	0.89	0.71, 1.11	0.3
	Age	1.00	0.99, 1.01	0.7
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.62	0.51, 0.76	<0.001
	Hemoglobin	0.98	0.93, 1.03	0.4
	altitude_cat			
	Low altitude	—	—	
	Moderate altitude	0.97	0.82, 1.16	0.7
	s(BMI)			<0.001
Atelectasis percent + OSA + age + sex + hb + altitude	Sex			
	Man	—	—	
	Woman	0.97	0.82, 1.15	0.7
	Age	1.00	1.0, 1.00	0.9
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.91	0.77, 1.07	0.2
	Hemoglobin	1.01	0.97, 1.05	0.7
	altitude_cat			
	Low altitude	—	—	

| Group	Characteristic	OR	95% CI	p-value
	Moderate altitude	0.95	0.83, 1.09	0.4
	s(atelectasis_percent)			<0.001
Fully adjusted model	Sex			
	Man	—	—	
	Woman	0.93	0.78, 1.11	0.5
	Age	1.00	0.99, 1.00	0.7
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.90	0.76, 1.06	0.2
	Hemoglobin	1.00	0.96, 1.04	0.9
	altitude_cat			
	Low altitude	—	—	
	Moderate altitude	0.97	0.84, 1.11	0.6
	s(BMI)			0.4
	s(atelectasis_percent)			<0.001

Figure SpO2:

sBMI

Now, take the inverse logit function to assess partial effect on mean SpO2.

Partial effect on mean SpO2:

A

BMI only
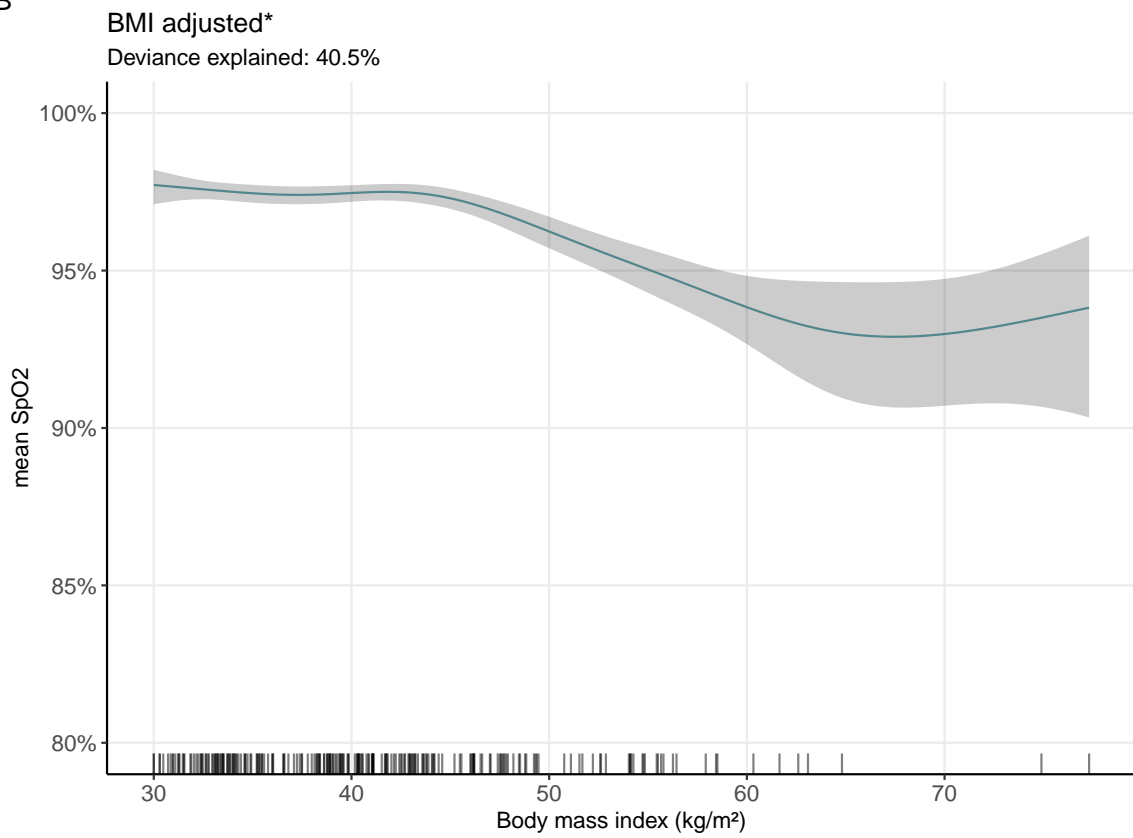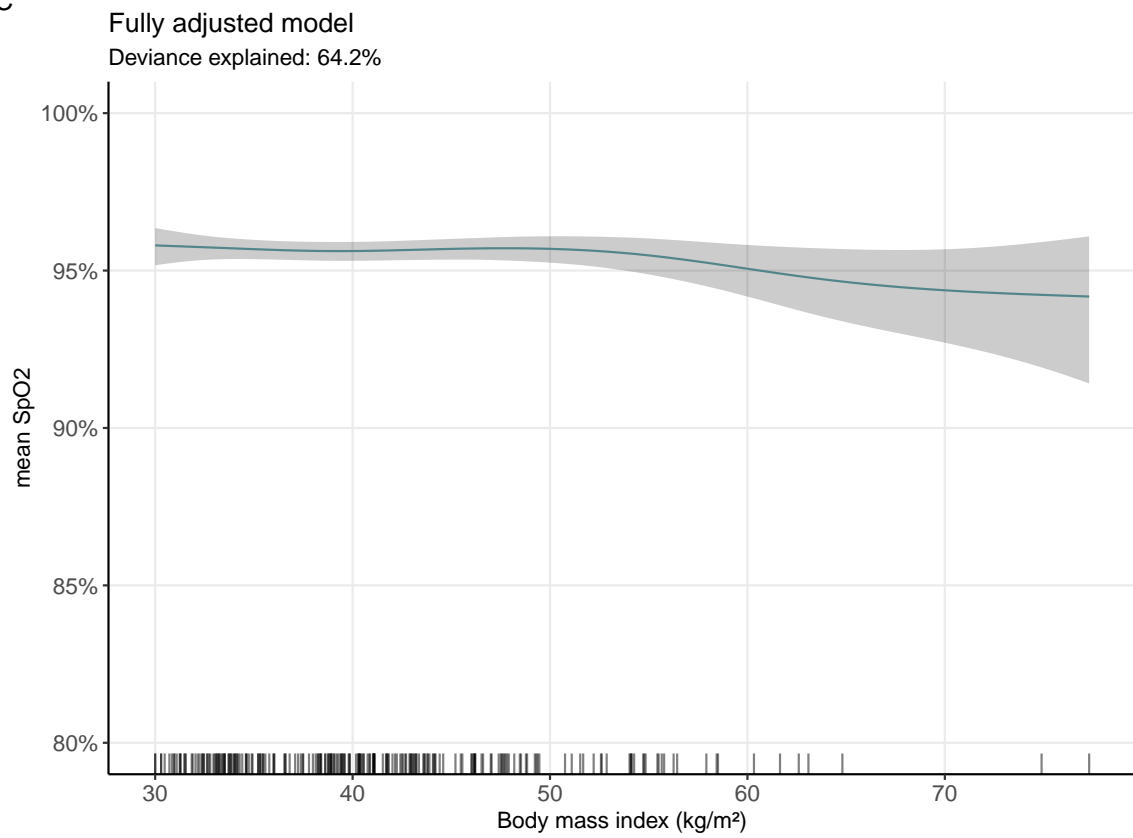Deviance explained: 34.1%

sBMI_adjusted

Partial effect on mean SpO2:

B
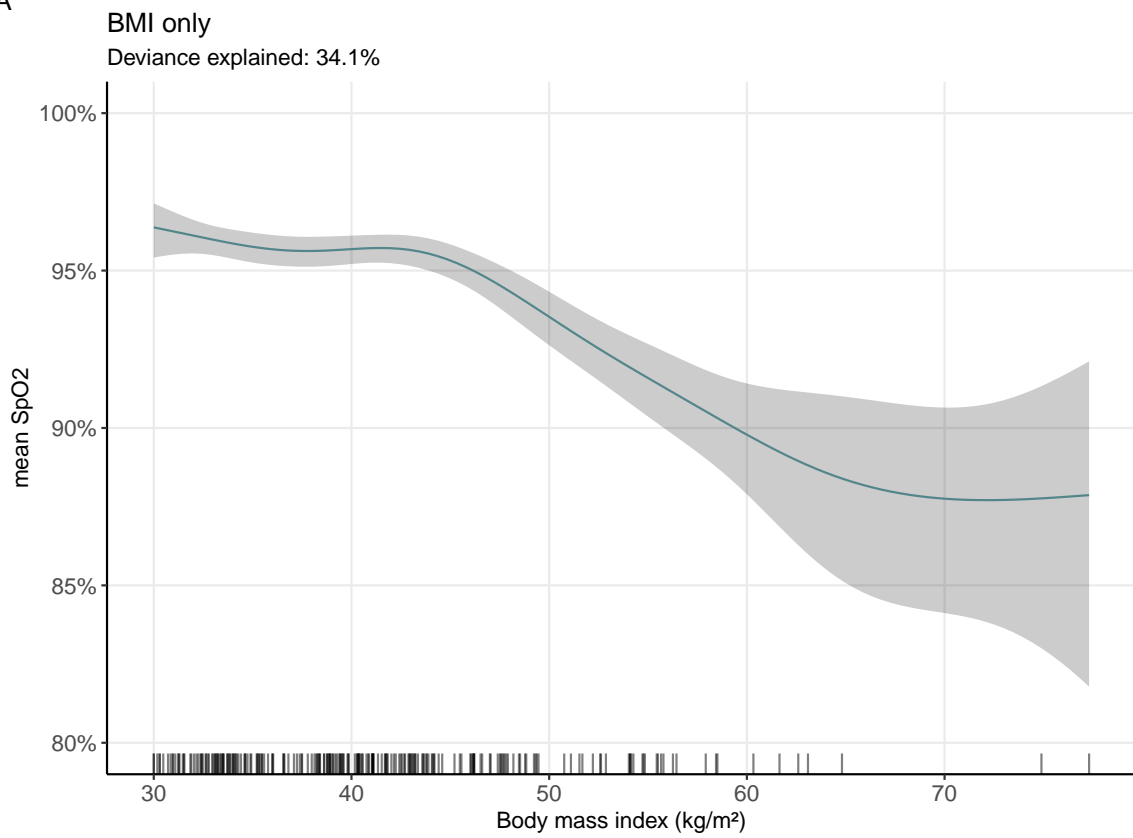
BMI adjusted*
Deviance explained: 40.5%

Fully adjusted BMI

Partial effect on mean SpO2:

C

Fully adjusted model

Deviance explained: 64.2%

sAtelectasis percent

Check residuals:

Draw a personalized plot:

D

Atelectasis percent

Deviance explained: 62.3%
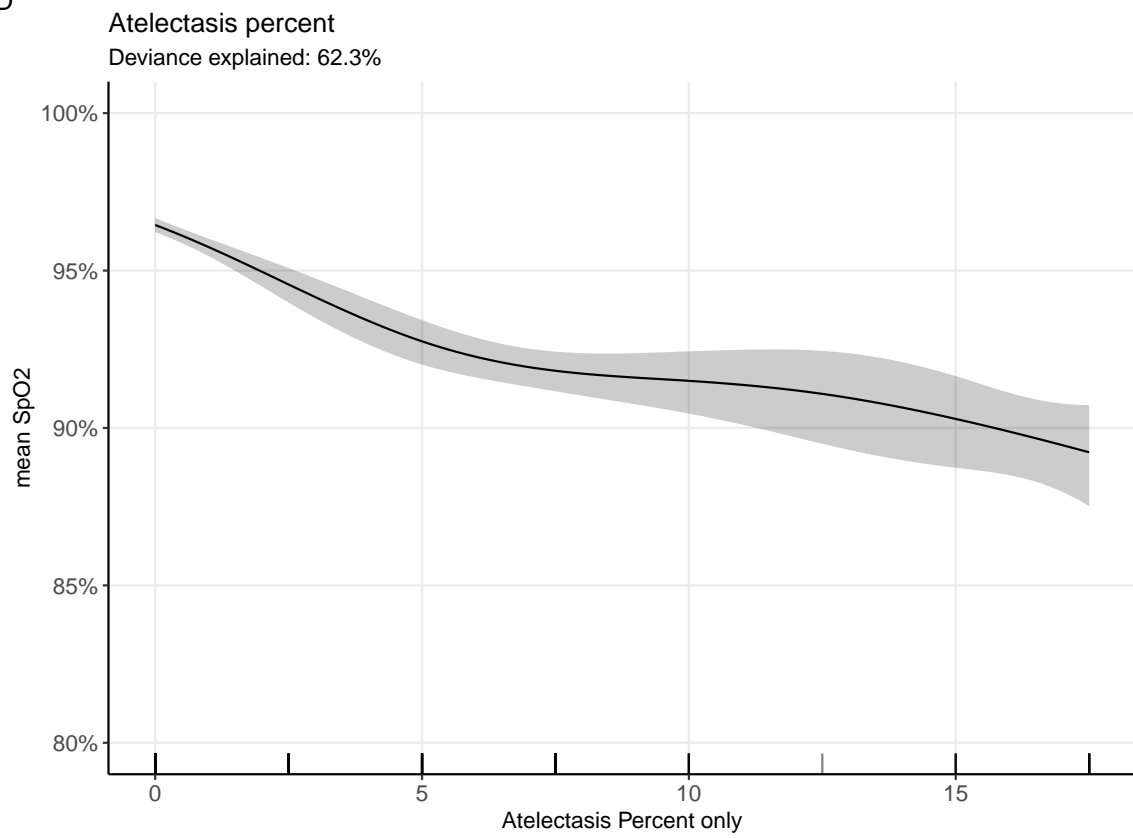
Atelectasis percent adjusted

Partial effect on mean SpO2:

E

Atelectasis percent adjusted*
Deviance explained: 63.2%
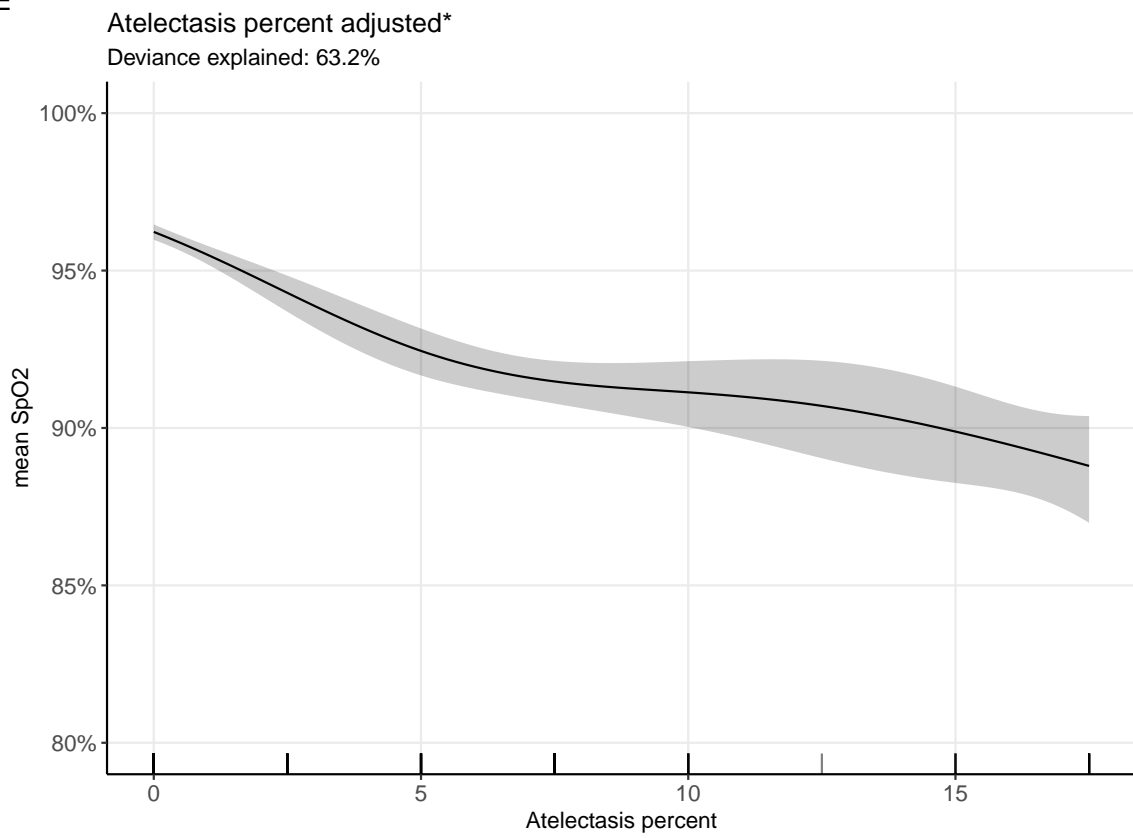
Fully adjusted Atelectasis Percent

Partial effect on mean SpO2:

F

Fully adjusted model
Deviance explained: 64.2%

Figure 3

A BMI only
Deviance explained: 34.1%

B BMI adjusted*
Deviance explained: 40.5%

C Fully adjusted model
Deviance explained: 64.2%

D Atelectasis percent
Deviance explained: 62.3%

E Atelectasis percent adjusted*
Deviance explained: 63.2%

F Fully adjusted model
Deviance explained: 64.2%