

# Preoperative Atelectasis

## Part 6: Statistical Modelling of SpO2

Javier Mancilla Galindo

2023-11-30

### Table of contents

<b>Setup</b>	<b>1</b>
<b>Model SpO2</b>	<b>3</b>
Fractional regression model . . . . .	3
OSA + BMI . . . . .	14
sBMI + OSA + atelectasis percent . . . . .	14
Adjusted model BMI . . . . .	15
Adjusted model Atelectasis Percent . . . . .	17
Fully adjusted model . . . . .	17
Table S2 . . . . .	18
<b>Figure SpO2 models</b>	<b>21</b>
<b>Predictions SpO2</b>	<b>29</b>
Figure 3 . . . . .	31

### Setup

#### Packages used

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

```

```

tidyverse, # Used for basic data handling and visualization.
RColorBrewer, #Color palettes for data visualization.
table1, #Used to add labes 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.
metR # Used to plot predictions of SpO2.
)

```

## 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] metR_0.14.1      gratia_0.8.1      gtsummary_1.7.2    gt_0.10.0
[5] mgcv_1.9-0       nlme_3.1-164      gridExtra_2.3      table1_1.4.3
[9] RColorBrewer_1.1-3 lubridate_1.9.3    forcats_1.0.0      stringr_1.5.1
[13] dplyr_1.1.4      purrr_1.0.2       readr_2.1.4        tidyr_1.3.0
[17] tibble_3.2.1     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.

As a last note, I first assessed the relationships between variables without removing any outliers. Examination of residuals showed that there were some influential outliers having an impact on the models. Thus, I decided to remove a total of 4 outliers only for the SpO2 models shown here (3 for the SpO2 ~ BMI relationship and 1 for SpO2 ~ atelectasis percent). This document presents the results of analyses after removing outliers. Results of models for the total sample before exclusion of outliers are presented in *Part 8*.

### Fractional regression model

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

```
data <- data %>% mutate(spo2_fraction = spo2_VP0/100)
```

### Empty model

First, I will fit an empty model

### BMI residuals

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:

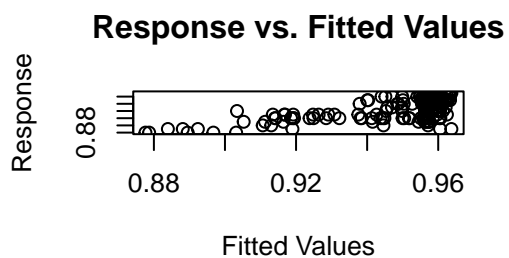
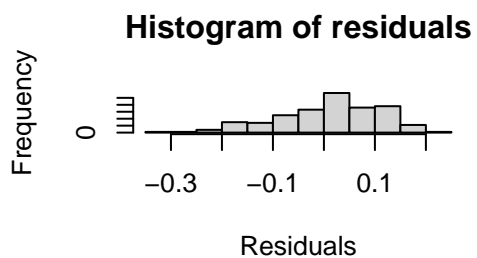
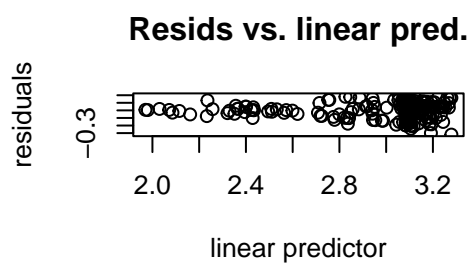
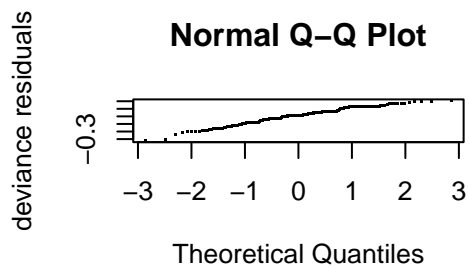
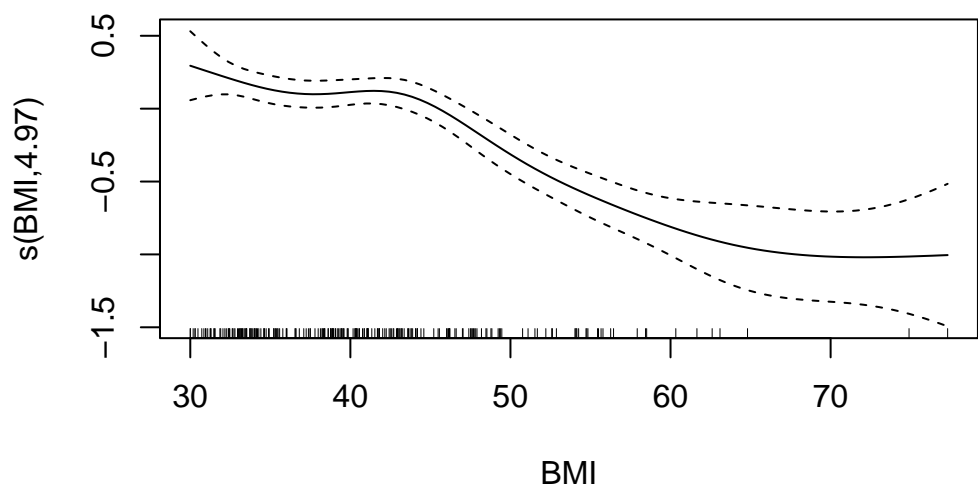
	edf	Ref.df	F	p-value
s(BMI)	4.968	5.829	21.79	<2e-16 ***

---

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



Method: GCV    Optimizer: outer newton  
 full convergence after 4 iterations.  
 Gradient range  $[-4.075644\text{e-}09, -4.075644\text{e-}09]$

```
(score 0.01071767 & scale 0.0108264).
Hessian positive definite, eigenvalue range [2.025108e-05,2.025108e-05].
Model rank = 8 / 8
```

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

```
      k'   edf k-index p-value
s(BMI) 7.00 4.97    1.11    0.97
```

There are influential residuals. Will assess which of these could be removed according to Cook's distance.

```
data %>%
  mutate(
    cooks_d = cooks.distance(model_BMI),
    outlier = ifelse(cooks_d < 4/nrow(data), "keep", "delete")
  ) %>%
  filter(outlier=="delete") %>%
  dplyr::select(ID,BMI,spo2_VPO,cooks_d,outlier) %>%
  gt()
```

ID	Body mass index	Oxygen saturation (SpO2)	Oxygen saturation (SpO2)	outlier
71	30.00	99	0.01698986	delete
122	30.00	89	0.13348733	delete
166	32.46	89	0.03196280	delete

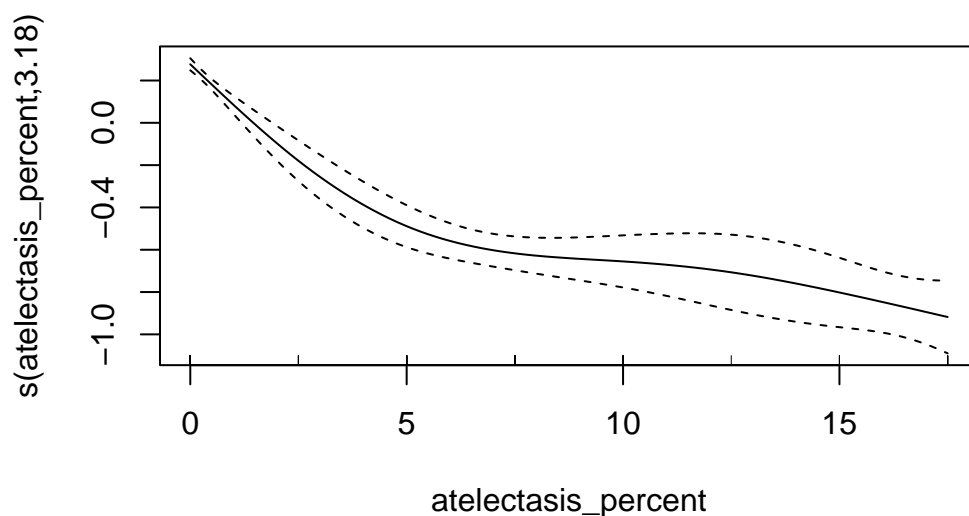
I will remove such residuals:

```
data <- data %>% filter(!ID %in% c("71","122","166"))
```

Now, I will examine atelectasis percentage:

### Atelectasis percent residuals

Greater explanation on why a smooth term for atelectasis percentage was kept instead of a categorical term can be found in **Part 8**, as well as an explanation of the k value selected.



Family: quasibinomial

Link function: logit

Formula:

spo2\_fraction ~ s(atelectasis\_percent, k = 5)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.0345	0.0241	125.9	<2e-16 ***

---

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

Approximate significance of smooth terms:

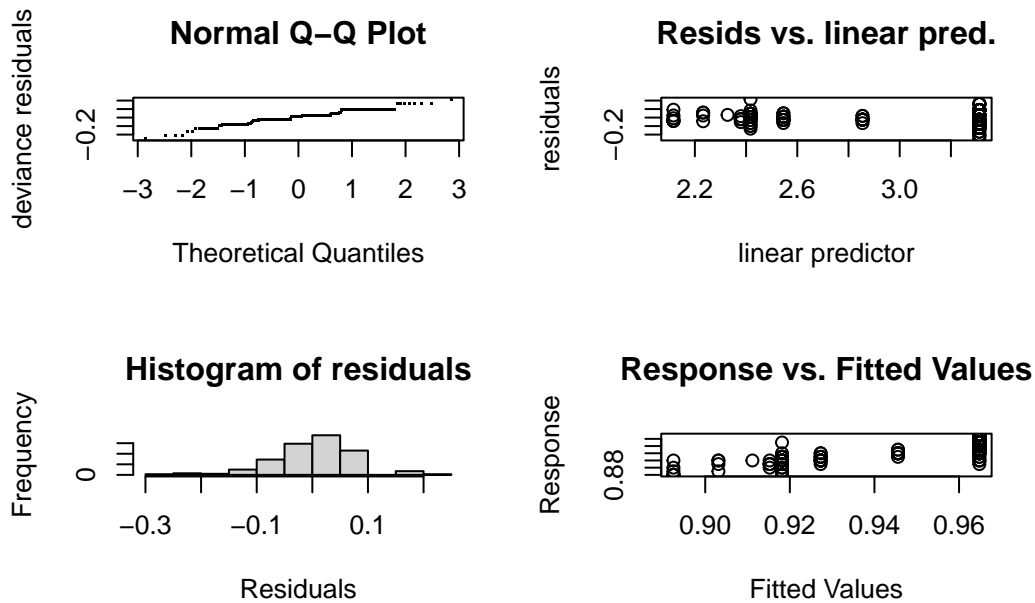
	edf	Ref.df	F	p-value
s(atelectasis_percent)	3.176	3.63	115.2	<2e-16 ***

---

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

R-sq.(adj) = 0.697 Deviance explained = 64.3%

GCV = 0.00553 Scale est. = 0.0055429 n = 233



Method: GCV Optimizer: outer newton  
 full convergence after 3 iterations.  
 Gradient range [2.584867e-11,2.584867e-11]  
 (score 0.005529978 & scale 0.005542931).  
 Hessian positive definite, eigenvalue range [2.314762e-05,2.314762e-05].  
 Model rank = 5 / 5

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(atelectasis_percent)	4.00	3.18	0.94	0.2

Outliers:

```
data %>%
  mutate(
    cooks_d = cooks.distance(model_atel_smooth),
    outlier = ifelse(cooks_d < 4/nrow(data), "keep","delete")
  ) %>%
  filter(outlier=="delete") %>%
  dplyr::select(ID,BMI,spo2_VP0,atelectasis_percent,cooks_d,outlier) %>%
```



gt()

ID	Body mass index	Oxygen saturation (SpO2)	Percentage of atelectasis	Oxygen saturation (SpO2)
4	46.16	92	15.0	0.01853617
39	35.47	92	15.0	0.01853617
107	55.44	92	17.5	0.05640899
129	60.33	88	7.5	0.01947016
164	41.10	91	0.0	0.02412831
204	48.78	97	7.5	0.03597710

I will remove these outliers.

### BMI only model

Family: quasibinomial

Link function: logit

Formula:

spo2\_fraction ~ s(BMI, k = 8)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.01363	0.03096	97.34	<2e-16 ***

---

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

Approximate significance of smooth terms:

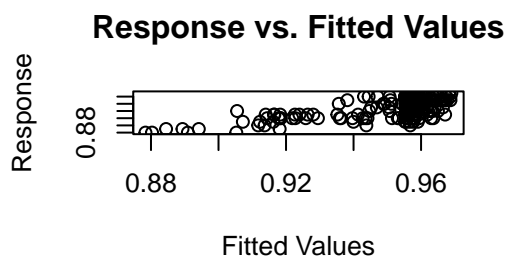
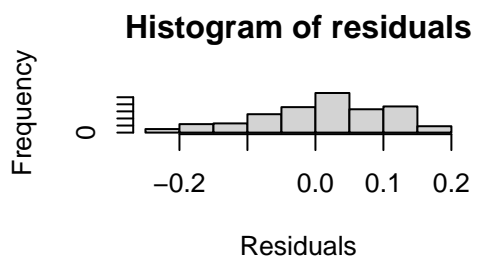
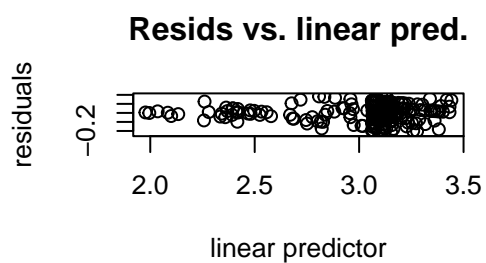
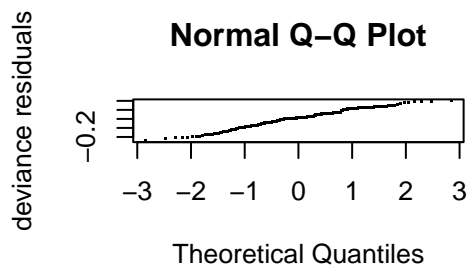
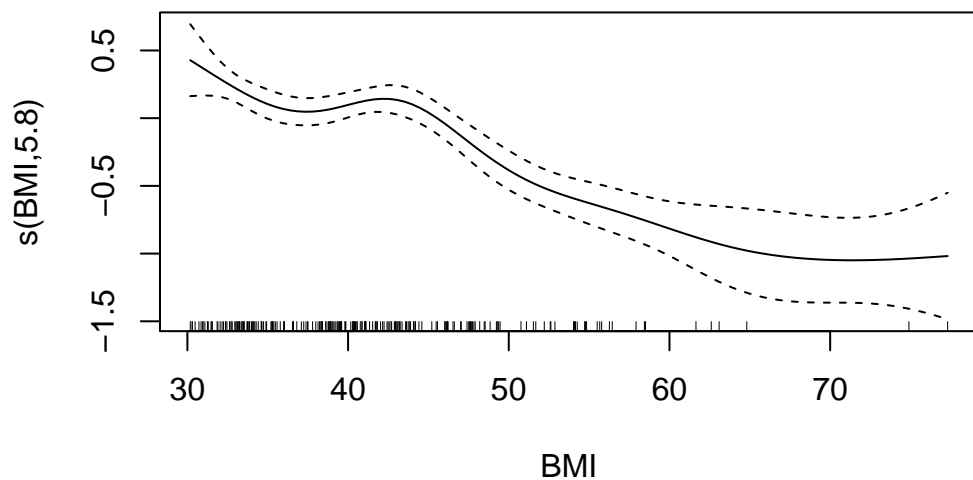
	edf	Ref.df	F	p-value
s(BMI)	5.799	6.518	22.99	<2e-16 ***

---

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

R-sq.(adj) = 0.43 Deviance explained = 38.7%

GCV = 0.009523 Scale est. = 0.0093856 n = 227



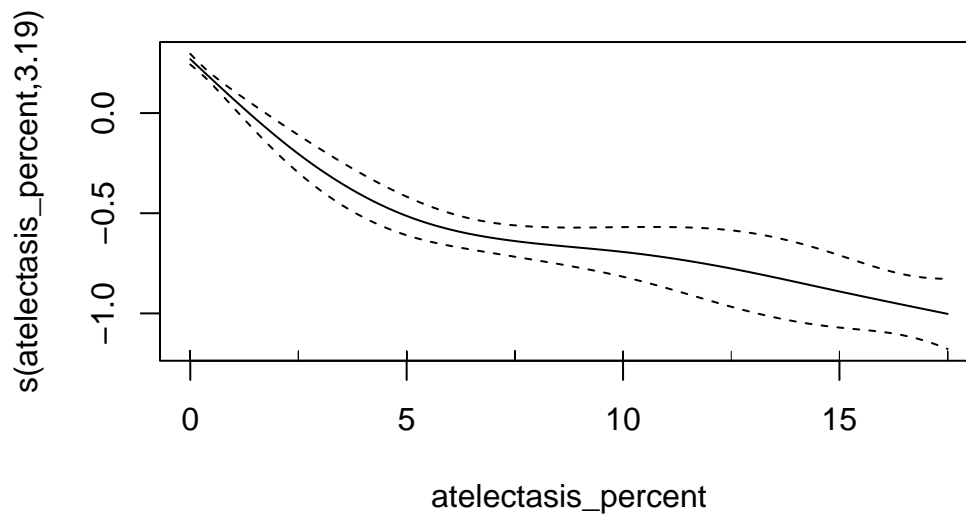
Method: GCV    Optimizer: outer newton  
 full convergence after 5 iterations.  
 Gradient range [7.189245e-09,7.189245e-09]

```
(score 0.009522991 & scale 0.009385629).
Hessian positive definite, eigenvalue range [3.738925e-05,3.738925e-05].
Model rank = 8 / 8
```

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(BMI)	7.0	5.8	1.14	0.98

### Atelectasis percent model



```
Family: quasibinomial
Link function: logit
```

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

```
Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.05346    0.02339   130.5  <2e-16 ***
---

```

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

Approximate significance of smooth terms:

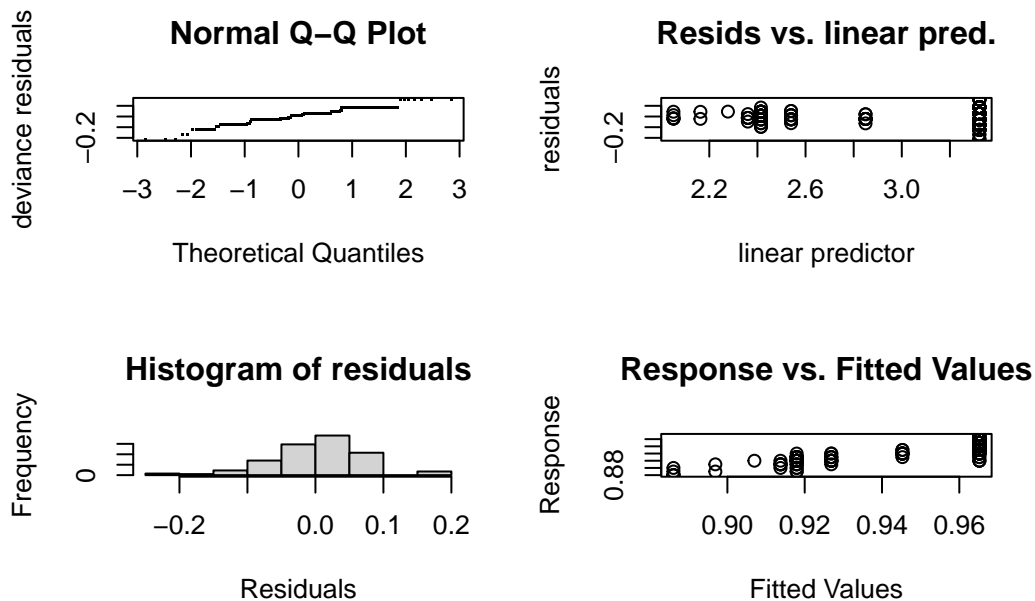
	edf	Ref.df	F	p-value
s(atelectasis_percent)	3.194	3.63	127.9	<2e-16 ***

---

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

R-sq.(adj) = 0.735 Deviance explained = 67%

GCV = 0.00501 Scale est. = 0.0050014 n = 227



Method: GCV Optimizer: outer newton

full convergence after 2 iterations.

Gradient range [5.019533e-09,5.019533e-09]

(score 0.005010037 & scale 0.005001372).

Hessian positive definite, eigenvalue range [2.054912e-05,2.054912e-05].

Model rank = 5 / 5

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(atelectasis_percent)	4.00	3.19	0.92	0.14

## 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)	3.0325	0.0382	79.377	< 2e-16 ***
sleep_apneaYes	-0.6273	0.1061	-5.914	1.23e-08 ***

---

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

R-sq.(adj) = 0.139 Deviance explained = 12.4%

GCV = 0.013031 Scale est. = 0.013381 n = 227

## 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)	3.05416	0.02326	131.3	<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)	3.702	4.519	0.995	0.402
s(atelectasis_percent)	3.150	3.587	62.816	<2e-16 ***

---

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

R-sq.(adj) = 0.742 Deviance explained = 67.9%

GCV = 0.0050285 Scale est. = 0.0049413 n = 227

## 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)	3.05800	0.03086	99.086	< 2e-16 ***
sleep_apneaYes	-0.46539	0.08750	-5.319	2.57e-07 ***

---

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)	5.828	6.537	21.57	<2e-16 ***

---

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

R-sq.(adj) = 0.481 Deviance explained = 45.2%

GCV = 0.0085874 Scale est. = 0.0083706 n = 227

## 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)	3.05992	0.02413	126.784	<2e-16 ***
sleep_apneaYes	-0.07722	0.07069	-1.092	0.276

---

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)	1.000	1.000	1.41	0.236
s(atelectasis_percent)	3.165	3.605	54.52	<2e-16 ***

---

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

R-sq.(adj) = 0.736 Deviance explained = 67.3%

GCV = 0.00505 Scale est. = 0.0049995 n = 227

## 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)	3.505012	0.404492	8.665	1.14e-15 ***
sexWoman	-0.090059	0.103764	-0.868	0.386
age	-0.001313	0.002927	-0.449	0.654
sleep_apneaYes	-0.478262	0.094269	-5.073	8.49e-07 ***
hb	-0.021080	0.025152	-0.838	0.403
altitude_catModerate altitude	-0.057652	0.085521	-0.674	0.501

---

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

Approximate significance of smooth terms:

edf	Ref.df	F	p-value
-----	--------	---	---------

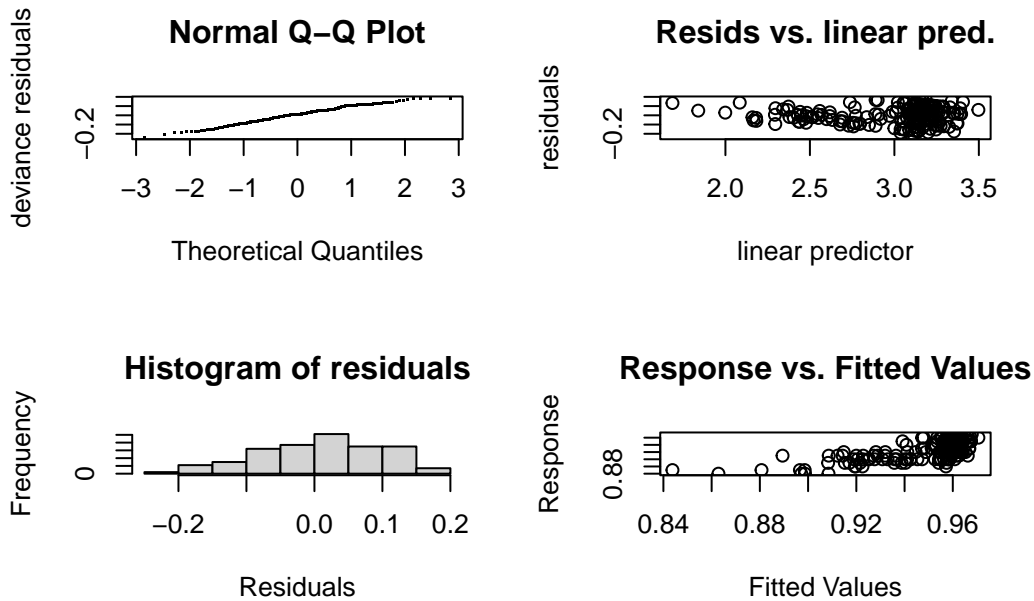
s(BMI) 5.879 6.569 21.13 <2e-16 \*\*\*

---

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

R-sq.(adj) = 0.475 Deviance explained = 45.8%

GCV = 0.0087677 Scale est. = 0.0083922 n = 225



Method: GCV Optimizer: outer newton

full convergence after 5 iterations.

Gradient range [6.767106e-11,6.767106e-11]

(score 0.008767665 & scale 0.008392233).

Hessian positive definite, eigenvalue range [3.145143e-05,3.145143e-05].

Model rank = 13 / 13

Basis dimension (k) checking results. Low p-value (k-index<1) may indicate that k is too low, especially if edf is close to k'.

	k'	edf	k-index	p-value
s(BMI)	7.00	5.88	1.19	0.99



## 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)	2.9604830	0.3006343	9.847	<2e-16 ***
sexWoman	-0.0226450	0.0784683	-0.289	0.773
age	-0.0001248	0.0022581	-0.055	0.956
sleep_apneaYes	-0.0734390	0.0760344	-0.966	0.335
hb	0.0088852	0.0186168	0.477	0.634
altitude_catModerate altitude	-0.0526134	0.0639735	-0.822	0.412

---

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)	3.167	3.604	107.3	<2e-16 ***

---

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

R-sq.(adj) = 0.736 Deviance explained = 68%

GCV = 0.0050499 Scale est. = 0.0049432 n = 225

## 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)	3.0193878	0.3062344	9.860	<2e-16 ***
sexWoman	-0.0349358	0.0794760	-0.440	0.661
age	-0.0005502	0.0022964	-0.240	0.811
sleep_apneaYes	-0.0857308	0.0770947	-1.112	0.267
hb	0.0068415	0.0187388	0.365	0.715
altitude_catModerate altitude	-0.0514186	0.0640281	-0.803	0.423

---

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)	1.000	1.000	1.043	0.308
s(atelectasis_percent)	3.148	3.588	54.963	<2e-16 ***

---

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

R-sq.(adj) = 0.737 Deviance explained = 68.1%

GCV = 0.0050733 Scale est. = 0.0049473 n = 225

Build a dataframe to compare models:

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

Model	aR2	dev
Fully_adjusted	73.7	68.1
adjusted_atelectasis	73.6	68.0
sBMI_atel	74.2	67.9
sBMI_atel_OSA	73.6	67.3
atel_smooth	73.5	67.0
adjusted_BMI	47.5	45.8
sBMI_OSA	48.1	45.2
sBMI	43.0	38.7
OSA_only	13.9	12.4
empty	0.0	0.0

## Table S2

Create table for models:

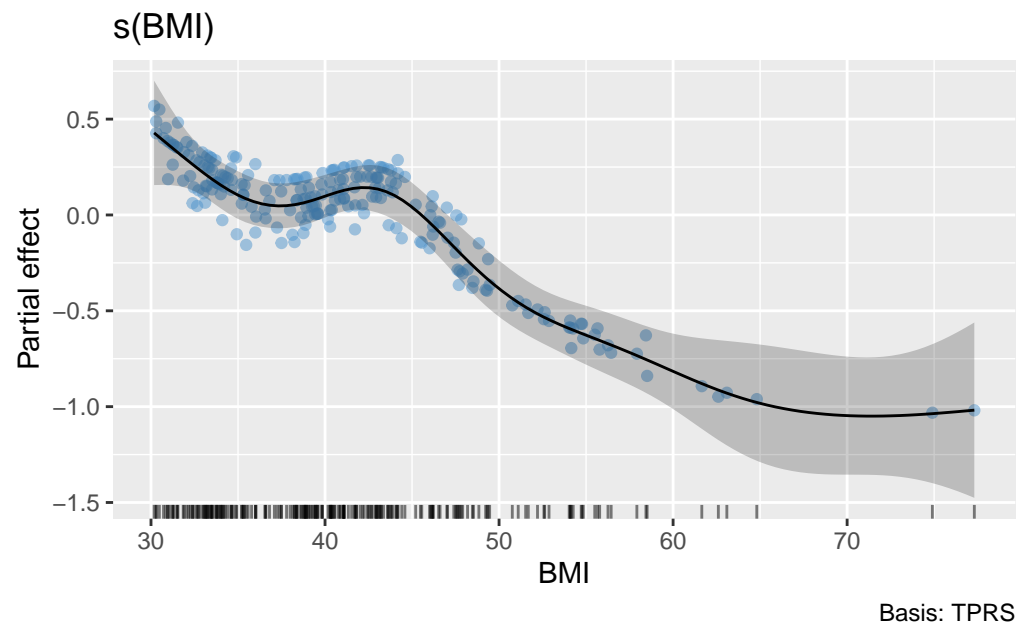
Group	Characteristic	OR	95% CI	p-value
BMI only	s(BMI)			<0.001
OSA only	Obstructive sleep apnea			
	No	—	—	
	Yes	0.53	0.43, 0.66	<0.001
Atelectasis percent only	s(atelectasis_percent)			<0.001
BMI + OSA	Obstructive sleep apnea			
	No	—	—	
	Yes	0.63	0.53, 0.75	<0.001
BMI + Atelectasis percent	s(BMI)			<0.001
	s(BMI)			0.4
	s(atelectasis_percent)			<0.001
BMI + OSA + Atelectasis percent	Obstructive sleep apnea			
	No	—	—	
	Yes	0.93	0.81, 1.06	0.3
	s(BMI)			0.2
BMI + OSA + age + sex + hb + altitude	s(atelectasis_percent)			<0.001
	Sex			
	Man	—	—	
	Woman	0.91	0.75, 1.12	0.4
	Age	1.00	0.99, 1.00	0.7
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.62	0.52, 0.75	<0.001
	Hemoglobin	0.98	0.93, 1.03	0.4
	altitude_cat			
	Low altitude	—	—	
	Moderate altitude	0.94	0.80, 1.12	0.5
	s(BMI)			<0.001

Group	Characteristic	OR	95% CI	p-value
Atelectasis percent + OSA + age + sex + hb + altitude	Sex			
	Man	—	—	
	Woman	0.98	0.84, 1.14	0.8
	Age	1.00	1.00, 1.00	>0.9
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.93	0.80, 1.08	0.3
	Hemoglobin	1.01	0.97, 1.05	0.6
	altitude_cat			
	Low altitude	—	—	
	Moderate altitude	0.95	0.84, 1.08	0.4
	s(atelectasis_percent)			<0.001
Fully adjusted model	Sex			
	Man	—	—	
	Woman	0.97	0.83, 1.13	0.7
	Age	1.00	1.0, 1.00	0.8
	Obstructive sleep apnea			
	No	—	—	
	Yes	0.92	0.79, 1.07	0.3
	Hemoglobin	1.01	0.97, 1.04	0.7
	altitude_cat			
	Low altitude	—	—	
	Moderate altitude	0.95	0.84, 1.08	0.4
	s(BMI)			0.3
	s(atelectasis_percent)			<0.001

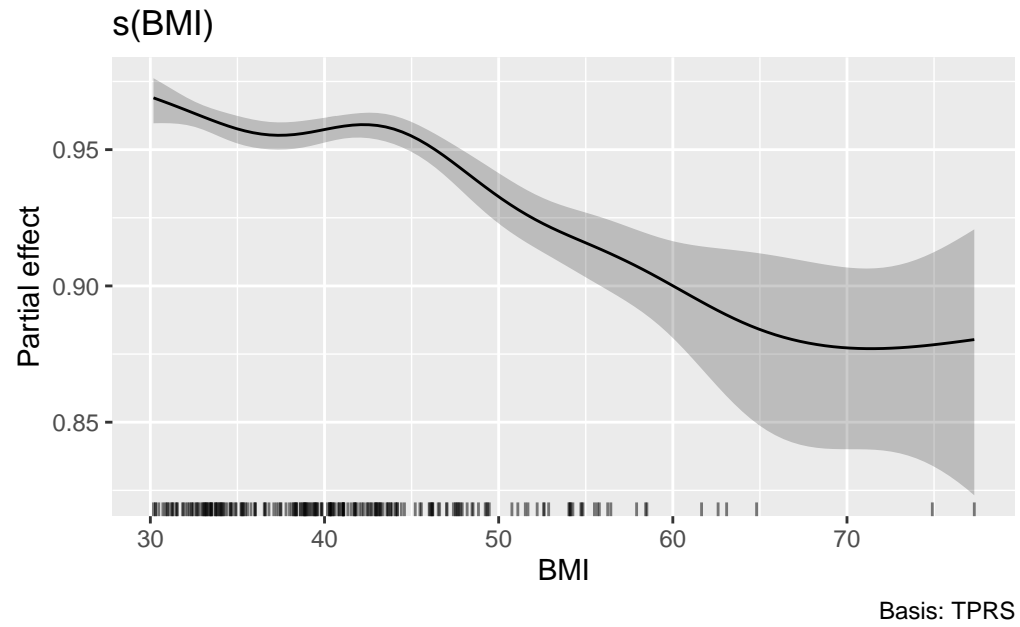
## Figure SpO2 models

Figure 2a: sBMI

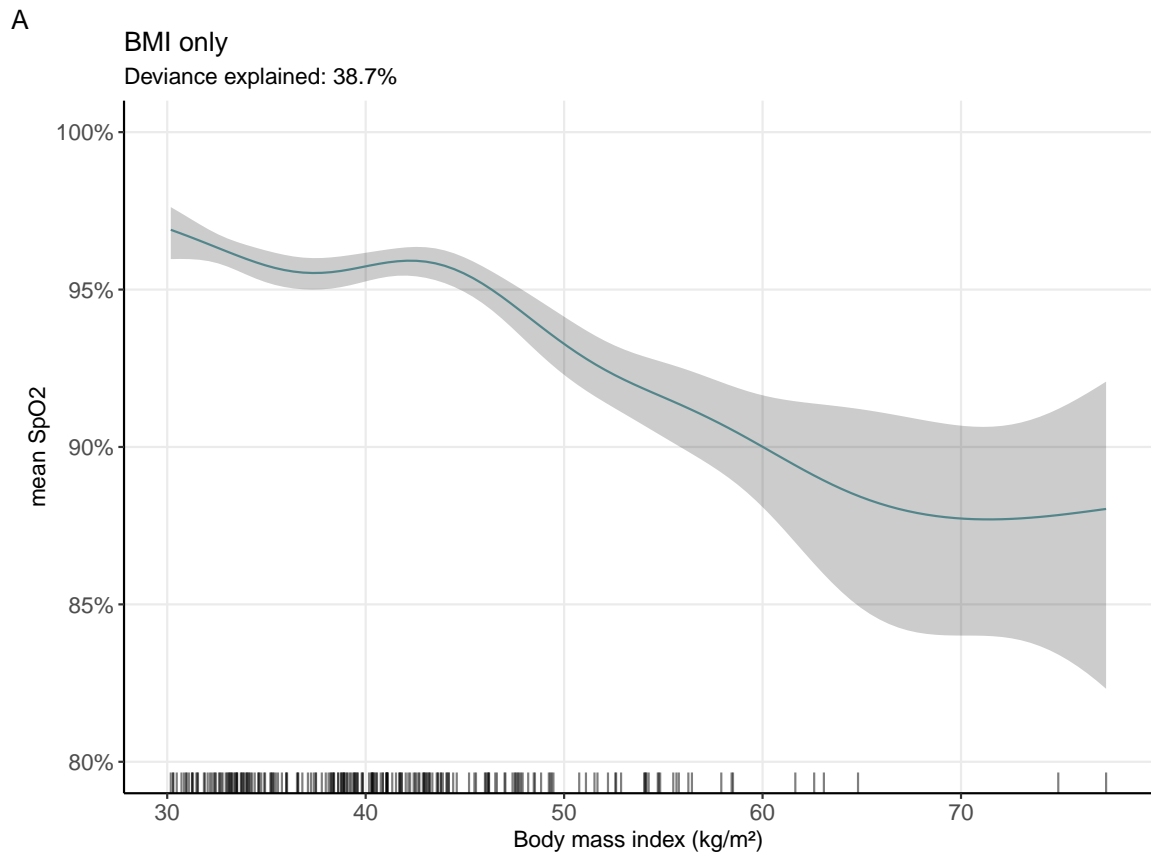
Assessment of residuals. This was done for all models.



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

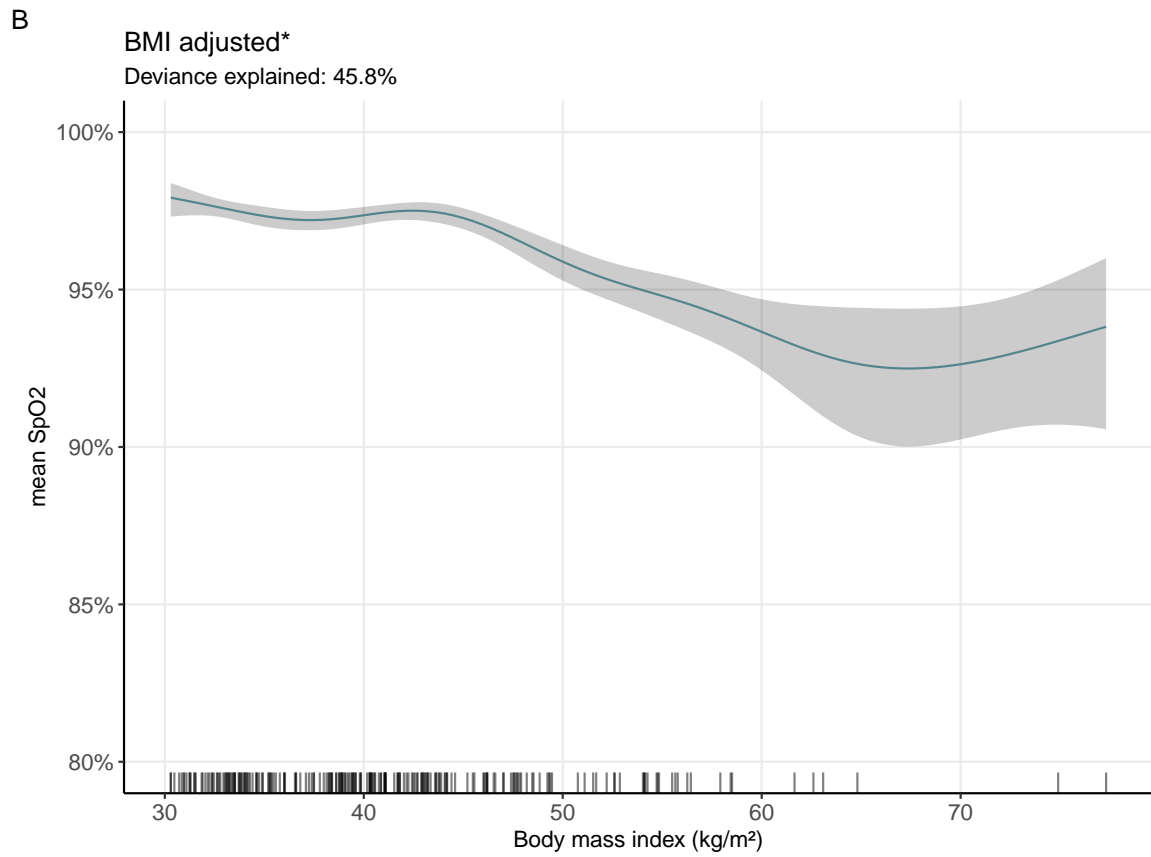


Partial effect on mean SpO2:



**Figure 2b: sBMI\_adjusted**

Partial effect on mean SpO2:

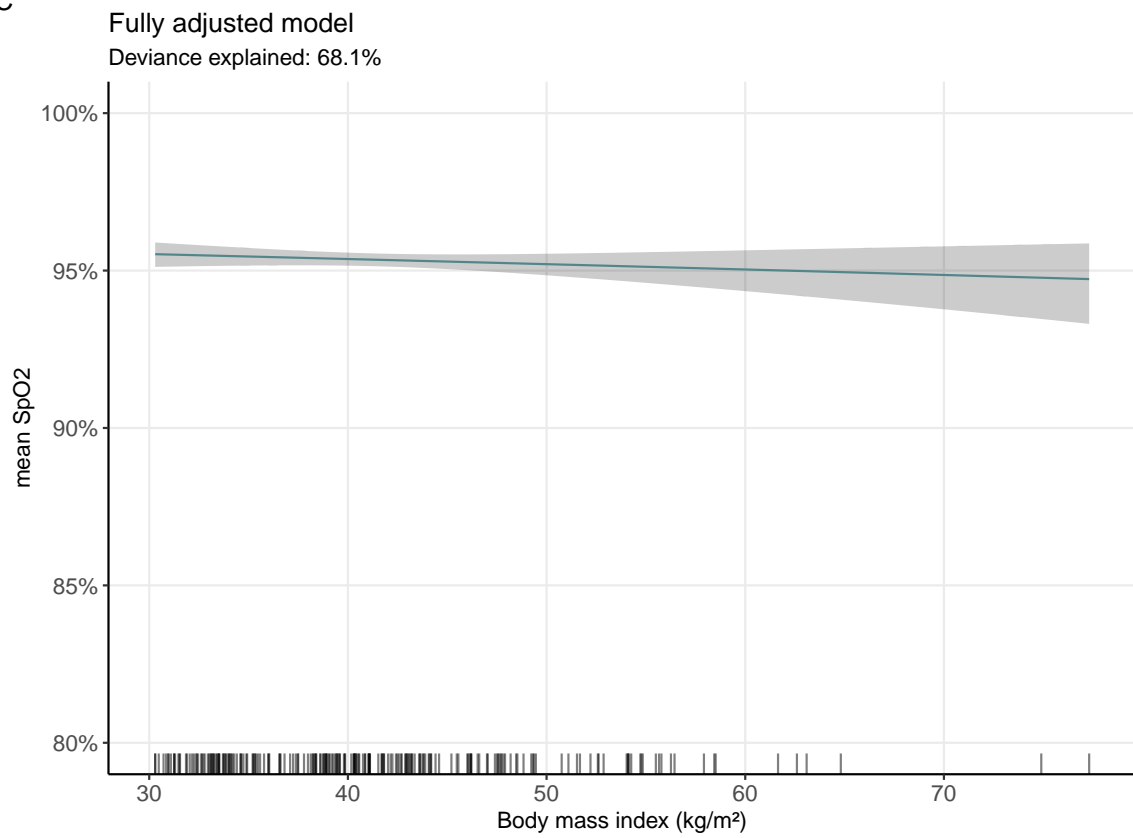


**Figure 2c: Fully adjusted BMI**

Partial effect on mean SpO2:



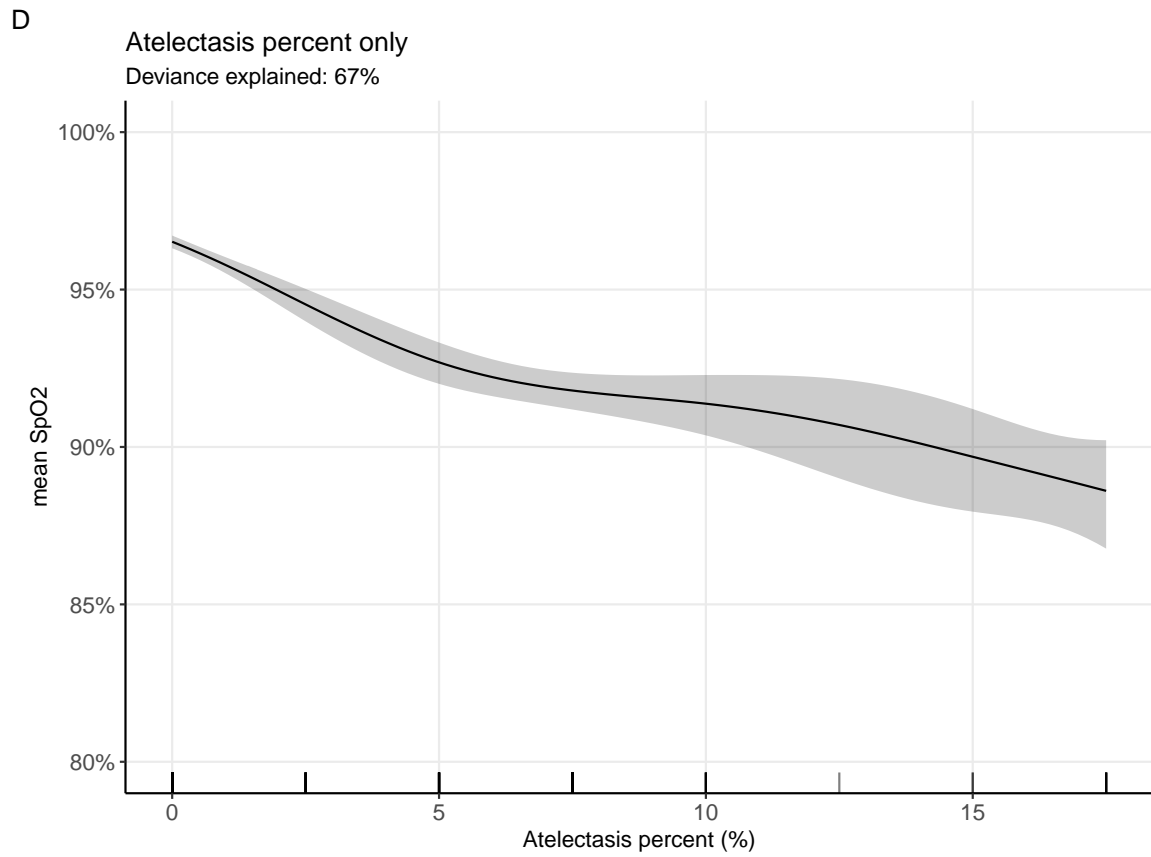
C



**Figure 2d: sAtelectasis percent**

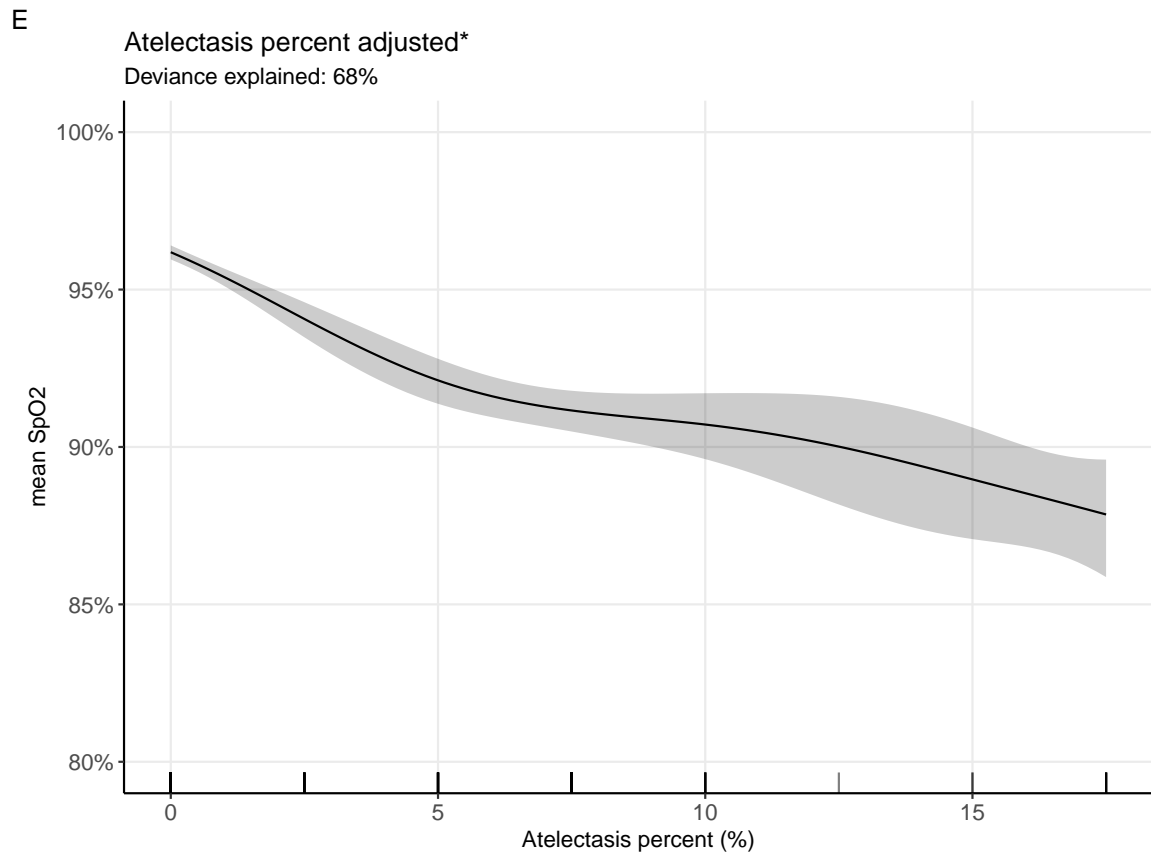
Check residuals:

Draw a personalized plot:



**Figure 2e: Atelectasis percent adjusted**

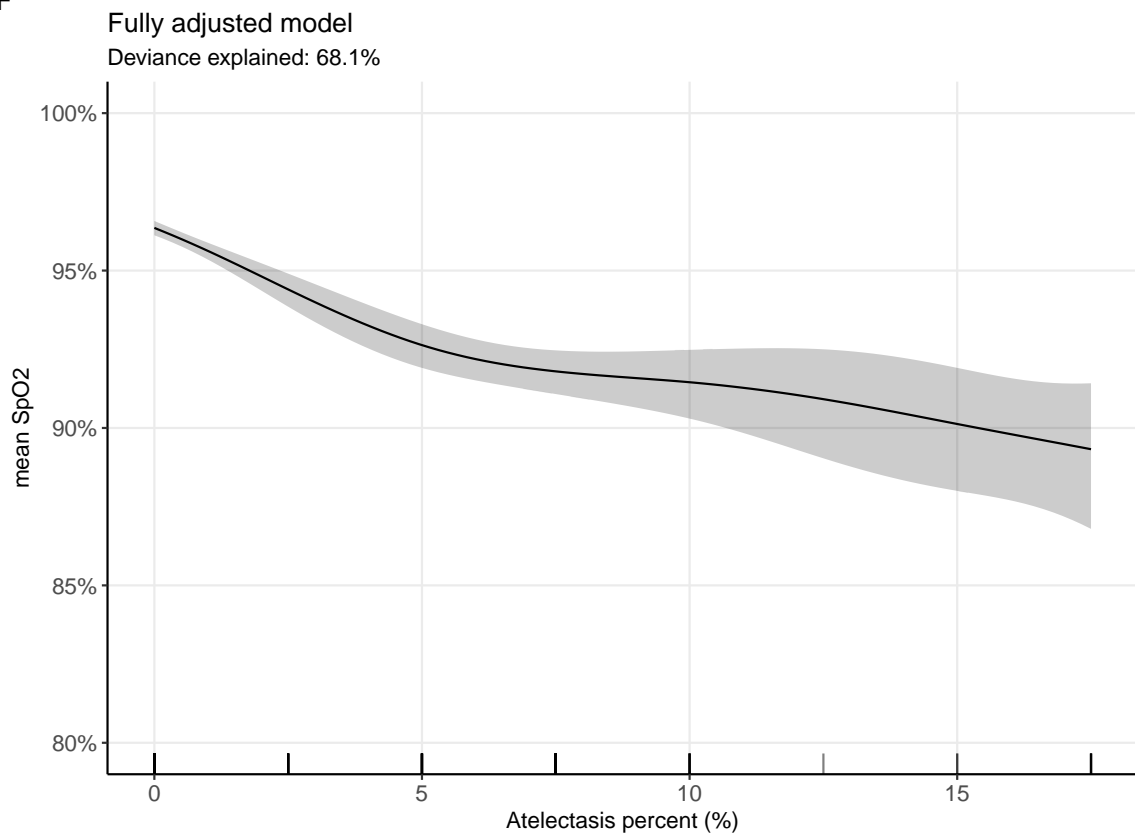
Partial effect on mean SpO2:



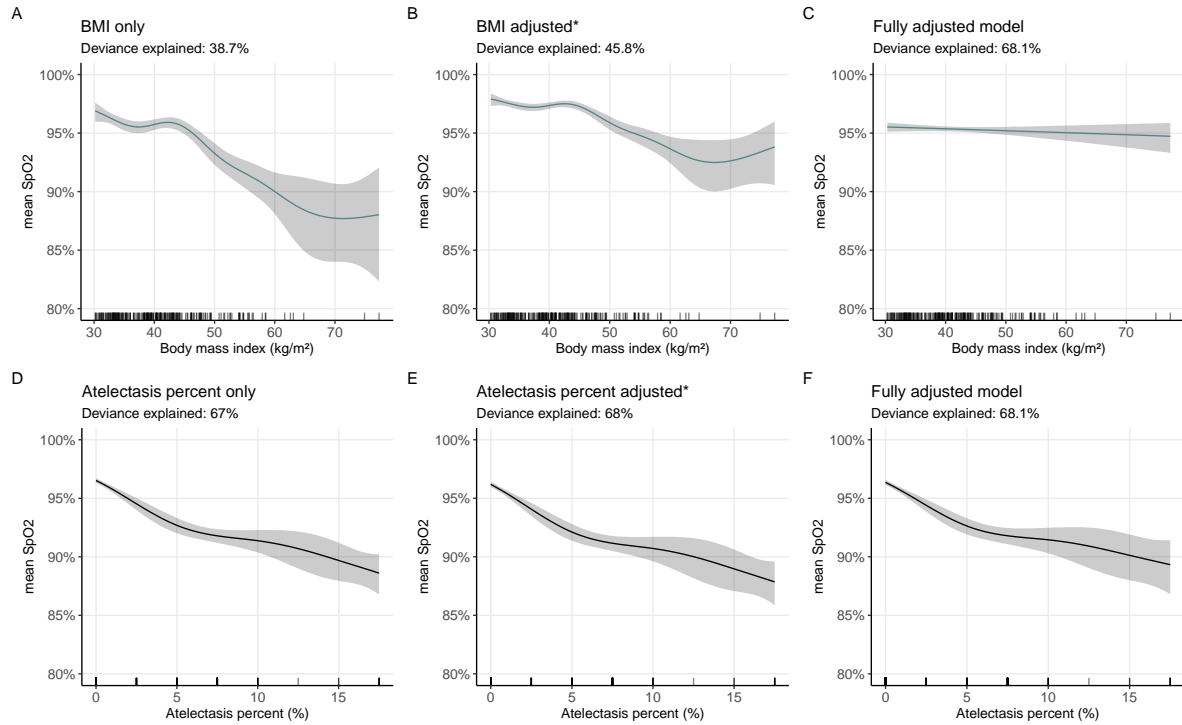
**Figure 2f: Fully adjusted Atelectasis Percent**

Partial effect on mean SpO2:

F



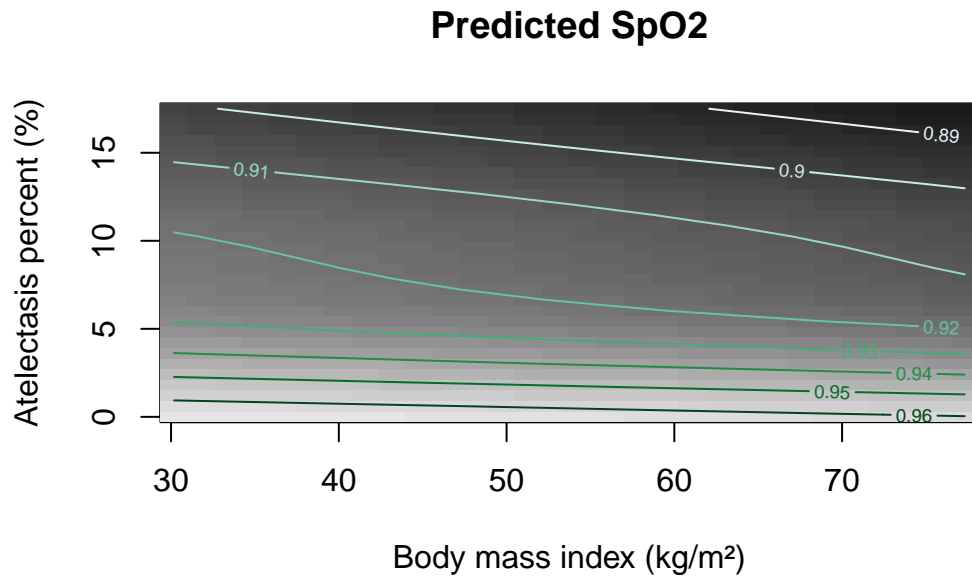
**Figure 2**



## Predictions SpO2

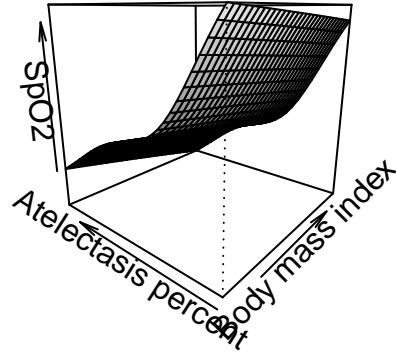
These are the predicted SpO2 values in the fully adjusted model (adjusted\_plus):

```
vis.gam(model_plus,
  view=c("BMI","atelectasis_percent"),
  color = "gray",
  type = "response",
  plot.type = "contour",
  contour.col = brewer.pal(9, "BuGn"),
  nlevels=10,
  ylab = "Atelectasis percent (%)",
  xlab = "Body mass index (kg/m2)",
  main = "Predicted SpO2"
)
```



This figure shows that this model is not able to predict SpO2 values above 96%. The range of predicted values of SpO2 that can be predicted with the data and model created in this study are within 88-96%. Lines correspond to a level of SpO2, so it can be seen that most of these are almost perpendicular to the y axis, meaning that most of the decrease in SpO2 is driven by increasing atelectasis percentage. Nonetheless, lines are not perfectly horizontal, which reflects that there is some residual effect of BMI on SpO2. Furthermore, this model shows that drops in SpO2 are more accentuated at the lower part of atelectasis percentage extension (93% to 96% mostly occur at atelectasis percentage lower than 5%). At SpO2 92% and lower, jumps are not as accentuated and there is a greater effect of increasing BMI as the lines tend to be more inclined. A 3D plot could perhaps allow to visualize these patterns if this is not clear enough from the 2D plot:

## Predicted SpO2



**Figure 3**

The 2D plot was recreated with the accompanying sourced script ***Figure3.R*** which also saves the 3D plot as FigureS4.

