

Preoperative Atelectasis

Part 5: Statistical Modelling of Atelectasis

Javier Mancilla Galindo

2023-11-28

Table of contents

Setup	1
DAG	3
Testing of conditional independencies in DAG:	3
Prevalence Ratio	4
Table 2	4
Ordinal Logistic Regression Model	5
Check proportional odds assumption for main variable of interest:	5
Univariate models for covariates:	8
Multivariable model	10

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.  
  dagitty, #Used in conjunction with https://www.dagitty.net/ to create
```

```

    #directed acyclic graph to inform statistical modelling.
lavaan, #Used to create correlation matrix to assess conditional independencies.
broom, #Used to exponentiate coefficients of regression models.
sandwich, #Used to calculate robust standard errors for prevalence ratios.
flextable, #Used to export tables.
rms, #Used to model ordinal outcome (atelectasis percent) and
    #test proportional odds assumptions.
VGAM, #Used to model partial proportional odds model.
gt #Used to present a summary of the results of regression 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] splines    stats4    stats      graphics  grDevices  utils      datasets
[8] methods    base

```

other attached packages:

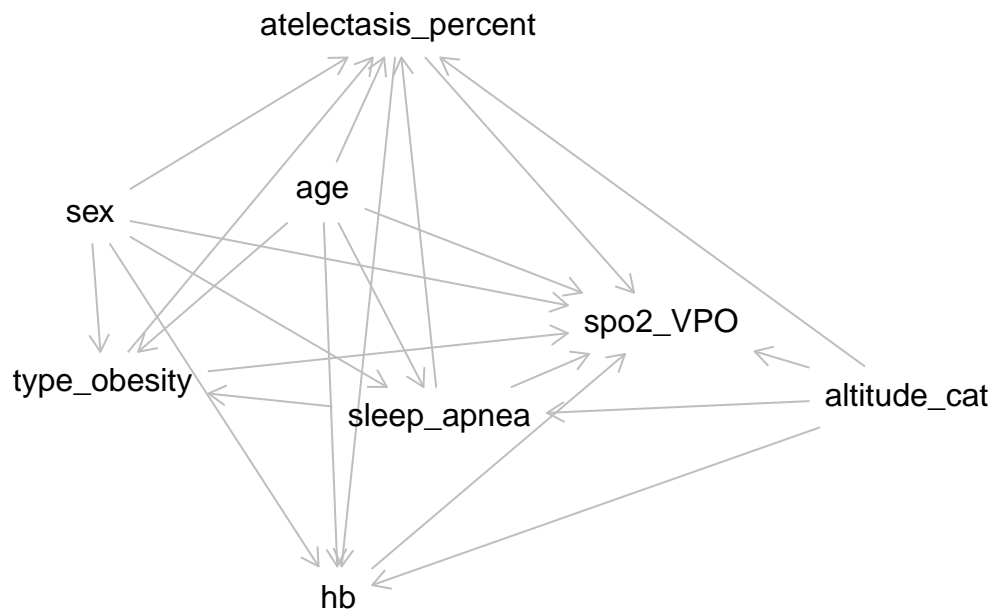
```

[1] gt_0.10.0          VGAM_1.1-9          rms_6.7-1           Hmisc_5.1-1
[5] flextable_0.9.4    sandwich_3.0-2      broom_1.0.5         lavaan_0.6-16
[9] dagitty_0.3-1      table1_1.4.3        RColorBrewer_1.1-3  lubridate_1.9.3
[13] forcats_1.0.0      stringr_1.5.1       dplyr_1.1.4         purrr_1.0.2
[17] readr_2.1.4        tidyr_1.3.0         tibble_3.2.1        ggplot2_3.4.4
[21] tidyverse_2.0.0    pacman_0.5.1

```

DAG

DAG generated in the [DAGitty website](#) and sourced from the accompanying script *DAG.R*



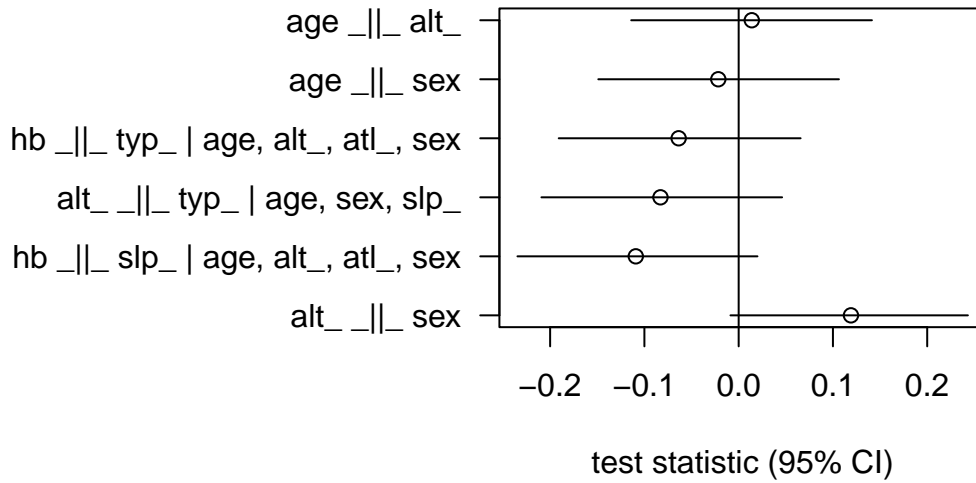
Testing of conditional independencies in DAG:

This procedure was performed as suggested in [this article](#).

Implied conditional independencies:

```
age _||_ alt_  
age _||_ sex  
alt_ _||_ sex  
alt_ _||_ typ_ | age, sex, slp_  
hb _||_ slp_ | age, alt_, atl_, sex  
hb _||_ typ_ | age, alt_, atl_, sex
```

Local tests results plot:



Conditional independence assumption OK as all confidence intervals contain 0.

The minimal set of adjustment for models is *age*, *sex*, and *sleep_apnea*.

Prevalence Ratio

This [paper](#) and accompanying code were used to calculate prevalence ratios.

A modified Poisson regression model with robust errors will be applied to obtain prevalence ratios.

Prevalence ratios were calculated with the accompanying sourced script *Prevalence_Ratio.R*

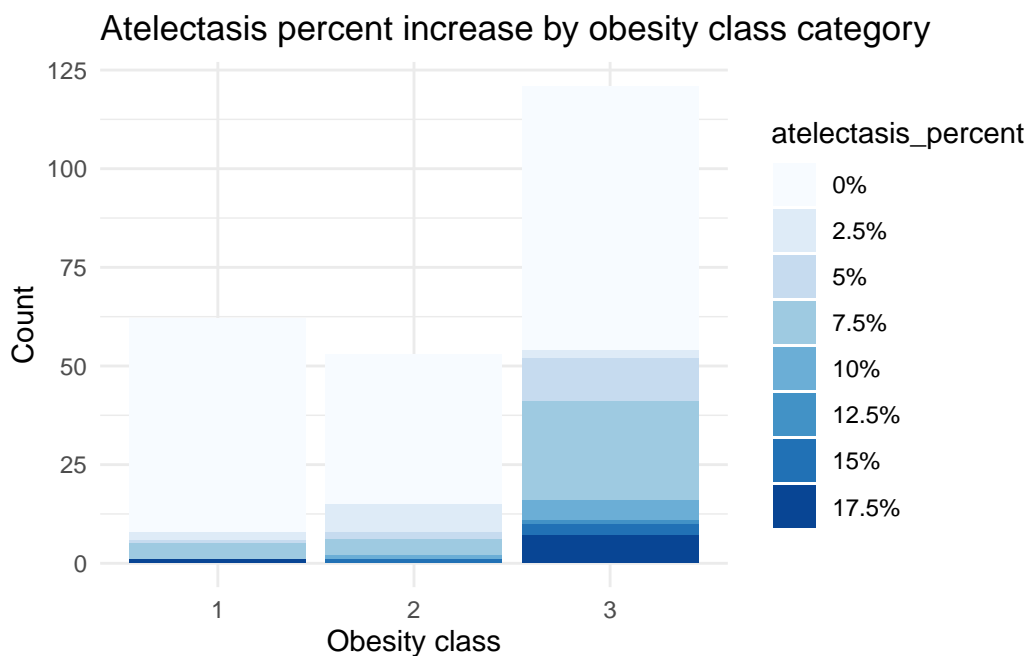
Table 2

Category	PR	SE	95%CI	aPR	aSE	a95%CI
Class 2 Obesity	2.19	0.40	1.01-4.76	2.13	0.37	1.03-4.41
Class 3 Obesity	3.46	0.35	1.76-6.8	3.10	0.33	1.63-5.9

Ordinal Logistic Regression Model

This modelling strategy was performed according to:

- Harrel, Frank. March, 2022. "Assessing the Proportional Odds Assumption and its Impact". Statistical Thinking. March 9, 2022.



Check proportional odds assumption for main variable of interest:

		Model Likelihood Ratio Test	Discrimination Indexes	Rank Discrim. Indexes
Obs	236	LR ² 24.41	R^2 0.109	0.308
Distinct Y	8	d.f. 2	$R^2_{2,236}$ 0.091	
$Y_{0.5}$	1	$\Pr(>^2)$ <0.0001	$R^2_{2,163.1}$ 0.128	
max log		Score ² 22.96	$ \Pr(Y$	
$L/$	3×10^{-7}	$\Pr(>^2)$ <0.0001	median) $^{-1/2}$ 0.170	

		S.E.	Wald Z	Pr(> Z)
y 2.5%	-1.9123	0.3778	-5.06	<0.0001
y 5%	-2.1553	0.3832	-5.62	<0.0001

		S.E.	Wald Z	Pr(> Z)
y 7.5%	-2.5045	0.3918	-6.39	<0.0001
y 10%	-3.7397	0.4386	-8.53	<0.0001
y 12.5%	-4.1576	0.4666	-8.91	<0.0001
y 15%	-4.2438	0.4736	-8.96	<0.0001
y 17.5%	-4.6737	0.5162	-9.05	<0.0001
type_obesity=2	0.8682	0.4805	1.81	0.0708
type_obesity=3	1.7702	0.4187	4.23	<0.0001

Odds ratio for type obesity in an univariable model:

Effects							
Response:							
atelectasis_percent							
	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
type_obesity --- 2:1	1	2		0.8682	0.4805	-0.07355	1.810
<i>Odds Ratio</i>	1	2		2.3830		0.92910	6.110
type_obesity --- 3:1	1	3		1.7700	0.4187	0.94970	2.591
<i>Odds Ratio</i>	1	3		5.8720		2.58500	13.340

Proportional odds assumption:

Wald			
Statistics for			
atelectasis_percent			
	²	d.f.	P
type_obesity	20.79	2	<0.0001
TOTAL	20.79	2	<0.0001

This shows that the proportional odds assumption is not met since $p < 0.05$ in the ANOVA test.

There are a couple of alternatives for modelling. One would be to fit a full multinomial model, although this would be expected to be unoptimal due to loss of statistical power, less parsimonious, and difficult interpretation compared to ordinal. A second approach would be to fit a partial proportional odds model allowing nominal effects for obesity class categories.

However, it is known that violations of the proportional odds assumption may not be as serious in some cases, as explained in the reference provided before. Thus, I will test how these 2 alternative modelling strategies would compare against a proportional odds model.

As a note, it is known that having few observations per category does not affect the results of ordinal regression, and that some categories may need to be combined to assess proportional odds assumption. [REF](#)

Thus, I will create atelectasis percent categories by collapsing non-integer atelectasis percentage categories (i.e., 2.5%, 7.5%) against the immediate lower category, resulting in 5% jumps (0-5%, 5-10%, 10-15%, and 15%) which meet the assumption of being equi-distant categories for ordinal regression:

0%	5%	10%	15%
170	47	7	12

Are subgroups better represented now?

	1	2	3
0%	56	45	69
5%	5	6	36
10%	0	1	6
15%	1	1	10

Some improvement.

Will now test the impact of not meeting the proportional odds assumption in a model adjusted for covariates:

Comparison of proportional odds (PO), partial proportional odds (PPO), and multinomial model:

	PO	PPO	Multinomial
Deviance	416.6654	412.1552	401.0035
d.f.	8	12	18
AIC	432.6654	436.1552	437.0035
p	5	9	15
LR χ^2	46.74975	51.25994	62.41161
LR - p	41.74975	42.25994	47.41161
LR χ^2 test for PO		4.510197	15.661861
d.f.		4	10
Pr(> χ^2)		0.3413401	0.1097298
MCS R2	0.1797057	0.1952335	0.2323765

MCS R2 adj	0.1621412	0.1639506	0.1820030
McFadden R2	0.1008809	0.1106134	0.1346775
McFadden R2 adj	0.07930200	0.07177138	0.06994077
Mean difference from PO		0.01698003	0.03155911

Lowest AIC is for the proportional odds (PO) model. Likewise, the McFadden adjusted R2 is the highest for the PO model. Thus, I will present the PO model despite proportional odds assumption not met as this is not causing serious problems and seems to be the best model according to the results shown and discussed.

Univariate models for covariates:

	Model Likelihood Ratio Test	Discrimination Indexes	Rank Discrim. Indexes
Obs 236	LR ² 26.05	R^2 0.116	0.377
Distinct Y 8	d.f. 1	$R^2_{1,236}$ 0.101	
$Y_{0.5}$ 1	Pr(> ²) <0.0001	$R^2_{1,163.1}$ 0.142	
max log	Score ² 32.43	Pr(Y	
L/ 8×10^{-7}		median)- ¹ / ₂ 0.224	
	Pr(> ²) <0.0001		

		S.E.	Wald Z	Pr(> Z)
y 2.5%	-0.9434	0.1518	-6.21	<0.0001
y 5%	-1.1940	0.1611	-7.41	<0.0001
y 7.5%	-1.5483	0.1768	-8.76	<0.0001
y 10%	-2.8408	0.2697	-10.53	<0.0001
y 12.5%	-3.2688	0.3145	-10.39	<0.0001
y 15%	-3.3563	0.3250	-10.33	<0.0001
y 17.5%	-3.7933	0.3856	-9.84	<0.0001
sleep_apnea=Yes	2.1919	0.4287	5.11	<0.0001

Effects

Response:

atelectasis_percent

	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
sleep_apnea --- Yes:No	1	2		2.192	0.4287	1.352	3.032
<i>Odds Ratio</i>	1	2		8.952		3.864	20.740

	Model Likelihood Ratio Test	Discrimination Indexes	Rank Discrim. Indexes
Obs 236	LR ² 2.08	R^2 0.010	0.096
Distinct Y 8	d.f. 1	$R^2_{1,236}$ 0.005	
$Y_{0.5}$ 1	Pr(> ²) 0.1490	$R^2_{1,163.1}$ 0.007	
max log	Score ² 2.24	Pr(Y	
L/ 2×10^{-7}		median)- $\frac{1}{2}$ 0.173	
	Pr(> ²) 0.1347		

		S.E.	Wald Z	Pr(> Z)
y 2.5%	-0.1503	0.4093	-0.37	0.7134
y 5%	-0.3731	0.4100	-0.91	0.3628
y 7.5%	-0.6938	0.4119	-1.68	0.0921
y 10%	-1.8759	0.4418	-4.25	<0.0001
y 12.5%	-2.2836	0.4676	-4.88	<0.0001
y 15%	-2.3679	0.4745	-4.99	<0.0001
y 17.5%	-2.7909	0.5164	-5.40	<0.0001
sex=Woman	-0.6380	0.4315	-1.48	0.1392

Effects

Response:

atelectasis_percent

	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
sex --- Woman:Man	1	2		-0.6380	0.4315	-1.4840	0.2076
<i>Odds Ratio</i>	1	2		0.5284		0.2268	1.2310

	Model Likelihood Ratio Test	Discrimination Indexes	Rank Discrim. Indexes
Obs 236	LR ² 0.63	R^2 0.003	0.049
Distinct Y 8	d.f. 1	$R^2_{1,236}$ 0.000	
$Y_{0.5}$ 1	Pr(> ²) 0.4273	$R^2_{1,163.1}$ 0.000	
max log	Score ² 0.63	Pr(Y	
L/ 0.003		median)- $\frac{1}{2}$ 0.173	
	Pr(> ²) 0.4274		

		S.E.	Wald Z	Pr(> Z)
y 2.5%	-0.2840	0.5705	-0.50	0.6186
y 5%	-0.5054	0.5715	-0.88	0.3765
y 7.5%	-0.8240	0.5733	-1.44	0.1506
y 10%	-1.9975	0.5986	-3.34	0.0008
y 12.5%	-2.4040	0.6188	-3.89	0.0001
y 15%	-2.4885	0.6239	-3.99	<0.0001
y 17.5%	-2.9121	0.6559	-4.44	<0.0001
age	-0.0110	0.0138	-0.79	0.4278

Effects

Response:

atelectasis_percent

	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
age	32.75	48.25	15.5	-0.1702	0.2146	-0.5908	0.2504
<i>Odds Ratio</i>	32.75	48.25	15.5	0.8435		0.5539	1.2850

Multivariable model

	Model Likelihood Ratio Test	Discrimination Indexes	Rank Discrim. Indexes
Obs 236	LR ² 47.02	R^2 0.201	0.421
Distinct Y 8	d.f. 5	$R^2_{5,236}$ 0.163	
$Y_{0.5}$ 1	Pr(> ²) <0.0001	$R^2_{5,163.1}$ 0.227	
max log	Score ² 50.97	Pr(Y	
$L/ 6 \times 10^{-5}$		median)- ^{1/2} 0.220	
	Pr(> ²) <0.0001		

		S.E.	Wald Z	Pr(> Z)
y 2.5%	-2.1985	0.9201	-2.39	0.0169
y 5%	-2.4694	0.9232	-2.67	0.0075
y 7.5%	-2.8511	0.9268	-3.08	0.0021
y 10%	-4.1901	0.9497	-4.41	<0.0001
y 12.5%	-4.6295	0.9666	-4.79	<0.0001
y 15%	-4.7201	0.9709	-4.86	<0.0001
y 17.5%	-5.1689	0.9948	-5.20	<0.0001
type_obesity=2	0.8102	0.4921	1.65	0.0996

		S.E.	Wald Z	Pr(> Z)
type_obesity=3	1.6660	0.4282	3.89	<0.0001
sleep_apnea=Yes	2.1999	0.4783	4.60	<0.0001
sex=Woman	0.3976	0.5132	0.77	0.4385
age	-0.0054	0.0149	-0.36	0.7160

Effects							
Response:							
atelectasis_percent							
	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
age	32.75	48.25	15.5	-0.08401	0.2309	-0.5366	0.3686
<i>Odds Ratio</i>	32.75	48.25	15.5	0.91940		0.5847	1.4460
type_obesity --- 2:1	1.00	2.00		0.81020	0.4921	-0.1542	1.7750
<i>Odds Ratio</i>	1.00	2.00		2.24800		0.8571	5.8980
type_obesity --- 3:1	1.00	3.00		1.66600	0.4282	0.8268	2.5050
<i>Odds Ratio</i>	1.00	3.00		5.29100		2.2860	12.2500
sleep_apnea --- Yes:No	1.00	2.00		2.20000	0.4783	1.2630	3.1370
<i>Odds Ratio</i>	1.00	2.00		9.02400		3.5340	23.0400
sex --- Woman:Man	1.00	2.00		0.39760	0.5132	-0.6082	1.4030
<i>Odds Ratio</i>	1.00	2.00		1.48800		0.5443	4.0690