Evans Cohort Analysis

Boaz

2025-02-06

Introduction

This is a secondary analysis of the Multiparameter Intelligent Monitoring in Intensive Care II database, consisting of 24,000 patients admitted to the Beth Israel Deaconess Medical Center ICU between 2001 and 2008. Patients requiring mechanical ventilation who did not require vasopressors or have a diagnosis of sepsis were identified (1776 patients), and the primary outcome was 28-day mortality. Our objective was to investigate the **association between the use of indwelling arterial catheters (IAC) and mortality in hemodynamically stable patients with respiratory failure**, using several methods like multivariate regression and stratification to adjust for confounding.

Loading useful packages

```
library(tidyverse) # for data handling
library(EpiStats)  # Provides functions for epidemiological analysis
library(knitr)     # to get well formatted tables
library(rstatix)   # Provides a simple and intuitive pipe-friendly framework
library(gtsummary) # Creates publication-ready summary tables for models
library(report)    # Facilitates the creation of reports for statistical models
library(finalfit)  # For final results tables and plots for publication
```

Load in data

Let's load in the dataset from the project directory using the read.csv() function and call it catheter. The dataset has 46 variables and 1776 selected participants who met the inclusion criteria.

```
catheter <- read.csv("full_cohort_data.csv")

glimpse(catheter)
#> Rows: 1,776
#> Columns: 46
#> $ aline_flg         <int> 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, ~
#> $ icu_los_day       <dbl> 7.63, 1.14, 2.86, 0.58, 1.75, 1.38, 7.06, 15.30, 3.~
#> $ hospital_los_day  <int> 13, 1, 5, 3, 5, 9, 27, 33, 4, 7, 2, 6, 11, 18, 14, ~
#> $ age               <dbl> 72.36841, 64.92076, 36.50000, 44.49191, 23.74217, 3~
#> $ gender_num        <int> 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, ~
#> $ weight_first      <dbl> 75.0, 55.0, 70.0, NA, 95.2, 72.0, 90.0, 69.7, 52.6,~
#> $ bmi               <dbl> 29.91279, 20.12131, 27.11827, NA, 28.46456, 23.9824~
#> $ sapsi_first       <int> 15, NA, 16, 21, 18, 14, 15, 16, 9, 13, 16, 18, 14, ~
#> $ sofa_first        <int> 9, 5, 5, 7, 7, 5, 6, 8, 5, 9, 5, 7, 12, 6, 6, 5, 7,~
#> $ service_unit      <chr> "SICU", "MICU", "MICU", "SICU", "SICU", "SICU", "SI~
#> $ service_num       <int> 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, ~
#> $ day_icu_intime    <chr> "Friday   ", "Saturday ", "Friday   ", "Saturday ",~
#> $ day_icu_intime_num <int> 6, 7, 6, 7, 7, 1, 7, 5, 6, 6, 5, 4, 2, 6, 5, 6, 4, ~
```

```
#> $ hour_icu_intime   <int> 6, 17, 3, 4, 7, 12, 22, 14, 21, 10, 20, 3, 12, 20, ~
#> $ hosp_exp_flg      <int> 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ~
#> $ icu_exp_flg       <int> 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ~
#> $ day_28_flg        <int> 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ~
#> $ mort_day_censored <dbl> 11.92, 731.00, 731.00, 0.00, 731.00, 731.00, 731.00~
#> $ censor_flg        <int> 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, ~
#> $ sepsis_flg        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
#> $ chf_flg           <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, ~
#> $ afib_flg          <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
#> $ renal_flg         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
#> $ liver_flg         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, ~
#> $ copd_flg          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
#> $ cad_flg           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
#> $ stroke_flg        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
#> $ mal_flg           <int> 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, ~
#> $ resp_flg          <int> 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, ~
#> $ map_1st           <dbl> 92.0000, 86.6667, 69.6667, 101.0000, 105.0000, 104.~
#> $ hr_1st            <int> 86, 85, 135, 125, 107, 90, 94, 105, 85, 114, 66, 92~
#> $ temp_1st          <dbl> 95.90, 97.60, 96.30, 100.10, 96.30, 98.40, 100.40, ~
#> $ spo2_1st          <int> 100, 100, 99, 100, 100, 100, 100, 100, 100, 93, 100~
#> $ abg_count         <int> 22, 1, 3, 4, 9, 0, 18, 40, 3, 11, 2, 4, 22, 27, 16,~
#> $ wbc_first         <dbl> 8.1, NA, 27.0, 7.1, 4.8, 12.1, 21.6, 19.9, 11.1, 7.~
#> $ hgb_first         <dbl> 14.1, NA, 13.1, 12.6, 10.7, 14.4, 13.4, 7.8, 11.4, ~
#> $ platelet_first    <int> 354, NA, 295, 262, 22, 182, 130, 20, 238, 137, 423,~
#> $ sodium_first      <int> 138, NA, 144, 139, 146, 145, 143, 140, 143, 143, 14~
#> $ potassium_first   <dbl> 4.6, NA, 3.9, 4.2, 3.4, 3.6, 3.8, 3.7, 4.0, 3.7, 3.~
#> $ tco2_first        <dbl> 15, NA, 17, 31, 19, 26, 32, 20, 25, 28, 32, 22, 26,~
#> $ chloride_first    <int> 109, NA, 101, 100, 110, 110, 104, 105, 107, 104, 11~
#> $ bun_first         <int> 41, NA, 16, 16, 10, 10, 17, 30, 15, 2, 15, 43, 12, ~
#> $ creatinine_first  <dbl> 1.6, NA, 0.8, 0.5, 1.0, 0.7, 1.3, 1.2, 0.7, 0.3, 0.~
#> $ po2_first         <int> 196, NA, 298, 146, 134, NA, 38, 57, 212, 284, 76, 1~
#> $ pco2_first        <int> 39, NA, 30, 23, 30, NA, 62, 28, 41, 33, 56, 32, 32,~
#> $ iv_day_1          <dbl> 2230.87500, 600.00000, 2086.80029, NA, 2358.24414, ~
```

Data handling

Let's select and rename variables of interest: the exposure (Use of IAC), the outcome (death) and potential confounders (age; sex; sofa score; service unit; coincident diseases like COPD and other respiratory diseases, congestive heart failure (CHF), Atrial fibrillation (AF), chronic renal disease, chronic liver diseases, coronary artery disease, stroke and malignancy; and laboratory tests such as wbc, hemoglobin, platelet, sodium, potassium, bicarbonate, chloride, urea, creatinine, po2 and pco2 .

lack of pneumonia, DNR at admission and Change in code status during ICU admission.

```
variables.of.int <- catheter %>%
  select(IAC = aline_flg,
         death = day_28_flg,
         age,
         sex = gender_num,
         sofa = sofa_first,
         service_unit,
         COPD = copd_flg,
         Other.resp.D = resp_flg,
         CHF = chf_flg,
```

```r
        AF = afib_flg,
        chr.renal.D = renal_flg,
        chr.liver.D = liver_flg,
        coronary.D = cad_flg,
        stroke = stroke_flg,
        malignancy = mal_flg,
        wbc = wbc_first,
        hemoglobin = hgb_first,
        platelet = platelet_first,
        sodium = sodium_first,
        potassium = potassium_first,
        bicarbonate = tco2_first,
        chloride = chloride_first,
        urea = bun_first,
        creat = creatinine_first,
        po2 = po2_first,
        pco2 = pco2_first)
```

```r
head(variables.of.int)
#>   IAC death      age sex sofa service_unit COPD Other.resp.D CHF AF chr.renal.D
#> 1   1     1 72.36841   1    9         SICU    0            0   0  0           0
#> 2   0     0 64.92076   0    5         MICU    0            0   0  0           0
#> 3   0     0 36.50000   0    5         MICU    0            0   0  0           0
#> 4   1     1 44.49191   0    7         SICU    0            0   0  0           0
#> 5   1     0 23.74217   1    7         SICU    0            0   0  0           0
#> 6   0     0 36.54657   1    5         SICU    0            0   0  0           0
#>   chr.liver.D coronary.D stroke malignancy  wbc hemoglobin platelet sodium
#> 1           0          0      0          1  8.1       14.1      354    138
#> 2           0          0      0          0   NA         NA       NA     NA
#> 3           0          0      0          0 27.0       13.1      295    144
#> 4           0          0      0          1  7.1       12.6      262    139
#> 5           0          0      0          0  4.8       10.7       22    146
#> 6           0          0      0          0 12.1       14.4      182    145
#>   potassium bicarbonate chloride urea creat po2 pco2
#> 1       4.6          15      109   41   1.6 196   39
#> 2        NA          NA       NA   NA    NA  NA   NA
#> 3       3.9          17      101   16   0.8 298   30
#> 4       4.2          31      100   16   0.5 146   23
#> 5       3.4          19      110   10   1.0 134   30
#> 6       3.6          26      110   10   0.7  NA   NA
```

Let's recode categorical variables to get summary statistics table

```r
variables.of.int.recoded <- variables.of.int %>%
  mutate(IAC = case_when(IAC == 1 ~ "yes" , IAC == 0 ~ "no" )) %>%
  mutate(death = case_when(death == 1 ~ "yes" , death == 0 ~ "no" )) %>%
  mutate(sex = case_when(sex == 1 ~ "male" , sex == 0 ~ "female" )) %>%
  mutate(COPD = case_when(COPD == 1 ~ "yes" , COPD == 0 ~ "no" )) %>%
  mutate(Other.resp.D = case_when(Other.resp.D == 1 ~ "yes" , Other.resp.D == 0 ~ "no" )) %>%
  mutate(CHF = case_when(CHF == 1 ~ "yes" , CHF == 0 ~ "no" )) %>%
  mutate(AF = case_when(AF == 1 ~ "yes" , AF == 0 ~ "no" )) %>%
  mutate(chr.renal.D = case_when(chr.renal.D == 1 ~ "yes" , chr.renal.D == 0 ~ "no" )) %>%
  mutate(chr.liver.D = case_when(chr.liver.D == 1 ~ "yes" , chr.liver.D == 0 ~ "no" )) %>%
  mutate(coronary.D = case_when(coronary.D == 1 ~ "yes" , coronary.D == 0 ~ "no" )) %>%
```

```
  mutate(stroke = case_when(stroke == 1 ~ "yes" , stroke == 0 ~ "no" )) %>%
  mutate(malignancy = case_when(malignancy == 1 ~ "yes" , malignancy == 0 ~ "no" ))

head(variables.of.int.recoded)
#>   IAC death      age    sex sofa service_unit COPD Other.resp.D CHF AF
#> 1 yes   yes 72.36841   male    9         SICU   no           no  no no
#> 2  no    no 64.92076 female    5         MICU   no           no  no no
#> 3  no    no 36.50000 female    5         MICU   no           no  no no
#> 4 yes   yes 44.49191 female    7         SICU   no           no  no no
#> 5 yes    no 23.74217   male    7         SICU   no           no  no no
#> 6  no    no 36.54657   male    5         SICU   no           no  no no
#>   chr.renal.D chr.liver.D coronary.D stroke malignancy  wbc hemoglobin platelet
#> 1          no          no         no     no        yes  8.1       14.1      354
#> 2          no          no         no     no         no   NA         NA       NA
#> 3          no          no         no     no         no 27.0       13.1      295
#> 4          no          no         no     no        yes  7.1       12.6      262
#> 5          no          no         no     no         no  4.8       10.7       22
#> 6          no          no         no     no         no 12.1       14.4      182
#>   sodium potassium bicarbonate chloride urea creat po2 pco2
#> 1    138       4.6          15      109   41   1.6 196   39
#> 2     NA        NA          NA       NA   NA    NA  NA   NA
#> 3    144       3.9          17      101   16   0.8 298   30
#> 4    139       4.2          31      100   16   0.5 146   23
#> 5    146       3.4          19      110   10   1.0 134   30
#> 6    145       3.6          26      110   10   0.7  NA   NA
```

Let's get summary statistics with the tbl_summary() function from the "gtsummary" package. This function is handy when dealing with both numerical and categorical data: - Count and Percentage for categorical data - Median and Quartiles for numeric data

```
tbl_summary(variables.of.int.recoded,
            missing = "no")
```

Unmatched Cohorts Analysis

1. Crude estimate of the association between IAC and hospital mortality

Let's compute the crude estimates first using the CS() function from the EpiStats package. This function provides us with two tables: - A 2x2 contingency table illustrates the exposure and outcome of interest. - A table presents all possible estimates of the association.

```
result.1 <- CS(variables.of.int, death, IAC, exact = TRUE, full = FALSE)

kable(result.1$df1, align = result.1$df1.align)
```

	Cases	Non Cases	Total	Risk
Exposed	170	814	984	0.17
Unexposed	113	679	792	0.14
Total	283	1493	1776	0.16

```
kable(result.1$df2, align = result.1$df2.align)
```

Characteristic	N = 1,776[1]
IAC	984 (55%)
death	283 (16%)
age	54 (38, 73)
sex	
female	750 (42%)
male	1,025 (58%)
sofa	6 (4, 7)
service_unit	
FICU	62 (3.5%)
MICU	732 (41%)
SICU	982 (55%)
COPD	157 (8.8%)
Other.resp.D	565 (32%)
CHF	213 (12%)
AF	207 (12%)
chr.renal.D	60 (3.4%)
chr.liver.D	99 (5.6%)
coronary.D	123 (6.9%)
stroke	222 (13%)
malignancy	256 (14%)
wbc	11.3 (8.2, 15.0)
hemoglobin	12.70 (11.10, 14.15)
platelet	239 (182, 297)
sodium	140 (137, 142)
potassium	4.00 (3.60, 4.40)
bicarbonate	24.0 (22.0, 27.0)
chloride	104 (101, 107)
urea	15 (11, 22)
creat	0.90 (0.70, 1.10)
po2	195 (108, 323)
pco2	41 (36, 47)

[1]n (%); Median (Q1, Q3)

	Point estimate	95%CI.ll	95%CI.ul
Risk difference	0.03	-0.00	0.06
Risk ratio	1.21	0.97	1.51
Attr. frac. ex.	0.17	-0.03	0.34
Attr. frac. pop	0.10	NA	NA
chi2(1)	2.97	NA	NA
Pr>chi2	0.085	NA	NA
Fisher p.value	0.090	NA	NA

Patients with an indwelling arterial catheter faced a **21%** (95%CI of -3% to 51%) to higher mortality risk

compared to those without it. Additionally, **10%** of deaths can be attributed to the use of indwelling arterial catheters in ICU patients who require mechanical ventilation and are hemodynamically stable. However, this association is not significant (chi2 p = 0.085)

When working with multiple exposures, the CSTable() function can be quite useful providing with a single table.

Let's say sex, COPD, Other.resp.D, CHF, chr.renal.D and AF are exposures too.

```
result.2 <- CSTable(variables.of.int, "death",
       exposure = c("IAC", "sex", "COPD", "Other.resp.D", "CHF", "chr.renal.D", "AF"),
       exact = TRUE,
       sort = "rr",
       full = TRUE)
kable(result.2$df, align = result.2$align)
```

	Tot.Exp	Cases.Exp	AR.Exp%	Tot.Unexp	Cases.Unexp	AR.Unexp%	RR	CI.ll	CI.ul	p(Fisher)
AF	207	86	41.55	1569	197	12.56	3.31	2.69	4.07	0.000
CHF	213	68	31.92	1563	215	13.76	2.32	1.84	2.93	0.000
Other.resp.D	565	128	22.65	1211	155	12.80	1.77	1.43	2.19	0.000
COPD	157	41	26.11	1619	242	14.95	1.75	1.31	2.33	0.001
chr.renal.D	60	14	23.33	1716	269	15.68	1.49	0.93	2.39	0.148
IAC	984	170	17.28	792	113	14.27	1.21	0.97	1.51	0.090
sex	1025	139	13.56	750	144	19.20	0.71	0.57	0.87	0.002

2. Adjusting for confounding variables using stratification

Next, we will stratify the cohort by service unit type (MICU, SICU, FICU) and compute the estimates within each stratum using the function CSInter().

```
result.3 <- CSInter(variables.of.int, "death", "IAC", by = "service_unit", full = TRUE)

kable(result.3$df1, align = result.3$align)
```

CSInter death - IAC by(service_unit)	Total	Cases	Risk %	P.est.	Stats	95%CI.ll	95%CI.ul
service_unit = SICU	982	NA	NA	Risk difference	0.02	-0.02	0.07
Exposed	694	113	16.28	Risk Ratio	1.17	0.84	1.64
Unexposed	288	40	13.89	Attrib.risk.exp	0.15	-0.19	0.39
	NA	NA	NA	Attrib.risk.pop	0.11	NA	NA
service_unit = MICU	732	NA	NA	Risk difference	0.07	0.01	0.13
Exposed	252	55	21.83	Risk Ratio	1.46	1.06	2.00
Unexposed	480	72	15.00	Attrib.risk.exp	0.31	0.06	0.50
	NA	NA	NA	Attrib.risk.pop	0.14	NA	NA
service_unit = FICU	62	NA	NA	Risk difference	0.01	-0.10	0.12
Exposed	38	2	5.26	Risk Ratio	1.26	0.12	13.19
Unexposed	24	1	4.17	Attrib.risk.exp	0.21	-7.26	0.92
	NA	NA	NA	Attrib.risk.pop	0.14	NA	NA
Missing / Missing %	0	0.0%	NA	NA	NA	NA	NA

```
kable(result.3$df2, align = result.3$align)
```

Point Estimate	Chi2	p.value	Stats	95%CI.ll	95%CI.ul
Woolf test of homogeneity	0.85	0.654	NA	NA	NA
Crude RR for IAC	NA	NA	1.21	0.97	1.51
MH RR IAC adjusted for service_unit	NA	NA	1.30	1.04	1.64
Adjusted/crude relative change	NA	NA	7.68	NA	NA

The association between IAC and mortality exists in both strata appears to be higher in the MICU, with a risk ratio of 1.46. Furthermore, the *Mantel Heaenszel risk ratio* was **1.30** (1.04-1.64) meaning that the effect (1.21) was underestimated by the service unit type.

3. Multivariable model (Logistic regression) to adjust for confunding variables

Let's build our model by including all variables of interest in the data frame using the `glm()` function. We will specify the family as "binomial" since we are performing logistic regression; our response variable consists of binary data, where death is coded as 1 for yes and 0 for no.

```
model.1 <- glm(death ~ .,
               na.omit(variables.of.int),
                family = "binomial")
summary(model.1)
#>
#> Call:
#> glm(formula = death ~ ., family = "binomial", data = na.omit(variables.of.int))
#>
#> Coefficients:
#>                    Estimate Std. Error z value Pr(>|z|)
#> (Intercept)      -5.348e+00  2.712e+00  -1.972  0.04865 *
#> IAC               1.416e-02  2.038e-01   0.069  0.94460
#> age               5.451e-02  6.399e-03   8.519  < 2e-16 ***
#> sex               2.249e-01  1.780e-01   1.264  0.20631
#> sofa              1.803e-01  4.416e-02   4.084 4.43e-05 ***
#> service_unitMICU  1.082e+00  6.866e-01   1.575  0.11518
#> service_unitSICU  9.509e-01  6.777e-01   1.403  0.16060
#> COPD              3.046e-02  2.530e-01   0.120  0.90416
#> Other.resp.D      6.063e-01  1.962e-01   3.091  0.00200 **
#> CHF               2.978e-01  2.367e-01   1.258  0.20848
#> AF                4.014e-01  2.127e-01   1.887  0.05917 .
#> chr.renal.D      -7.242e-01  4.134e-01  -1.752  0.07981 .
#> chr.liver.D       2.841e-01  3.625e-01   0.784  0.43326
#> coronary.D       -1.503e-01  2.837e-01  -0.530  0.59624
#> stroke            2.124e+00  2.251e-01   9.438  < 2e-16 ***
#> malignancy        4.113e-01  2.184e-01   1.884  0.05960 .
#> wbc               2.669e-02  1.194e-02   2.235  0.02543 *
#> hemoglobin       -1.270e-01  4.542e-02  -2.796  0.00517 **
#> platelet          3.118e-04  8.312e-04   0.375  0.70755
#> sodium            5.269e-02  2.842e-02   1.854  0.06375 .
#> potassium        -7.686e-02  1.093e-01  -0.703  0.48190
#> bicarbonate      -4.391e-02  2.140e-02  -2.052  0.04021 *
#> chloride         -7.494e-02  2.437e-02  -3.076  0.00210 **
#> urea              1.531e-02  6.714e-03   2.280  0.02259 *
#> creat            -2.309e-01  1.276e-01  -1.809  0.07044 .
#> po2              -4.928e-04  6.556e-04  -0.752  0.45221
#> pco2              9.618e-05  6.594e-03   0.015  0.98836
```

```
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
#>     Null deviance: 1366.92  on 1580  degrees of freedom
#> Residual deviance:  935.05  on 1554  degrees of freedom
#> AIC: 989.05
#>
#> Number of Fisher Scoring iterations: 6
```

To obtain the final model, we will be following the **Mixed Stepwise Selection** approach based the AIC criterion using the `step()` function. In fact, Stepwise selection is a method for fitting regression models that involves the iterative selection of independent variables based on the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). It can be achieved through: - **Forward selection**: starts with no variables and adds each new variable incrementally, testing for statistical significance.
- **Backward selection**: begins with a full model and then removes the least statistically significant variables one at a time. - **Mixed selection**: combines both forward and backward selection to determine the best model.

```
# Select best model according to AIC using mixed selection
model.1.best <- step(model.1,
                     direction = "both",
                     trace = FALSE)

summary(model.1.best)
#>
#> Call:
#> glm(formula = death ~ age + sofa + Other.resp.D + AF + chr.renal.D +
#>     stroke + malignancy + wbc + hemoglobin + sodium + bicarbonate +
#>     chloride + urea + creat, family = "binomial", data = na.omit(variables.of.int))
#>
#> Coefficients:
#>               Estimate Std. Error z value Pr(>|z|)
#> (Intercept)  -4.866279   2.451014  -1.985  0.04710 *
#> age           0.052962   0.006077   8.716  < 2e-16 ***
#> sofa          0.188385   0.038932   4.839 1.31e-06 ***
#> Other.resp.D  0.708084   0.175419   4.037 5.42e-05 ***
#> AF            0.431849   0.208613   2.070  0.03844 *
#> chr.renal.D  -0.645694   0.394262  -1.638  0.10148
#> stroke        2.036510   0.212757   9.572  < 2e-16 ***
#> malignancy    0.418905   0.215164   1.947  0.05155 .
#> wbc           0.023410   0.011229   2.085  0.03709 *
#> hemoglobin   -0.131784   0.043028  -3.063  0.00219 **
#> sodium        0.057571   0.027597   2.086  0.03697 *
#> bicarbonate  -0.039559   0.018806  -2.104  0.03542 *
#> chloride     -0.078155   0.023825  -3.280  0.00104 **
#> urea          0.016133   0.006468   2.494  0.01262 *
#> creat        -0.233303   0.122958  -1.897  0.05777 .
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
```

Characteristic	OR[1]	95% CI[1]	p-value
age	1.05	1.04, 1.07	<0.001
sofa	1.21	1.12, 1.30	<0.001
Other.resp.D	2.03	1.44, 2.87	<0.001
AF	1.54	1.02, 2.31	0.038
chr.renal.D	0.52	0.23, 1.11	0.10
stroke	7.66	5.07, 11.7	<0.001
malignancy	1.52	0.99, 2.31	0.052
wbc	1.02	1.00, 1.05	0.037
hemoglobin	0.88	0.80, 0.95	0.002
sodium	1.06	1.00, 1.12	0.037
bicarbonate	0.96	0.93, 1.00	0.035
chloride	0.92	0.88, 0.97	0.001
urea	1.02	1.00, 1.03	0.013
creat	0.79	0.61, 0.98	0.058

[1]OR = Odds Ratio, CI = Confidence Interval

```
#>    Null deviance: 1366.92  on 1580  degrees of freedom
#> Residual deviance:  943.81  on 1566  degrees of freedom
#> AIC: 973.81
#>
#> Number of Fisher Scoring iterations: 6
```

As it can be seen, the output is a bit messy. Therefore, we will be using the `tbl_regression()` from the "gtsummary" package to get a tidy table.

```
tbl_regression(model.1.best, exponentiate = TRUE)
```

Let's now get an automatic report of the model using the "report" package.

```
report(model.1.best)
```

We fitted a logistic model (estimated using ML) to predict death with age, sofa, Other.resp.D, AF, chr.renal.D, stroke, malignancy, wbc, hemoglobin, sodium, bicarbonate, chloride, urea and creat (formula: death ~ age + sofa + Other.resp.D + AF + chr.renal.D + stroke + malignancy + wbc + hemoglobin + sodium + bicarbonate + chloride + urea + creat). The model's explanatory power is substantial (Tjur's R2 = 0.30). The model's intercept, corresponding to age = 0, sofa = 0, Other.resp.D = 0, AF = 0, chr.renal.D = 0, stroke = 0, malignancy = 0, wbc = 0, hemoglobin = 0, sodium = 0, bicarbonate = 0, chloride = 0, urea = 0 and creat = 0, is at -4.87 (95% CI [-9.76, -0.14], p = 0.047). Within this model:

- The effect of age is statistically significant and positive (beta = 0.05, 95% CI [0.04, 0.07], p < .001; Std. beta = 1.10, 95% CI [0.86, 1.35])
- The effect of sofa is statistically significant and positive (beta = 0.19, 95% CI [0.11, 0.27], p < .001; Std. beta = 0.43, 95% CI [0.26, 0.61])
- The effect of Other resp D is statistically significant and positive (beta = 0.71, 95% CI [0.37, 1.05], p < .001; Std. beta = 0.33, 95% CI [0.17, 0.50])
- The effect of AF is statistically significant and positive (beta = 0.43, 95% CI [0.02, 0.84], p = 0.038; Std. beta = 0.14, 95% CI [6.67e-03, 0.27])
- The effect of chr renal D is statistically non-significant and negative (beta = -0.65, 95% CI [-1.45, 0.11], p = 0.101; Std. beta = -0.12, 95% CI [-0.26, 0.02])

- The effect of stroke is statistically significant and positive (beta = 2.04, 95% CI [1.62, 2.46], p < .001; Std. beta = 0.66, 95% CI [0.53, 0.80])
- The effect of malignancy is statistically non-significant and positive (beta = 0.42, 95% CI [-8.91e-03, 0.84], p = 0.052; Std. beta = 0.15, 95% CI [-3.15e-03, 0.30])
- The effect of wbc is statistically significant and positive (beta = 0.02, 95% CI [7.74e-04, 0.05], p = 0.037; Std. beta = 0.16, 95% CI [5.24e-03, 0.31])
- The effect of hemoglobin is statistically significant and negative (beta = -0.13, 95% CI [-0.22, -0.05], p = 0.002; Std. beta = -0.29, 95% CI [-0.47, -0.10])
- The effect of sodium is statistically significant and positive (beta = 0.06, 95% CI [3.43e-03, 0.11], p = 0.037; Std. beta = 0.27, 95% CI [0.02, 0.53])
- The effect of bicarbonate is statistically significant and negative (beta = -0.04, 95% CI [-0.08, -2.38e-03], p = 0.035; Std. beta = -0.20, 95% CI [-0.39, -0.01])
- The effect of chloride is statistically significant and negative (beta = -0.08, 95% CI [-0.12, -0.03], p = 0.001; Std. beta = -0.45, 95% CI [-0.73, -0.18])
- The effect of urea is statistically significant and positive (beta = 0.02, 95% CI [3.57e-03, 0.03], p = 0.013; Std. beta = 0.23, 95% CI [0.05, 0.42])
- The effect of creat is statistically non-significant and negative (beta = -0.23, 95% CI [-0.50, -0.02], p = 0.058; Std. beta = -0.24, 95% CI [-0.52, -0.02])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald z-distribution approximation.

The final model identified 11 predictor variables that significantly impacted hospital mortality: **age, sofa, other.resp.D, AF, stroke, wbc, hemoglobin, sodium, bicarbonate, chloride, and urea**. The presence of stroke increased the odds of death by approximately 7.66 times (95% CI: 5.07 to 11.7), while respiratory diseases (excluding COPD) increased the odds by about 2.03 times (95% CI: 1.44 to 2.87).

4. Summary table

I would like to create a summary table combining : - all univariable logistic regressions that are possible with the variables available in our dataframe, - a multivariable logistic regression that includes all variables available in our dataframe, and - a multivariable logistic regression that includes only some of variables available in our dataframe.

We start with all possible univariable logistic regressions

```
# Set response and predictor variables
response <- "death"
predictor <- c("IAC", "age", "sex", "sofa", "service_unit", "COPD", "Other.resp.D",
               "CHF", "AF", "chr.renal.D", "chr.liver.D", "coronary.D", "stroke",
               "malignancy", "wbc", "hemoglobin", "platelet", "sodium", "potassium",
               "bicarbonate", "chloride", "urea", "creat", "po2", "pco2")

# Save results
glmuni <- variables.of.int %>%
  glmuni(response, predictor) %>%
  fit2df(explanatory_name = "Variables",
         estimate_name = "Crude OR",
         estimate_suffix = " (95% CI, p value)")

kable(glmuni)
```

Variables	Crude OR (95% CI, p value)
IAC	1.25 (0.97-1.63, p=0.085)
age	1.07 (1.06-1.08, p<0.001)
sex	0.66 (0.51-0.85, p=0.001)

Variables	Crude OR (95% CI, p value)
sofa	1.18 (1.12-1.24, p<0.001)
service_unitMICU	4.13 (1.50-17.10, p=0.018)
service_unitSICU	3.63 (1.32-15.00, p=0.031)
COPD	2.01 (1.36-2.92, p<0.001)
Other.resp.D	2.00 (1.54-2.59, p<0.001)
CHF	2.94 (2.12-4.04, p<0.001)
AF	4.95 (3.61-6.77, p<0.001)
chr.renal.D	1.64 (0.86-2.94, p=0.115)
chr.liver.D	1.27 (0.74-2.09, p=0.363)
coronary.D	2.06 (1.33-3.10, p=0.001)
stroke	8.87 (6.52-12.08, p<0.001)
malignancy	1.64 (1.18-2.26, p=0.003)
wbc	1.01 (0.99-1.03, p=0.298)
hemoglobin	0.86 (0.81-0.91, p<0.001)
platelet	1.00 (1.00-1.00, p=0.464)
sodium	0.98 (0.95-1.00, p=0.075)
potassium	1.14 (0.97-1.32, p=0.095)
bicarbonate	1.03 (1.00-1.06, p=0.024)
chloride	0.96 (0.94-0.98, p<0.001)
urea	1.03 (1.02-1.04, p<0.001)
creat	1.09 (0.99-1.20, p=0.074)
po2	1.00 (1.00-1.00, p=0.019)
pco2	1.00 (0.99-1.01, p=0.641)

Here is the multivariable logistic regression with all variables

```
glmmulti_full <- variables.of.int %>%
  glmmulti(response, predictor) %>%
  fit2df(explanatory_name = "Variables",
         estimate_name = "Adjusted OR - full model")

kable(glmmulti_full)
```

Variables	Adjusted OR - full model
IAC	1.01 (0.68-1.51, p=0.945)
age	1.06 (1.04-1.07, p<0.001)
sex	1.25 (0.88-1.78, p=0.206)
sofa	1.20 (1.10-1.31, p<0.001)
service_unitMICU	2.95 (0.88-13.93, p=0.115)
service_unitSICU	2.59 (0.79-12.07, p=0.161)
COPD	1.03 (0.62-1.68, p=0.904)
Other.resp.D	1.83 (1.25-2.70, p=0.002)
CHF	1.35 (0.84-2.14, p=0.208)
AF	1.49 (0.98-2.26, p=0.059)
chr.renal.D	0.48 (0.21-1.07, p=0.080)
chr.liver.D	1.33 (0.64-2.66, p=0.433)
coronary.D	0.86 (0.49-1.48, p=0.596)
stroke	8.37 (5.41-13.08, p<0.001)
malignancy	1.51 (0.98-2.30, p=0.060)
wbc	1.03 (1.00-1.05, p=0.025)
hemoglobin	0.88 (0.81-0.96, p=0.005)

Variables	Adjusted OR - full model
platelet	1.00 (1.00-1.00, p=0.708)
sodium	1.05 (1.00-1.11, p=0.064)
potassium	0.93 (0.75-1.15, p=0.482)
bicarbonate	0.96 (0.92-1.00, p=0.040)
chloride	0.93 (0.88-0.97, p=0.002)
urea	1.02 (1.00-1.03, p=0.023)
creat	0.79 (0.60-0.99, p=0.070)
po2	1.00 (1.00-1.00, p=0.452)
pco2	1.00 (0.99-1.01, p=0.988)

Here is the multivariable logistic regression with variables of the best model identified by the stepwise selection.

```r
# Set predictor variables to be included in the final model
predictor_final <- c("age", "sofa", "Other.resp.D", "AF", "stroke", "wbc",
                     "hemoglobin", "sodium", "bicarbonate", "chloride", "urea")

# Save results
glmmulti_final <- variables.of.int %>%
  glmuni(response, predictor_final) %>%
  fit2df(explanatory_name = "Variables",
         estimate_name = "Adjusted OR - final model")

kable(glmmulti_final)
```

Variables	Adjusted OR - final model
age	1.07 (1.06-1.08, p<0.001)
sofa	1.18 (1.12-1.24, p<0.001)
Other.resp.D	2.00 (1.54-2.59, p<0.001)
AF	4.95 (3.61-6.77, p<0.001)
stroke	8.87 (6.52-12.08, p<0.001)
wbc	1.01 (0.99-1.03, p=0.298)
hemoglobin	0.86 (0.81-0.91, p<0.001)
sodium	0.98 (0.95-1.00, p=0.075)
bicarbonate	1.03 (1.00-1.06, p=0.024)
chloride	0.96 (0.94-0.98, p<0.001)
urea	1.03 (1.02-1.04, p<0.001)

Here is how we can merge all those 3 tables:

```r
glmuni$index <- 1:nrow(glmuni) # to keep track of the original order

# We then merge the data frames
finaltab <- glmuni %>%
  merge(glmmulti_full) %>%
  left_join(glmmulti_final, by = "Variables")

# Sort the merged data frame based on the original order
finaltab <- finaltab[order(finaltab$index), ] %>%
  select(-index)
```

```r
rownames(finaltab) <- NULL
kable(finaltab)
```

Variables	Crude OR (95% CI, p value)	Adjusted OR - full model	Adjusted OR - final model
IAC	1.25 (0.97-1.63, p=0.085)	1.01 (0.68-1.51, p=0.945)	NA
age	1.07 (1.06-1.08, p<0.001)	1.06 (1.04-1.07, p<0.001)	1.07 (1.06-1.08, p<0.001)
sex	0.66 (0.51-0.85, p=0.001)	1.25 (0.88-1.78, p=0.206)	NA
sofa	1.18 (1.12-1.24, p<0.001)	1.20 (1.10-1.31, p<0.001)	1.18 (1.12-1.24, p<0.001)
service_unitMICU	4.13 (1.50-17.10, p=0.018)	2.95 (0.88-13.93, p=0.115)	NA
service_unitSICU	3.63 (1.32-15.00, p=0.031)	2.59 (0.79-12.07, p=0.161)	NA
COPD	2.01 (1.36-2.92, p<0.001)	1.03 (0.62-1.68, p=0.904)	NA
Other.resp.D	2.00 (1.54-2.59, p<0.001)	1.83 (1.25-2.70, p=0.002)	2.00 (1.54-2.59, p<0.001)
CHF	2.94 (2.12-4.04, p<0.001)	1.35 (0.84-2.14, p=0.208)	NA
AF	4.95 (3.61-6.77, p<0.001)	1.49 (0.98-2.26, p=0.059)	4.95 (3.61-6.77, p<0.001)
chr.renal.D	1.64 (0.86-2.94, p=0.115)	0.48 (0.21-1.07, p=0.080)	NA
chr.liver.D	1.27 (0.74-2.09, p=0.363)	1.33 (0.64-2.66, p=0.433)	NA
coronary.D	2.06 (1.33-3.10, p=0.001)	0.86 (0.49-1.48, p=0.596)	NA
stroke	8.87 (6.52-12.08, p<0.001)	8.37 (5.41-13.08, p<0.001)	8.87 (6.52-12.08, p<0.001)
malignancy	1.64 (1.18-2.26, p=0.003)	1.51 (0.98-2.30, p=0.060)	NA
wbc	1.01 (0.99-1.03, p=0.298)	1.03 (1.00-1.05, p=0.025)	1.01 (0.99-1.03, p=0.298)
hemoglobin	0.86 (0.81-0.91, p<0.001)	0.88 (0.81-0.96, p=0.005)	0.86 (0.81-0.91, p<0.001)
platelet	1.00 (1.00-1.00, p=0.464)	1.00 (1.00-1.00, p=0.708)	NA
sodium	0.98 (0.95-1.00, p=0.075)	1.05 (1.00-1.11, p=0.064)	0.98 (0.95-1.00, p=0.075)
potassium	1.14 (0.97-1.32, p=0.095)	0.93 (0.75-1.15, p=0.482)	NA
bicarbonate	1.03 (1.00-1.06, p=0.024)	0.96 (0.92-1.00, p=0.040)	1.03 (1.00-1.06, p=0.024)
chloride	0.96 (0.94-0.98, p<0.001)	0.93 (0.88-0.97, p=0.002)	0.96 (0.94-0.98, p<0.001)
urea	1.03 (1.02-1.04, p<0.001)	1.02 (1.00-1.03, p=0.023)	1.03 (1.02-1.04, p<0.001)
creat	1.09 (0.99-1.20, p=0.074)	0.79 (0.60-0.99, p=0.070)	NA
po2	1.00 (1.00-1.00, p=0.019)	1.00 (1.00-1.00, p=0.452)	NA
pco2	1.00 (0.99-1.01, p=0.641)	1.00 (0.99-1.01, p=0.988)	NA