

# Prediction of in-hospital mortality

Erik van Zwet

18 juli, 2024

## Contents

|          |                                         |          |
|----------|-----------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                     | <b>1</b> |
| <b>2</b> | <b>Read the data and pre-processing</b> | <b>2</b> |
| <b>3</b> | <b>Casemix model</b>                    | <b>2</b> |
| 3.1      | Calibration . . . . .                   | 4        |
| 3.2      | Bootstrap validation . . . . .          | 5        |
| 3.3      | Funnel plot . . . . .                   | 6        |
| <b>4</b> | <b>Stratify on presenting complaint</b> | <b>7</b> |
| 4.1      | Calibration . . . . .                   | 8        |
| 4.2      | Funnel plots . . . . .                  | 9        |

## 1 Introduction

Load packages.

```
suppressPackageStartupMessages({  
  library(dplyr)  
  library(haven)  
  library(ggplot2)  
  library(cowplot)  
  library(ggrepel)  
  library(rms)}  
)
```

## 2 Read the data and pre-processing

Read the data. Wouter: “half jaar met ontbrekende vitale parameters in locatie 105 eruit filteren.”

```
data=read_sav("K:\\PRESENT-ED2\\Nieuwe dataset 7EDs\\Uitgifte_NEED_nov2023_exCOVID.sav")
data=as_factor(data,only_labelled = TRUE)
data=subset(data,! (ID==105 & YEARS == 2019))

d=select(data,ID,MORT_FINAL_DIC,PRESENTINGCOMPLAINT3, LEEFTIJD, GESLACHT,
          TRIAGECAT_2, GCS_CAT, VITALSCORE_CAT,
          MEDICATIE_CAT, CONSULTEN_CAT, VOCHT_TOE,
          BLOEDONDERZOEK_TOT, RADIOLOZ)

d$LEEFTIJD=as.numeric(as.character(d$LEEFTIJD))
d$MORT_FINAL_DIC=factor(d$MORT_FINAL_DIC,labels=c("Levend", "Dood"))
levels(d$ID)=LETTERS[1:7]
d$PRESENTINGCOMPLAINT3=recode(d$PRESENTINGCOMPLAINT3,
                              Palpitations="Other",
                              Headache="Other",
                              Collapse="Other",
                              "Urinary tract problems"="Other",
                              Wounds="Other")

d=na.omit(d)
d=droplevels(d)
```

## 3 Casemix model

```
fit=glm(MORT_FINAL_DIC=="Dood" ~ PRESENTINGCOMPLAINT3 +
        LEEFTIJD + GESLACHT + TRIAGECAT_2 + GCS_CAT +
        VITALSCORE_CAT + CONSULTEN_CAT +
        BLOEDONDERZOEK_TOT + RADIOLOZ + MEDICATIE_CAT,
        family=binomial, na.action = na.exclude, data=d)
```

```
summary(fit)
```

Call:

```
glm(formula = MORT_FINAL_DIC == "Dood" ~ PRESENTINGCOMPLAINT3 +  
  LEEFTIJD + GESLACHT + TRIAGECAT_2 + GCS_CAT + VITALSCORE_CAT +  
  CONSULTEN_CAT + BLOEDONDERZOEK_TOT + RADIOLOZ + MEDICATIE_CAT,  
  family = binomial, data = d, na.action = na.exclude)
```

Coefficients:

|                                                         | Estimate   | Std. Error | z value | Pr(> z ) |
|---------------------------------------------------------|------------|------------|---------|----------|
| (Intercept)                                             | -1.025e+01 | 1.144e-01  | -89.647 | < 2e-16  |
| PRESENTINGCOMPLAINT3Feeling unwell                      | 9.139e-01  | 6.906e-02  | 13.232  | < 2e-16  |
| PRESENTINGCOMPLAINT3Abdominal pain                      | 6.685e-01  | 8.064e-02  | 8.289   | < 2e-16  |
| PRESENTINGCOMPLAINT3Dyspnea                             | 1.022e+00  | 7.129e-02  | 14.341  | < 2e-16  |
| PRESENTINGCOMPLAINT3Chest pain                          | -7.056e-01 | 1.023e-01  | -6.896  | 5.36e-12 |
| PRESENTINGCOMPLAINT3Trauma                              | 4.884e-01  | 8.623e-02  | 5.664   | 1.48e-08 |
| PRESENTINGCOMPLAINT3Other                               | 6.381e-01  | 6.948e-02  | 9.184   | < 2e-16  |
| LEEFTIJD                                                | 4.970e-02  | 9.237e-04  | 53.813  | < 2e-16  |
| GESLACHTMan                                             | 1.206e-01  | 2.573e-02  | 4.685   | 2.80e-06 |
| TRIAGECAT_2Geel                                         | 4.441e-01  | 5.598e-02  | 7.934   | 2.11e-15 |
| TRIAGECAT_2Oranje                                       | 1.228e+00  | 5.658e-02  | 21.711  | < 2e-16  |
| TRIAGECAT_2Rood                                         | 2.696e+00  | 6.211e-02  | 43.403  | < 2e-16  |
| GCS_CATGCS<15                                           | 8.990e-01  | 4.564e-02  | 19.696  | < 2e-16  |
| GCS_CATGCS=15                                           | -3.726e-01 | 3.385e-02  | -11.009 | < 2e-16  |
| VITALSCORE_CATOne or multiple vital paramaters measured | -6.354e-02 | 6.612e-02  | -0.961  | 0.33655  |
| VITALSCORE_CATAAll vital parameters measured            | -1.772e-01 | 6.524e-02  | -2.717  | 0.00659  |
| CONSULTEN_CAT1 consult                                  | -1.529e-01 | 3.011e-02  | -5.076  | 3.85e-07 |
| CONSULTEN_CAT2 consulten                                | -9.927e-02 | 3.889e-02  | -2.553  | 0.01069  |
| CONSULTEN_CAT3 of meer consulten                        | 1.218e-01  | 6.735e-02  | 1.808   | 0.07058  |
| BLOEDONDERZOEK_TOTJa                                    | 1.215e+00  | 7.043e-02  | 17.250  | < 2e-16  |
| RADIOLOZJa                                              | 6.400e-01  | 3.852e-02  | 16.613  | < 2e-16  |
| MEDICATIE_CAT00 "JA"                                    | 3.955e-01  | 2.682e-02  | 14.746  | < 2e-16  |

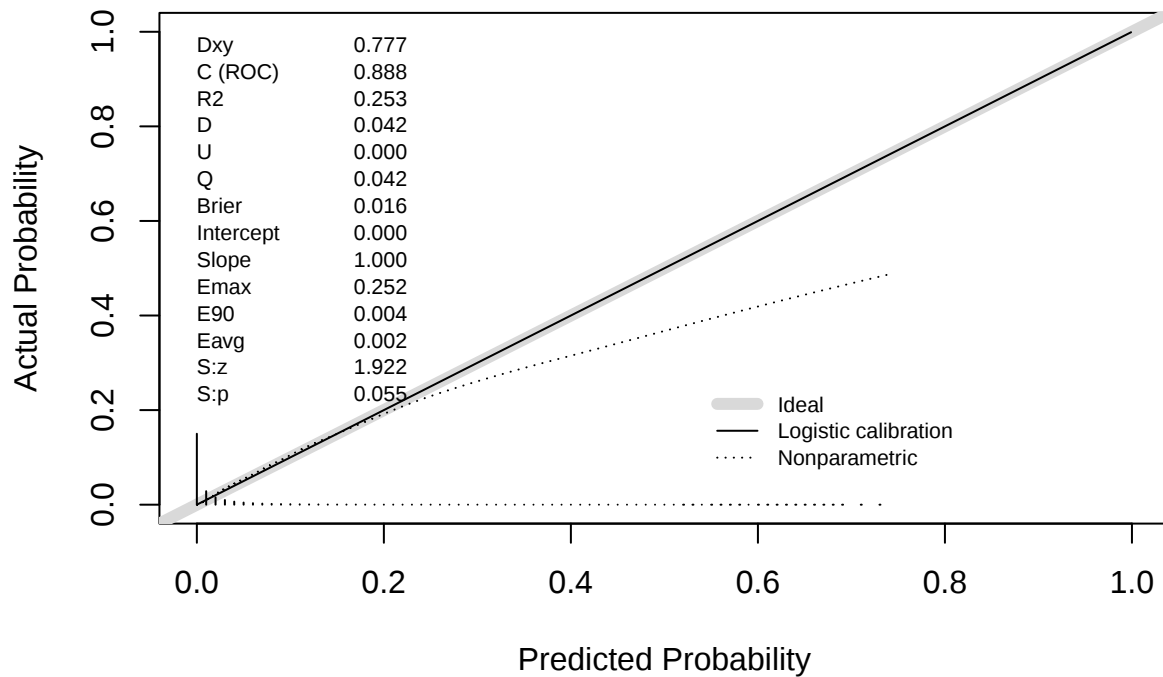
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 69132 on 390066 degrees of freedom  
Residual deviance: 52735 on 390045 degrees of freedom  
AIC: 52779

Number of Fisher Scoring iterations: 9

### 3.1 Calibration

```
d$predicted=predict(fit,type="response")
val.prob(d$predicted, d$MORT_FINAL_DIC=="Dood")
```



| Dxy              | C (ROC)       | R2           | D                | D:Chi-sq        |
|------------------|---------------|--------------|------------------|-----------------|
| 7.767950e-01     | 8.883975e-01  | 2.534561e-01 | 4.203366e-02     | 1.639695e+04    |
| D:p              | U             | U:Chi-sq     | U:p              | Q               |
| NA               | -5.127324e-06 | 1.527951e-10 | 1.000000e+00     | 4.203879e-02    |
| Brier            | Intercept     | Slope        | E <sub>max</sub> | E <sub>90</sub> |
| 1.605688e-02     | 3.125601e-11  | 1.000000e+00 | 2.518817e-01     | 4.301094e-03    |
| E <sub>avg</sub> | S:z           | S:p          |                  |                 |
| 1.737585e-03     | 1.922052e+00  | 5.459921e-02 |                  |                 |

We see some over-prediction when the predicted probabilities are larger than 40% (and under-prediction otherwise):

```
group_by(d,predicted > 0.2) %>%
  summarise(n=n(),
            predicted=sum(predicted),
            observed = sum(MORT_FINAL_DIC=="Dood"))
```

```
# A tibble: 2 x 4
  `predicted > 0.2`      n predicted observed
  <lgl>              <int>    <dbl>    <int>
1 FALSE           386333    5719.     5920
2 TRUE             3734     1153.      952
```

### 3.2 Bootstrap validation

The data set is quite large, so we expect little overfitting/optimism. The `rms` package has functions for bootstrapping the AUC. We re-fit the model with `rms::lrm`

```
fit.lrm <- lrm(MORT_FINAL_DIC=="Dood" ~ PRESENTINGCOMPLAINT3 +
  LEEFTIJD + GESLACHT +
  TRIAGECAT_2 + GCS_CAT + VITALSCORE_CAT +
  CONSULTEN_CAT +
  BLOEDONDERZOEK_TOT + RADIOLOZ + MEDICATIE_CAT,
  data=d, x=TRUE, y=TRUE,maxit=1000)
```

Then use bootstrap validation. I'm using relatively few replicates to save time.

```
val=validate(fit.lrm, B=10,maxit=1000)
val
```

|           | index.orig | training | test    | optimism | index.corrected | n  |
|-----------|------------|----------|---------|----------|-----------------|----|
| Dxy       | 0.7768     | 0.7771   | 0.7763  | 0.0008   | 0.7760          | 10 |
| R2        | 0.2535     | 0.2539   | 0.2531  | 0.0008   | 0.2526          | 10 |
| Intercept | 0.0000     | 0.0000   | -0.0056 | 0.0056   | -0.0056         | 10 |
| Slope     | 1.0000     | 1.0000   | 0.9992  | 0.0008   | 0.9992          | 10 |
| Emax      | 0.0000     | 0.0000   | 0.0014  | 0.0014   | 0.0014          | 10 |
| D         | 0.0420     | 0.0423   | 0.0420  | 0.0003   | 0.0417          | 10 |
| U         | 0.0000     | 0.0000   | 0.0000  | 0.0000   | 0.0000          | 10 |
| Q         | 0.0420     | 0.0423   | 0.0420  | 0.0003   | 0.0417          | 10 |
| B         | 0.0161     | 0.0161   | 0.0161  | 0.0001   | 0.0160          | 10 |
| g         | 2.3004     | 2.2961   | 2.2936  | 0.0025   | 2.2979          | 10 |
| gp        | 0.0262     | 0.0264   | 0.0262  | 0.0001   | 0.0261          | 10 |

The output includes the optimism corrected values. As expected, there is very little optimism.

The output does not provide the AUC, but only shows Somers' Dxy. However, there is a simple relation:

$$AUC = 0.5 \times D_{xy} + 0.5$$

So, the optimism corrected AUC is

```
0.5*val[1,5] + 0.5
```

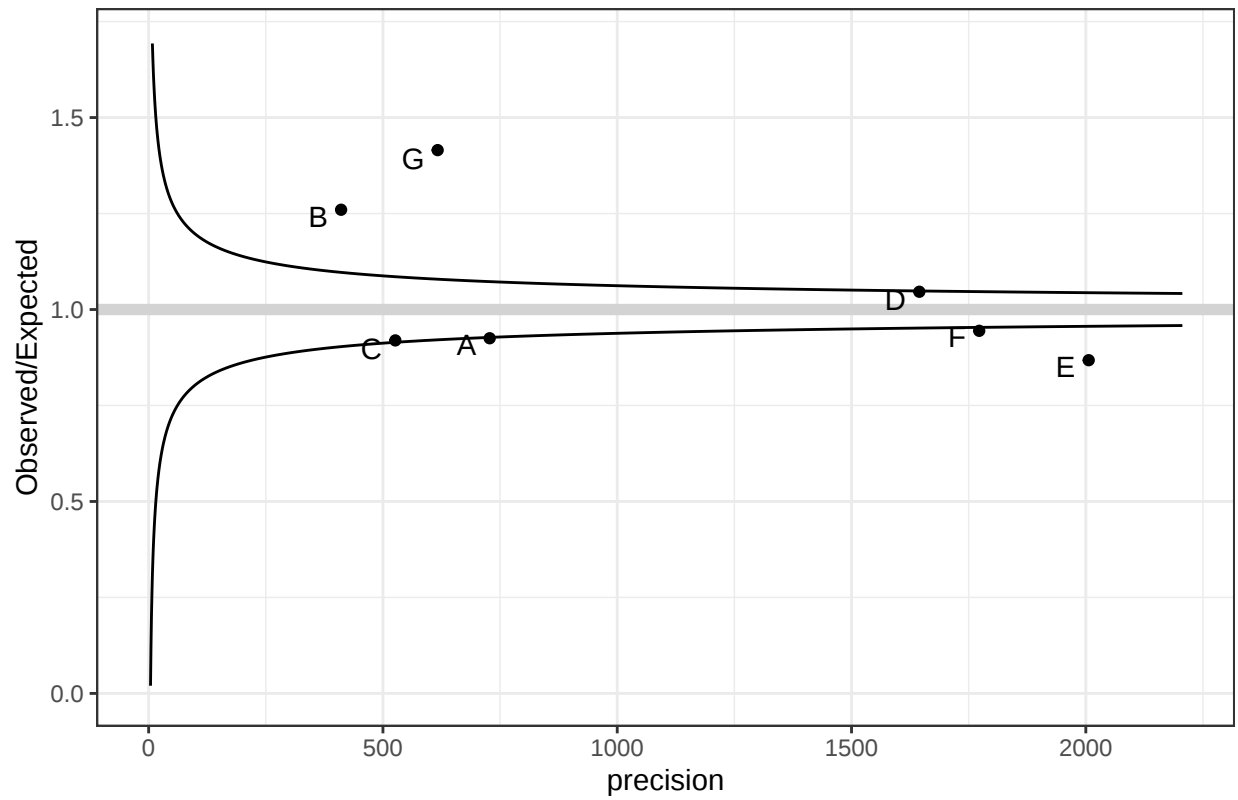
```
[1] 0.8880043
```

### 3.3 Funnel plot

# A tibble: 7 x 6

|   | ID    | n     | E     | O     | Variance | OE    |
|---|-------|-------|-------|-------|----------|-------|
|   | <fct> | <int> | <dbl> | <int> | <dbl>    | <dbl> |
| 1 | A     | 46388 | 659.  | 610   | 597.     | 0.925 |
| 2 | B     | 31385 | 382.  | 481   | 355.     | 1.26  |
| 3 | C     | 31941 | 484.  | 445   | 445.     | 0.920 |
| 4 | D     | 83803 | 1492. | 1561  | 1353.    | 1.05  |
| 5 | E     | 92714 | 1787. | 1551  | 1592.    | 0.868 |
| 6 | F     | 60804 | 1493. | 1410  | 1257.    | 0.945 |
| 7 | G     | 43032 | 575.  | 814   | 536.     | 1.42  |

all



Saving 6.5 x 4.5 in image

## 4 Stratify on presenting complaint

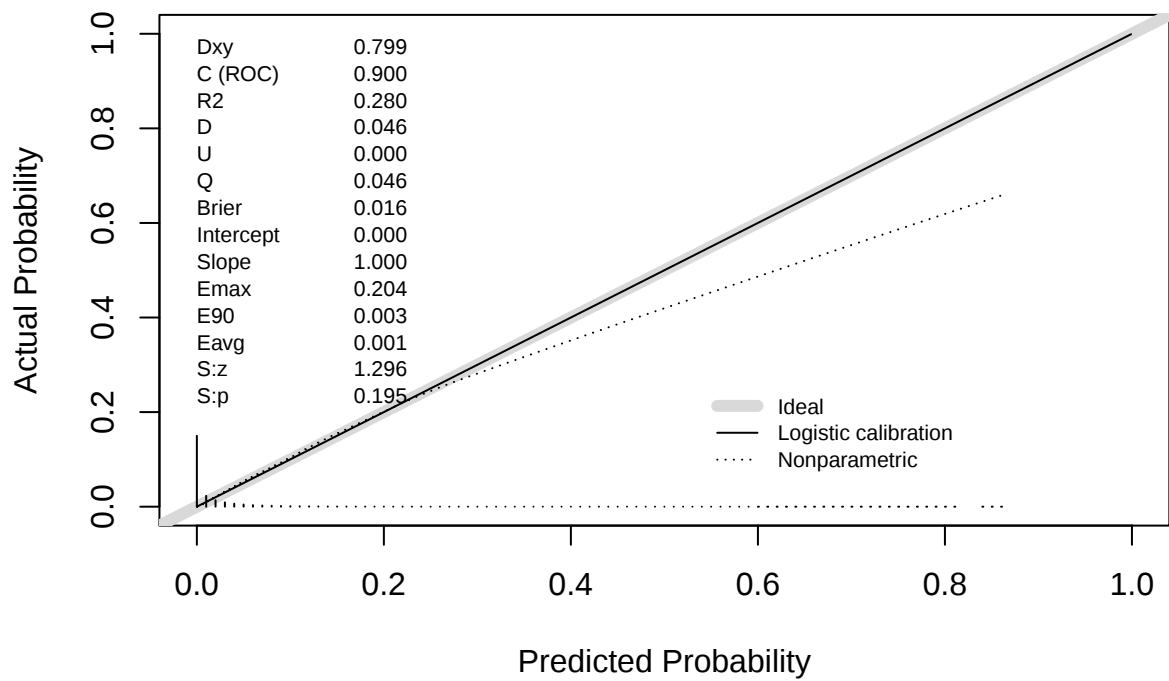
We stratify by forming the interaction between PRESENTINGCOMPLAINT3 and all the other variables in the model.

```
fit=glm(MORT_FINAL_DIC=="Dood" ~ PRESENTINGCOMPLAINT3*(
  LEEFTIJD + GESLACHT + TRIAGECAT_2 + GCS_CAT +
  VITALSCORE_CAT + CONSULTEN_CAT + VOCHT_TOE +
  BLOEDONDERZOEK_TOT + RADIOLOZ + MEDICATIE_CAT),
  family=binomial, na.action = na.exclude, data=d)

# summary(fit)
```

4.1 Calibration

```
d$predicted=predict(fit,type="response")
val.prob(d$predicted, d$MORT_FINAL_DIC=="Dood")
```



|                  |               |              |                  |                 |
|------------------|---------------|--------------|------------------|-----------------|
| Dxy              | C (ROC)       | R2           | D                | D:Chi-sq        |
| 7.992895e-01     | 8.996448e-01  | 2.795246e-01 | 4.645913e-02     | 1.812317e+04    |
| D:p              | U             | U:Chi-sq     | U:p              | Q               |
| NA               | -5.127324e-06 | 4.511094e-10 | 1.000000e+00     | 4.646425e-02    |
| Brier            | Intercept     | Slope        | E <sub>max</sub> | E <sub>90</sub> |
| 1.576479e-02     | 6.482869e-09  | 1.000000e+00 | 2.038477e-01     | 2.955980e-03    |
| E <sub>avg</sub> | S:z           | S:p          |                  |                 |
| 9.866685e-04     | 1.296078e+00  | 1.949487e-01 |                  |                 |

## 4.2 Funnel plots

``summarise()`` has grouped output by 'ID'. You can override using the ``groups`` argument.

