Prediction of in-hospital mortality

Erik van Zwet

08 juli, 2024

Contents

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.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

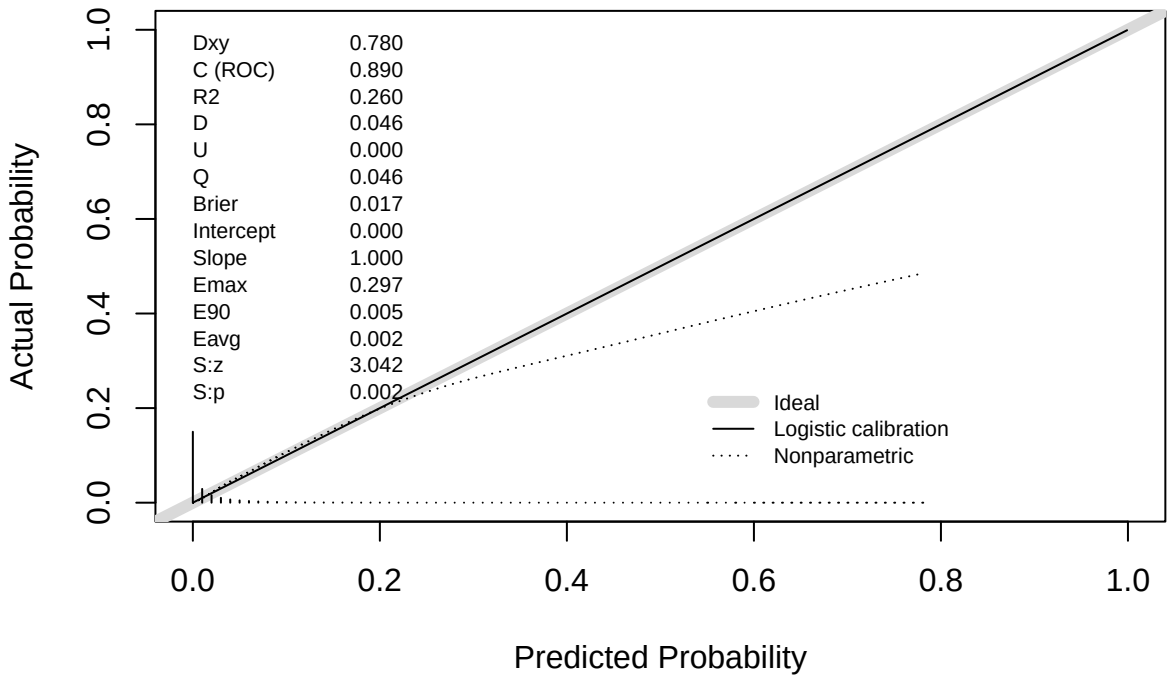
```
fit=glm(MORT_FINAL_DIC=="Dood" ~ PRESENTINGCOMPLAINT3 +
  LEEFTIJD + GESLACHT + TRIAGECAT_2 + GCS_CAT +
    VITALSCORE_CAT + CONSULTEN_CAT +
    BLOEDONDERZOEK_TOT + RADIOLOZ + MEDICATIE_CAT,
  family=binomial, na.action = na.exclude, data=d)
```

```r
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)                                          -1.035e+01  8.437e-02 -122.667  < 2e-16
PRESENTINGCOMPLAINT3Feeling unwell                    9.521e-01  5.105e-02   18.649  < 2e-16
PRESENTINGCOMPLAINT3Abdominal pain                    6.606e-01  5.953e-02   11.097  < 2e-16
PRESENTINGCOMPLAINT3Dyspnea                           1.187e+00  5.209e-02   22.792  < 2e-16
PRESENTINGCOMPLAINT3Chest pain                       -6.073e-01  7.375e-02   -8.235  < 2e-16
PRESENTINGCOMPLAINT3Trauma                            4.041e-01  6.404e-02    6.310 2.78e-10
PRESENTINGCOMPLAINT3Other                             5.967e-01  5.146e-02   11.595  < 2e-16
LEEFTIJD                                              5.095e-02  6.835e-04   74.542  < 2e-16
GESLACHTMan                                           1.461e-01  1.861e-02    7.854 4.02e-15
TRIAGECAT_2Geel                                       4.229e-01  3.837e-02   11.023  < 2e-16
TRIAGECAT_2Oranje                                     1.256e+00  3.904e-02   32.157  < 2e-16
TRIAGECAT_2Rood                                       2.886e+00  4.491e-02   64.266  < 2e-16
GCS_CATGCS<15                                         8.473e-01  3.446e-02   24.589  < 2e-16
GCS_CATGCS=15                                        -3.388e-01  2.386e-02  -14.198  < 2e-16
VITALSCORE_CATOne or multiple vital paramaters measured -1.231e-01  4.745e-02   -2.595  0.00946
VITALSCORE_CATAll vital parameters measured           -2.050e-01  4.629e-02   -4.428 9.53e-06
CONSULTEN_CAT1 consult                                -8.988e-02  2.144e-02   -4.193 2.75e-05
CONSULTEN_CAT2 consulten                              -5.387e-02  2.907e-02   -1.853  0.06390
CONSULTEN_CAT3 of meer consulten                      1.115e-01  5.252e-02    2.124  0.03370
BLOEDONDERZOEK_TOTJa                                  1.225e+00  5.268e-02   23.260  < 2e-16
RADIOLOZJa                                            6.679e-01  2.876e-02   23.227  < 2e-16
MEDICATIE_CAT00 "JA"                                  4.425e-01  1.932e-02   22.910  < 2e-16

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 131465  on 697185  degrees of freedom
Residual deviance:  99655  on 697164  degrees of freedom
AIC: 99699

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
 7.801423e-01  8.900712e-01  2.595283e-01  4.562452e-02  3.180977e+04
          D:p             U      U:Chi-sq          U:p             Q
           NA -2.868675e-06  8.949428e-09  1.000000e+00  4.562739e-02
        Brier     Intercept         Slope          Emax           E90
 1.725578e-02  2.018238e-11  1.000000e+00  2.968008e-01  4.829391e-03
         Eavg           S:z           S:p
 1.949890e-03  3.042331e+00  2.347537e-03
```

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             689981    10977.    11367
2 TRUE                7205     2298.     1908
```

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           0.7802   0.7801  0.7800   0.0000          0.7802 10
R2            0.2595   0.2597  0.2594   0.0003          0.2593 10
Intercept     0.0000   0.0000 -0.0025   0.0025         -0.0025 10
Slope         1.0000   1.0000  0.9994   0.0006          0.9994 10
Emax          0.0000   0.0000  0.0007   0.0007          0.0007 10
D             0.0456   0.0456  0.0456   0.0000          0.0456 10
U             0.0000   0.0000  0.0000   0.0000          0.0000 10
Q             0.0456   0.0456  0.0456   0.0000          0.0456 10
B             0.0173   0.0172  0.0173   0.0000          0.0173 10
g             2.3489   2.3498  2.3484   0.0013          2.3476 10
gp            0.0284   0.0284  0.0284   0.0000          0.0284 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:

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

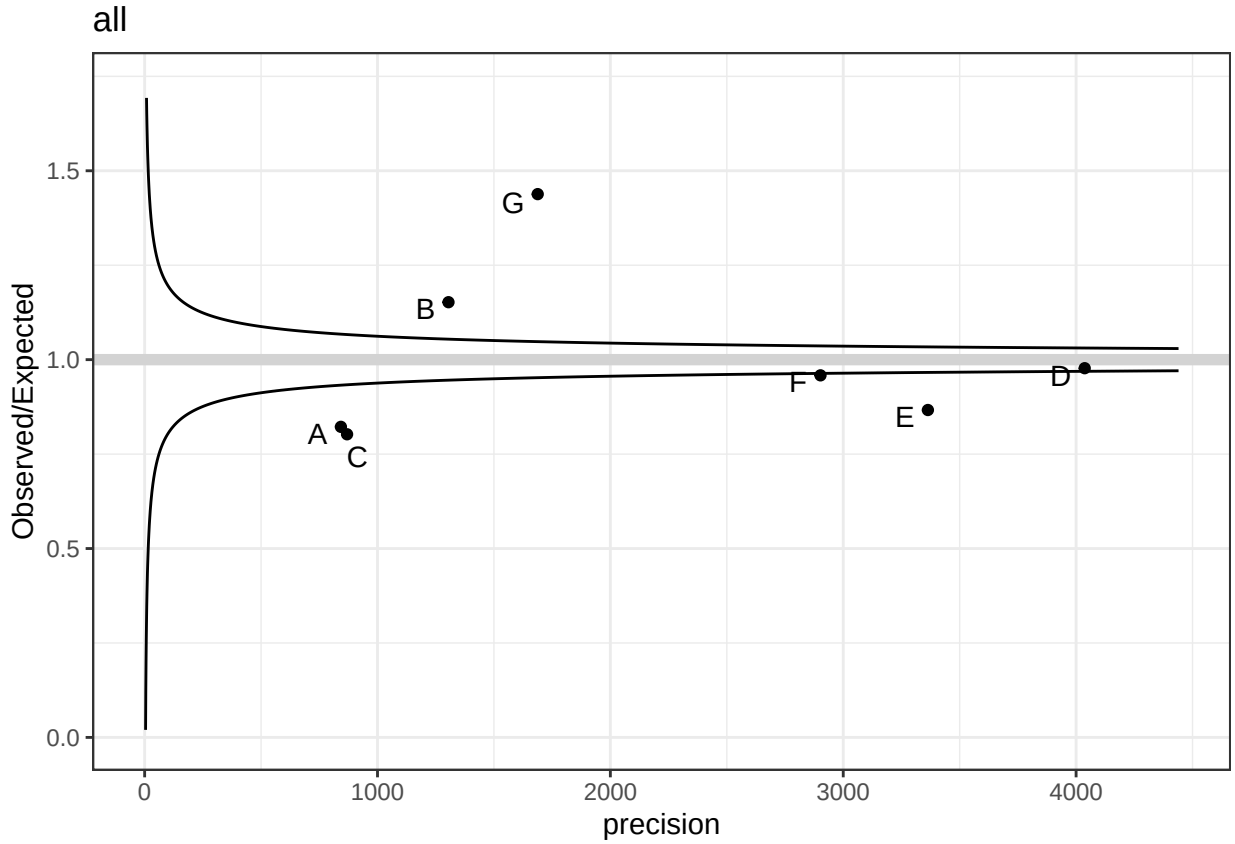
So, the optimism corrected AUC is

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

```
[1] 0.8900756
```

3.3 Funnel plot

```
# A tibble: 7 x 6
  ID         n     E     O Variance    OE
  <fct>  <int> <dbl> <int>    <dbl> <dbl>
1 A      46388  742.   610     653. 0.822
2 B      88076 1199.  1381    1101. 1.15
3 C      48062  789.   633     716. 0.802
4 D     169857 3603.  3521    3215. 0.977
5 E     148467 2959.  2565    2604. 0.867
6 F      93922 2427.  2326    2029. 0.958
7 G     102414 1557.  2239    1436. 1.44
```




```
Saving 6.5 x 4.5 in image
```

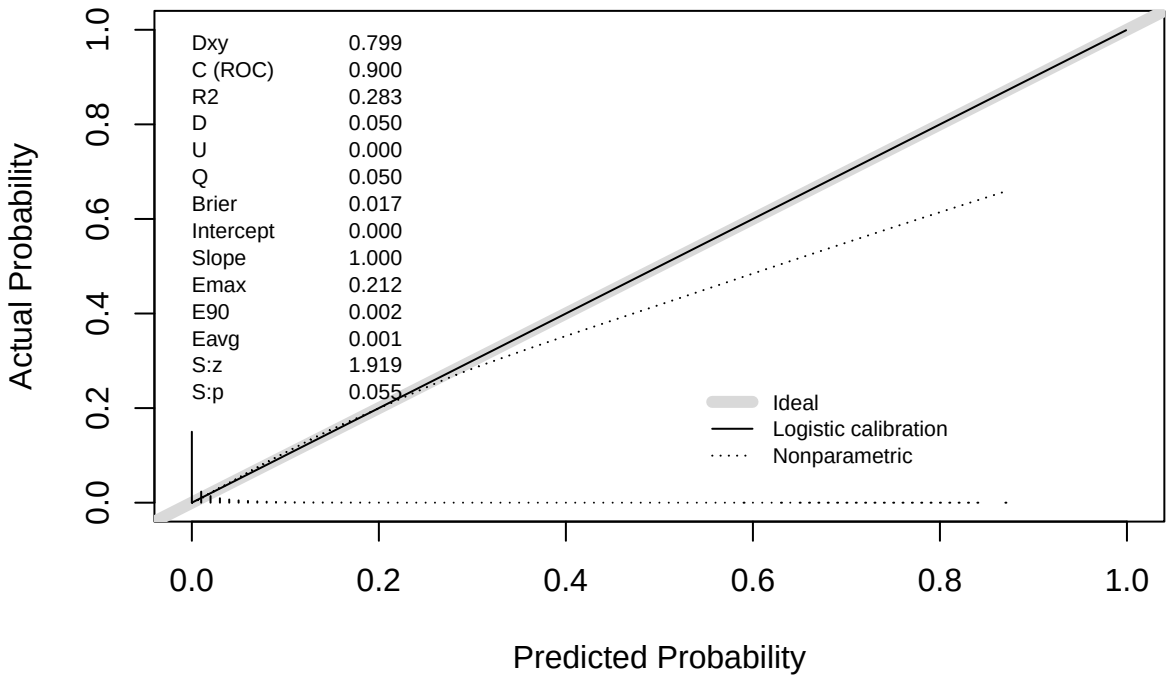
4 Stratify on presenting complaint

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
7.992113e-01  8.996056e-01  2.829987e-01  4.985520e-02  3.475935e+04
         D:p             U      U:Chi-sq          U:p             Q
          NA -2.868675e-06 -2.401066e-09  1.000000e+00  4.985807e-02
       Brier     Intercept         Slope          Emax           E90
1.694588e-02  7.998748e-10  1.000000e+00  2.124826e-01  2.431053e-03
        Eavg           S:z           S:p
1.143398e-03  1.919115e+00  5.496983e-02
```

4.2 Funnel plots

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