phia_lca2_jsg2145

Jared Garfinkel

8/9/2020

```
# Read public-release datasets
#Biomarker dataset
biomarker <- read_csv(file = "./data/Shims22016adultbio.csv",
                      col_names = TRUE,
                      col_types = NULL,
                      locale = default_locale(),
                      na = c("", "NA"),
                      quoted_na = TRUE,
                      quote = "\"",
                      comment = "",
                      trim_ws = TRUE,
                      skip = 0,
                      n_max = Inf,
                      progress = show_progress(),
                      skip_empty_rows = TRUE)
```

```
bio_dat <- biomarker %>%
  filter(hiv1statusfinalsurvey == 1) %>%
  select(personid,
         hiv1statusfinalsurvey,
         awareselfreported,
         gender,
         arvstatus,
         artselfreported,
         resultvlc,
         vls,
         tri90,
         tri90aware,
         tri90art,
         tri90vls) %>%
  mutate(resultvlc = recode(resultvlc,
                    "< LLOD" = "1",
                    "< LLOQ: 20" = "20",
                    "< LLOQ: 40" = "40",
                    "< LLOQ: 400" = "400",
                    "< LLOQ: 839" = "839",
                    "> ULOQ 10000000" = "10000000"),
         resultvlc = as.numeric(resultvlc),
         vlunder200 = if_else(resultvlc < 200, 1, 2),
         across(.funs = as_factor))
```

```r
str(bio_dat)
```

```
## tibble [3,055 x 13] (S3: tbl_df/tbl/data.frame)
##  $ personid             : chr [1:3055] "SW00000000000502" "SW00000000000803" "SW00000000000806" "SW0
##  $ hiv1statusfinalsurvey: chr [1:3055] "1" "1" "1" "1" ...
##  $ awareselfreported    : num [1:3055] 1 2 1 1 1 1 1 1 1 1 ...
##  $ gender               : num [1:3055] 2 2 2 2 1 2 2 2 2 1 ...
##  $ arvstatus            : num [1:3055] 1 2 2 1 1 1 1 2 1 2 ...
##  $ artselfreported      : num [1:3055] 1 99 99 1 1 1 1 2 1 2 ...
##  $ resultvlc            : num [1:3055] 20 99137 1468618 1 20 ...
##  $ vls                  : num [1:3055] 1 2 2 1 1 1 1 2 1 2 ...
##  $ tri90                : num [1:3055] 1 1 2 1 1 1 1 1 1 1 ...
##  $ tri90aware           : num [1:3055] 1 2 99 1 1 1 1 1 1 1 ...
##  $ tri90art             : num [1:3055] 1 3 99 1 1 1 1 2 1 2 ...
##  $ tri90vls             : num [1:3055] 1 2 99 1 1 1 1 2 1 2 ...
##  $ vlunder200           : num [1:3055] 1 2 2 1 1 1 1 2 1 2 ...
```

```r
# aware:                1 - Aware or considered aware because ARVs detectable
#                       2 - Unaware and ARVs not detectable, or unaware and ARV testing results missing
#                       99 - Missing
#
# art:                  1 - ARVs detectable, self-reported on ART, or both ARVs detectable and self-rep
#                       2 - Unaware or aware, ARVs not detectable and self-reported not on ART, or awar
#                       99 - Missing
# awareselfreported:    1 - Self-report aware of HIV + status
#                       2 - Self-report not aware of HIV + status
#                       99 - missing
# artselfreported:      1 - On ART
#                       2 - Not on ART
#                       99 - Missing
# awareartselfreported: 1 - Self-report as not previously diagnosed
#                       2 - Self-report as previously diagnosed, not on ART
#                       3 - Previously diagnosed, on ART
#                       99 - Missing, including incomplete tri90 information
# arvstatus:            1 - ARV detected
#                       2 - ARV not detected
#                       99 - Missing
# resultvlc:            > ULOQ 10000000 - Upper limit of quantification 10000000
#                       < LLOD - less than lower limit of detection
#                       < LLOQ: 839 - less than lower limit of quantification of 839
#                       < LLOQ: 400 - less than lower limit of quantification of 400
#                       < LLOQ: 40 - less than lower limit of quantification of 40
#                       < LLOQ: 20 - less than lower limit of quantification of 20
# hivselfreport:        1 - Self-reported positive
#                       2 - Self-reported negative
#                       3 - Self-reported never tested or never received test result
#                       99 - Missing
# tri90art              1 - ARVs detectable, self-reported on ART, or both ARVs detectable and self-rep
#                       2 - ARVs not detectable and self-reported not on ART or missing ARV data and se
#                       3 - Recoded as not on ART (unaware and ARVs not detectable or unaware and ARVs
#                       99 - Incomplete Tri90 information
#
```

```r
bio_dat_lca = bio_dat %>%
  select(awareselfreported,
         arvstatus,
         artselfreported,
         gender,
         vls,
         tri90,
         tri90aware,
         tri90art,
         tri90vls,
         vlunder200) %>%
  mutate(awareselfreported = recode(awareselfreported, "1" = "1", "2" = "0", "99" = NULL),
         arvstatus = recode(arvstatus, "1" = "1", "2" = "0", "99" = NULL),
         artselfreported = recode(artselfreported, "1" = "1", "2" = "0", "99" = NULL),
         gender = recode(gender, "1" = "1", "2" = "0", "99" = NULL),
         vls = recode(vls, "1" = "1", "2" = "0", "99" = NULL),
         tri90 = recode(tri90, "1" = "1", "2" = "0", "99" = NULL),
         tri90aware = recode(tri90aware, "1" = "1", "2" = "0", "99" = NULL),
         tri90art = recode(tri90art, "1" = "1", "2" = "0", "99" = NULL),
         tri90vls = recode(tri90vls, "1" = "1", "2" = "0", "99" = NULL),
         vlunder200 = recode(vlunder200, "1" = "1", "2" = "0", "99" = NULL))
```

```r
bio_dat_lca_full = bio_dat %>%
  select(awareselfreported,
         arvstatus,
         gender,
         vlunder200) %>%
  mutate(across(everything(), ~recode(.x, "1" = "1", "2" = "0", "99" = NULL)))
```

```r
bio_dat1 = bio_dat %>%
  mutate(awareselfreported = recode(awareselfreported, "1" = "1", "2" = "0", "99" = NULL))

lca_prepper = function(.) {
  mutate(across(.funs = recode(., "1" = "1", "2" = "0", "99" = NULL))
}
```

"Full Model"

```r
# full model
bio_lca_full = bio_dat_lca %>%
  select(awareselfreported, arvstatus, vlunder200)



set.seed(22)
bio_lca = randomLCA::randomLCA(bio_lca_full, calcSE = TRUE)
summary(bio_lca)
```

```
##   Classes      AIC      BIC     AIC3    logLik penlogLik
##         2 6777.542 6819.714 6784.542 -3381.771 -3381.818
```

```
## Class probabilities
## Class  1 Class  2
##   0.2347   0.7653
## Outcome probabilities
##          awareselfreported arvstatus vlunder200
## Class  1            0.5015    0.0251     0.0464
## Class  2            0.9773    0.9872     0.9118
```

```
probs = outcomeProbs(bio_lca)
postClassProbs(bio_lca)
```

```
##    awareselfreported arvstatus vlunder200 Freq      Class 1      Class 2
## 1                  0         0          0  329 0.9998197716 0.0001802284
## 2                  0         0          1   16 0.9630959474 0.0369040526
## 3                  0         1          0   13 0.6488982251 0.3511017749
## 4                  0         1          1   46 0.0086194775 0.9913805225
## 5                  0        NA          0    3 0.9864386751 0.0135613249
## 6                  0        NA          1    3 0.2549479739 0.7450520261
## 7                  1         0          0  332 0.9923495506 0.0076504494
## 8                  1         0          1   42 0.3789605022 0.6210394978
## 9                  1         1          0  204 0.0414237541 0.9585762459
## 10                 1         1          1 2015 0.0002032504 0.9997967496
## 11                 1        NA          0    7 0.6297360897 0.3702639103
## 12                 1        NA          1   43 0.0079374908 0.9920625092
## 13                NA         0          0    1 0.9960598053 0.0039401947
## 14                NA         1          0    1 0.0776775598 0.9223224402
```

```
BIC(bio_lca)
```

```
## [1] 6819.714
```

```
diseased <- ifelse(probs[[1]]$Outcome[1] < probs[[2]]$Outcome[1], 2, 1)
notdiseased <- 3 - diseased
sens <- apply(probs[[diseased]], 1, function(x)

  sprintf("%3.2f (%3.2f, %3.2f)",
          x[1],
          x[2],
          x[3])
)
spec <- apply(probs[[notdiseased]], 1, function(x)
  sprintf("%3.2f (%3.2f, %3.2f)",
          1 - x[1],
          1 - x[3],
          1 - x[2])
)

stable <- data.frame(sens, spec)
names(stable) <- c("Sensitivity", "Specificity")
print(stable, row.names = TRUE)
```

```
stable = function(probs = probs){

  diseased <- ifelse(probs[[1]]$Outcome[1] < probs[[2]]$Outcome[1], 2, 1)

  notdiseased <- 3 - diseased

  sens <- apply(probs[[diseased]], 1, function(x)
    sprintf("%3.2f (%3.2f, %3.2f)",
            x[1],
            x[2],
            x[3])
  )

  spec <- apply(probs[[notdiseased]], 1, function(x)
    sprintf("%3.2f (%3.2f, %3.2f)",
            1 - x[1],
            1 - x[3],
            1 - x[2])
  )

  stable <- data.frame(sens, spec)

  names(stable) <- c("Sensitivity", "Specificity")

  return(print(stable, row.names = TRUE))
}

# stable(probs2)
```

"Fuller model"

```
# full model
set.seed(22)
bio_lca_fuller = randomLCA::randomLCA(bio_dat_lca_full, random = FALSE, calcSE = TRUE)

probs_full = outcomeProbs(bio_lca_fuller)

stable(probs_full)
```

```
##                        Sensitivity       Specificity
## awareselfreported 0.98 (0.97, 0.98) 0.50 (0.46, 0.54)
## arvstatus         0.99 (0.98, 0.99) 0.97 (0.94, 0.99)
## gender            0.31 (0.29, 0.33) 0.62 (0.58, 0.66)
## vlunder200        0.91 (0.90, 0.92) 0.96 (0.93, 0.97)
```

"Simple Model"

A model of sensitivity and specificity for just one variable should be compared to a true (known) outcome. As such, a contingency table should be sufficient to calculate the sensitivity and specificity of each variable.

awareselfreported

```
xtabs(~ awareselfreported + tri90aware, data = bio_dat_lca)
```

```
##                  tri90aware
## awareselfreported    0    1
##                 0  346   59
##                 1    0 2591
```

Assuming the final classifications are true for tri90aware, there are 59 false negatives and no false positives. This corresponds to a sensitivity of 2591/2650, or 0.978, and a specificity of 346/ 346, or 1.00. The prevalance (of awareness of HIV status) is therefore 0.885. The PPV can be calculated as 2591/2591, or 1.00, and the NPV is 346/405 = 0.854.

arvstatus

```
xtabs(~ arvstatus + tri90aware, data = bio_dat_lca)
```

```
##          tri90aware
## arvstatus    0    1
##         0  345  369
##         1    0 2279
```

Assuming the final classifications are true for tri90aware, there are 369 false negatives and no false positives. This corresponds to a sensitivity of 2279/2648, or 0.861, and a specificity of 345/345, or 1.00. The prevalance (of awareness of HIV status) is therefore 0.885. The PPV can be calculated as 2279/2279, or 1.00, and the NPV is 345/714 = 0.483.

vlunder200

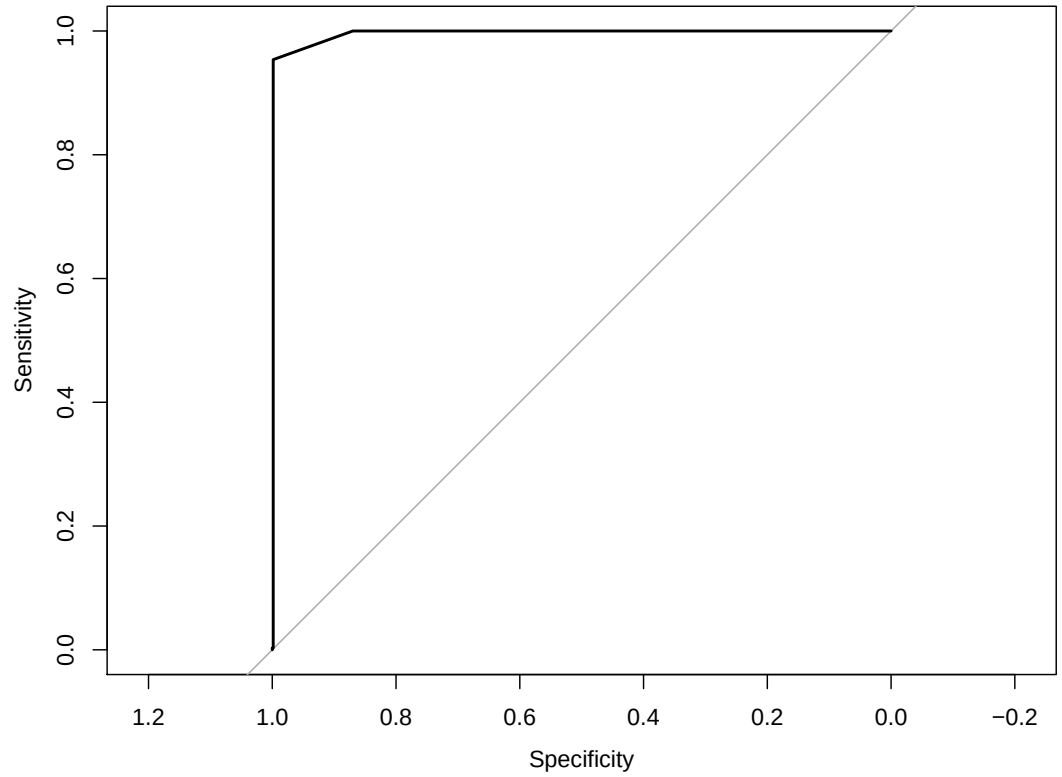
```
xtabs(~ vlunder200 + tri90aware, data = bio_dat_lca)
```

```
##           tri90aware
## vlunder200    0    1
##          0  330  546
##          1   16 2105
```

Assuming the final classifications are true for tri90aware, there are 546 false negatives and 16 false positives. This corresponds to a sensitivity of 2105/2651, or 1, and a specificity of 330/346, or 0.954. The prevalance (of awareness of HIV status) is therefore 0.885. The PPV can be calculated as 2105/2121, or 0.992, and the NPV is 330/876 = 0.377.

ROC

```
roc(bio_dat$tri90aware, c(bio_dat$awareselfreported + bio_dat$arvstatus + bio_dat$vlunder200), data = bi
  plot()
```




```
test1 = roc(tri90aware ~ awareselfreported, bio_dat)
test2 = roc(tri90aware ~ arvstatus, bio_dat)
roc.test(test1, test2)
```

```
##
##  DeLong's test for two correlated ROC curves
##
## data:  test1 and test2
## Z = 15.526, p-value < 2.2e-16
## alternative hypothesis: true difference in AUC is not equal to 0
## sample estimates:
## AUC of roc1 AUC of roc2
##   0.9884949   0.9294748
```

```
roc.test(bio_dat$tri90aware, bio_dat$awareselfreported, bio_dat$arvstatus)
```

```
##
##  DeLong's test for two correlated ROC curves
##
## data:  bio_dat$awareselfreported and bio_dat$arvstatus by bio_dat$tri90aware (1, 2)
## Z = 15.526, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in AUC is not equal to 0
## sample estimates:
## AUC of roc1 AUC of roc2
##   0.9884949   0.9294748
```