

phia_lca_jsg2145

Jared Garfinkel

8/2/2020

```
# 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)

#Individual interview dataset
individual <- read_csv(file = "./data/Shims22016adultind.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)

ind_dat <-
  filter(individual, bt_status == 1) %>% #
  select(personid,
         gender,
         age,
         hivstatusfinal,
         known_hiv_status,
         arvscurrent) %>%
  mutate(gender = factor(gender,
                        labels = c("1" = "male",
```

```

                                "2" = "female")),
  across(.funs = as_factor))

bio_dat <- biomarker %>%
  filter(hiv1statusfinalsurvey == 1) %>%
  select(personid,
         hiv1statusfinalsurvey,
         aware,
         art,
         awareselfreported,
         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))

```

```

# 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
#

```

First 90 (awareness)

pre-processed data

```

# This table shows the respondents/informants who reported that they were aware of their hiv status or
tbl1 = xtabs(~ awareselfreported + tri90art, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl2 = xtabs(~ awareselfreported + arvstatus, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl3 = xtabs(~ awareselfreported + artselfreported, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl4 = xtabs(~ tri90aware + awareselfreported, data = bio_dat)

```

```

# Grand total
nrow(bio_dat) # 3,055

```

```
## [1] 3055
```

```

# This table shows the respondents/informants who reported that they were aware of their hiv status or
tbl1_tri90 = xtabs(~ tri90aware + tri90art, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl2_tri90 = xtabs(~ tri90aware + arvstatus, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl3_tri90 = xtabs(~ tri90aware + artselfreported, data = bio_dat)

# This contingency table shows the respondents/informants who reported that they were aware of their hiv
tbl4_tri90 = xtabs(~ tri90aware + awareselfreported, data = bio_dat)

# This table shows the respondents processed tri90aware status and their viral load suppression (under
xtabs(~ tri90aware + vlunder200, data = bio_dat)

```

```

##          vlunder200
## tri90aware    1    2

```

```
##      1  2105  546
##      2    16  330
##     99    44   14
```

tri90art

This combined chart shows the contingency table of awareselfreported vs. tri90art along with tri90aware vs. tri90art. While there are missing values in awareselfreported, there are none in tri90aware. The 57 missing values from tri90art were classified as missing for tri90aware data. The 60 respondents who were on art were reclassified as tri90aware.

```
g1 <- tableGrob(tbl1)
g2 <- tableGrob(tbl1_tri90)

haligned <- gtable_combine(g1,g2, along=1)
valigned <- gtable_combine(g1,g2, along=2)
grid.arrange(haligned, valigned, ncol=2, top = "tri90art, awareselfreported vs. tri90aware")
```

tri90art, awareselfreported vs. tri90aware

	1	2	3	99
1	2309	282	0	52
2	59	0	346	5
99	1	0	0	1

	1	2	3	99
1	2369	282	0	0
2	0	0	346	0
99	0	0	0	58

	1	2	3	99
1	2309	282	0	52
2	59	0	346	5
99	1	0	0	1

	1	2	3	99
1	2369	282	0	0
2	0	0	346	0
99	0	0	0	58

arvstatus

This chart shows contingency tables for awareselfreported alongside those for tri90aware with respect to arvstatus. As before, those with detectable arvs in their blood sample were reclassified as tri90aware regardless of self-reported aware status. Those missing arvstatus data appear to have been treated or classified differently. 3 respondents with missing arvstatus data were classified as tri90aware compared to 50 self-reported aware who were missing arvstatus data. Meanwhile, there was 1 respondent missing arvstatus classified as not

tri90aware compared to 6 of those who reported that they were unaware and had missing arvstatus data. Although 52 respondents with missing arvstatus data were considered null under tri90aware, none who were missing awareselfreported data were also missing arvstatus data. Finally, it appears that 6 respondents who had no detectable arvs were considered missing tri90aware data, even though only 1 who had no detectable arvs were missing awareselfreported data. So, overall the biggest change was between those missing arvstatus data who said they were aware, most of whom it appears were reclassified as missing data in tri90aware. $50/3055 = 1.6\%$. This would effectively reduce the prevalence of awareness by removing about 50 individuals from the analysis.

Why were subjects removed from the analysis?

Generally an observation was removed if they had incomplete tri90 data, including blood-test data. Specifically, observations to be excluded in the 90 90 90 analysis with ARV data due to incomplete data defined as: aware missing and ARV data missing, aware and both ARV and self-reported ART data missing, or ARVs detectable / self-reported on ART and VLS missing.

```
h1 <- tableGrob(tbl2)
h2 <- tableGrob(tbl2_tri90)

haligned <- gtable_combine(h1,h2, along=1)
valigned <- gtable_combine(h1,h2, along=2)
grid.arrange(haligned, valigned, ncol=2, top = "arvstatus")
```

arvstatus

	1	2	99		1	2	99
1	2219	374	50	1	2279	369	3
2	59	345	6	2	0	345	1
99	1	1	0	99	0	6	52

	1	2	99
1	2219	374	50
2	59	345	6
99	1	1	0
	1	2	99
1	2279	369	3
2	0	345	1
99	0	6	52

artselfreported

44 respondents were removed from the aware group if they reported that they were aware and on art. 3 respondents were removed from the aware group if they were awareselfreported and had no detectable arvs. 64 respondents who reported that they were unaware were reclassified. Of those reclassified, it appears that 55 of them were classified as aware and 9 were classified as missing data and not included in the tri90 analysis. So, as about 50 respondents were reclassified as aware from artselfreported data, almost 50 respondents were removed from the tri90 analysis who had previously been considered aware. This is a differential misclassification and deserves further study.

```
j1 <- tableGrob(tbl3)
j2 <- tableGrob(tbl3_tri90)

haligned <- gtable_combine(j1,j2, along=1)
valigned <- gtable_combine(j1,j2, along=2)
grid.arrange(haligned, valigned, ncol=2, top = "artselfreported")
```

artselfreported

	1	2	99		1	2	99
1	2325	309	9	1	2281	306	64
2	0	0	410	2	0	0	346
99	0	0	2	99	44	3	11

	1	2	99
1	2325	309	9
2	0	0	410
99	0	0	2
	1	2	99
1	2281	306	64
2	0	0	346
99	44	3	11

awareselfreported

tri90 (processed) data

```
bio_cong_dat = biomarker %>%
  filter(tri90 == 1) %>%
  select(personid,
```

```

hiv1statusfinalsurvey,
aware,
art,
awareselfreported,
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",
  "> ULQ 10000000" = "10000000"),
  resultvlc = as.numeric(resultvlc),
  vlunder200 = if_else(resultvlc < 200, 1, 2),
  across(.funs = as_factor))

```

```

# Grand total
nrow(bio_cong_dat)

```

```
## [1] 2997
```

```

# while all the respondents classified as unaware under tri90aware were unaware under awareselfreported
xtabs(~tri90aware + awareselfreported, bio_cong_dat)

```

```

##           awareselfreported
## tri90aware    1    2    99
##           1 2591    59    1
##           2     0   346    0

```

```

# While no respondents classified as unaware under tri90aware were asked for artselfreported data, 306
xtabs(~tri90aware + artselfreported, bio_cong_dat)

```

```

##           artselfreported
## tri90aware    1    2    99
##           1 2281   306    64
##           2     0    0   346

```

```

# While no respondents who tested positive for art were classified as aware, 282 respondents who tested
xtabs(~ tri90aware + tri90art, data = bio_cong_dat)

```

```

##           tri90art
## tri90aware    1    2    3
##           1 2369   282    0
##           2     0    0   346

```

```
# There were 546 respondents with viral load above 200 who were classified as aware. This may be indica
xtabs(~tri90aware + vlunder200, bio_cong_dat)
```

```
##          vlunder200
## tri90aware    1    2
##          1 2105  546
##          2   16  330
```

```
# There were 369 respondents with no detectable arvs reclassified as tri90aware and 3 respondents with
xtabs(~tri90aware + arvstatus, bio_cong_dat)
```

```
##          arvstatus
## tri90aware    1    2   99
##          1 2279  369    3
##          2    0  345    1
```

```
# This table shows that 444 respondents with viral load above 1000 copies/ milliliter were reclassified
xtabs(~tri90aware + vls, bio_cong_dat)
```

```
##          vls
## tri90aware    1    2
##          1 2207  444
##          2   35  311
```

```
# These variables are in agreement
xtabs(~tri90aware + aware, data = bio_cong_dat)
```

```
##          aware
## tri90aware    1    2
##          1 2651    0
##          2    0  346
```

Differential misclassification

Differential misclassification occurs when misclassification of exposure is not equal between subjects that have or do not have the health outcome, or when misclassification of the health outcome is not equal between exposed and unexposed subjects.

Assume the tri90 classifications represent the truth. It is possible to compare the classifications of respondents using pre-processed measurements (i.e. of blood samples and survey responses) to the “true” values and assess whether the misclassification is differential or non-differential.

```
nrow(bio_cong_dat) # 2,997
```

```
## [1] 2997
```

```
# The variable aware is useful here because it mirrors tri90aware, yet it shows that reclassification t
xtabs(~tri90aware + awareselfreported + aware, bio_cong_dat) %>%
  ftable(.)
```



```
##                                aware      1      2
## tri90aware awareselfreported
## 1          1                    2591      0
##          2                    59        0
##          99                    1        0
## 2          1                    0        0
##          2                    0      346
##          99                    0        0
```

```
# This table shows that 4 respondents missing artselfreported data were classified as tri90aware who re
xtabs(~tri90aware + awareselfreported + artselfreported, bio_cong_dat) %>%
  ftable(.)
```

```
##                                artselfreported      1      2      99
## tri90aware awareselfreported
## 1          1                    2281    306      4
##          2                    0        0     59
##          99                    0        0      1
## 2          1                    0        0      0
##          2                    0        0     346
##          99                    0        0      0
```

```
# Some respondents marked that they were not on art and had no detectable arvs in their blood. All of t
xtabs(~tri90aware + awareselfreported + tri90art, bio_cong_dat) %>%
  ftable(.)
```

```
##                                tri90art      1      2      3
## tri90aware awareselfreported
## 1          1                    2309    282      0
##          2                    59        0      0
##          99                    1        0      0
## 2          1                    0        0      0
##          2                    0        0     346
##          99                    0        0      0
```

```
# Of those with viral loads above 200 copies/milliliter, 532 reported that they were aware and were sub
xtabs(~tri90aware + awareselfreported + vlunder200, bio_cong_dat) %>%
  ftable(.)
```

```
##                                vlunder200      1      2
## tri90aware awareselfreported
## 1          1                    2059    532
##          2                    46      13
##          99                    0        1
## 2          1                    0        0
##          2                    16     330
##          99                    0        0
```

```
# There was 1 respondent with missing arvstatus information who selfreported that they were unaware and
xtabs(~tri90aware + awareselfreported + arvstatus, bio_cong_dat) %>%
  ftable(.)
```

```
##               arvstatus      1      2      99
## tri90aware awareselfreported
## 1           1               2219  369      3
##           2               59      0      0
##           99               1      0      0
## 2           1               0      0      0
##           2               0     345      1
##           99               0      0      0
```

```
# This table shows that there were reclassifications based on vls in each category. 48 respondents were
xtabs(~tri90aware + awareselfreported + vls, bio_cong_dat) %>%
  ftable(.)
```

```
##               vls      1      2
## tri90aware awareselfreported
## 1           1           2159  432
##           2            48     11
##           99            0      1
## 2           1            0      0
##           2            35     311
##           99            0      0
```

```
# This table shows that of 59 individuals who self reported unaware, 46 were viral load suppressed (und
xtabs(~tri90aware + awareselfreported + vlunder200 + arvstatus, bio_cong_dat) %>%
  ftable(., col.vars = c("vlunder200", "arvstatus"))
```

```
##               vlunder200      1      2
##               arvstatus      1      2      99      1      2      99
## tri90aware awareselfreported
## 1           1           2015  41      3     204  328      0
##           2           46      0      0      13      0      0
##           99           0      0      0       1      0      0
## 2           1           0      0      0       0      0      0
##           2           0     16      0       0     329      1
##           99           0      0      0       0      0      0
```

```
xtabs(~tri90aware + awareselfreported + vls + arvstatus, bio_cong_dat) %>%
  ftable(., col.vars = c("vls", "arvstatus"))
```

```
##               vls      1      2
##               arvstatus      1      2      99      1      2      99
## tri90aware awareselfreported
## 1           1           2096  60      3     123  309      0
##           2           48      0      0      11      0      0
##           99           0      0      0       1      0      0
## 2           1           0      0      0       0      0      0
##           2           0     35      0       0     310      1
##           99           0      0      0       0      0      0
```

```
# No respondents who self-reported that they were unaware were asked about art. So, respondents could o
xtabs(~tri90aware + awareselfreported + artselfreported + tri90art, bio_cong_dat) %>%
  ftable(., col.vars = c("artselfreported", "tri90art"))
```

```
##               artselfreported  1      2      99
##               tri90art        1      2      3      1      2      3      1      2      3
## tri90aware awareselfreported
## 1          1          2281    0    0    24  282    0    4    0    0
##          2          0    0    0    0    0    0    59    0    0
##          99          0    0    0    0    0    0    1    0    0
## 2          1          0    0    0    0    0    0    0    0    0
##          2          0    0    0    0    0    0    0    0  346
##          99          0    0    0    0    0    0    0    0    0
```

```
# Of those self-reported on art, 90/2281 or 3.9% of cases had no detectable arvs or were missing data.
xtabs(~tri90aware + awareselfreported + artselfreported + arvstatus, bio_cong_dat) %>%
  ftable(., col.vars = c("artselfreported", "arvstatus"))
```

```
##               artselfreported  1      2      99      1      2      99      1      2      99
##               arvstatus        1      2      99      1      2      99      1      2      99
## tri90aware awareselfreported
## 1          1          2191   87    3    24  282    0    4    0    0
##          2          0    0    0    0    0    0    59    0    0
##          99          0    0    0    0    0    0    1    0    0
## 2          1          0    0    0    0    0    0    0    0    0
##          2          0    0    0    0    0    0    0  345    1
##          99          0    0    0    0    0    0    0    0    0
```

```
# 60 respondents who had detectable arvs in their blood were reclassified as tri90aware from self-report
xtabs(~tri90aware + awareselfreported + arvstatus + tri90art, bio_cong_dat) %>%
  ftable(., col.vars = c("arvstatus", "tri90art"))
```

```
##               arvstatus  1      2      99
##               tri90art   1      2      3      1      2      3      1      2      3
## tri90aware awareselfreported
## 1          1          2219    0    0   87  282    0    3    0    0
##          2          59    0    0    0    0    0    0    0    0
##          99          1    0    0    0    0    0    0    0    0
## 2          1          0    0    0    0    0    0    0    0    0
##          2          0    0    0    0    0  345    0    0    1
##          99          0    0    0    0    0    0    0    0    0
```

```
#This table shows that somehow 122 individuals with viral loads under 1000 copies/milliliter were class
xtabs(~vlunder200 + vls, bio_dat)
```

```
##          vls
## vlunder200  1    2
##          1 2165    0
##          2  122  768
```

“vlunder200”

The variable, vlunder200, was coded as `ifelse(vlunder200 < 200, 1, 2)`, however, the variable was still a character when it ran. So, hundreds of observations had been misclassified. It should be fine now.

Second 90 (arv status)

If a person self reports that they are unaware and has detectable arvs in their blood, what is the probability of them being tri90aware?

If a person self reports that they are unaware and is viral load suppressed (under 200 copies/ milliliter), what is the probability of them being tri90aware?

If a person self reports that they are aware and has detectable arvs in their blood, what is the probability of them self-reporting that they are on arvs?

We cannot test if a person has arvs in their blood nor if they self-report their arv status because they are not asked these questions if the respondent is self-reported unaware

lca with pre-processed variables

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

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

# simple model
# bio_lca_prep = bio_dat_lca %>%
#   select(awareselfreported)

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
```

```
outcomeProbs(bio_lca)
```

```
## Class 1
##          Outcome p      2.5 %      97.5 %
## awareselfreported 0.50145245 0.46319656 0.53969135
## arvstatus         0.02508350 0.01048525 0.05879867
## vlunder200        0.04636324 0.02781268 0.07631544
## Class 2
##          Outcome p      2.5 %      97.5 %
## awareselfreported 0.9772817 0.9698269 0.9829271
## arvstatus         0.9872169 0.9768865 0.9929634
## vlunder200        0.9117746 0.8982707 0.9236383
```

```
BIC(bio_lca)
```

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

```
outcome = NULL
for (i in 1:10) {
  outcome[[i]] = bio_dat_lca %>%
    randomLCA(., freq=.$freq, nclass = i)
}

BIC = NULL
for(i in 1:10) {
  BIC[[i]] = BIC(outcome[[i]])
}

bic_data = NULL
for (i in 1:10) {
  bic_data[[i]] =
    data.frame(nclasses = i,
              bic = BIC[[i]])
}

tibble(bic_data) %>%
  unnest() %>%
  knitr::kable()
```

nclasses	bic
1	27917.625
2	13162.807
3	10673.623
4	9143.888
5	9140.159
6	9151.988
7	9242.600
8	9329.248
9	9411.271
10	9501.205

There are NAs in all the observations that are tri90aware == 0.

```
# assume that tri90aware is the truth

bio_mod_lca = bio_dat_lca %>%
  mutate(across(.funs = as.numeric)) %>%
  na.exclude()

nrow(bio_mod_lca)

bio_mod_lca %>%
  group_by(tri90aware) %>%
  summarize(sum = n())

bio_unr = bio_dat_lca %>%
  filter(tri90aware == "0")

bio_r = bio_dat_lca %>%
  filter(tri90aware == "1")

bio_mod = glm(tri90aware ~ . -aware -awareselfreported -tri90 -tri90aware, data = bio_mod_lca, family =

skimr::skim(bio_cong_dat)
skimr::skim(bio_mod_lca)
```