

# R Notebook

```
# prevalence paper analysis
sr_c <- readRDS("prevalance.rds") #place data in working directory
sr_c <- sr_c |>
  dplyr::mutate_at(c("age_questionnaire", "dentist_12months", "rate"), ~as.numeric())

library(Hmisc)

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##   format.pval, units

label(sr_c$age_questionnaire) = "Age"
label(sr_c$gender_questionnaire) = "Gender"
label(sr_c$marital_status) = "Marital Status"
label(sr_c$religion) = "Religion"
label(sr_c$level_of_education) = "Level of Education"
# label(sr_c$number_of_years_on_treatment)='Start of treatment Date'
label(sr_c$dentist_12months) = "5. How often did you go to the dentist during the past 12
  ↪ months? "
label(sr_c$rate) = "Salivary flowrate per minute"
label(sr_c$hpv_nos) = "Number of HPV types detected"
label(sr_c$hpv_nos2) = "Combinations of HPV types identified"
label(sr_c$hpvst) = "Result of HPV testing"

sr_c$gender_questionnaire.factor = factor(sr_c$gender_questionnaire, levels = c("1",
  "2"))
sr_c$marital_status.factor = factor(sr_c$marital_status, levels = c("1", "2", "3",
  "4", "5"))
sr_c$religion.factor = factor(sr_c$religion, levels = c("1", "2", "3", "4", "5",
  "6", "7", "8"))
sr_c$level_of_education.factor = factor(sr_c$level_of_education, levels = c("1",
  "2", "3", "4", "5", "6", "7"))
sr_c$dentist_12months.factor = factor(sr_c$dentist_12months, levels = c("1", "2",
  "3", "4", "5", "6", "7", "99"))
sr_c$cigarettes.factor = factor(sr_c$cigarettes, levels = c("1", "2", "3", "4", "5",
  "6"))
sr_c$drinks_per_day.factor = factor(sr_c$drinks_per_day, levels = c("0", "1", "2",
  "3", "4", "5", "99"))

levels(sr_c$gender_questionnaire.factor) = c("Male", "Female")
levels(sr_c$marital_status.factor) = c("Single", "Married", "Divorced/separated",
  "Cohabiting", "Widowed")
```

```

levels(sr_c$religion.factor) = c("Catholic", "Protestant/Anglican", "Pentecostal/Born
  ↪ Again",
  "Muslim", "Adventist", "Atheist", "Traditionalist", "Other")
levels(sr_c$level_of_education.factor) = c("Postgraduate Degree", "College/University
  ↪ completed",
  "A level completed/ Tertiary institution after O level (S6)", "Secondary school (O
  ↪ level) completed (S4-S5)",
  "Primary school completed (P7-P3)", "Less than primary school (P1-P6)", "No formal
  ↪ schooling (No Education)")
levels(sr_c$dentist_12months.factor) = c("Once", "Twice", "Three times", "Four times",
  "More than four times", "I had no visit to dentist during the past 12 months",
  "I have never received dental care/visited a dentist", "I dont know/dont remember")
levels(sr_c$cigarettes.factor) = c("Never (1)", "Seldom (2)", "Several times a month
  ↪ (3)",
  "Once a week (4)", "Several times a week (5)", "Every day (6)")
levels(sr_c$drinks_per_day.factor) = c("Less than 1 drink", "1 drink", "2 drinks",
  "3 drinks", "4 drinks", "5 or more drinks", "Did not drink alcohol during the past 30
  ↪ days")

label(sr_c$gender_questionnaire.factor) = "Gender"
label(sr_c$marital_status.factor) = "Marital Status"
label(sr_c$religion.factor) = "Religion"
label(sr_c$level_of_education.factor) = "Level of Education"
colnames(sr_c)[39] <- "How often did you go to the dentist during the past 12 months?"
label(sr_c$cigarettes.factor) <- "Cigarette consumption"
label(sr_c$drinks_per_day.factor) <- "Alcohol consumption"
library(lubridate)

```

```

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

```

```
library(dplyr)
```

```

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:Hmisc':
##
##   src, summarize

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

```

```

sr_c |>
  dplyr::select(25, 35, 36, 37, 38, 39, 1, 23, 32, 40, 41) |>
  tidyr::drop_na() |>
  gtsummary::tbl_summary(statistic = list(gtsummary::all_continuous() ~ "{mean}
→ ({sd})")) |>
  gtsummary::modify_caption("**Table 1. Participant Characteristics**") |>
  # for nicer display on website
gtsummary::as_flex_table() #|>flextable::save_as_docx(path = './table1.docx')

```

```

## Warning: fonts used in `flextable` are ignored because the `pdflatex` engine is
## used and not `xelatex` or `lualatex`. You can avoid this warning by using the
## `set_flextable_defaults(fonts_ignore=TRUE)` command or use a compatible engine
## by defining `latex_engine: xelatex` in the YAML header of the R Markdown
## document.

```

Table 1: **Table 1. Participant Characteristics**

| Characteristic               | N = 1,712 <sup>1</sup> |
|------------------------------|------------------------|
| Age                          | 45 (10)                |
| Gender                       |                        |
| Male                         | 534 (31%)              |
| Female                       | 1,178 (69%)            |
| Marital Status               |                        |
| Single                       | 112 (6.5%)             |
| Married                      | 700 (41%)              |
| Divorced/separated           | 540 (32%)              |
| Cohabiting                   | 140 (8.2%)             |
| Widowed                      | 220 (13%)              |
| Religion                     |                        |
| Catholic                     | 567 (33%)              |
| Protestant/Anglican          | 464 (27%)              |
| Pentecostal/Born Again       | 318 (19%)              |
| Muslim                       | 320 (19%)              |
| Adventist                    | 36 (2.1%)              |
| Atheist                      | 2 (0.1%)               |
| Traditionalist               | 3 (0.2%)               |
| Other                        | 2 (0.1%)               |
| Level of Education           |                        |
| Postgraduate Degree          | 3 (0.2%)               |
| College/University completed | 73 (4.3%)              |

Table 1: \*\*Table 1. Participant Characteristics\*\*

| Characteristic                                                 | N = 1,712 <sup>1</sup> |
|----------------------------------------------------------------|------------------------|
| A level completed/ Tertiary institution after O level (S6)     | 145 (8.5%)             |
| Secondary school (O level) completed (S4-S5)                   | 265 (15%)              |
| Primary school completed (P7-S3)                               | 566 (33%)              |
| Less than primary school (P1-P6)                               | 548 (32%)              |
| No formal schooling (No Education)                             | 112 (6.5%)             |
| How often did you go to the dentist during the past 12 months? |                        |
| Once                                                           | 221 (13%)              |
| Twice                                                          | 94 (5.5%)              |
| Three times                                                    | 20 (1.2%)              |
| Four times                                                     | 5 (0.3%)               |
| More than four times                                           | 16 (0.9%)              |
| I had no visit to dentist during the past 12 months            | 1,121 (65%)            |
| I have never received dental care/visited a dentist            | 232 (14%)              |
| I dont know/dont remember                                      | 3 (0.2%)               |
| Salivary flowrate per minute                                   | 1.20 (0.59)            |
| Number of HPV types detected                                   |                        |
| negative                                                       | 561 (33%)              |
| No_reaction                                                    | 478 (28%)              |
| Single                                                         | 199 (12%)              |
| Multiple                                                       | 474 (28%)              |
| Result of HPV testing                                          |                        |
| negative                                                       | 561 (33%)              |
| positive                                                       | 1,151 (67%)            |
| Cigarette consumption                                          |                        |
| Never (1)                                                      | 1,634 (95%)            |
| Seldom (2)                                                     | 15 (0.9%)              |
| Several times a month (3)                                      | 4 (0.2%)               |
| Once a week (4)                                                | 6 (0.4%)               |
| Several times a week (5)                                       | 27 (1.6%)              |
| Every day (6)                                                  | 26 (1.5%)              |
| Alcohol consumption                                            |                        |
| Less than 1 drink                                              | 5 (0.3%)               |
| 1 drink                                                        | 118 (6.9%)             |

Table 1: \*\*Table 1. Participant Characteristics\*\*

| Characteristic                                | N = 1,712 <sup>1</sup> |
|-----------------------------------------------|------------------------|
| 2 drinks                                      | 136 (7.9%)             |
| 3 drinks                                      | 85 (5.0%)              |
| 4 drinks                                      | 40 (2.3%)              |
| 5 or more drinks                              | 127 (7.4%)             |
| Did not drink alcohol during the past 30 days | 1,201 (70%)            |

<sup>1</sup>Mean (SD); n (%)

prevalence paper analysis continued

```

broom::tidy(glm(as.factor(hpvst) ~ age_questionnaire, data = sr_c, family = binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```

## # A tibble: 2 x 7
##   term                estimate std.error statistic p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>   <dbl>   <dbl>    <dbl>
## 1 (Intercept)          1.57    0.226      2.00  0.0452    1.01    2.46
## 2 age_questionnaire    1.01    0.00498    1.21  0.225    0.996    1.02

```

```

broom::tidy(glm(as.factor(hpvst) ~ gender_questionnaire, data = sr_c, family = binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```

## # A tibble: 2 x 7
##   term                estimate std.error statistic p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>   <dbl>   <dbl>    <dbl>
## 1 (Intercept)          1.85    0.193      3.20  0.00139    1.27    2.72
## 2 gender_questionnaire  1.06    0.111    0.555  0.579    0.855    1.32

```

```

broom::tidy(glm(as.factor(hpvst) ~ level_of_education.factor, data = sr_c, family =
  ↪ binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```

## # A tibble: 7 x 7
##   term                estimate std.error statistic p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>   <dbl>   <dbl>    <dbl>
## 1 (Intercept)          0.500    1.22    -0.566  0.571    0.0232    5.22
## 2 level_of_education.fa~  3.29    1.25    0.953  0.340    0.301    72.6
## 3 level_of_education.fa~  3.62    1.24    1.04    0.299    0.339    78.9
## 4 level_of_education.fa~  4.39    1.23    1.20    0.230    0.415    95.2
## 5 level_of_education.fa~  3.85    1.23    1.10    0.273    0.366    83.1
## 6 level_of_education.fa~  4.56    1.23    1.24    0.217    0.434    98.6
## 7 level_of_education.fa~  4.40    1.24    1.19    0.233    0.409    96.5

```

```

broom::tidy(glm(as.factor(hpvst) ~ religion.factor, data = sr_c, family = binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

```

```
## # A tibble: 8 x 7
```

| ##   | term                  | estimate | std.error | statistic | p.value  | conf.low | conf.high |
|------|-----------------------|----------|-----------|-----------|----------|----------|-----------|
| ##   | <chr>                 | <dbl>    | <dbl>     | <dbl>     | <dbl>    | <dbl>    | <dbl>     |
| ## 1 | (Intercept)           | 2.02e+0  | 0.0892    | 7.89      | 2.97e-15 | 1.70e+ 0 | 2.41      |
| ## 2 | religion.factorProte~ | 1.07e+0  | 0.134     | 0.484     | 6.29e- 1 | 8.21e- 1 | 1.39      |
| ## 3 | religion.factorPente~ | 9.02e-1  | 0.147     | -0.702    | 4.83e- 1 | 6.76e- 1 | 1.20      |
| ## 4 | religion.factorMuslim | 1.13e+0  | 0.150     | 0.787     | 4.31e- 1 | 8.40e- 1 | 1.51      |
| ## 5 | religion.factorAdven~ | 8.75e-1  | 0.358     | -0.372    | 7.10e- 1 | 4.40e- 1 | 1.81      |
| ## 6 | religion.factorAthei~ | 4.95e-1  | 1.42      | -0.497    | 6.19e- 1 | 1.95e- 2 | 12.6      |
| ## 7 | religion.factorTradi~ | 9.89e-1  | 1.23      | -0.00862  | 9.93e- 1 | 9.42e- 2 | 21.4      |
| ## 8 | religion.factorOther  | 3.86e+5  | 379.      | 0.0340    | 9.73e- 1 | 5.92e-21 | NA        |

```

broom::tidy(glm(as.factor(hpvst) ~ marital_status.factor, data = sr_c, family =
  ↪ binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```
## # A tibble: 5 x 7
```

| ##   | term                   | estimate | std.error | statistic | p.value | conf.low | conf.high |
|------|------------------------|----------|-----------|-----------|---------|----------|-----------|
| ##   | <chr>                  | <dbl>    | <dbl>     | <dbl>     | <dbl>   | <dbl>    | <dbl>     |
| ## 1 | (Intercept)            | 1.65     | 0.193     | 2.60      | 0.00945 | 1.14     | 2.43      |
| ## 2 | marital_status.factor~ | 1.32     | 0.210     | 1.33      | 0.184   | 0.871    | 1.99      |
| ## 3 | marital_status.factor~ | 1.31     | 0.214     | 1.25      | 0.212   | 0.854    | 1.98      |
| ## 4 | marital_status.factor~ | 1.17     | 0.263     | 0.609     | 0.542   | 0.701    | 1.96      |
| ## 5 | marital_status.factor~ | 1.08     | 0.239     | 0.326     | 0.745   | 0.674    | 1.72      |

```

broom::tidy(glm(as.factor(hpvst) ~ cigarettes.factor, data = sr_c, family = binomial),
  exponentiate = TRUE, conf.int = TRUE)

```

```
## # A tibble: 6 x 7
```

| ##   | term                  | estimate | std.error | statistic | p.value  | conf.low | conf.high |
|------|-----------------------|----------|-----------|-----------|----------|----------|-----------|
| ##   | <chr>                 | <dbl>    | <dbl>     | <dbl>     | <dbl>    | <dbl>    | <dbl>     |
| ## 1 | (Intercept)           | 2.05     | 0.0527    | 13.7      | 1.59e-42 | 1.85     | 2.28      |
| ## 2 | cigarettes.factorSel~ | 0.974    | 0.550     | -0.0485   | 9.61e- 1 | 0.344    | 3.14      |
| ## 3 | cigarettes.factorSev~ | 0.162    | 1.16      | -1.57     | 1.16e- 1 | 0.00802  | 1.27      |
| ## 4 | cigarettes.factorUnc~ | 2.43     | 1.10      | 0.811     | 4.17e- 1 | 0.391    | 46.7      |
| ## 5 | cigarettes.factorSev~ | 1.16     | 0.425     | 0.342     | 7.33e- 1 | 0.520    | 2.82      |
| ## 6 | cigarettes.factorEve~ | 1.10     | 0.428     | 0.213     | 8.32e- 1 | 0.489    | 2.68      |

```
broom::tidy(glm(as.factor(hpvst) ~ drinks_per_day.factor, data = sr_c, family =  
  ↪ binomial),  
  exponentiate = TRUE, conf.int = TRUE)
```

```
## # A tibble: 7 x 7
##   term                estimate std.error statistic p.value conf.low conf.high
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        1.50      0.913     0.444    0.657    0.249    11.4
## 2 drinks_per_day.factor~ 1.35    0.934     0.322    0.748    0.173     8.47
## 3 drinks_per_day.factor~ 1.30    0.931     0.285    0.775    0.167     8.14
## 4 drinks_per_day.factor~ 1.29    0.941     0.268    0.788    0.163     8.19
## 5 drinks_per_day.factor~ 1.38    0.973     0.334    0.738    0.167     9.37
## 6 drinks_per_day.factor~ 1.30    0.932     0.283    0.777    0.167     8.14
## 7 drinks_per_day.factor~ 1.40    0.915     0.365    0.715    0.183     8.46
```

```
print(broom::tidy(glm(as.factor(hpvst) ~ age_questionnaire + gender_questionnaire +
  marital_status.factor + religion.factor + level_of_education.factor +
  ↪ cigarettes.factor +
  drinks_per_day.factor, data = sr_c, family = binomial), exponentiate = TRUE,
  conf.int = TRUE), n = 30)
```

[illegible]

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```

```
## # A tibble: 31 x 7
##   term                estimate std.error statistic p.value conf.low conf.high
##   <chr>                <dbl>      <dbl>      <dbl>   <dbl>   <dbl>   <dbl>
## 1 (Intercept)         1.40e-1    1.70      -1.16    0.247 3.00e- 3    3.64
## 2 age_questionnaire   1.01e+0    0.00589    1.24    0.214 9.96e- 1    1.02
## 3 gender_questionnaire 1.13e+0    0.127     0.986   0.324 8.83e- 1    1.45
## 4 marital_status.facto~ 1.13e+0    0.235     0.512   0.608 7.07e- 1    1.78
## 5 marital_status.facto~ 1.06e+0    0.244     0.227   0.820 6.52e- 1    1.70
## 6 marital_status.facto~ 9.97e-1    0.280    -0.0115 0.991 5.74e- 1    1.73
## 7 marital_status.facto~ 8.13e-1    0.286    -0.722   0.471 4.63e- 1    1.42
## 8 religion.factorProte~ 1.08e+0    0.136     0.585   0.558 8.30e- 1    1.41
## 9 religion.factorPente~ 9.20e-1    0.156    -0.535   0.593 6.79e- 1    1.25
## 10 religion.factorMuslim 1.10e+0    0.156     0.620   0.536 8.12e- 1    1.50
## 11 religion.factorAdven~ 9.09e-1    0.367    -0.262   0.794 4.50e- 1    1.92
## 12 religion.factorAthei~ 5.31e-1    1.48     -0.427   0.669 1.91e- 2    14.7
## 13 religion.factorTradi~ 1.99e+0    1.42     0.486   0.627 1.52e- 1    69.3
## 14 religion.factorOther 4.31e+5 375.      0.0346   0.972 1.16e-20    NA
## 15 level_of_education.f~ 4.33e+0    1.38     1.06    0.289 3.34e- 1    135.
## 16 level_of_education.f~ 4.73e+0    1.38     1.13    0.259 3.69e- 1    147.
## 17 level_of_education.f~ 5.79e+0    1.37     1.28    0.201 4.55e- 1    179.
## 18 level_of_education.f~ 5.04e+0    1.37     1.18    0.238 3.98e- 1    155.
## 19 level_of_education.f~ 5.90e+0    1.37     1.30    0.195 4.67e- 1    182.
## 20 level_of_education.f~ 5.81e+0    1.38     1.27    0.203 4.47e- 1    182.
## 21 cigarettes.factorSel~ 1.02e+0    0.559    0.0346   0.972 3.54e- 1    3.33
## 22 cigarettes.factorSev~ 1.72e-1    1.17    -1.51    0.131 8.41e- 3    1.37
## 23 cigarettes.factorOnc~ 2.34e+0    1.11     0.770   0.441 3.68e- 1    45.4
## 24 cigarettes.factorSev~ 1.29e+0    0.442     0.573   0.567 5.61e- 1    3.26
## 25 cigarettes.factorEve~ 1.15e+0    0.445     0.308   0.758 4.95e- 1    2.90
## 26 drinks_per_day.facto~ 1.49e+0    0.946     0.422   0.673 1.87e- 1    9.56
## 27 drinks_per_day.facto~ 1.41e+0    0.943     0.361   0.718 1.77e- 1    8.97
## 28 drinks_per_day.facto~ 1.34e+0    0.955     0.307   0.759 1.66e- 1    8.74
## 29 drinks_per_day.facto~ 1.41e+0    0.986     0.346   0.729 1.66e- 1    9.75
## 30 drinks_per_day.facto~ 1.43e+0    0.946     0.378   0.706 1.79e- 1    9.17
## # i 1 more row
```

```
# test for multicollinearity using VIF
car::vif((glm(as.factor(hpvst) ~ age_questionnaire + gender_questionnaire +
  ↳ marital_status.factor +
  ↳ religion.factor + level_of_education.factor + cigarettes.factor +
  ↳ drinks_per_day.factor,
  data = sr_c, family = binomial))) # A VIF value greater than 10 is often considered
  ↳ indicative of high multicollinearity.
```

```
##                               GVIF Df GVIF^(1/(2*Df))
```



```
## age_questionnaire      1.401448  1      1.183828
## gender_questionnaire   1.298267  1      1.139415
## marital_status.factor  1.676008  4      1.066681
## religion.factor        1.586645  7      1.033523
## level_of_education.factor 1.408288  6      1.028942
## cigarettes.factor      1.198606  5      1.018281
## drinks_per_day.factor  1.300861  6      1.022161
```

```
performance::check_collinearity((glm(as.factor(hpvst) ~ age_questionnaire +
  ↳ gender_questionnaire +
  marital_status.factor + religion.factor + level_of_education.factor +
  ↳ cigarettes.factor +
  drinks_per_day.factor, data = sr_c, family = binomial)))
```

```
## # Check for Multicollinearity
```

```
##
```

```
## Low Correlation
```

```
##
```

```
##           Term  VIF  VIF 95% CI Increased SE Tolerance
## age_questionnaire 1.40 [1.33, 1.49]          1.18      0.71
## gender_questionnaire 1.30 [1.23, 1.38]          1.14      0.77
## marital_status.factor 1.68 [1.58, 1.79]          1.29      0.60
## religion.factor 1.59 [1.49, 1.70]          1.26      0.63
## level_of_education.factor 1.41 [1.33, 1.50]          1.19      0.71
## cigarettes.factor 1.20 [1.14, 1.28]          1.09      0.83
## drinks_per_day.factor 1.30 [1.24, 1.39]          1.14      0.77
## Tolerance 95% CI
## [0.67, 0.75]
## [0.72, 0.81]
## [0.56, 0.63]
## [0.59, 0.67]
## [0.67, 0.75]
## [0.78, 0.87]
## [0.72, 0.81]
```

```
# prevalence paper second table
```

```
sr_c |>
```

```
  dplyr::select(3:20, 27, 28) |>
```

```
  tidyr::drop_na() |>
```

```
  tidyr::pivot_longer(cols = tidyselect::starts_with("Hpv_"), names_to = "HPV",
    values_to = "number", values_drop_na = TRUE) |>
```

```
  gtsummary::tbl_cross(row = HPV, col = number, percent = "column") |>
```

```
  gtsummary::modify_caption("**Table 2. HPV types**") |>
```

```
  # for nicer display on website
```

```
gtsummary::as_flex_table() #/>flextable::save_as_docx(path =
```

```
  ↳ 'F:/buwembo/ameded/data/table2.docx')
```

```
## Warning: fonts used in `flextable` are ignored because the `pdflatex` engine is
## used and not `xelatex` or `lualatex`. You can avoid this warning by using the
## `set_flextable_defaults(fonts_ignore=TRUE)` command or use a compatible engine
## by defining `latex_engine: xelatex` in the YAML header of the R Markdown
## document.
```

Table 2: **Table 2. HPV types**

|        | number        |              |               |
|--------|---------------|--------------|---------------|
|        | 0             | 1            | Total         |
| HPV    |               |              |               |
| Hpv_11 | 1,701 (6.3%)  | 11 (0.6%)    | 1,712 (5.9%)  |
| Hpv_18 | 1,614 (5.9%)  | 98 (5.1%)    | 1,712 (5.9%)  |
| Hpv_31 | 1,577 (5.8%)  | 135 (7.1%)   | 1,712 (5.9%)  |
| Hpv_33 | 1,634 (6.0%)  | 78 (4.1%)    | 1,712 (5.9%)  |
| Hpv_35 | 1,665 (6.1%)  | 47 (2.5%)    | 1,712 (5.9%)  |
| Hpv_39 | 1,650 (6.1%)  | 62 (3.3%)    | 1,712 (5.9%)  |
| Hpv_42 | 1,423 (5.2%)  | 289 (15%)    | 1,712 (5.9%)  |
| Hpv_43 | 1,505 (5.5%)  | 207 (11%)    | 1,712 (5.9%)  |
| Hpv_44 | 1,655 (6.1%)  | 57 (3.0%)    | 1,712 (5.9%)  |
| Hpv_45 | 1,406 (5.2%)  | 306 (16%)    | 1,712 (5.9%)  |
| Hpv_51 | 1,642 (6.0%)  | 70 (3.7%)    | 1,712 (5.9%)  |
| Hpv_52 | 1,647 (6.1%)  | 65 (3.4%)    | 1,712 (5.9%)  |
| Hpv_56 | 1,638 (6.0%)  | 74 (3.9%)    | 1,712 (5.9%)  |
| Hpv_58 | 1,634 (6.0%)  | 78 (4.1%)    | 1,712 (5.9%)  |
| Hpv_59 | 1,663 (6.1%)  | 49 (2.6%)    | 1,712 (5.9%)  |
| Hpv_66 | 1,564 (5.8%)  | 148 (7.8%)   | 1,712 (5.9%)  |
| Hpv_68 | 1,581 (5.8%)  | 131 (6.9%)   | 1,712 (5.9%)  |
| Total  | 27,199 (100%) | 1,905 (100%) | 29,104 (100%) |

```

# sub analysis for possible vaccine effect using gender and age group
sr_c |>
  dplyr::select(1, 3:20, 25, 26, 28) |>
  tibble::rowid_to_column("UID") |>
  tidyr::drop_na() |>
  tidyr::pivot_longer(cols = tidyselct::starts_with("Hpv_"), names_to = "HPV",
    values_to = "number", values_drop_na = TRUE) -> srcag
srcag <- srcag[!(srcag$number %in% 0), ] # reduce the number of rows

srcag <- srcag |>
  dplyr::mutate(vacc = case_when(HPV == "Hpv_16" ~ 1, HPV == "Hpv_18" ~ 1)) |>
  dplyr::mutate(age2 = case_when(age_questionnaire <= 24 ~ "positive",
  ↪ age_questionnaire >
    24 ~ "negative")) |>
  dplyr::mutate_at("vacc", ~tidyr::replace_na(., 0)) |>
  dplyr::mutate(sex2 = case_when(gender_questionnaire == "1" ~ "Male",
  ↪ gender_questionnaire ==
    "2" ~ "Female"))
broom.mixed::tidy((lme4::glmer(vacc ~ sex2 + (1 | UID), family = "binomial", srcag)),

```

```
effects = "fixed", conf.int = TRUE, exponentiate = TRUE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## # A tibble: 2 x 8
##   effect term          estimate std.error statistic    p.value conf.low conf.high
##   <chr> <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 fixed (Intercept)  0.0523  0.00660   -23.4  8.82e-121  0.0408  0.0670
## 2 fixed sex2Male     1.12    0.249     0.523 6.01e- 1  0.727  1.73
```

```
broom.mixed::tidy((lme4::glmer(vacc ~ age2 + (1 | UID), family = "binomial", srcag)),
  effects = "fixed", conf.int = TRUE, exponentiate = TRUE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## # A tibble: 2 x 8
##   effect term          estimate std.error statistic    p.value conf.low conf.high
##   <chr> <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 fixed (Intercept)  0.0543  0.00575   -27.5  1.43e-166  0.0441  0.0668
## 2 fixed age2positive  0.969   0.508    -0.0597 9.52e- 1  0.347  2.71
```

```
broom.mixed::tidy((lme4::glmer(vacc ~ sex2 + age2 + (1 | UID), family = "binomial",
  srcag)), effects = "fixed", conf.int = TRUE, exponentiate = TRUE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## # A tibble: 3 x 8
##   effect term          estimate std.error statistic    p.value conf.low conf.high
##   <chr> <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 fixed (Intercept)  0.0524  0.00670   -23.0  1.88e-117  0.0407  0.0673
## 2 fixed sex2Male     1.12    0.249     0.523 6.01e- 1  0.727  1.73
## 3 fixed age2positive  0.968   0.507    -0.0620 9.51e- 1  0.347  2.70
```

```
# test for multicollinearity
car::vif((lme4::glmer(vacc ~ sex2 + age2 + (1 | UID), family = "binomial", srcag))) # A
↪ VIF value greater than 10 is often considered indicative of high multicollinearity
```

```
## boundary (singular) fit: see help('isSingular')
```

```
##   sex2   age2
## 1.00002 1.00002
```

```
performance::check_collinearity((lme4::glmer(vacc ~ sex2 + age2 + (1 | UID), family =
  ↪ "binomial",
  srcag)))
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
## Term   VIF   VIF 95% CI Increased SE Tolerance Tolerance 95% CI
## sex2 1.00 [1.00, Inf]          1.00      1.00      [0.00, 1.00]
## age2 1.00 [1.00, Inf]          1.00      1.00      [0.00, 1.00]
```