HPV Project
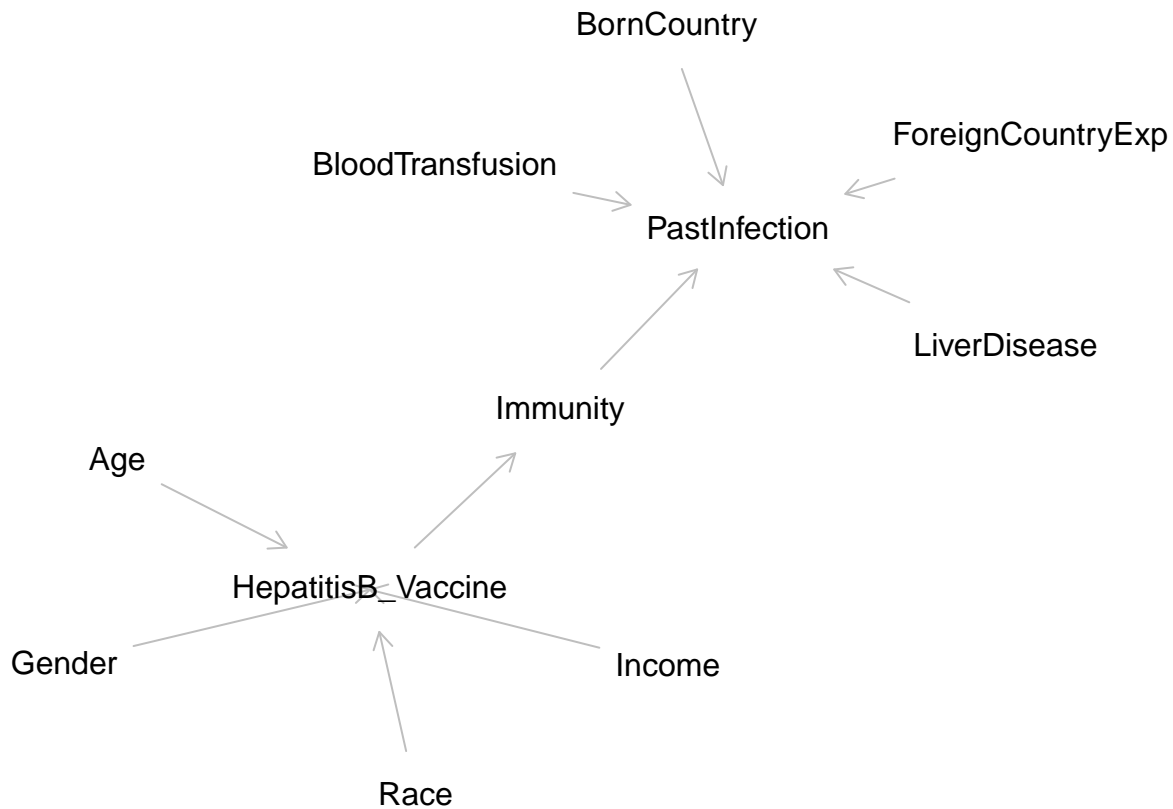
Hong and Moiyyad

2024-10-28

Our Assumption

Predictive Model

We will use *Immunity, PastInfection, ForeignCountryExp, BornCountry, BloodTransfusion, LiverDisease* as predictors to predict whether *PastInfection* or not.

- Immunity: we will use predicted immunity to replace the missing values in **LBXHBS** (Hepatitis B Surface Antibody);the hepatitis B surface antibody (anti-HBs) demonstrates immunity acquired through vaccination.

- BloodTransfusion: indicated by **MCQ092** (Ever received blood transfusion);

- ForeignCountryExp: indicated by **DMQADFC** (Whether served in a foreign country).

- BornCountry: indicated by **DMDBORN4** (Born Country).

- LiverDisease: indicated by **MCQ160L** (Ever told you had any liver condition (for age>-20))

- PastInfection: indicated by **LBDHBC** (Hepatitis B Core Antigen) or **MCQ160L** (Ever told you had any liver condition)

- CurrentInfection: indicated by **LBDHBG** (Hepatitis B Surface Antigen). Number of Non-Missing is 402. of Missing observations. For these participants with non-missing LBDHBG (31 Positive, 375 Negative), they all had past infection experience indicated by **LBXHBC** (Hepatitis B Core Antibody).

BornCountry

BloodTransfusion

PastInfection ⟵ Immunity

ForeignCountryExp

LiverDisease
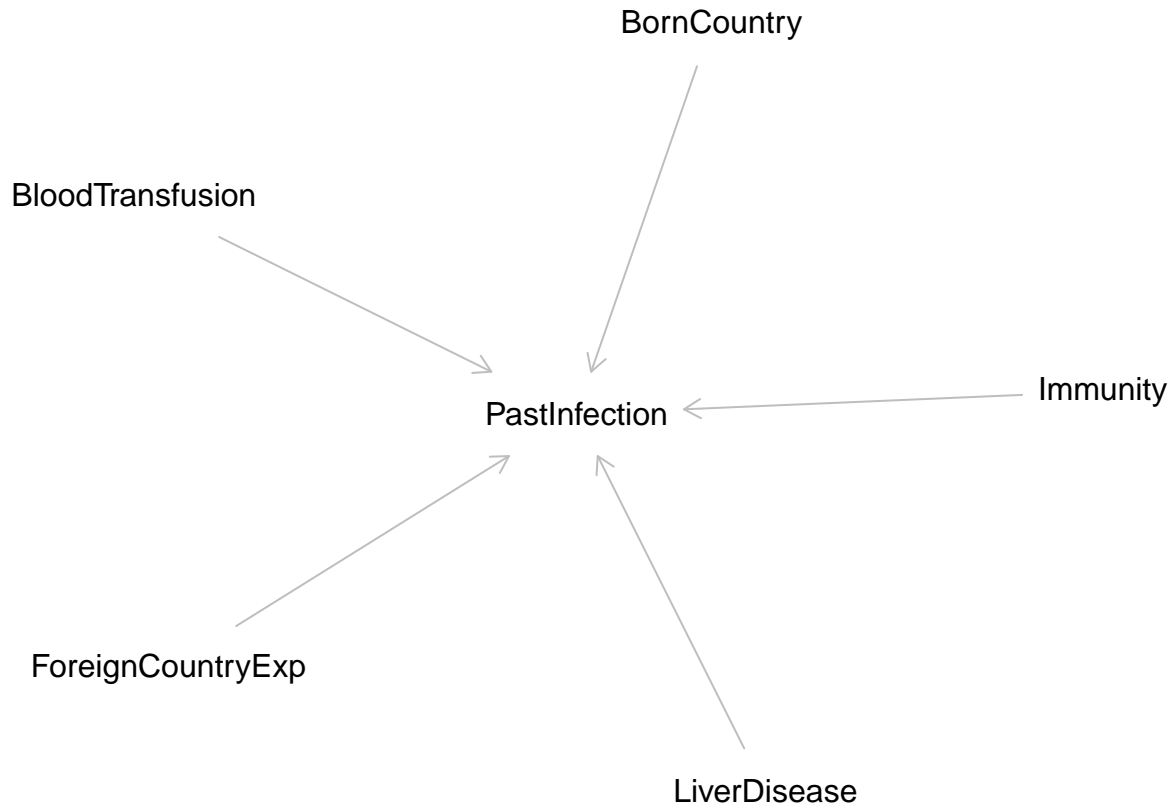
Our DataSet

We merged several dataset:

- DEMO_J.XPT (Demographic dataset);

- HEPB_S_J.XPT (immunization dataset);
- HEPBD_J.XPT (Blood test dataset);
- HEQ_J.XPT (Hepatitis questionare dataset);
- IMQ_J.XPT (vaccine dataset);
- MCQ_J.XPT (Medical condition dataset).

Survey Design

```
HepatitisB_svyd<-svydesign(data = FullDat,ids = ~`SDMVPSU:Masked variance pseudo-PSU`,
                           strata = ~`SDMVSTRA:Masked variance pseudo-stratum`,
                           weights = ~`WTMEC2YR:Full Sample 2 year MEC exam weight`,
                           nest = TRUE)
Means<-svymean(~`LBXHBC:Hepatitis B Core Antibody`,HepatitisB_svyd)
barplot(Means)
```
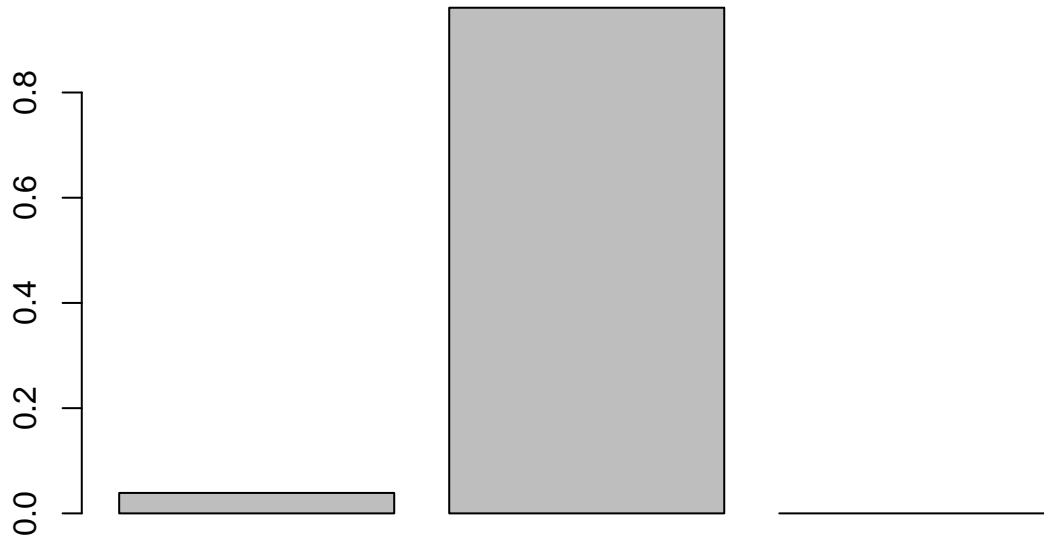

'LBXHBC:Hepatitis B Core Antibody'Positive

## Step 1: The probability for a participant taking the HBV

$$log \frac{P(Vaccine=Yes|Age,Race,Gender,Income)}{P(Vaccine=No|Age,Race,Gender,Income)} = b_0 + b_1 * Age + b_2 * Race + b_3 * Gender + b_4 * Income$$

```
##
## Call:
## svyglm(formula = WhetherVaccine ~ `RIDAGEYR:Age(0-80)` + `RIAGENDR:Gender` +
##     `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)` +
##     FamilyIncome_Low + FamilyIncome_Medium, design = P1, family = quasibinomial())
##
## Survey design:
## svydesign(ids = ~`SDMVPSU:Masked variance pseudo-PSU`, strata = ~`SDMVSTRA:Masked variance pseudo-st
##     weights = ~`WTMEC2YR:Full Sample 2 year MEC exam weight`,
##     data = FullDat, nest = TRUE)
##
## Coefficients:
##                                                                               Estimate
## (Intercept)                                                                   2.155390
## `RIDAGEYR:Age(0-80)`                                                         -0.060893
## `RIAGENDR:Gender`Female                                                       0.410076
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Other Hispanic    0.519918
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic White 0.627495
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Black 0.076320
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Asian 0.831346
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Others 0.329411
```

```
## FamilyIncome_Low                                                             -0.319285
## FamilyIncome_Medium                                                          -0.087242
##                                                                              Std. Error
## (Intercept)                                                                    0.136605
## `RIDAGEYR:Age(0-80)`                                                           0.002192
## `RIAGENDR:Gender`Female                                                        0.090164
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Other Hispanic     0.148452
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic White 0.128692
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Black 0.091350
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Asian 0.136816
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Others 0.240545
## FamilyIncome_Low                                                               0.248825
## FamilyIncome_Medium                                                            0.099666
##                                                                              t value
## (Intercept)                                                                    15.778
## `RIDAGEYR:Age(0-80)`                                                          -27.777
## `RIAGENDR:Gender`Female                                                         4.548
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Other Hispanic      3.502
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic White  4.876
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Black  0.835
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Asian  6.076
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Others 1.369
## FamilyIncome_Low                                                               -1.283
## FamilyIncome_Medium                                                            -0.875
##                                                                              Pr(>|t|)
## (Intercept)                                                                    4.11e-06
## `RIDAGEYR:Age(0-80)`                                                           1.44e-07
## `RIAGENDR:Gender`Female                                                        0.003900
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Other Hispanic     0.012791
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic White 0.002778
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Black 0.435452
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Asian 0.000902
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Others 0.219905
## FamilyIncome_Low                                                               0.246764
## FamilyIncome_Medium                                                            0.415048
##
## (Intercept)                                                                    ***
## `RIDAGEYR:Age(0-80)`                                                           ***
## `RIAGENDR:Gender`Female                                                        **
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Other Hispanic     *
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic White **
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Black
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Asian ***
## `RIDRETH3:Race(Six Categories Including Non-Hispanic Asian)`Non-Hispanic Others
## FamilyIncome_Low
## FamilyIncome_Medium
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.000745)
##
## Number of Fisher Scoring iterations: 4
```
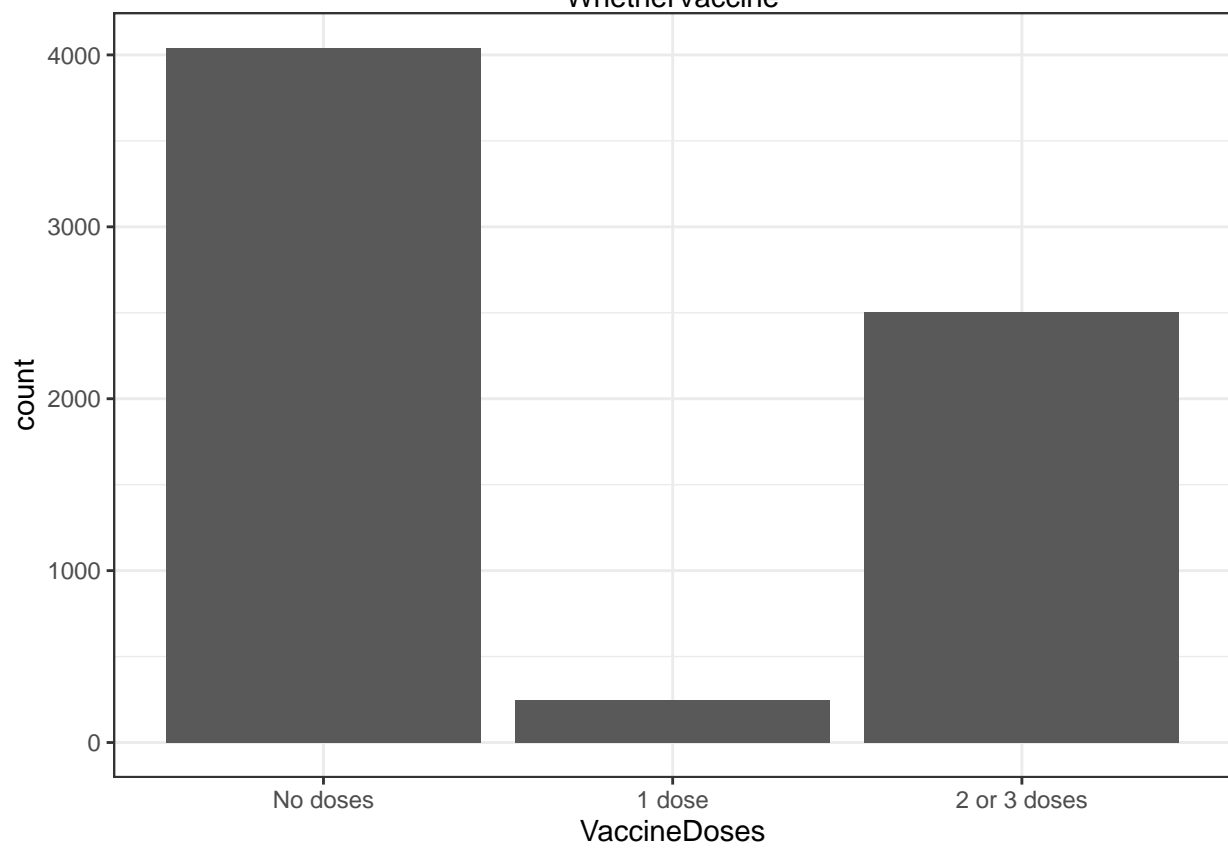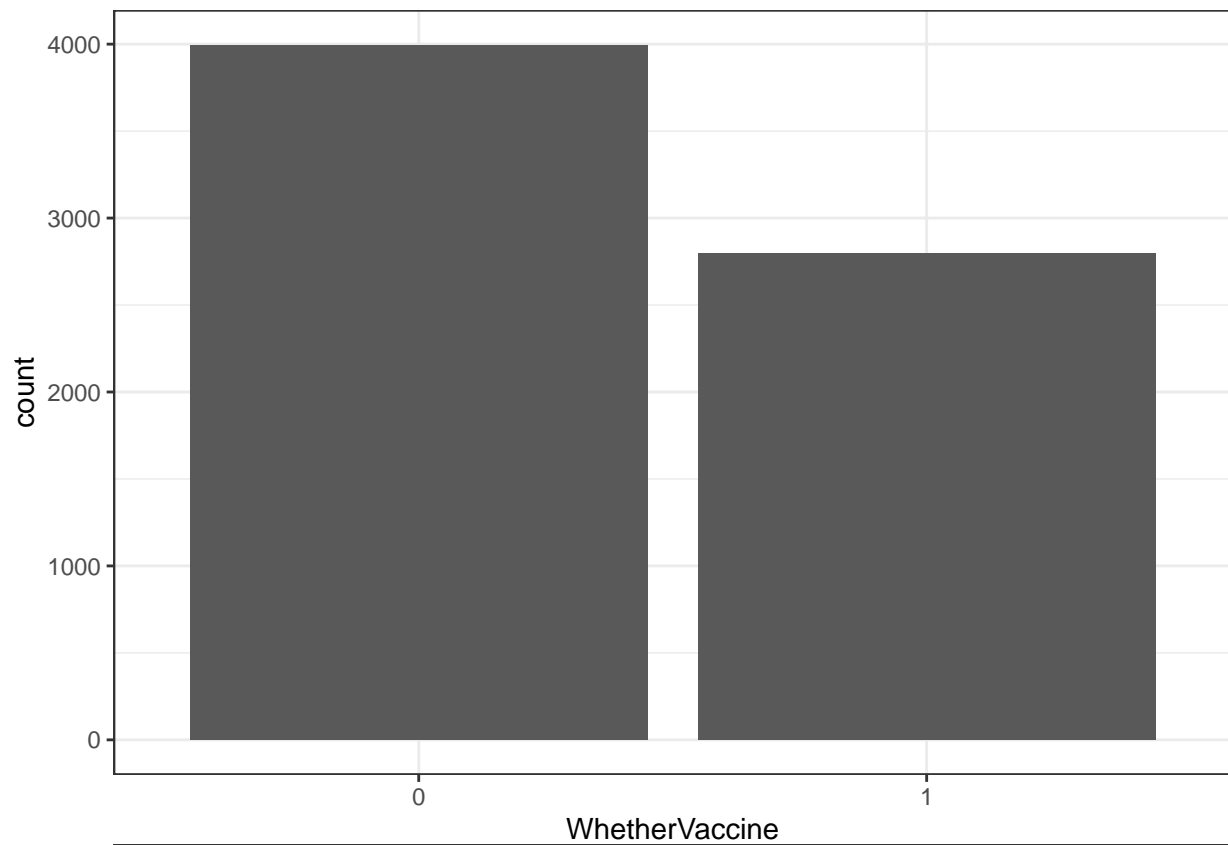
The sample size used to predict WhetherVaccineOrNot is 6793. Masked Variance weights (SDMVPSU and

SDMVSTRA) and Full Sample 2 year MEC weight (WTMEC2YR) are used in svyglm function.

Original Variable "IMQ020: Hepatitis B 3 dose", if "Refused" or "Dont Know" or "Missing", use predicted_probability to impute values. If the predicted_probability in Step 1 is larger than 0.5, classify the participants to "1 dose", below than 0.5, classify the participants to "0 dose".

Step 2:

Step 2: The probability for a participant getting immunity through the HBV.

$log \frac{P(Immunity=Yes|VaccineDoses))}{P(Immunity=No|VaccineDoses)} = \beta_0 + \beta_1 * VaccineDoses$

```
##
## Call:
## svyglm(formula = 'LBXHBS:Hepatitis B Surface Antibody' ~ VaccineDoses,
##     design = P2, family = quasibinomial())
##
## Survey design:
## svydesign(ids = ~'SDMVPSU:Masked variance pseudo-PSU', strata = ~'SDMVSTRA:Masked variance pseudo-st:
##     weights = ~'WTMEC2YR:Full Sample 2 year MEC exam weight',
##     data = FullDat, nest = TRUE)
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.49829    0.08251   18.16 1.28e-10 ***
## VaccineDoses1 dose      -0.61814    0.23590   -2.62   0.0212 *
## VaccineDoses2 or 3 doses -1.23097    0.10666  -11.54 3.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.001404)
##
## Number of Fisher Scoring iterations: 4


##                    response       min       max      mean        sd
## 1 Predicted_Pro_Immunity 0.5664365 0.8173196 0.7206227 0.1197933
```

The sample size used to predict WhetherImmunityOrNot is 6774. Masked Variance weights (SDMVPSU and SDMVSTRA) and Full Sample 2 year MEC weight (WTMEC2YR) are used in svyglm function.

Step 3: Predictive Model

$log(\frac{P(PastInfection=Yes)}{P(PastInfection=No)} = B_0 + B_1 * ImmunityProb + B_2 * ForeignCountryExp + B_3 * BornCountry + B_4 * BloodsTranfusion$

```
##
## Call:
## svyglm(formula = 'LBXHBC:Hepatitis B Core Antibody' ~ Predicted_Pro_Immunity +
##     'DMQADFC:Sever in a foreign country' + 'DMDBORN4:Born Country' +
##     'MCQ092:Ever received blood transfusion' + 'MCQ160L:Ever told you had any liver condition (for a
##     design = P3, family = quasibinomial())
##
```

```
## Survey design:
## svydesign(ids = ~'SDMVPSU:Masked variance pseudo-PSU', strata = ~'SDMVSTRA:Masked variance pseudo-st:
##     weights = ~'WTMEC2YR:Full Sample 2 year MEC exam weight',
##     data = FullDat, nest = TRUE)
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                    5.3138
## Predicted_Pro_Immunity                                        -1.9458
## 'DMQADFC:Sever in a foreign country'Yes                       -0.8589
## 'DMDBORN4:Born Country'Others                                 -1.5039
## 'MCQ092:Ever received blood transfusion'Yes                   -0.5336
## 'MCQ160L:Ever told you had any liver condition (for age>=20)'Yes  -1.5197
##                                                              Std. Error
## (Intercept)                                                    0.6122
## Predicted_Pro_Immunity                                         0.9109
## 'DMQADFC:Sever in a foreign country'Yes                        0.5090
## 'DMDBORN4:Born Country'Others                                  0.2995
## 'MCQ092:Ever received blood transfusion'Yes                    0.1810
## 'MCQ160L:Ever told you had any liver condition (for age>=20)'Yes   0.3051
##                                                              t value
## (Intercept)                                                    8.680
## Predicted_Pro_Immunity                                        -2.136
## 'DMQADFC:Sever in a foreign country'Yes                       -1.687
## 'DMDBORN4:Born Country'Others                                 -5.022
## 'MCQ092:Ever received blood transfusion'Yes                   -2.948
## 'MCQ160L:Ever told you had any liver condition (for age>=20)'Yes  -4.981
##                                                              Pr(>|t|)
## (Intercept)                                                  5.73e-06 ***
## Predicted_Pro_Immunity                                       0.058403 .
## 'DMQADFC:Sever in a foreign country'Yes                      0.122427
## 'DMDBORN4:Born Country'Others                                0.000520 ***
## 'MCQ092:Ever received blood transfusion'Yes                  0.014581 *
## 'MCQ160L:Ever told you had any liver condition (for age>=20)'Yes 0.000553 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.8903627)
##
## Number of Fisher Scoring iterations: 6
```