/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

PROGRAM: RFI.LR\_on\_Time\_since\_Vax.sas

AUTHOR: Eric Bush

CREATED: September 29, 2021

MODIFIED:

PURPOSE: RFI for logistic regression on vaccine breakthrough cases

INPUT:

OUTPUT:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

options ps=**65** ls=**110** ; \* Portrait pagesize settings \*;

/\*options ps=50 ls=150 ; \* Landscape pagesize settings \*;\*/

options pageno=**1**;

\*\* PROC contents of starting dataset \*\*;

**PROC** **contents** data= TimeSinceVax varnum ; title1 'TimeSinceVax'; **run**;

\*\*\* STEP 2 - Univariate analysis \*\*\*;

\*\*\*--------------------------------\*\*\*;

**PROC** **freq** data= TimeSinceVax;

tables Age\_Group Gender\_Code Vaccination\_Code Breakthrough ;

title2 'Univariate analysis of selected predictor variables';

\*\* excluded continuous variables: Age AND Followup\_Time AND Time\_Since\_Vax \*\*;

**run**;

\*\* STEP 3 - Logit plots of continuous and ordinal variables \*\*;

\*\*------------------------------------------------------------\*\*;

**DATA** TSV; set TimeSinceVax;

AgeGrp=**0**;

If Age\_Group = '18-29' then AgeGrp=**1**;

If Age\_Group = '30-39' then AgeGrp=**2**;

If Age\_Group = '40-49' then AgeGrp=**3**;

If Age\_Group = '50-59' then AgeGrp=**4**;

If Age\_Group = '60-69' then AgeGrp=**5**;

If Age\_Group = '70-79' then AgeGrp=**6**;

If Age\_Group = '80+' then AgeGrp=**7**;

Gender=**0**;

If Gender\_code = 'F' then Gender=**1**;

If Gender\_code = 'M' then Gender=**2**;

If Gender\_code = 'U' then Gender=**3**;

VxType=**0**;

If Vaccination\_Code = 'Janssen' then VxType=**1**;

If Vaccination\_Code = 'Moderna' then VxType=**2**;

If Vaccination\_Code = 'Pfizer' then VxType=**3**;

**run**;

/\* proc freq data=TSV; tables Age\_Group \* AgeGrp / list; run;\*/

title2 'Logit plots of continuous and ordinal variables'; options pageno=**1**;

\*\* for ordinal variables \*\*;

**%macro** BinO(dso, outcomeo, varo);

proc means data=&dso nway noprint;

class &varo; var &outcomeo &varo;

output out=Obins sum(&outcomeo)=&outcomeo ;

data Obins; set Obins;

logit = log((&outcomeo+**1**) / (\_freq\_-&outcomeo+**1**));

proc gplot data=Obins;

plot logit\*&varo; symbol v=star i=none;

title4 "Estimated logit plot of &varo";

run; quit;

**%mend**;

%***BinO***(TSV, Breakthrough, AgeGrp)

\*\* for continuous variables \*\*;

**%macro** BinC(ds, outcome, var);

proc rank data=&ds groups=**10** out=rank;

var &var; ranks bin;

proc means data=rank nway noprint;

class bin; var &outcome &var;

output out=bins sum(&outcome)=&outcome mean(&var)=&var;

data bins; set bins;

logit = log((&outcome+**1**) / (\_freq\_-&outcome+**1**));

proc gplot data=bins;

plot logit\*&var; symbol v=star i=none;

title4 "Estimated logit plot of &var";

run; quit;

**%mend**;

%***BinC***(TimeSinceVax, Breakthrough, Age)

%***BinC***(TimeSinceVax, Breakthrough, Followup\_Time)

%***BinC***(TimeSinceVax, Breakthrough, Time\_Since\_Vax)

\*\* Adding quadratic term for Age \*\*;

**PROC** **univariate** data=TSV plot normal; var AgeGrp; **run**;

**DATA** TSV\_age; set TSV;

Age\_neg = -**1**\*Age;

Age\_inv = **1**/Age;

Age\_sq = Age\*Age;

Age\_cube = Age\*Age\*Age;

Age\_log = log10(Age);

Age\_ln = log(Age);

**run**;

**PROC** **logistic** desc data=TSV\_age ;

/\* model Breakthrough =Age\_neg / clodds=pl ;\*/

/\* model Breakthrough =Age\_inv / clodds=pl ;\*/

/\* model Breakthrough =Age\_sq / clodds=pl ;\*/

/\* model Breakthrough =Age\_cube / clodds=pl ;\*/

/\* model Breakthrough =Age\_log / clodds=pl ;\*/

/\* model Breakthrough =Age\_ln / clodds=pl ;\*/

model Breakthrough =Age Age\_sq / clodds=pl ;

/\* class AgeGrp / param=ref;\*/

/\* model Breakthrough = AgeGrp / clodds=pl ;\*/

**run**;

/\*----------------------------------------------------------------------------------------\*

|FINDINGS:

| Age\_Group should be put on the CLASS statment since it is non-linear in logit

| So it will NOT be treated as an ordinal variable.

| Follow-up time is linear in logit so can keep as a continuous variable.

| Time since vaccination is linear in logit so can keep as a continuous variable.

| Age is NOT linear in the logit; several transformations were tested.

| Age squared resulted in the lowest AIC.

\*----------------------------------------------------------------------------------------\*/

/\*ods trace on / listing;\*/

ods listing close;

ods output nobs=denom bestsubsets=score;

**proc** **logistic** desc data=TSV ;

/\* class Gender\_Code Vaccination\_Code / param=ref;\*/

model Breakthrough = Gender VxType Followup\_Time Time\_Since\_Vax Age Age\*Age

/ selection =score best=**3** ;

title4 'Model selection - best subsets';

**run**;

/\*ods trace off;\*/

ods listing;

/\*proc print data= denom; proc print data= score; run;\*/

**data** denom; set denom;

if label='Number of Observations Used';

call symput('obs', N);

**data** subset; set score;

sc = -scorechisq + log(&obs) \* (numberofvariables+**1**);

aic = -scorechisq+**2** \* (numberofvariables+**1**);

**proc** **sort** data=subset; by sc;

**proc** **print** data=subset; var sc variablesinmodel ;

**proc** **sort** data=subset; by aic;

**proc** **print** data=subset; var aic variablesinmodel ;

**run**;

\*\*\* STEP 5 Model building (TAKE 2) \*\*\*;

\*\*\*------------------------------------------\*\*\*;

**proc** **logistic** desc data=TimeSinceVax ;

class Gender\_Code Vaccination\_Code / param=ref;

model Breakthrough = Gender\_Code Vaccination\_Code Followup\_Time Time\_Since\_Vax Age Age\*Age

Gender\_Code | Vaccination\_Code | Followup\_Time | Time\_Since\_Vax | Age @**3**

/ selection =forward slentry=**.001** hierarchy=single include=**6** ;

title4 'Model selection - forward (single) ';

**run**;

/\*--------------------------------------------------------------\*

|FINDINGS: Four 2-way interaction terms are significant:

| 1. Followup\_Time \* Time\_Since\_Vax

| 2. Time\_Since\_Vax \* Age

| 3. Time\_Since\_Vax \* Vaccination\_Code

| 4. Followup\_Time \* Vaccination\_Code

| AND one 3-way interaction term was significant:

| 5. Followup\_Time \* Time\_Since\_Vax \* Vaccination\_Code

\*--------------------------------------------------------------\*/

title; options pageno=**1**;

**proc** **logistic** desc data=TimeSinceVax ;

class Gender\_Code Vaccination\_Code / param=ref;

model Breakthrough = Gender\_Code Vaccination\_Code Followup\_Time Time\_Since\_Vax Age Age\*Age

/ clodds=pl ;

title2 'Full model - main terms only ';

**run**;

options pageno=**1**;

**proc** **logistic** desc data=TimeSinceVax ;

class Gender\_Code Vaccination\_Code / param=ref;

model Breakthrough = Followup\_Time Time\_Since\_Vax Age Age\*Age

/ clodds=pl ;

title2 'Best model - no interaction terms ';

**run**;

options pageno=**1**;

**proc** **logistic** desc data=TimeSinceVax ;

class Gender\_Code Vaccination\_Code / param=ref;

model Breakthrough = Followup\_Time Time\_Since\_Vax Age Age\*Age Vaccination\_Code

Followup\_Time\*Time\_Since\_Vax

Time\_Since\_Vax \* Age

Time\_Since\_Vax \* Vaccination\_Code

Followup\_Time \* Vaccination\_Code

Followup\_Time \* Time\_Since\_Vax \* Vaccination\_Code

/ clodds=pl ;

title2 'Best model - main terms and 2 and 3 way interactions ';

**run**;