/\* Set libname as r \*/

libname r "C:\Users\bk658\Desktop\ILE project";

**run**;

/\* set libname for actual XPT file combined dataset \*/

libname XPTfile xport "C:\Users\bk658\Desktop\ILE project\Final datasts\LLCP2021.XPT" ;

**run**;

/\*data step reads it and unpacks it into the libname mapped to r\*/

**data** r.BRFSS21\_a;

set XPTfile.LLCP2021;

**run**;

/\* 438693 observations and 303 variables\*/

/\* set libname for actual XPT file combined dataset V1 \*/

libname XPTfile xport "C:\Users\bk658\Desktop\ILE project\Final datasts\LLCP21V1.XPT" ;

**run**;

/\*data step reads it and unpacks it into the libname mapped to r\*/

**data** r.BRFSS21V1\_a;

set XPTfile.LLCP21V1;

**run**;

/\*66124 observations and 303 variable\*/

/\* set libname for actual XPT file combined dataset V2 \*/

libname XPTfile xport "C:\Users\bk658\Desktop\ILE project\Final datasts\LLCP21V2.XPT" ;

**run**;

**data** r.BRFSS21V2\_a;

set XPTfile.LLCP21V2;

**run**;

/\*41752 observations and 304 variables\*/

/\* set libname for actual XPT file combined dataset V3\*/

libname XPTfile xport "C:\Users\bk658\Desktop\ILE project\Final datasts\LLCP21V3.XPT";

**run**;

**data** r.BRFSS21V3\_a;

set XPTfile.LLCP21V3;

**run**;

/\*23586 observations and 304 variables\*/

/\*Generating a new working datset by creating a new and uniform final weight variable \*/

/\*Extract the data for states that used common module of Cancer Survivorship for Arkansas(5),

Delaware(10), Hawaii(15), Michigan(26),NewJersey(34), Montana(30), New Mexico(35), Puerto Rico(72), Utah(49), Wisconsin(55)\*/

**Data** r.LLCP;

SET r.BRFSS21\_a(WHERE=(\_STATE IN(**5**,**10**,**15**,**26**,**30**,**34**,**35**,**49**,**55**,**72**)));

\_FINALWT=\_LLCPWT;

**run**;

/\*67535 observations and 304 variables\*/

/\* Version 1 was used by Nebraska(31) and Okhlahoma(40)\*/

**Data** r.LLCPV1;

SET r.BRFSS21V1\_a(WHERE=(\_STATE IN(**31**,**40**)));

\_FINALWT=\_LCPWTV1;

**RUN**; /\*9898 observations and 304 variables\*/

/\* Version 2 was used by Kansas(20)\*/

**Data** r.LLCPV2;

SET r.BRFSS21V2\_a(WHERE=(\_STATE IN(**20**)));

\_FINALWT=\_LCPWTV2;

**RUN**;/\*8697 observations and 304 variable\*/

/\* Version 3 was used by Kansas(20)\*/

**Data** r.LLCPV3;

SET r.BRFSS21V3\_a(WHERE=(\_STATE IN(**24**)));

\_FINALWT=\_LCPWTV3;

**RUN**;/\*5119 observations and 304 variable\*/

**Data** r.COMBINEDVERSION;

SET r.LLCP r.LLCPV1 r.LLCPV2 r.LLCPV3;

**run**;

/\* After excluding the states who didnt opted for optional modules there are 91249 observations and 310 variables \*/

/\*Now further subsetting the datset to the required variables as per research proposal\*/

**Data** r.COMBINEDVERSION\_a;

SET r.COMBINEDVERSION(keep=

\_SEX

\_STATE

\_AGE\_G

PRIMINSR

EXERANY2

CHCOCNCR

\_EDUCAG

INCOME3

CNCRTYP1

\_RACE

\_BMI5CAT

\_PSU

\_STSTR

\_FINALWT

);

**run**;

**data** r.combinedversion\_b;

set r.combinedversion\_a;

/\*Apply exclusion criteria of removing patients who are not having cancer\*/

if CHCOCNCR= **2** then delete;

/\*Missing data\*/

\* set refused/dont know to missing;

if PRIMINSR in (**77**,**99**) then call missing(PRIMINSR);

if EXERANY2 GE **7** then call missing(EXERANY2);

if CHCOCNCR GE **7** then call missing(CHCOCNCR);

if CSRVPAIN GE **7** then call missing(CSRVPAIN);

if INCOME3 GE **77** then call missing(INCOME3);

if CNCRTYP1 GE **77** then call missing(CNCRTYP1);

if \_RACE GE **9** then call missing(\_RACE);

if \_EDUCAG GE **9** then call missing (\_EDUCAG);

/\*Recoding the variables as needed for outcome and exposure \*/

/\* Age, gender,race, BMI,cancer,education, physical pain doesnot require recoding\*/

/\* variables that require recoding are Income, Cancer type, Insurance \*/

/\* Income \*/

if missing(INCOME3)then call missing(income\_cat);

else if INCOME3 < **6** then income\_cat = **1**;

else income\_cat = **2**;

/\* Income \*/

if missing(\_RACE)then call missing(race\_cat);

else if \_RACE = **1** then race\_cat = **1**;

else if \_RACE = **2** then race\_cat = **2**;

else if \_RACE = **3** then race\_cat = **3**;

else if \_RACE = **4** then race\_cat = **4**;

else if \_RACE in (**5**,**6**,**7**) then race\_cat = **5**;

else if \_RACE = **8** then race\_cat = **6**;

else race\_cat = **7**;

/\*Cancer type\*/

if missing(CNCRTYP1) then call missing(cancer\_cat);

else if CNCRTYP1 = **1** then cancer\_cat = **1**;

else if CNCRTYP1 in (**2**, **4**, **19** , **20** ) then cancer\_cat =**2**;

else if CNCRTYP1 in (**3**, **10**, **11**, **12**, **13** ,**14**) then cancer\_cat= **3**;

else if CNCRTYP1 in (**5**,**6**,**7**,**8**,**9**, **15**, **16**, **17**,**18**,**21**,**22**,**23**,**24**,**25**,**26**,**27**,**28**,**29**,**30**) then cancer\_cat= **4**;

else cancer\_cat =**5** ;

/\*Insurance\*/

if missing(PRIMINSR)then call missing(Insurance\_cat) ;

else if PRIMINSR < **11** then Insurance\_cat = **1**;

else Insurance\_cat = **2**;

**run**;

/\* Formats \*/

**proc** **format**;

value \_SEX

**1**= "Male"

**2**= "Female";

value \_age\_g

**1**= "18-24"

**2**= "25-34"

**3**= "35-44"

**4**= "45-54"

**5**= "55-64"

**6**= "65 or older";

value race\_cat

**1**= "NH White"

**2**= "NH Black"

**3**= "NH American Indian/Alaskan Native"

**4**= "NH Asian"

**5**= "NH Other race"

**6**= "Hispanic";

value income\_cat

**1**= "<$50k"

**2**= ">$50k";

value \_EDUCAG

**1**="Did not graduate HS"

**2**="Graduated HS"

**3**= "Attended College/Technical School"

**4**= "Graduated from College/Technical School";

value \_BMI5CAT

**1**= "Underweight"

**2**= "Normal Weight"

**3**= "Overweight"

**4**= "Obese";

value Insurance\_cat

**1**= "Insured"

**2**="Not Insured";

value cancer\_cat

**1**= "Breast"

**2**= "Reproductive"

**3**= "Endometrial"

**4**= "Other cancers";

value Pain

**1**= "Yes"

**2**= "No";

value EXERANY

**1**= "Yes"

**2**= "No";

Value CHCOCNCR

**1**= "Yes"

**2**= "No";

**run**;

/\* Summary Statistics \*/

\*race/ethnicity ;

**proc** **freq** data = r.COMBINEDVERSION\_b;

tables race\_cat/missing;

format race\_cat race\_cat.;

**run**;

\*Gender;

**proc** **freq** data= r.combinedversion\_b;

tables \_SEX/missing;

format \_SEX \_sex.;

**run**;

\*Education;

**proc** **freq** data = r.combinedversion\_b;

tables \_educag/missing;

format \_educag \_educag.;

**run**;

\*Age;

**proc** **freq** data = r.combinedversion\_b;

tables \_age\_g/ missing;

format \_age\_g \_age\_g.;

**run**;

\*Income;

**proc** **freq** data =r.combinedversion\_b;

tables income\_cat/missing;

format income\_cat income\_cat.;

**run**;

\*Cancer (Y/N);

**proc** **freq** data= r.combinedversion\_b;

tables CHCOCNCR/missing;

format CHCOCNCR CHCOCNCR.;

**run**;

\*Insurance type (Y/N);

**proc** **freq** data= r.combinedversion\_b;

tables Insurance\_cat/missing;

format Insurance\_cat Insurance\_cat.;

**run**;

\*Cancer type;

**proc** **freq** data =r.combinedversion\_b;

tables cancer\_cat/missing;

format cancer\_cat cancer\_cat.;

**run**;

**data** r.combinedversion\_d;

set r.combinedversion\_b;

if \_SEX=**.** OR \_AGE\_G=**.** OR Insurance\_cat=**.** OR EXERANY2=**.** OR \_EDUCAG=**.** OR income\_cat=**.** OR cancer\_cat=**.** OR race\_cat=**.** OR \_BMI5CAT=**.** OR CHCOCNCR=**.**

then delete;

**run**;

**proc** **sort** data=r.combinedversion\_d;

by EXERANY2;

**run**;

/\*Combined version d has 5640 observations and b has 8817 obs and 19 variables\*/

/\*---Table 1-- \*/

/\*Weighted estimates \*/

\*Unweighted categorical variables by outcome;

**proc** **freq** data=r.combinedversion\_d;

table race\_cat\*EXERANY2 Insurance\_cat\*EXERANY2 \_educag\*EXERANY2 \_SEX\*EXERANY2 income\_cat\*EXERANY2 \_age\_g\*EXERANY2 cancer\_cat\*EXERANY2

\_BMI5CAT\*EXERANY2/chisq;

format race\_cat race\_cat. Insurance\_cat Insurance\_cat. \_educag \_educag. \_SEX \_SEX. income\_cat income\_cat. \_age\_g \_age\_g. cancer\_cat cancer\_cat.

\_BMI5CAT \_BMI5CAT. EXERANY2 EXERANY2.;

**run**;

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

/\* LOGISTIC REGRESSION

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

/\* Model 1 - Unadjusted \*/

**proc** **logistic** data=r.combinedversion\_d;

class race\_cat / param= ref;

model EXERANY2(EVENT='Ye')=race\_cat / clodds=wald orpvalue;

format race\_cat race\_cat. Insurance\_cat Insurance\_cat. \_educag \_educag. \_SEX \_SEX. income\_cat income\_cat. \_age\_g \_age\_g. cancer\_cat cancer\_cat.

\_BMI5CAT \_BMI5CAT. EXERANY2 EXERANY2.;

**run**;

/\*model 2 with age and sex\*/

/\* Ref for age 1 is 18-24 years nad Sex '1" is male and race "1" is white\*/

**proc** **logistic** data=r.combinedversion\_d;

class race\_cat \_age\_g (ref="65 or older") \_SEX / param = ref;

model EXERANY2(EVENT='Ye')=race\_cat \_age\_g \_SEX / clodds=wald orpvalue;

format race\_cat race\_cat. Insurance\_cat Insurance\_cat. \_educag \_educag. \_SEX \_SEX. income\_cat income\_cat. \_age\_g \_age\_g. cancer\_cat cancer\_cat.

\_BMI5CAT \_BMI5CAT. EXERANY2 EXERANY2.;

**run**;

/\*model 3: Model 2\* with BMI and Cancer type. Here reference for BMI is 4 which is obese and cancer type reference is 1 which is breast cancer \*/

**proc** **logistic** data=r.combinedversion\_d;

class race\_cat \_age\_g (ref="65 or older") \_SEX \_BMI5CAT(ref='Obese') cancer\_cat / param = ref;

model EXERANY2(EVENT='Ye')=race\_cat \_age\_g \_SEX \_BMI5CAT cancer\_cat / clodds=wald orpvalue;

format race\_cat race\_cat. Insurance\_cat Insurance\_cat. \_educag \_educag. \_SEX \_SEX. income\_cat income\_cat. \_age\_g \_age\_g. cancer\_cat cancer\_cat.

\_BMI5CAT \_BMI5CAT. EXERANY2 EXERANY2.;**run**;

/\*model 4: Model 3 and rest of all covariates. Education Ref 1 is did not graduate HS, income cat 1 is less than $50k\*/

**proc** **logistic** data=r.combinedversion\_d;

class race\_cat \_age\_g (ref="65 or older") \_SEX \_BMI5CAT (ref= "Obese") cancer\_cat income\_cat \_educag/ param = ref;

model EXERANY2(EVENT='Ye')=race\_cat \_age\_g \_SEX \_BMI5CAT cancer\_cat income\_cat \_educag / clodds=wald orpvalue;

format race\_cat race\_cat. Insurance\_cat Insurance\_cat. \_educag \_educag. \_SEX \_SEX. income\_cat income\_cat. \_age\_g \_age\_g. cancer\_cat cancer\_cat.

\_BMI5CAT \_BMI5CAT. EXERANY2 EXERANY2.;

**run**;