/*Buchanan. Causal Inference in HIV/AIDS Research*/
/*SAMPLE SAS CODE FOR CHAPTER 3*/

/*newid: participant id (this has been changed from original WIHS study id)
taidsdth: time to AIDS/Death
tarv: time to ART initiation
ageatfda: age in years at baseline
BASEIDU: 0-no history of IDU; 1-history of IDU
black: 0-not African-American; 1-African-American
cd4nadir: CD4 count nadir prior to baseline
dth: Indicator for AIDS/Death event*/

*Read data into SAS and prepare data;
proc import datafile="H:\Survival_Tutorial\WIHS\WIHS_IDU.csv"
   out=prea dbms=csv replace;
   getnames=yes;
   datarow=2;
run;

data a; set prea(drop=t);
id= _n_;
if black=1 then aa=1; else if black=0 then aa=0;
if dth=1 then y=1; else if dth=0 then y=0; *death/AIDS indicator;
label y="AIDS/death=1";
if idu=baseidu;
age=ageatfda/10; *decades of age at baseline;
cd4n=cd4nadir/100; *centiles of nadir CD4 at baseline;
t=round(taidsdth,.0001);
if t>10 then do; t=10; y=0; end; *censor at 10 years;
if t<9 and y=0 then drop=1; else drop=0; *indicator for drop out;
label aa="african-am" age="age decades" cd4n="CD4 centiles"
y="AIDS/death" t="years" drop="drop out";
/*time-varying ARV exposure*/
/*if they initiated arv before aids/death/censor then arv=1*/
tarv=round(tarv,.0001);
if tarv<t then arv=1; else arv=0;
label arv="arv before AIDS/death";
keep pid aa age cd4n idu y t drop arv tarv;
run;

*Unadjusted model;
proc phreg data=a;
model t*y(0)=idu/ties=efron rl;
title1 "Unadjusted Cox model";
run;

*IP weighted Cox model and curves;
*Read spline macro code into SAS;
*Fit restricted quadratic splines on continuous variables;
*See Howe, et al. Splines for trend analysis and continuous confounder control. Epidemiology. 2011;
%inc "H:\Survival_Tutorial\WIHS\qspline4.sas";
%qspline4(a,age);
%qspline4(a,cd4n);
/*Model for exposure weights*/
/*Use this model to obtain Pr(Xi=x)*/
/*Since IDU is at baseline, we do not include time-varying ARV variable*/
/*All pairwise interactions included*/
proc sort data=a; by pid;
proc logistic data=a noprint desc;
   model idu=;
   output out=n p=n;
   title1 "IP exposure model/numerator";
proc logistic data=a desc;
   model idu=aa age age1-age3 cd4n cd4n1-cd4n3
       aa*age aa*age1 aa*age2 aa*age3
       aa*cd4n aa*cd4n1 aa*cd4n2 aa*cd4n3
       age*cd4n age*cd4n1 age*cd4n2 age*cd4n3
       age*cd4n1 age*cd4n2 age*cd4n3/lackfit;
   output out=d p=d;
   title1 "IP exposure model/denominator";
run;
*Define the exposure weights;
data a;
merge a n d; by pid;
   if idu=1 then ew=n/d; /*Exposure weight*/
   else if idu=0 then ew=(1-n)/(1-d);
   label n= d=;
   drop _level_; run;
/*Model for drop out weights*/
*Create dataset with monthly records for each subject (Hernan, 2000);
data b; set a;
array j{121} j1-j121 (0, 0.08, 0.17, 0.25, 0.33, 0.42, 0.5, 0.58, 0.67, 0.75, 0.83, 0.92,
1, 1.08, 1.17, 1.25, 1.33, 1.42, 1.5, 1.58, 1.67, 1.75, 1.83, 1.92,
2, 2.08, 2.17, 2.25, 2.33, 2.42, 2.5, 2.58, 2.67, 2.75, 2.83, 2.92,
3, 3.08, 3.17, 3.25, 3.33, 3.42, 3.5, 3.58, 3.67, 3.75, 3.83, 3.92,
4, 4.08, 4.17, 4.25, 4.33, 4.42, 4.5, 4.58, 4.67, 4.75, 4.83, 4.92,
5, 5.08, 5.17, 5.25, 5.33, 5.42, 5.5, 5.58, 5.67, 5.75, 5.83, 5.92,
6, 6.08, 6.17, 6.25, 6.33, 6.42, 6.5, 6.58, 6.67, 6.75, 6.83, 6.92,
7, 7.08, 7.17, 7.25, 7.33, 7.42, 7.5, 7.58, 7.67, 7.75, 7.83, 7.92,
8, 8.08, 8.17, 8.25, 8.33, 8.42, 8.5, 8.58, 8.67, 8.75, 8.83, 8.92,
10);
   do k=1 to 120; in=j(k);
      if j(k)<t<=j(k+1) then do; out=t; yy=y; drop2=drop; output; end;
      else if j(k+1)<t then do; out=j(k+1); yy=0; drop2=0; output; end;
   end;
   keep pid in out yy idu aa age cd4n drop2 ew arv tarv;
run;
/*Add restricted quadratic splines to continuous variables in dataset*/
%qspline4(b,age);
%qspline4(b,cd4n);
%qspline4(b,in);
**Correct time-varying covariate (arv=1 only after initiation)**

```sas
if tarv>=out then arv=0;
else if tarv<out then arv=1;
run;
```

*Use this model to obtain Pr(Di>t);
*Time is modeled as continuous with a spline;

```sas
proc sort data=b; by pid in;
proc logistic data=b;
model drop2=idu in in1 in2 in3;
output out=dn p=dn;
title1 "IP drop out model/numerator"
```

```sas
proc logistic data=b;
model drop2= in in1 in2 in3 idu
  aa age age1-age3 cd4n cd4n1-cd4n3 arv
  idu*aa idu*age idu*age1 idu*age2 idu*age3
  idu*cd4n idu*cd4n1 idu*cd4n2 idu*cd4n3
  idu*arv
  aa*age aa*age1 aa*age2 aa*age3
  aa*cd4n aa*cd4n1 aa*cd4n2 aa*cd4n3 aa*arv
  age*cd4n
  age1*cd4n1 age1*cd4n2 age1*cd4n3
  age2*cd4n1 age2*cd4n2 age2*cd4n3
  age3*cd4n1 age3*cd4n2 age3*cd4n3
  arv*age arv*age1 arv*age2 arv*age3
  arv*cd4n arv*cd4n1 arv*cd4n2 arv*cd4n3 /lackfit;
output out=dd p=dd;
title1 "IP drop out model/denominator"
run;
```

*Define the drop out weights and combined weight;

```sas
data b; merge b dn dd; by pid in;
retain num den;
  *Multiply drop out weights for each participant;
  if first.pid then do; num=1; den=1; end;
  num=num*dn;
  den=den*dd;
dw=num/den; *drop out weight;
fw=ew*dw; *combined weight;
label dn= dd=; drop _level_
drop num den;
```

```sas
proc means data=b n mean std min max sum fw=8;
var dn dd dw ew fw;
run;
```

*Standardized model;
*Robust estimate of variance employed with COVS option;

```sas
proc phreg data=b covs;
  model out*yy(0)=idu/entry=in ties=efron r1;
  weight fw; title1 "IP weighted Cox model"
run;
```