**SAS Code – HB BMI change**

**Table 1.**

libname raw '/data\_source/user\_data';

libname d '/data\_source/user\_data/data';

**data** cst\_cq\_rst\_2010;

set raw.cst\_cq\_rst\_2010;

CST\_date=input(HME\_DT,yymmdd8.);

format CST\_date yymmdd10.;

if CST\_DATE=**.** then CST\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** g1eq\_2010;

set raw.g1eq\_2010;

INDEX\_DATE=input(HME\_DT,yymmdd8.);

format INDEX\_DATE yymmdd10.;

if INDEX\_DATE=**.** then INDEX\_DATE=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** cbr\_cq\_rst\_2010;

set raw.cbr\_cq\_rst\_2010;

cbr\_date=input(HME\_DT,yymmdd8.);

format cbr\_date yymmdd10.;

if cbr\_date=**.** then cbr\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** ccr\_cq\_rst\_2010;

set raw.ccr\_cq\_rst\_2010;

CCR\_date=input(HME\_DT,yymmdd8.);

format CCR\_date yymmdd10.;

if CCR\_date=**.** then CCR\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** ccx\_cq\_rst\_2010;

set raw.ccx\_cq\_rst\_2010;

CCX\_date=input(HME\_DT,yymmdd8.);

format CCX\_date yymmdd10.;

if CCX\_date=**.** then CCX\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** clv\_cq\_rst\_2010;

set raw.clv\_cq\_rst\_2010;

CLV\_date=input(HME\_DT,yymmdd8.);

format CLV\_date yymmdd10.;

if CLV\_date=**.** then CLV\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**proc** **sql**;

create table g1eq\_2010\_1 as

select a.\*,b.indi\_dscm\_no,b.CST\_date, abs(CST\_date-INDEX\_DATE) as dist\_1

from g1eq\_2010 as a

left join cst\_cq\_rst\_2010 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no;

**quit**;

**proc** **sort** data=g1eq\_2010\_1;

by indi\_dscm\_no dist\_1;**run**;

**proc** **sort** data=g1eq\_2010\_1 nodupkey;

by indi\_dscm\_no;**run**;

**proc** **sql**;

create table g1eq\_2010\_2 as

select a.\*,b.indi\_dscm\_no,b.cbr\_date, abs(cbr\_date-INDEX\_DATE) as dist\_2

from g1eq\_2010\_1 as a

left join cbr\_cq\_rst\_2010 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no;

**quit**;

**proc** **sort** data=g1eq\_2010\_2;

by indi\_dscm\_no dist\_1;**run**;

**proc** **sort** data=g1eq\_2010\_2 nodupkey;

by indi\_dscm\_no;**run**;

**proc** **sql**;

create table g1eq\_2010\_3 as

select a.\*,b.indi\_dscm\_no,b.CCR\_date, abs(CCR\_date-INDEX\_DATE) as dist\_3

from g1eq\_2010\_2 as a

left join CCR\_cq\_rst\_2010 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no;

**quit**;

**proc** **sort** data=g1eq\_2010\_3;

by indi\_dscm\_no dist\_1;**run**;

**proc** **sort** data=g1eq\_2010\_3 nodupkey;

by indi\_dscm\_no;**run**;

**proc** **sql**;

create table g1eq\_2010\_4 as

select a.\*,b.indi\_dscm\_no,b.CCX\_date, abs(CCX\_date-INDEX\_DATE) as dist\_4

from g1eq\_2010\_3 as a

left join ccx\_cq\_rst\_2010 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no;

**quit**;

**proc** **sort** data=g1eq\_2010\_4;

by indi\_dscm\_no dist\_1;**run**;

**proc** **sort** data=g1eq\_2010\_4 nodupkey;

by indi\_dscm\_no;**run**;

**proc** **sql**;

create table g1eq\_2010\_5 as

select a.\*,b.indi\_dscm\_no,b.CLV\_date, abs(CLV\_date-INDEX\_DATE) as dist\_5

from g1eq\_2010\_4 as a

left join clv\_cq\_rst\_2010 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no;

**quit**;

**proc** **sort** data=g1eq\_2010\_5;

by indi\_dscm\_no dist\_1;**run**;

**proc** **sort** data=g1eq\_2010\_5 nodupkey;

by indi\_dscm\_no;**run**;

**data** d.g1eq\_2010\_NEW;

set g1eq\_2010\_5;

if dist\_1 ne **.** or dist\_2 ne **.** or dist\_3 ne **.** or dist\_4 ne **.** or dist\_5 ne **.**;

**run**;

**data** d.T20\_C;

set raw.t20\_201001 raw.t20\_201002 raw.t20\_201003 raw.t20\_201004 raw.t20\_201005 raw.t20\_201006 raw.t20\_201007 raw.t20\_201008 raw.t20\_201009 raw.t20\_201010 raw.t20\_201011 raw.t20\_201012 raw.t20\_201101 raw.t20\_201102 raw.t20\_201103 raw.t20\_201104 raw.t20\_201105 raw.t20\_201106 raw.t20\_201107 raw.t20\_201108 raw.t20\_201109 raw.t20\_201110 raw.t20\_201111 raw.t20\_201112 raw.t20\_201201 raw.t20\_201202 raw.t20\_201203 raw.t20\_201204 raw.t20\_201205 raw.t20\_201206 raw.t20\_201207 raw.t20\_201208 raw.t20\_201209 raw.t20\_201210 raw.t20\_201211 raw.t20\_201212 raw.t20\_201301 raw.t20\_201302 raw.t20\_201303 raw.t20\_201304 raw.t20\_201305 raw.t20\_201306 raw.t20\_201307 raw.t20\_201308 raw.t20\_201309 raw.t20\_201310 raw.t20\_201311 raw.t20\_201312 raw.t20\_201401 raw.t20\_201402 raw.t20\_201403 raw.t20\_201404 raw.t20\_201405 raw.t20\_201406 raw.t20\_201407

raw.t20\_201408 raw.t20\_201409 raw.t20\_201410 raw.t20\_201411 raw.t20\_201412 raw.t20\_201501 raw.t20\_201502 raw.t20\_201503 raw.t20\_201504 raw.t20\_201505 raw.t20\_201506 raw.t20\_201507 raw.t20\_201508 raw.t20\_201509 raw.t20\_201510 raw.t20\_201511 raw.t20\_201512 raw.t20\_201601 raw.t20\_201602 raw.t20\_201603

raw.t20\_201604 raw.t20\_201605 raw.t20\_201606 raw.t20\_201607 raw.t20\_201608 raw.t20\_201609 raw.t20\_201610 raw.t20\_201611 raw.t20\_201612 raw.t20\_201701 raw.t20\_201702 raw.t20\_201703 raw.t20\_201704 raw.t20\_201705 raw.t20\_201706 raw.t20\_201707 raw.t20\_201708 raw.t20\_201709 raw.t20\_201710 raw.t20\_201711

raw.t20\_201712 raw.t20\_201801 raw.t20\_201802 raw.t20\_201803 raw.t20\_201804 raw.t20\_201805 raw.t20\_201806 raw.t20\_201807 raw.t20\_201808 raw.t20\_201809 raw.t20\_201810 raw.t20\_201811 raw.t20\_201812 raw.t20\_201901 raw.t20\_201902 raw.t20\_201903 raw.t20\_201904 raw.t20\_201905 raw.t20\_201906 raw.t20\_201907

raw.t20\_201908 raw.t20\_201909 raw.t20\_201910 raw.t20\_201911 raw.t20\_201912 raw.t20\_202001 raw.t20\_202002 raw.t20\_202003 raw.t20\_202004 raw.t20\_202005 raw.t20\_202006 raw.t20\_202007 raw.t20\_202008 raw.t20\_202009 raw.t20\_202010 raw.t20\_202011 raw.t20\_202012 raw.t20\_202101 raw.t20\_202102 raw.t20\_202103

raw.t20\_202104 raw.t20\_202105 raw.t20\_202106 raw.t20\_202107 raw.t20\_202108 raw.t20\_202109 raw.t20\_202110 raw.t20\_202111 raw.t20\_202112 ;

where upcase(substr(SICK\_SYM1,**1**,**1**))="C";

recu\_date=input(MDCARE\_STRT\_DT,yymmdd8.);

format recu\_date yymmdd10.;

drop MDCARE\_STRT\_DT;

**run**;

**proc** **sql**;

create table B as

select distinct a.\*,case when b.CMN\_KEY ne '' then **1** else **0** end as CANCER\_1YR

from d.g1eq\_2010\_NEW as a

left join d.t20\_c as b

on a.indi\_dscm\_no=b.indi\_dscm\_no and b.recu\_date>**0** and **0**<=b.recu\_date-a.INDEX\_DATE<**365**;

**quit**;

**proc** **sql**;

create table B\_1 as

select distinct a.\*,input(c.DTH\_ASSMD\_DT,yymmdd8.) as DEATH\_DATE

from B as a

left join raw.tg\_dth as c

on a.indi\_dscm\_no=c.indi\_dscm\_no ;

**quit**;

**data** B\_2;

set B\_1;

if DEATH\_DATE>**0** and **0**<=DEATH\_DATE-INDEX\_DATE<**365** then DEATH\_1YR=**1**;

format INDEX\_DATE DEATH\_DATE yymmdd10.;

**run**;

**proc** **freq** data=B\_2;

table CANCER\_1YR /norow nocol nopercent;

**run**;

**data** B\_3;

set B\_2;

if CANCER\_1YR=**1** or DEATH\_1YR=**1** then delete;

**run**;

**proc** **sql**;

create table C as

select distinct a.\*,a.G1E\_BMI as G1E\_BMI\_2010, b.G1E\_BMI as G1E\_BMI\_2014,a.G1E\_WSTC as G1E\_WSTC\_2010,b.G1E\_WSTC as G1E\_WSTC\_2014,case when b.HME\_DT ne '' then **1** else **0** end as CON1

from B\_3 as a

left join raw.g1eq\_2014 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO;

**quit**;

**proc** **sql**;

create table C\_1 as

select distinct a.\*,b.SEX\_TYPE,b.BYEAR,b.CALC\_CTRB\_VTILE\_FD

from C as a

left join raw.bfc\_2010 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO;

**quit**;

**data** C\_2;

set C\_1;

if CON1=**1**;

**run**;

**data** C\_3;

set C\_2;

if G1E\_BMI\_2010 ne **.** and G1E\_BMI\_2014 ne **.** and G1E\_WSTC\_2010 ne **.** and G1E\_WSTC\_2014 ne **.**;

**run**;

**data** d.DATESET\_BMI;

set C\_3;

if SEX\_TYPE="" then delete;

**run**;

**data** cst\_cq\_rst\_2010;

set raw.cst\_cq\_rst\_2010;

CST\_date=input(HME\_DT,yymmdd8.);

format CST\_date yymmdd10.;

if CST\_DATE=**.** then CST\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** ccr\_cq\_rst\_2010;

set raw.ccr\_cq\_rst\_2010;

CCR\_date=input(HME\_DT,yymmdd8.);

format CCR\_date yymmdd10.;

if CCR\_date=**.** then CCR\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** ccx\_cq\_rst\_2010;

set raw.ccx\_cq\_rst\_2010;

CCX\_date=input(HME\_DT,yymmdd8.);

format CCX\_date yymmdd10.;

if CCX\_date=**.** then CCX\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** cbr\_cq\_rst\_2010;

set raw.cbr\_cq\_rst\_2010;

cbr\_date=input(HME\_DT,yymmdd8.);

format cbr\_date yymmdd10.;

if cbr\_date=**.** then cbr\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** clv\_cq\_rst\_2010;

set raw.clv\_cq\_rst\_2010;

CLV\_date=input(HME\_DT,yymmdd8.);

format CLV\_date yymmdd10.;

if CLV\_date=**.** then CLV\_date=input(substr(HME\_DT,**1**,**6**)||"01",yymmdd8.);

**run**;

**data** cst\_cq\_rst\_2010\_1;

set cst\_cq\_rst\_2010;

keep INDI\_DSCM\_NO CST\_date FAM

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN;

/\*QC\_PFHX\_CST\_YN=2 X QC\_PFHX\_CST\_YN=1 O\*/

if QC\_PFHX\_CST\_YN=**0** then FAM=**9**;

else if QC\_PFHX\_CST\_YN=**2** then FAM=**0**;

else if QC\_PFHX\_CST\_YN=**1** then FAM=**1**;

else if QC\_PFHX\_CST\_YN=**.** then FAM=**9**;

**run**;

**proc** **freq** data=cst\_cq\_rst\_2010\_1;

table FAM/norow nocol nopercent;

**run**;

**data** ccr\_cq\_rst\_2010\_1;

set ccr\_cq\_rst\_2010;

keep INDI\_DSCM\_NO CCR\_date FAM

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN;

if QC\_PFHX\_CCR\_YN=**0** then FAM=**9**;

else if QC\_PFHX\_CCR\_YN=**2** then FAM=**0**;

else if QC\_PFHX\_CCR\_YN=**1** then FAM=**1**;

else if QC\_PFHX\_CCR\_YN=**.** then FAM=**9**;

**run**;

**proc** **freq** data=ccr\_cq\_rst\_2010\_1;

table FAM/norow nocol nopercent;

**run**;

**data** CCX\_cq\_rst\_2010\_1;

set CCX\_cq\_rst\_2010;

keep INDI\_DSCM\_NO CCX\_date FAM

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN;

if QC\_PFHX\_CCX\_YN=**0** then FAM=**9**;

else if QC\_PFHX\_CCX\_YN=**2** then FAM=**0**;

else if QC\_PFHX\_CCX\_YN=**1** then FAM=**1**;

else if QC\_PFHX\_CCX\_YN=**.** then FAM=**9**;

**run**;

**proc** **freq** data=CCX\_cq\_rst\_2010\_1;

table FAM/norow nocol nopercent;

**run**;

**data** CBR\_cq\_rst\_2010\_1;

set CBR\_cq\_rst\_2010;

keep INDI\_DSCM\_NO CBR\_date FAM

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN;

if QC\_PFHX\_CBR\_YN=**0** then FAM=**9**;

else if QC\_PFHX\_CBR\_YN=**2** then FAM=**0**;

else if QC\_PFHX\_CBR\_YN=**1** then FAM=**1**;

else if QC\_PFHX\_CBR\_YN=**.** then FAM=**9**;

**run**;

**proc** **freq** data=CBR\_cq\_rst\_2010\_1;

table FAM/norow nocol nopercent;

**run**;

**data** CLV\_cq\_rst\_2010\_1;

set CLV\_cq\_rst\_2010;

keep INDI\_DSCM\_NO CLV\_date FAM

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN;

if QC\_PFHX\_CLV\_YN=**0** then FAM=**9**;

else if QC\_PFHX\_CLV\_YN=**2** then FAM=**0**;

else if QC\_PFHX\_CLV\_YN=**1** then FAM=**1**;

else if QC\_PFHX\_CLV\_YN=**.** then FAM=**9**;

**run**;

**proc** **freq** data=CLV\_cq\_rst\_2010\_1;

table FAM/norow nocol nopercent;

**run**;

**proc** **sort** data=cst\_cq\_rst\_2010\_1 out= cst\_cq\_rst\_2010\_2 ;

by INDI\_DSCM\_NO CST\_date descending QC\_PHX\_LV\_YN descending QC\_PHX\_CHB\_YN descending QC\_PHX\_CHC\_YN descending QC\_PHX\_LC\_YN descending QC\_MNS\_YN descending QC\_BRFD\_DRT descending QC\_DLV\_FRQ descending QC\_OPLL\_YN ;**run**;

**proc** **sort** data=cst\_cq\_rst\_2010\_2 out=cst\_cq\_rst\_2010\_3 nodupkey;

by INDI\_DSCM\_NO CST\_date;**run**;

**proc** **sort** data=CCR\_cq\_rst\_2010\_1 out= CCR\_cq\_rst\_2010\_2 ;

by INDI\_DSCM\_NO CCR\_date descending QC\_PHX\_LV\_YN descending QC\_PHX\_CHB\_YN descending QC\_PHX\_CHC\_YN descending QC\_PHX\_LC\_YN descending QC\_MNS\_YN descending QC\_BRFD\_DRT descending QC\_DLV\_FRQ descending QC\_OPLL\_YN ;**run**;

**proc** **sort** data=CCR\_cq\_rst\_2010\_2 out=CCR\_cq\_rst\_2010\_3 nodupkey;

by INDI\_DSCM\_NO CCR\_date;**run**;

**proc** **sort** data=CCX\_cq\_rst\_2010\_1 out= CCX\_cq\_rst\_2010\_2 ;

by INDI\_DSCM\_NO CCX\_date descending QC\_PHX\_LV\_YN descending QC\_PHX\_CHB\_YN descending QC\_PHX\_CHC\_YN descending QC\_PHX\_LC\_YN descending QC\_MNS\_YN descending QC\_BRFD\_DRT descending QC\_DLV\_FRQ descending QC\_OPLL\_YN ;**run**;

**proc** **sort** data=CCX\_cq\_rst\_2010\_2 out=CCX\_cq\_rst\_2010\_3 nodupkey;

by INDI\_DSCM\_NO CCX\_date;**run**;

**proc** **sort** data=CBR\_cq\_rst\_2010\_1 out= CBR\_cq\_rst\_2010\_2 ;

by INDI\_DSCM\_NO CBR\_date descending QC\_PHX\_LV\_YN descending QC\_PHX\_CHB\_YN descending QC\_PHX\_CHC\_YN descending QC\_PHX\_LC\_YN descending QC\_MNS\_YN descending QC\_BRFD\_DRT descending QC\_DLV\_FRQ descending QC\_OPLL\_YN ;**run**;

**proc** **sort** data=CBR\_cq\_rst\_2010\_2 out=CBR\_cq\_rst\_2010\_3 nodupkey;

by INDI\_DSCM\_NO CBR\_date;**run**;

**proc** **sort** data=CLV\_cq\_rst\_2010\_1 out= CLV\_cq\_rst\_2010\_2 ;

by INDI\_DSCM\_NO CLV\_date descending QC\_PHX\_LV\_YN descending QC\_PHX\_CHB\_YN descending QC\_PHX\_CHC\_YN descending QC\_PHX\_LC\_YN descending QC\_MNS\_YN descending QC\_BRFD\_DRT descending QC\_DLV\_FRQ descending QC\_OPLL\_YN ;**run**;

**proc** **sort** data=CLV\_cq\_rst\_2010\_2 out=CLV\_cq\_rst\_2010\_3 nodupkey;

by INDI\_DSCM\_NO CLV\_date;**run**;

**proc** **sql**;

create table BMI\_1 as

select distinct a.\*,b.\*

from d.DATESET\_BMI as a

left join cst\_cq\_rst\_2010\_3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.CST\_date=b.CST\_date;

**quit**;

**data** BMI\_11 BMI\_12(drop= QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN FAM);

set BMI\_1;

if FAM ne **.** then output BMI\_11;

else if FAM =**.** then output BMI\_12;

**run**;

**proc** **sql**;

create table BMI\_2 as

select distinct a.\*,b.\*

from BMI\_12 as a

left join CCR\_cq\_rst\_2010\_3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.CCR\_date=b.CCR\_date;

**quit**;

**data** BMI\_21 BMI\_22(drop= QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN FAM);

set BMI\_2;

if FAM ne **.** then output BMI\_21;

else if FAM =**.** then output BMI\_22;

**run**;

**proc** **sql**;

create table BMI\_3 as

select distinct a.\*,b.\*

from BMI\_22 as a

left join CCX\_cq\_rst\_2010\_3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.CCX\_date=b.CCX\_date;

**quit**;

**data** BMI\_31 BMI\_32(drop= QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN FAM);

set BMI\_3;

if FAM ne **.** then output BMI\_31;

else if FAM =**.** then output BMI\_32;

**run**;

**proc** **sql**;

create table BMI\_4 as

select distinct a.\*,b.\*

from BMI\_32 as a

left join CBR\_cq\_rst\_2010\_3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.CBR\_date=b.CBR\_date;

**quit**;

**data** BMI\_41 BMI\_42(drop= QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN FAM);

set BMI\_4;

if FAM ne **.** then output BMI\_41;

else if FAM =**.** then output BMI\_42;

**run**;

**proc** **sql**;

create table BMI\_5 as

select distinct a.\*,b.\*

from BMI\_42 as a

left join CLV\_cq\_rst\_2010\_3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.CLV\_date=b.CLV\_date;

**quit**;

**data** BMI\_51 BMI\_52(drop= QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN FAM);

set BMI\_5;

if FAM ne **.** then output BMI\_51;

else if FAM =**.** then output BMI\_52;

**run**;

**data** D.BMI\_C;

set BMI\_11 BMI\_21 BMI\_31 BMI\_41 BMI\_51;

**run**;

**proc** **sql**;

create table d.T20\_C\_INDEX\_1 as

select a.\*,b.INDEX\_DATE

from d.T20\_C as a

left join D.BMI\_C as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO

having INDEX\_DATE>**0** and INDEX\_DATE<=recu\_date;

**quit**;

**data** d.T20\_SH;

set raw.t20\_201001 raw.t20\_201002 raw.t20\_201003 raw.t20\_201004 raw.t20\_201005 raw.t20\_201006 raw.t20\_201007 raw.t20\_201008 raw.t20\_201009 raw.t20\_201010 raw.t20\_201011 raw.t20\_201012 raw.t20\_201101 raw.t20\_201102 raw.t20\_201103 raw.t20\_201104 raw.t20\_201105 raw.t20\_201106;

where "I05"<=substr(SICK\_SYM1,**1**,**3**)<="I09" or

"I20"<=substr(SICK\_SYM1,**1**,**3**)<="I27" or

"I30"<=substr(SICK\_SYM1,**1**,**3**)<="I52" or

"I60"<=substr(SICK\_SYM1,**1**,**3**)<="I69" ;

if "I60"<=substr(SICK\_SYM1,**1**,**3**)<="I69" then STK\_20=**1**;

if "I05"<=substr(SICK\_SYM1,**1**,**3**)<="I09" or

"I20"<=substr(SICK\_SYM1,**1**,**3**)<="I27" or

"I30"<=substr(SICK\_SYM1,**1**,**3**)<="I52" then HTDZ\_20=**1**;

recu\_date=input(MDCARE\_STRT\_DT,yymmdd8.);

format recu\_date yymmdd10.;

drop MDCARE\_STRT\_DT;

**run**;

**proc** **sql**;

create table d.T20\_SH\_INDEX\_1 as

select a.\*,b.INDEX\_DATE

from d.T20\_SH as a

left join D.BMI\_C as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO

having INDEX\_DATE>**0** and INDEX\_DATE<=recu\_date<=INDEX\_DATE+**6**\***30.5**;

**quit**;

**data** T20\_SH\_INDEX;

set d.T20\_SH\_INDEX\_1;

where "I20"<=substr(SICK\_SYM1,**1**,**3**)<="I25" or

"I60"<=substr(SICK\_SYM1,**1**,**3**)<="I64"

or substr(SICK\_SYM1,**1**,**3**)="I69";

**run**;

**proc** **sql**;

create table STK\_20 as

select distinct INDI\_DSCM\_NO, STK\_20 , min(recu\_date) as STK\_20\_DATE format yymmdd10.

from T20\_SH\_INDEX where STK\_20=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table HTDZ\_20 as

select distinct INDI\_DSCM\_NO, HTDZ\_20 , min(recu\_date) as HTDZ\_20\_DATE format yymmdd10.

from T20\_SH\_INDEX where HTDZ\_20=**1**

group by INDI\_DSCM\_NO;**quit**;

**data** CANCER;

set d.T20\_C\_INDEX\_1;

if upcase(substr(SICK\_SYM1,**1**,**1**))="C" then AC=**1**;else AC=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C16" then C16=**1**; else C16=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C22" then C22=**1**; else C22=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C23" then C23=**1**; else C23=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C24" then C24=**1**; else C24=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C25" then C25=**1**; else C25=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C18" then C18=**1**; else C18=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C19" then C19=**1**; else C19=**0**;

if upcase(substr(SICK\_SYM1,**1**,**3**))="C20" then C20=**1**; else C20=**0**;

**run**;

**proc** **sql**;

create table AC as

select distinct INDI\_DSCM\_NO, AC , min(recu\_date) as AC\_DATE format yymmdd10.

from CANCER where AC=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C16 as

select distinct INDI\_DSCM\_NO, C16 , min(recu\_date) as C16\_DATE format yymmdd10.

from CANCER where C16=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C22 as

select distinct INDI\_DSCM\_NO, C22 , min(recu\_date) as C22\_DATE format yymmdd10.

from CANCER where C22=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C23 as

select distinct INDI\_DSCM\_NO, C23 , min(recu\_date) as C23\_DATE format yymmdd10.

from CANCER where C23=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C24 as

select distinct INDI\_DSCM\_NO, C24 , min(recu\_date) as C24\_DATE format yymmdd10.

from CANCER where C24=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C25 as

select distinct INDI\_DSCM\_NO, C25 , min(recu\_date) as C25\_DATE format yymmdd10.

from CANCER where C25=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C18 as

select distinct INDI\_DSCM\_NO, C18 , min(recu\_date) as C18\_DATE format yymmdd10.

from CANCER where C18=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C19 as

select distinct INDI\_DSCM\_NO, C19 , min(recu\_date) as C19\_DATE format yymmdd10.

from CANCER where C19=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table C20 as

select distinct INDI\_DSCM\_NO, C20 , min(recu\_date) as C20\_DATE format yymmdd10.

from CANCER where C20=**1**

group by INDI\_DSCM\_NO;**quit**;

**proc** **sql**;

create table D1 as

select a.\*,aa.\*,b.\*,c.\*,d.\*,e.\*,f.\*,g.\*,h.\*,i.\*,j.\*,k.\*

from D.BMI\_C as a

left join C16 as b on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO

left join C22 as c on a.INDI\_DSCM\_NO=c.INDI\_DSCM\_NO

left join C23 as d on a.INDI\_DSCM\_NO=d.INDI\_DSCM\_NO

left join C24 as e on a.INDI\_DSCM\_NO=e.INDI\_DSCM\_NO

left join C25 as f on a.INDI\_DSCM\_NO=f.INDI\_DSCM\_NO

left join C18 as g on a.INDI\_DSCM\_NO=g.INDI\_DSCM\_NO

left join C19 as h on a.INDI\_DSCM\_NO=h.INDI\_DSCM\_NO

left join C20 as i on a.INDI\_DSCM\_NO=i.INDI\_DSCM\_NO

left join STK\_20 as j on a.INDI\_DSCM\_NO=j.INDI\_DSCM\_NO

left join HTDZ\_20 as k on a.INDI\_DSCM\_NO=k.INDI\_DSCM\_NO

left join AC as aa on a.INDI\_DSCM\_NO=aa.INDI\_DSCM\_NO;

**quit**;

**data** D5;

set D1;

if DEATH\_DATE ne **.** then DEATH=**1**; else DEATH=**0**;

CRC=max(C18,C19,C20);

if CRC=**1** then CRC\_DATE=min(C18\_DATE,C19\_DATE,C20\_DATE);

format CRC\_DATE yymmdd10.;

drop C18 C19 C20 C18\_DATE C19\_DATE C20\_DATE;

if AC=**.** then do;

AC=**0**; AC\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<AC\_DATE then AC\_DATE=DEATH\_DATE;end;

if C16=**.** then do;

C16=**0**; C16\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<C16\_DATE then C16\_DATE=DEATH\_DATE;end;

if C22=**.** then do;

C22=**0**; C22\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<C22\_DATE then C22\_DATE=DEATH\_DATE;end;

if C23=**.** then do;

C23=**0**; C23\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<C23\_DATE then C23\_DATE=DEATH\_DATE;end;

if C24=**.** then do;

C24=**0**; C24\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<C24\_DATE then C24\_DATE=DEATH\_DATE;end;

if C25=**.** then do;

C25=**0**; C25\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<C25\_DATE then C25\_DATE=DEATH\_DATE;end;

if CRC=**.** then do;

CRC=**0**; CRC\_DATE=input("2021-12-31",yymmdd10.);

if DEATH=**1** and DEATH\_DATE<CRC\_DATE then CRC\_DATE=DEATH\_DATE;end;

**run**;

**data** K;

set d.T20\_K\_2010;

where substr(SICK\_SYM1,**1**,**4**) in ("K758","K760") or SICK\_SYM1="K76";

**run**;

**proc** **sql**;

create table K1 as

select a.\*,b.INDEX\_DATE

from K as a

left join D5 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO and a.recu\_date>=b.INDEX\_DATE

having INDEX\_DATE>**0**;

**quit**;

**proc** **sort** data=K1 out=K2 nodupkey;

by INDI\_DSCM\_NO;

**quit**;

**data** K3;

set K2;K=**1**;keep INDI\_DSCM\_NO K;

**run**;

**proc** **sql**;

create table B5 as

select a.\*,case when b.K=**1** then **1** else **0** end as K

from D5 as a

left join K3 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO;

**quit**;

**data** D.BMI\_ALL;

set B5;

drop a;

AC\_DAY=AC\_DATE-INDEX\_DATE;

AC\_YEAR=AC\_DAY/**365**;

C16\_DAY=C16\_DATE-INDEX\_DATE;

C22\_DAY=C22\_DATE-INDEX\_DATE;

C23\_DAY=C23\_DATE-INDEX\_DATE;

C24\_DAY=C24\_DATE-INDEX\_DATE;

C25\_DAY=C25\_DATE-INDEX\_DATE;

CRC\_DAY=CRC\_DATE-INDEX\_DATE;

Age= **2010**-byear;

if Q\_PHX\_DX\_STK=**1** then STK=**1**;

else if Q\_PHX\_DX\_STK ne **1** and STK\_20=**1** then STK=**1**;

else STK=**0**;

if Q\_PHX\_DX\_HTDZ=**1** then HTDZ=**1**;

else if Q\_PHX\_DX\_HTDZ ne **1** and HTDZ\_20=**1** then HTDZ=**1**;

else HTDZ=**0**;

if Q\_PHX\_DX\_HTN=**1** then HTN=**1**;

else if Q\_PHX\_DX\_HTN ne **1** and (G1E\_BP\_SYS>=**140** or G1E\_BP\_DIA>=**90**) then HTN=**1**;

else HTN=**0**;

if Q\_PHX\_DX\_DM=**1** then DM=**1**;

else if Q\_PHX\_DX\_DM ne **1** and (G1E\_FBS>=**126** ) then DM=**1**;

else DM=**0**;

if Q\_PHX\_DX\_DLD=**1** then DLD=**1**;

else DLD=**0**;

if Q\_DRK\_FRQ\_V09N=**0** then DRK=**0**;

else if Q\_DRK\_FRQ\_V09N=**1** then DRK=**1**;

else if Q\_DRK\_FRQ\_V09N in (**2**,**3**) then DRK=**2**;

else if Q\_DRK\_FRQ\_V09N in (**4**,**5**) then DRK=**3**;

else if Q\_DRK\_FRQ\_V09N in (**6**,**7**) then DRK=**4**;

if Q\_PA\_MD=**0** then MD=**0**;

else if Q\_PA\_MD in (**1**,**2**) then MD=**1**;

else if Q\_PA\_MD in (**3**,**4**,**5**) then MD=**2**;

else if Q\_PA\_MD in (**6**,**7**) then MD=**3**;

if G1E\_BMI<**18.5** then BMI\_2010='1';

else if G1E\_BMI<**23** then BMI\_2010='2';

else if G1E\_BMI<**25** then BMI\_2010='3';

else if G1E\_BMI<**30** then BMI\_2010='4';

else if G1E\_BMI>=**30** then BMI\_2010='5';

if g1e\_ldl>=**130** then LDL=**1**; else LDL=**0**;

if g1e\_tg>=**150** then TG=**1**; else TG=**0**;

DRK\_A=Q\_DRK\_FRQ\_V09N\*Q\_DRK\_AMT\_V09N;

if DRK\_A=**.** then DRK\_A=**0**;

if DRK\_A=**0** then DRK\_CAT=**0**;

else if DRK\_A<**7** then DRK\_CAT=**1**;

else if DRK\_A<**14** then DRK\_CAT=**2**;

else if DRK\_A<**21** then DRK\_CAT=**3**;

else if DRK\_A>=**21** then DRK\_CAT=**4**;

**run**;

**proc** **means** data=D.BMI\_ALL mean std median q1 q3;

class BMI\_2010;

var age calc\_ctrb\_vtile\_fd G1E\_WSTC g1e\_bmi G1E\_FBS G1E\_TG G1E\_SGOT G1E\_SGPT g1e\_hdl G1E\_LDL Q\_PA\_MD;

**run**;

**proc** **freq** data=D.BMI\_ALL;

table (SEX\_TYPE HTN HTDZ STK DM K

DRK Q\_SMK\_YN FAM DLD

QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN QC\_MNS\_YN QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN )\*BMI\_2010/norow nopercent nocol;

**run**;

**Table 2. Hepatobiliary cancer risk according to BMI change**

**(ref 22)**

**Unadjusted**

**Adjusted**

C22: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

**data** BMI\_C234;

set D.BMI\_ALL;

if C22=**1** then do;C22\_COM=**1**;C22\_COM\_DAY=C22\_DAY;end;

else if DEATH\_DATE>**0** then do;C22\_COM=**2**;C22\_COM\_DAY=C22\_DAY;end;

else if C22=**0** and AC=**1** then do;C22\_COM=**2**;C22\_COM\_DAY=AC\_DAY;end;

else if C22=**0** and AC=**0** then do;C22\_COM=**0**;C22\_COM\_DAY=C22\_DAY;end;

if C23=**1** then do;C23\_COM=**1**;C23\_COM\_DAY=C23\_DAY;end;

else if DEATH\_DATE>**0** then do;C23\_COM=**2**;C23\_COM\_DAY=C23\_DAY;end;

else if C23=**0** and AC=**1** then do;C23\_COM=**2**;C23\_COM\_DAY=AC\_DAY;end;

else if C23=**0** and AC=**0** then do;C23\_COM=**0**;C23\_COM\_DAY=C23\_DAY;end;

if C24=**1** then do;C24\_COM=**1**;C24\_COM\_DAY=C24\_DAY;end;

else if DEATH\_DATE>**0** then do;C24\_COM=**2**;C24\_COM\_DAY=C24\_DAY;end;

else if C24=**0** and AC=**1** then do;C24\_COM=**2**;C24\_COM\_DAY=AC\_DAY;end;

else if C24=**0** and AC=**0** then do;C24\_COM=**0**;C24\_COM\_DAY=C24\_DAY;end;

if G1E\_BMI<**18.5** then BMI\_2010\_1='1';

else if **18.5**<=G1E\_BMI<**23** then BMI\_2010\_1='2';

else if **23**<=G1E\_BMI<**25** then BMI\_2010\_1='3';

else if **25**<=G1E\_BMI<**30** then BMI\_2010\_1='4';

else if G1E\_BMI>=**30** then BMI\_2010\_1='5';

if G1E\_BMI\_2014<**18.5** then BMI\_2014\_2='1';

else if **18.5**<=G1E\_BMI\_2014<**23** then BMI\_2014\_2='2';

else if **23**<=G1E\_BMI\_2014<**25** then BMI\_2014\_2='3';

else if **25**<=G1E\_BMI\_2014<**30** then BMI\_2014\_2='4';

else if G1E\_BMI\_2014>=**30** then BMI\_2014\_2='5';

if BMI\_2010\_1='1' and BMI\_2014\_2='1' then BMI\_C='11';

else if BMI\_2010\_1='1' and BMI\_2014\_2='2' then BMI\_C='12';

else if BMI\_2010\_1='1' and BMI\_2014\_2='3' then BMI\_C='13';

else if BMI\_2010\_1='1' and BMI\_2014\_2='4' then BMI\_C='14';

else if BMI\_2010\_1='1' and BMI\_2014\_2='5' then BMI\_C='15';

else if BMI\_2010\_1='2' and BMI\_2014\_2='1' then BMI\_C='21';

else if BMI\_2010\_1='2' and BMI\_2014\_2='2' then BMI\_C='22';

else if BMI\_2010\_1='2' and BMI\_2014\_2='3' then BMI\_C='23';

else if BMI\_2010\_1='2' and BMI\_2014\_2='4' then BMI\_C='24';

else if BMI\_2010\_1='2' and BMI\_2014\_2='5' then BMI\_C='25';

else if BMI\_2010\_1='3' and BMI\_2014\_2='1' then BMI\_C='31';

else if BMI\_2010\_1='3' and BMI\_2014\_2='2' then BMI\_C='32';

else if BMI\_2010\_1='3' and BMI\_2014\_2='3' then BMI\_C='33';

else if BMI\_2010\_1='3' and BMI\_2014\_2='4' then BMI\_C='34';

else if BMI\_2010\_1='3' and BMI\_2014\_2='5' then BMI\_C='35';

else if BMI\_2010\_1='4' and BMI\_2014\_2='1' then BMI\_C='41';

else if BMI\_2010\_1='4' and BMI\_2014\_2='2' then BMI\_C='42';

else if BMI\_2010\_1='4' and BMI\_2014\_2='3' then BMI\_C='43';

else if BMI\_2010\_1='4' and BMI\_2014\_2='4' then BMI\_C='44';

else if BMI\_2010\_1='4' and BMI\_2014\_2='5' then BMI\_C='45';

else if BMI\_2010\_1='5' and BMI\_2014\_2='1' then BMI\_C='51';

else if BMI\_2010\_1='5' and BMI\_2014\_2='2' then BMI\_C='52';

else if BMI\_2010\_1='5' and BMI\_2014\_2='3' then BMI\_C='53';

else if BMI\_2010\_1='5' and BMI\_2014\_2='4' then BMI\_C='54';

else if BMI\_2010\_1='5' and BMI\_2014\_2='5' then BMI\_C='55';

if FAM=**1** then FAM\_CAT=**2**;

else FAM\_CAT=**1**; **run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K/param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K/rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Table 3. Effect of weight reduction among obese or overweight**

**Among obese persons (ref 44)**

**Among overweighed persons (ref 33)**

**Adjusted**

C22: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

**data** BMI\_CRC;

set D.BMI\_ALL;

if CRC=**1** then do;CRC\_COM=**1**;CRC\_COM\_DAY=CRC\_DAY;end;

else if DEATH\_DATE>**0** then do;CRC\_COM=**2**;CRC\_COM\_DAY=CRC\_DAY;end;

else if CRC=**0** and AC=**1** then do;CRC\_COM=**2**;CRC\_COM\_DAY=AC\_DAY;end;

else if CRC=**0** and AC=**0** then do;CRC\_COM=**0**;CRC\_COM\_DAY=CRC\_DAY;end;

if G1E\_BMI<**18.5** then BMI\_2010\_1='1';

else if **18.5**<=G1E\_BMI<**23** then BMI\_2010\_1='2';

else if **23**<=G1E\_BMI<**25** then BMI\_2010\_1='3';

else if **25**<=G1E\_BMI<**30** then BMI\_2010\_1='4';

else if G1E\_BMI>=**30** then BMI\_2010\_1='5';

if G1E\_BMI\_2014<**18.5** then BMI\_2014\_2='1';

else if **18.5**<=G1E\_BMI\_2014<**23** then BMI\_2014\_2='2';

else if **23**<=G1E\_BMI\_2014<**25** then BMI\_2014\_2='3';

else if **25**<=G1E\_BMI\_2014<**30** then BMI\_2014\_2='4';

else if G1E\_BMI\_2014>=**30** then BMI\_2014\_2='5';

if BMI\_2010\_1='1' and BMI\_2014\_2='1' then BMI\_C='11';

else if BMI\_2010\_1='1' and BMI\_2014\_2='2' then BMI\_C='12';

else if BMI\_2010\_1='1' and BMI\_2014\_2='3' then BMI\_C='13';

else if BMI\_2010\_1='1' and BMI\_2014\_2='4' then BMI\_C='14';

else if BMI\_2010\_1='1' and BMI\_2014\_2='5' then BMI\_C='15';

else if BMI\_2010\_1='2' and BMI\_2014\_2='1' then BMI\_C='21';

else if BMI\_2010\_1='2' and BMI\_2014\_2='2' then BMI\_C='22';

else if BMI\_2010\_1='2' and BMI\_2014\_2='3' then BMI\_C='23';

else if BMI\_2010\_1='2' and BMI\_2014\_2='4' then BMI\_C='24';

else if BMI\_2010\_1='2' and BMI\_2014\_2='5' then BMI\_C='25';

else if BMI\_2010\_1='3' and BMI\_2014\_2='1' then BMI\_C='31';

else if BMI\_2010\_1='3' and BMI\_2014\_2='2' then BMI\_C='32';

else if BMI\_2010\_1='3' and BMI\_2014\_2='3' then BMI\_C='33';

else if BMI\_2010\_1='3' and BMI\_2014\_2='4' then BMI\_C='34';

else if BMI\_2010\_1='3' and BMI\_2014\_2='5' then BMI\_C='35';

else if BMI\_2010\_1='4' and BMI\_2014\_2='1' then BMI\_C='41';

else if BMI\_2010\_1='4' and BMI\_2014\_2='2' then BMI\_C='42';

else if BMI\_2010\_1='4' and BMI\_2014\_2='3' then BMI\_C='43';

else if BMI\_2010\_1='4' and BMI\_2014\_2='4' then BMI\_C='44';

else if BMI\_2010\_1='4' and BMI\_2014\_2='5' then BMI\_C='45';

else if BMI\_2010\_1='5' and BMI\_2014\_2='1' then BMI\_C='51';

else if BMI\_2010\_1='5' and BMI\_2014\_2='2' then BMI\_C='52';

else if BMI\_2010\_1='5' and BMI\_2014\_2='3' then BMI\_C='53';

else if BMI\_2010\_1='5' and BMI\_2014\_2='4' then BMI\_C='54';

else if BMI\_2010\_1='5' and BMI\_2014\_2='5' then BMI\_C='55';

if FAM=**1** then FAM\_CAT=**2**;

else FAM\_CAT=**1**;

if age<**60** then age\_cat\_60=**1**;

else if age>=**60** then age\_cat\_60=**2**;

**run**;

**data** BMI\_CRC\_2;

set BMI\_CRC;

if AC=**1** and AC\_DAY<**2**\***365** then delete;

**run**;

\*among obese person;

\*model2;

**proc** **phreg** data=BMI\_CRC\_2(where=(BMI\_C in ('42','43','44','45')));

class BMI\_C(ref="44") /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_CRC\_2(where=(BMI\_C in ('42','43','44','45')));

class BMI\_C(ref="44") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG CRC\_FAM\_CAT K /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl CRC\_FAM\_CAT K /rl eventcode=**1**;**run**;

\*among overweight person;

\*M1;

**proc** **phreg** data=BMI\_CRC(where=(BMI\_C in ('31','32','33','34','35')));

class BMI\_C(ref="33") /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_CRC(where=(BMI\_C in ('31','32','33','34','35')));

class BMI\_C(ref="33") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG CRC\_FAM\_CAT K /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl CRC\_FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Table 1. Women factors by baseline body mass index**

**data** BMI\_CRC\_3;

set BMI\_CRC;

if AC=**1** and AC\_DATE<input('20150101',yymmdd8.) then delete;

**run**;

**proc** **phreg** data=BMI\_CRC\_3(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**1**));

class BMI\_2010\_1(ref="1") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG CRC\_FAM\_CAT K QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_2010\_1 age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl CRC\_FAM\_CAT K QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN/rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_CRC\_3(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**3**));

class BMI\_2010\_1(ref="1") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG CRC\_FAM\_CAT K QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN /param=ref ref=first;

model CRC\_COM\_DAY\*CRC\_COM(**0**)=BMI\_2010\_1 age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl CRC\_FAM\_CAT K QC\_BRFD\_DRT QC\_DLV\_FRQ QC\_OPLL\_YN/rl eventcode=**1**;**run**;

**Supplementary Table 2. Sensitivity analysis (mode II, model III)**

**(ref 22)**

**Unadjusted**

**Adjusted**

C22: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

To conduct sensitivity analysis for the impact of the exposure duration, we excluded cancers developed within 24 months from the first BMI measurement (Model II) and up to second measurement of BMI (Model III).

BMI, body mass index; CI, confidence interval; HR, hazard ratio.

**proc** **sql**;

create table BMI\_CRC\_2014 as

select a.\*,input(b.HME\_DT,yymmdd8.) as DATE\_2014 format yymmdd10.

from BMI\_C234 as a

left join RAW.G1EQ\_2014 as b

on a.INDI\_DSCM\_NO=b.INDI\_DSCM\_NO;

**quit**;

**data** S1;

set BMI\_C234;

if AC=**1** and AC\_DATE-INDEX\_DATE<**365**\***2** then delete;

**run**;

**data** S2;

set BMI\_CRC\_2014;

if AC=**1** and AC\_DATE<DATE\_2014 then delete;

**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K/param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K/rl eventcode=**1**;**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S1;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K/param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K/rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C /rl eventcode=**1**;**run**;

**proc** **phreg** data=S2;

class BMI\_C(ref="22") SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Table 3. Interaction analysis**

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") Q\_SMK\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI13\_CAT Q\_SMK\_YN BMI13\_CAT\*Q\_SMK\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") Q\_SMK\_YN /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI13\_CAT Q\_SMK\_YN BMI13\_CAT\*Q\_SMK\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") Q\_SMK\_YN /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI13\_CAT Q\_SMK\_YN BMI13\_CAT\*Q\_SMK\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") sex\_type /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI13\_CAT sex\_type BMI13\_CAT\*sex\_type /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") sex\_type /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI13\_CAT sex\_type BMI13\_CAT\*sex\_type /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") sex\_type /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI13\_CAT sex\_type BMI13\_CAT\*sex\_type /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") age\_cat\_60 /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI13\_CAT age\_cat\_60 BMI13\_CAT\*age\_cat\_60 /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") age\_cat\_60 /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI13\_CAT age\_cat\_60 BMI13\_CAT\*age\_cat\_60 /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')));

class BMI13\_CAT(ref="22") age\_cat\_60 /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI13\_CAT age\_cat\_60 BMI13\_CAT\*age\_cat\_60 /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55') and SEX\_TYPE='2'));

class BMI13\_CAT(ref="22") QC\_MNS\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI13\_CAT QC\_MNS\_YN BMI13\_CAT\*QC\_MNS\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')and SEX\_TYPE='2'));

class BMI13\_CAT(ref="22") QC\_MNS\_YN /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI13\_CAT QC\_MNS\_YN BMI13\_CAT\*QC\_MNS\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI13\_CAT234(where=(BMI13\_CAT in ('11','22','33','44','55')and SEX\_TYPE='2'));

class BMI13\_CAT(ref="22") QC\_MNS\_YN /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI13\_CAT QC\_MNS\_YN BMI13\_CAT\*QC\_MNS\_YN /rl eventcode=**1**;**run**;

**Supplementary Table 4. Subgroup analysis by sex**

**(ref 22)**

**Adjusted**

C22: Adjusted for age, ~~sex,~~ income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='1'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='1'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='1'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2'));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Table 5. Subgroup analysis by smoking status**

**(ref 22)**

**Adjusted**

C22: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, ~~smoking status,~~ drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN in (**2**,**3**)));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN in (**2**,**3**)));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(Q\_SMK\_YN in (**2**,**3**)));

class BMI\_C(ref="22") HTN HTDZ STK DM SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Table 6. Subgroup analysis by age group**

**(ref 22)**

**Adjusted**

C22: Adjusted for ~~age,~~ sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, sex, income, hypertension, heart disease, stroke, diabetes mellitus, smoking status, drinking status, physical activity, lipid lowering drug, family history of cancer, triglyceride, HDL-C, and LDL-C.

**proc** **phreg** data=BMI\_C234(where=(age<**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(age>=**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(age<**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(age>=**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(age<**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(age>=**60**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN SEX\_TYPE MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C calc\_ctrb\_vtile\_fd Q\_SMK\_YN SEX\_TYPE HTN HTDZ STK DM MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Table 7. Subgroup analysis by menopausal status**

**(ref 22)**

**Adjusted**

C22: Adjusted for age, income, hypertension, heart disease, stroke, DM, drinking status, smoking status, physical activity, lipid lowering drug, family history of liver cancer, any liver disease, chronic hepatitis B, chronic hepatitis C, liver cirrhosis, non-alcoholic fatty liver disease, breast feeding, parity, use of oral contraceptives, triglyceride, HDL-C, and LDL-C.

C23, C24: Adjusted for age, income, hypertension, heart disease, stroke, DM, smoking status, physical activity, lipid lowering drug, breast feeding, parity, use of oral contraceptives, triglyceride, HDL-C, and LDL-C.

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**3**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /param=ref ref=first;

model C22\_COM\_DAY\*C22\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K QC\_PHX\_LV\_YN QC\_PHX\_CHB\_YN QC\_PHX\_CHC\_YN QC\_PHX\_LC\_YN /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**3**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C23\_COM\_DAY\*C23\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**1**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**proc** **phreg** data=BMI\_C234(where=(SEX\_TYPE='2' and QC\_MNS\_YN=**3**));

class BMI\_C(ref="22") HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG FAM\_CAT K /param=ref ref=first;

model C24\_COM\_DAY\*C24\_COM(**0**)=BMI\_C age calc\_ctrb\_vtile\_fd HTN HTDZ STK DM Q\_SMK\_YN MD DLD DRK TG g1e\_hdl FAM\_CAT K /rl eventcode=**1**;**run**;

**Supplementary Fig 1. KM curve by BMI**

<RCODE>

library(haven)

library(survival)

library(cmprsk)

library(survminer)

library(dplyr)

library(rms)

library(ggplot2)

C234\_1<-read.csv("C234\_1.csv",header=T)

C234\_2<-read.csv("C234\_2.csv",header=T)

C234\_3<-read.csv("C234\_3.csv",header=T)

C234<-rbind(C234\_1,C234\_2,C234\_3)

C22<-C234[,c(1,4,7)]

C23<-C234[,c(2,5,7)]

C24<-C234[,c(3,6,7)]

a<-survfit(Surv(C22\_YEAR,C22)~BMI\_C,data=C22)

ggsurvplot(fit=a,data=C22,break.time.by=1,title="C22",

fun=function(x){(1-x)},

xlab="Year", legend="right",censor=F,conf.int=F,legend.title="BMI",

legend.labs=c("Persistent underweight","Persistent normal BMI","Persistent overweight",

"Persistent obesity I","Persistent obesity II"),ylab="Cumulative Incidence"

,xlim=c(0,11), ylim=c(0,0.013),risk.table=T, risk.table.height=0.3, risk.table.title="",

tables.y.text.col=F, font.title=c(20) , font.y=c(18), font.x=c(18) , font.legend=c(17))

a<-survfit(Surv(C23\_YEAR,C23)~BMI\_C,data=C23)

ggsurvplot(fit=a,data=C23,break.time.by=1,title="C23",

fun=function(x){(1-x)},

xlab="Year", legend="right",censor=F,conf.int=F,legend.title="BMI",

legend.labs=c("Persistent underweight","Persistent normal BMI","Persistent overweight",

"Persistent obesity I","Persistent obesity II"),ylab="Cumulative Incidence"

,xlim=c(0,11), ylim=c(0,0.002),risk.table=T, risk.table.height=0.3, risk.table.title="",

tables.y.text.col=F, font.title=c(20) , font.y=c(18), font.x=c(18) , font.legend=c(17))

a<-survfit(Surv(C24\_YEAR,C24)~BMI\_C,data=C24)

ggsurvplot(fit=a,data=C24,break.time.by=1,title="C24",

fun=function(x){(1-x)},

xlab="Year", legend="right",censor=F,conf.int=F,legend.title="BMI",

legend.labs=c("Persistent underweight","Persistent normal BMI","Persistent overweight",

"Persistent obesity I","Persistent obesity II"),ylab="Cumulative Incidence"

,xlim=c(0,11), ylim=c(0,0.003),risk.table=T, risk.table.height=0.3, risk.table.title="",

tables.y.text.col=F, font.title=c(20) , font.y=c(18), font.x=c(18) , font.legend=c(17))