/\* 라이브러리 지정\*/

libname A "/userdata10/room583/data\_source/user\_data";

libname STORE "/userdata10/room583/data\_out/data\_store";

/\*20231012 환자군, 대조군 재구축

환자군: 05~17년도에 폐암인 케이스(C33 or C34 + V193 or V194) 이면서 02~04년도 폐암케이스 없는 대상자

대조군: 02~17년 폐암케이스 없는 대상자 \*/

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

/\*계획\*/

/\*환자군\*/

/\*1. 기준 -> 05~17년도에 폐암인 케이스(C33 or C34 + V193 or V194) N수 구하기\*/

/\*2. 기준 대상자에서 02~04년도 암 케이스(C33-, C34-) 제거한 N수 구하기\*/

/\*3. 기준 대상자에서 02~04년도 암 케이스(C-) 제거한 N수 구하기\*/

/\*모집단\*/

/\* + 위 환자군 중에서 이번 연구주제인 모집단 10~16년 폐암인 케이스 (C33 or C34 + V193 or V194) N수 구하기 \*/

/\*대조군\*/

/\*1. 기준 -> 02~17년도에 폐암케이스(C33 or C34 + V193 or V194)가 아닌 N수 구하기\*/

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

/\*1. 기준 -> 환자군 5~17년도에 폐암인 케이스(C33 or C34 + V193 or V194) N수 구하기\*/

/\*1-1\*/

/\*2005년~ 2017년 T20. T40 년도별 테이블 생성 (2005.01.01 ~ 2017.12.31)\*/

/\*T20\*/

**data** STORE.T20\_2006;

set A.T20\_2006: ;

STD\_YYYY = SUBSTR(MDCARE\_STRT\_DT,**1**,**4**);

**run**;

**data** STORE.T20\_2007;

set A.T20\_2007: ;

STD\_YYYY = SUBSTR(MDCARE\_STRT\_DT,**1**,**4**);

**run**;

/\*1-2\*/

/\*2005년~ 2017년 합친 테이블에서 + 상병코드 C33-,C34- \*/

/\*중복제거는 추 후 할 예정\*/

/\*T20\*/

**data** store.t20\_c33\_c34\_0517;

set STORE.t20\_2005 - STORE.t20\_2017;

if substr(sick\_sym1,**1**,**3**) in ('C33','C34') or

substr(sick\_sym2,**1**,**3**) in ('C33','C34') or

substr(sick\_sym3,**1**,**3**) in ('C33','C34') or

substr(sick\_sym4,**1**,**3**) in ('C33','C34') or

substr(sick\_sym5,**1**,**3**) in ('C33','C34') then cancer=**1**;

if cancer=**1**;

**run**;

/\*1-3\*/

/\* 위 테이블(05~17년 C33-,C34-) + V193, V194 특정기호구분\*/

/\*T20\*/

**data** store.t20\_c33\_c34\_0517\_v;

set store.t20\_c33\_c34\_0517;

if spcf\_sym\_type in ('V193','V194') then cancer\_v=**1**;

if cancer\_v=**1**;

**run**;

/\*1-4\*/

/\* 위 테이블(05~17년 C33-,C34- + V193, V194) + 중복 제거\*/

/\*중복 제거 방법 - 빠른 방문일자 id만 남겨두기\*/

/\*T20\*/

/\*① 정렬(sas는 정렬 먼저 해줘야함!! 아이디로 정렬한 후 날짜로 정렬)\*/

**proc** **sort** data=store.t20\_c33\_c34\_0517\_v;

by indi\_dscm\_no mdcare\_strt\_dt;

**run**;

/\*② 방문 횟수 세기(아이디별 방문 row수 찍어보기)\*/

**data** store.t20\_c\_0517\_v\_visit;

set store.t20\_c33\_c34\_0517\_v;

retain row\_number **0**;

by indi\_dscm\_no mdcare\_strt\_dt;

if first.indi\_dscm\_no then row\_number = **1**;

else row\_number+**1**;

**run**;

/\*③ 방문 횟수 첫번째만 남기기\*/

**data** store.t20\_c\_0517\_v\_visit\_1;

set store.t20\_c\_0517\_v\_visit;

if row\_number=**1**;

**run**;

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

/\* t20\_c\_0517\_v\_visit\_1 테이블 확인\*/

/\*환자군 => 2005년 ~ 2016년도 t20테이블 + C33-,C34- + v193, v194 & 첫 방문(중복제거) n 수 > 307,458명\*/

**proc** **contents** data=store.t20\_c\_0517\_v\_visit\_1;

**run**;

/\*------------------------------------------case1-------------------------------------------------------\*/

/\*2. 기준 대상자에서 02~04년도 암 케이스(C33-, C34-) 제거한 N수 구하기\*/

/\*2-1\*/

/\*02~04년도별 테이블 생성\*/

**data** STORE.T20\_2002;

set A.T20\_2002: ;

STD\_YYYY = SUBSTR(MDCARE\_STRT\_DT,**1**,**4**);

**run**;

**data** STORE.T20\_2003;

set A.T20\_2003: ;

STD\_YYYY = SUBSTR(MDCARE\_STRT\_DT,**1**,**4**);

**run**;

**data** STORE.T20\_2004;

set A.T20\_2004: ;

STD\_YYYY = SUBSTR(MDCARE\_STRT\_DT,**1**,**4**);

**run**;

/\*2-2\*/

/\*02~04년도 테이블 합치기 + 상병코드 C33-,C34- \*/

**data** store.t20\_c33\_c34\_0204;

set store.t20\_2002 - a.t20\_2004;

if substr(sick\_sym1,**1**,**3**) in ('C33','C34') or

substr(sick\_sym2,**1**,**3**) in ('C33','C34') or

substr(sick\_sym3,**1**,**3**) in ('C33','C34') or

substr(sick\_sym4,**1**,**3**) in ('C33','C34') or

substr(sick\_sym5,**1**,**3**) in ('C33','C34') then cancer=**1**;

if cancer=**1**;

**run**;

/\*2-2\*/

/\* 위 테이블 + 중복제거\*/

/\* ① 아이디별 방문 row수 찍어보기\*/

**data** store.t20\_c33\_c34\_0204\_visit;

set store.t20\_c33\_c34\_0204;

retain row\_number **0**;

by indi\_dscm\_no STD\_YYYY;

if first.indi\_dscm\_no then row\_number = **1**;

else row\_number+**1**;

**run**;

/\* ② 방문 횟수 첫번째만 남기기\*/

**data** store.t20\_c33\_c34\_0204\_visit\_1;

set store.t20\_c33\_c34\_0204\_visit;

if row\_number=**1**;

**run**;

/\*2-3\*/

/\*(2005년 ~ 2017년 + C33-. C34- + V193, V194) - (2002년 ~ 2004년 + C33-. C34-)

=> (t20\_c\_0517\_v\_visit\_1) - ( t20\_c\_0517\_v\_visit\_1 INNER JOIN t20\_c33\_c34\_0204\_visit\_1)\*/

/\*①indi\_dscm\_no으로 정렬\*/

**proc** **sort** data=store.t20\_c\_0517\_v\_visit\_1;

by indi\_dscm\_no;

**run**;

**proc** **sort** data=store.t20\_c33\_c34\_0204\_visit\_1;

by indi\_dscm\_no;

**run**;

/\*②INNER JOIN

t20\_c\_0517\_v\_visit\_1 & t20\_c33\_c34\_0204\_visit\_1 => 05~17년도 환자 중에서 02~04년도에 C33-, C34-진단 받은 환자를 알 수 있음\*/

**data** store.join\_0517\_c33\_c34\_0204;

merge store.t20\_c\_0517\_v\_visit\_1(in=a) store.t20\_c33\_c34\_0204\_visit\_1(in=b);

by indi\_dscm\_no;

if a and b;

**run**;

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

/\*=>t20\_c\_0517\_v\_visit\_1=>307,458명 -

t20\_c\_0517\_v\_visit\_1 & t20\_c33\_c34\_0204\_visit\_1 =>16,502명

= 290,956명 \*/

/\* <<<<<<<<<<<<<<<<<<<<<<<<patient\_group1\_all 환자군 정의>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

- 05~17 년도 폐암인 케이스(C33-,C34- + V193, V194) 이면서 02~04년도 폐암 케이스(C33-,C34-)가 없는 대상자 = 290,956명 \*/

**proc** **sql**;

create table store.patient\_group1 as

select indi\_dscm\_no

from store.t20\_c\_0517\_v\_visit\_1

except

select indi\_dscm\_no from store.join\_0517\_c33\_c34\_0204;

**run**;

**quit**;

**proc** **sql**;

create table store.patient\_group1\_all as

(select a.\* ,b.\* from store.patient\_group1 as a left join store.t20\_c\_0517\_v\_visit\_1 as b on a.indi\_dscm\_no=b.indi\_dscm\_no);

**quit**;

**run**;

**proc** **contents** data=store.patient\_group1\_all;

**run**;

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

/\*1. 기준 -> 02~17년도에 폐암케이스(C33 or C34 + V193 or V194)가 아닌 N수 구하기\*/

/\*02~17년도 테이블 합치기\*/

**data** store.t20\_0217;

set a.t20\_2002 - a.t20\_2017;

**run**;

/\*02~17년도 테이블에서 c33-, c34-. v193, v194인 테이블 생성\*/

**data** store.t20\_c\_0217\_except;

set store.t20\_0217;

if substr(sick\_sym1,**1**,**3**) in ('C33','C34') or

substr(sick\_sym2,**1**,**3**) in ('C33','C34') or

substr(sick\_sym3,**1**,**3**) in ('C33','C34') or

substr(sick\_sym4,**1**,**3**) in ('C33','C34') or

substr(sick\_sym5,**1**,**3**) in ('C33','C34') then cancer=**1**;

**run**;

**data** store.t20\_c\_0217\_v\_except;

set store.t20\_c\_0217\_except;

if spcf\_sym\_type in ('V193','V194') then cancer=**2**;

**run**;

/\*대조군 최종\*/

**proc** **sql**; create table store.t20\_0217\_cp\_final as (

select \* from store.t20\_c\_0217\_v\_except

where cancer not in (**2**) /\* c33-, c34-. v193, v194인사람 제거\*/

);**quit**;**run**;

**proc** **freq** data= store.t20\_0217\_cp\_final; table std\_yyyy exmd\_bz\_yyyy\_n term; **run**;

**proc** **sql**; select count(distinct indi\_dscm\_no) from store.t20\_0217\_cp\_final; **quit**; **run**;

**proc** **contents** data=store.t20\_0217\_cp\_final;

**run**;

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

/\* 대조군=>t20\_0217\_cp\_final=>\*/

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

/\* + 위 환자군(patient\_group1, 290,956명) 중에서 이번 연구주제인 모집단 10년~16년 폐암 케이스 (C33-,C34- + V193, V194)인 N수\*/

**proc** **sql**;

create table store.population\_final as select \* from store.patient\_group1\_all where std\_yyyy between '2010' and '2016';

**quit**;

**run**;

/\* <<<<<<<<<<<<<<<<<<<<<<<<모집단 정의>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

- patient\_group1 환자군 290,956명에서 모집단 10~16년 폐암인 케이스 (C33 or C34 + V193 or V194) N수 > population\_final (173.508명)\*/

**proc** **contents** data=store.population\_final;

**run**;

**proc** **freq** data=store.population\_final;

table std\_yyyy;

**run**;

/\* 년도별 n수 파악해보기\*/

**proc** **sql**;

create table conform\_yyyy as select std\_yyyy, count(\*) from store.population\_final

group by std\_yyyy;

**quit**;

**proc** **print**;

**run**;

/\*검진데이터 join (G1EQ\_2009~2017)\*/

**data** store.G1EQ\_0917;

set a.G1EQ\_2009 - a.G1EQ\_2017;

**run**;

**proc** **contents** data=store.G1EQ\_0917;

**run**;

/\*<<<<<<<<<<<<<폐암진단 기준 앞2년 뒤2년간 검진 여부확인>>>>>>>>>>>>\*/

/\*데이터 형태변환컬럼 추가 char -> num\*/

**data** store.population\_final\_1;

set store.population\_final;

std\_yyyy\_n = input(std\_yyyy, **4.**);

mdcare\_strt\_dt\_n = input(mdcare\_strt\_dt, **8.**);

**run**;

**data** store.G1EQ\_0917\_1;

set store.G1EQ\_0917;

exmd\_bz\_yyyy\_n = input(exmd\_bz\_yyyy, **4.**);

hme\_dt\_n = input(hme\_dt, **8.**);

**run**;

/\*

store.population\_final\_1

진단년도 : std\_yyyy\_n

진단일자 : mdcare\_strt\_dt\_n

store.G1EQ\_0917\_1

검진년도 : exmd\_bz\_yyyy\_n

검진일자 : hme\_dt\_n

\*/

/\*폐암 이전 검진 : 검진일이 진단 2년전(2009~) ~ 폐암진단일직전 case = 133,289 id=99,539\*/

/\*폐암 이전 검진 일 경우 : bf\_2 = 1, af\_2 = 0, term = -2,-1,0\*/

**proc** **sql**;

create table store.subject\_bf2 as

(select distinct a.\*, b.\* , '1' as bf\_2, '0' as af\_2, b.exmd\_bz\_yyyy\_n - a.std\_yyyy\_n as term

from store.population\_final\_1 as a

join store.G1EQ\_0917\_1 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no and exmd\_bz\_yyyy\_n>**2008** and exmd\_bz\_yyyy\_n<**2018** and

((a.std\_yyyy\_n - b.exmd\_bz\_yyyy\_n >**0** and a.std\_yyyy\_n - b.exmd\_bz\_yyyy\_n <**3**)

or (a.std\_yyyy\_n = b.exmd\_bz\_yyyy\_n and mdcare\_strt\_dt\_n > hme\_dt\_n)

));

**quit**;

**run**;

/\*폐암 이후 검진 : 검진일이 폐암진단일직후 ~ 진단 2년후(~2017) case = 36,131 id=31,433\*/

/\*폐암 이후 검진 일 경우 : bf\_2 = 0, af\_2 = 1, term = 0,1,2\*/

**proc** **sql**;

create table store.subject\_af2 as

(select distinct a.\*, b.\* , '0' as bf\_2, '1' as af\_2, b.exmd\_bz\_yyyy\_n - a.std\_yyyy\_n as term

from store.population\_final\_1 as a

join store.G1EQ\_0917\_1 as b

on a.indi\_dscm\_no=b.indi\_dscm\_no and exmd\_bz\_yyyy\_n>**2008** and exmd\_bz\_yyyy\_n<**2018** and

((b.exmd\_bz\_yyyy\_n - a.std\_yyyy\_n >**0** and b.exmd\_bz\_yyyy\_n - a.std\_yyyy\_n <**3**)

or (a.std\_yyyy\_n = b.exmd\_bz\_yyyy\_n and mdcare\_strt\_dt\_n < hme\_dt\_n)

));

**quit**;

**run**;

**proc** **freq** data=store.subject\_bf2; table std\_yyyy exmd\_bz\_yyyy\_n term bf\_2 af\_2; **run**;

**proc** **freq** data=store.subject\_af2; table std\_yyyy exmd\_bz\_yyyy\_n term bf\_2 af\_2; **run**;

**proc** **sql**; select count(distinct indi\_dscm\_no) from store.subject\_bf2; **quit**; **run**;

**proc** **sql**; select count(distinct indi\_dscm\_no) from store.subject\_af2; **quit**; **run**;

/\*b2 + af2 둘다 있는 사람만으로 해서 dataset\*/

**data** store.subject\_bfaf2; set store.subject\_bf2 store.subject\_af2; **run**;

**proc** **sql**;

create table store.subject\_bfaf2\_1 as

(select \*, max(bf\_2) as bf\_2\_ck, max(af\_2) as af\_2\_ck from store.subject\_bfaf2

group by indi\_dscm\_no); **quit**; **run**;

**proc** **sql**;

create table store.subject\_bfaf2\_2 as

(select \* from store.subject\_bfaf2\_1

where bf\_2\_ck='1' and af\_2\_ck='1'); **quit**; **run**;

**proc** **contents** data= store.subject\_bfaf2\_2; **run**;

**proc** **freq** data= store.subject\_bfaf2\_2; table std\_yyyy exmd\_bz\_yyyy\_n term bf\_2 af\_2; **run**;

**proc** **sql**; select count(distinct indi\_dscm\_no) from store.subject\_bfaf2\_2; **quit**; **run**;

**proc** **sql**; select count(indi\_dscm\_no) from store.subject\_bfaf2\_2 where term=**0** and bf\_2='1'; **quit**; **run**;

**proc** **sql**; create table work.check\_1 as (select distinct indi\_dscm\_no, std\_yyyy from store.subject\_bfaf2\_2); **quit**; **run**;

**proc** **freq** data= work.check\_1; table std\_yyyy; **run**;

/\*case=66,928, id=26,123\*/

/\*흡연, 음주, 운동 null 제거전 N수 파악\*/

**proc** **freq** data=store.subject\_bfaf2\_2; table q\_smk\_yn q\_smk\_drt q\_smk\_now\_drt q\_smk\_now\_amt\_v09n q\_smk\_now\_amt\_v0108; **run**;

**proc** **freq** data=store.subject\_bfaf2\_2; table q\_smk\_pre\_amt q\_smk\_pre\_drt; **run**;

/\*

q\_smk\_yn (1=34858, 2=21251, 3=10752, null=67)

q\_smk\_drt (null=all, 버전차이때문에 있는 변수)

q\_smk\_now\_drt (0~75=10707, null=56221)

q\_smk\_now\_amt\_v09n (1~80=10702, null=56226)

q\_smk\_now\_amt\_v0108 (null=all, 버전차이때문에 있는 변수

q\_smk\_pre\_amt (null=45790, 1~90=21138)

q\_smk\_pre\_drt (null=45744, 0~90=21184)

\*/

**proc** **freq** data=store.subject\_bfaf2\_2; table q\_drk\_amt\_v09n q\_drk\_frq\_v09n; **run**;

/\*

q\_drk\_amt\_v09n (0~87=28941, 0=8560, null=37987)

q\_drk\_frq\_v09n (0~7=66817, 0=46973, null=111)

\*/

**proc** **freq** data=store.subject\_bfaf2\_2; table q\_pa\_drt q\_pa\_frq q\_pa\_md q\_pa\_vd q\_pa\_walk; **run**;

/\*

q\_pa\_drt (null=all, 버전차이때문에 있는 변수)

q\_pa\_frq (null=all, 버전차이때문에 있는 변수)

q\_pa\_md (0~7=66804, null=124)

q\_pa\_vd (0~7=66834, null=94)

q\_pa\_walk (0~7=66809, null=119)

/\*-------------조건----------------\*/

/\*1. 흡연상태\*/

/\*2. 흡연갑년(하루평균의 갑량 \* 년수)\*/

/\*3. bmi\*/

/\*4. 알콜섭취량 ( 12g \* 주간횟수 \* 잔수)/7 = 하루평균섭취량\*/

/\*5. 운동 met - 시간이 없는데 어케구하지...?\*/

**proc** **freq** data= store.subject\_bfaf2\_2; table g1e\_bmi g1e\_wght g1e\_hght; **run**;

**proc** **sql**; select count(\*) from store.subject\_bfaf2\_2 where g1e\_wght>**0** and g1e\_hght>**0** and g1e\_bmi>**0** ; **quit**;**run**;

/\*g1e\_bmi (null=20, 11.1~48.7=66908)\*/

/\*g1e\_wght (null=18, 27~138=66910)\*/

/\*g1e\_hght (null=18, 111~196=66910)\*/

/\*키 몸무게 다있으면 bmi있음 확인\*/

**proc** **sql**; create table store.subject\_bfaf2\_3 as (

select \*,

case when q\_smk\_yn=**3** then q\_smk\_now\_amt\_v09n / **20** \* q\_smk\_now\_drt end as smk\_packyear\_now,

case when q\_smk\_yn=**2** then q\_smk\_pre\_amt /**20** \* q\_smk\_pre\_drt end as smk\_packyear\_pre,

case when q\_drk\_amt\_v09n>**0** and q\_drk\_frq\_v09n>**0** then q\_drk\_amt\_v09n \* q\_drk\_frq\_v09n \* **12** /**7** else **0** end as drk\_alcoholperday,

case when q\_pa\_vd is null then **0** end as q\_pa\_vd\_adj,

case when q\_pa\_md is null then **0** end as q\_pa\_md\_adj,

case when q\_pa\_walk is null then **0** end as q\_pa\_walk\_adj

from store.subject\_bfaf2\_2

where

(q\_smk\_yn=**1** or

(q\_smk\_yn=**2** and q\_smk\_pre\_amt is not null and q\_smk\_pre\_drt is not null) or

(q\_smk\_yn=**3** and q\_smk\_now\_drt is not null and q\_smk\_now\_amt\_v09n is not null)) and

g1e\_bmi is not null

);

**quit**; **run**;

**proc** **freq** data= store.subject\_bfaf2\_3; table q\_smk\_yn smk\_packyear\_now smk\_packyear\_pre g1e\_bmi drk\_alcoholperday ; **run**;

/\*n=66648\*/

/\*q\_smk\_yn (1=34845, 2=21122, 3=10681)\*/

/\*smk\_packyear\_now (0~240=10681, null=55967)\*/

/\*smk\_packyear\_pre (0~225=21122, null=45526) \*/

/\*g1e\_bmi (11.1~48.7=66648)\*/

/\*drk\_alcoholperday (0=47028, 1~720=66648)\*/

/\*1. 조건 완전히 정리되면 폐암 진단일과 더 가까운 건강검진일 뽑기(인당 2개의 검진자료)\*/

/\*2. 대상자들 기준으로 폐암진단일 기준 3년째인날 생존여부 확인\*/