OPTIONS VALIDVARNAME=ANY;

/\*libname user "D:\SNUlab\SAS\dump"; \* work로 대체; \*/

libname zio "/userdata08/room553/data\_source/user\_data/zio";

libname million "/userdata08/room553/data\_source/user\_data" inencoding="asciiany";

libname target "/userdata08/room553/data\_source/user\_data/0119";

/\* concatenated dataset list \*/

\* zio.nhid\_gj\_0208 : 검진 02-08;

\* zio.nhid\_gj\_0917 : 검진 09-17;

\* zio.nhid\_gj\_1820 : 검진 18-20;

\* zio.nhid\_gy20 : T20 02-20;

\* zio.nhid\_gy30 : T30 02-20;

\* zio.nhid\_gy40 : T40 02-20;

\* zio.nhid\_gy60 : T60 02-20;

\* zio.nhid\_jk : 자격 02-13 - 일단 생략;

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

/\* 1. T30 + T60 에서 antihistamine 복용 case만 뽑기

\* join key, 약제코드, 복용시작일자, 복용기간;

\*/

\* T30 - 완료;

\* 23.02.22 antihistamine 처방일 경우 anti = 1 만 표시하고 전체 인원 살려놓기;

**data** zio.nhid\_gy30\_AntiYesNoAll;

set zio.nhid\_gy30;

MEDI = substr(mcare\_div\_cd\_adj, **1**, **4**);

if MEDI in ("1305", "2599", "4636", "5248", "6706", "1318", "5308", "5309", "2674", "5327", "5328", "2680", "5326", "6772", "5329", "5330", "6840",

"6903", "1318", "1854", "5515", "3479", "3954", "1801") then anti = **1**;

else anti = **0**;

**run**;

**data** zio.nhid\_gy30\_AntiYesNoAll;

set zio.nhid\_gy30\_AntiYesNoAll;

keep cmn\_key tot\_mcnt mcare\_div\_cd\_adj MEDI anti;

**run**;

\* T60 - 완료;

\* 23.02.22 antihistamine 처방일 경우 anti = 1 만 표시하고 전체 인원 살려놓기;

**data** zio.nhid\_gy60\_AntiYesNoAll;

set zio.nhid\_gy60;

MEDI = substr(mcare\_div\_cd\_adj, **1**, **4**);

if MEDI in ("1305", "2599", "4636", "5248", "6706", "1318", "5308", "5309", "2674", "5327", "5328", "2680", "5326", "6772", "5329", "5330", "6840",

"6903", "1318", "1854", "5515", "3479", "3954", "1801") then anti = **1**;

else anti = **0**;

**run**;

**data** zio.nhid\_gy60\_AntiYesNoAll;

set zio.nhid\_gy60\_AntiYesNoAll;

keep cmn\_key tot\_mcnt mcare\_div\_cd\_adj MEDI anti;

**run**;

\* T30+T60;

**data** zio.nhid\_gy3060\_AntiYesNoAll;

set zio.nhid\_gy30\_AntiYesNoAll zio.nhid\_gy60\_AntiYesNoAll;

**run**;

\* T30+T60 & T20 에서 id, mdcare\_strt\_dt 가져와서 붙이기 - 1시간 반 이상 걸림;

**proc** **sql**;

create table zio.nhid\_gy3060\_AntiYesNoAll as

select p.\*, q.indi\_dscm\_no, q.cmn\_key, q.mdcare\_strt\_dt

from zio.nhid\_gy3060\_AntiYesNoAll as p left join zio.nhid\_gy20 as q

on p.cmn\_key = q.cmn\_key;

**quit**;

\* check N;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy3060\_AntiYesNoAll;

**quit**; \* 3,978,548;

\* 같은 날 처방받은 항히스타민제라면, 가장 많은 처방일수를 받은 처방이 맨 위로 오도록 하기 (중복으로 더해지는 것 방지);

**proc** **sort** data=zio.nhid\_gy3060\_AntiYesNoAll;

by indi\_dscm\_no mdcare\_strt\_dt descending tot\_mcnt;

**run**;

\* 같은 날 처방받은 항히스타민제 중, 가장 많은 처방일수만 남기고 drop duplicates;

**proc** **sort** data=zio.nhid\_gy3060\_AntiYesNoAll nodupkey;

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

**run**; \* 2/22 여기까지 돌려놓고 감;

\* (2/23) washout 빼자, 이하 내용 다 생략;

\* 약제 washout 기간 정의;

\* 사람별 2002~2004 사이의 복용량 합 계산 -> mdcn\_exec\_freq\_0204;

\* count distinct person id of one whose mdcn\_exec\_freq\_0204 < 30; \* N명 - 기재 필요;

\* 2002~2004 사이에 30일 미만으로 먹은 사람들을 대상으로, 2005~2020 의 복용량 합을 다시 계산해서 R06\_Nyear 변수 정의에 사용하면 방향성이 좀 더 정확해질 수 있을 것 같다;

\* 그러면 2002~2004 사이에 30일 이상으로 먹은 사람들은 삭제됨 (washout) ;

\* change to datetime;

**data** data\_mdcn\_exec\_freq\_0204;

set zio.nhid\_gy3060\_AntiYesNoAll;

mdcare\_strt\_dt\_new = INPUT(mdcare\_strt\_dt, yymmdd10.);

FORMAT mdcare\_strt\_dt\_new yymmdd10.;

**RUN**;

\* data\_mdcn\_exec\_freq\_0204 에 tot\_mcnt\_0204 생성;

\* 2002~2004년 복용시작인 처방만 뽑아서, 사람별로 sum 한 것;

\* 이 df에는 2002~2004에 대한 처방만 들어있음;

**proc** **sql**;

create table data\_mdcn\_exec\_freq\_0204 as

select \*, sum(tot\_mcnt) as tot\_mcnt\_0204

from data\_mdcn\_exec\_freq\_0204

where mdcare\_strt\_dt\_new >= input('20020101', yymmdd10.) AND mdcare\_strt\_dt\_new <= input('20021231', yymmdd10.)

group by indi\_dscm\_no

order by indi\_dscm\_no;

**quit**; \* washout 2002년 1년만으로 변경해보자;

\* 30일 미만 복용자만 남기기;

**data** data\_mdcn\_exec\_freq\_0204;

set data\_mdcn\_exec\_freq\_0204;

if tot\_mcnt\_0204 < **30**;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from data\_mdcn\_exec\_freq\_0204

where tot\_mcnt\_0204 < **30**;

**quit**; \* 2,307,304명; \* 기존 179만명이랑 별 차이가 없음 .. washout 기간을 줄여보자;

\* washout 기간 2002년만으로 잡았을 때 2,624,340;

\* washout 그냥 빼버리자;

\* 2002-2004 30일 미만 복용자 list 생성 (nodupkey로) ;

**proc** **sort** data=data\_mdcn\_exec\_freq\_0204 nodupkey;

by indi\_dscm\_no;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy3060\_antihistamine;

**quit**; \* 기존 xx명;

\* washout 넣어주는 작업 생략, 그냥 전기간으로 보자;

\* 2002~2004 사이에 30일 미만으로 먹은 179만명만 남긴 data\_mdcn\_exec\_freq\_0204 에서 작업 이어가기;

\* 2002~2004 사이의 처방 뿐만 아니라, 나머지 년도의 처방도 가져오기 위해 data\_mdcn\_exec\_freq\_0204 기준으로 zio.nhid\_gy3060\_antihistamine 랑 left join;

/\*proc sql; \*/

/\*create table zio.nhid\_gy3060\_antihis\_groupby as\*/

/\*select p.indi\_dscm\_no, q.\*\*/

/\*from data\_mdcn\_exec\_freq\_0204 as p left join zio.nhid\_gy3060\_antihistamine as q\*/

/\*on p.indi\_dscm\_no = q.indi\_dscm\_no; \*/

/\*quit; \*/

/\*\* 이제 zio.nhid\_gy3060\_antihis\_groupby 에는 2002~2004에 30일 미만으로 먹은 179만명에 대한 전체 처방이 다 들어있다; \*/

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy3060\_AntiYesNoAll;

**quit**; \* 3,978,548명 확인;

/\*\*/

/\*\* 마지막 처방일수 열 생성 (NPD\_final3 생성시 필요함); \*/

/\*proc sql; \*/

/\*create table zio.nhid\_gy3060\_antihis\_groupby as \*/

/\*select \*, case when mdcare\_strt\_dt = last\_MEDI\_date then tot\_mcnt end as last\_tot\_mcnt\*/

/\*from zio.nhid\_gy3060\_antihis\_groupby \*/

/\*group by indi\_dscm\_no\*/

/\*order by indi\_dscm\_no; \*/

/\*run; \*/

/\*\*/

/\*\*/

/\*\* 사람별 마지막 처방일자 - 첫 처방일자 열 생성; \*/

/\*\* last\_first\_gap + last\_MDCN\_EXEC\_FREQ (이렇게 해야 intensity 변수 생성할 때 한번 8일치 처방 받아간 사람의 intensity 가 8이 되는 걸 방지할 수 있음, 이런 경우는 1이 되어야 함); \*/

/\*\* last\_first\_gap; \*/

/\*proc sql; \*/

/\*create table zio.nhid\_gy3060\_antihis\_groupby as\*/

/\*select \*, last\_MEDI\_date\_new - first\_MEDI\_date\_new + last\_tot\_mcnt as last\_first\_gap\*/

/\*from zio.nhid\_gy3060\_antihis\_groupby\*/

/\*group by indi\_dscm\_no\*/

/\*order by indi\_dscm\_no; \*/

/\*quit; \*/

/\*\* 사람별 총 복용기간 열 생성; \*/

/\*\* (기존) 사람별, 약제별 총 복용기간 열 생성; \*/

/\*proc sql; \*/

/\*CREATE TABLE zio.nhid\_gy3060\_AntiYesNoAll\_groupby as\*/

/\*SELECT \*, SUM(tot\_mcnt) as tot\_mcnt\_final\*/

/\*FROM zio.nhid\_gy3060\_AntiYesNoAll\*/

/\*GROUP BY indi\_dscm\_no\*/

/\*ORDER BY indi\_dscm\_no; \*/

/\*quit; \*/

/\*\*/

/\*\* 사람별 평균 처방일수 변수 생성; \*/

/\*\* intensity 대체용 변수; \*/

/\*proc sql; \*/

/\*create table zio.nhid\_gy3060\_antihis\_groupby as\*/

/\*select \*, mean(tot\_mcnt) as tot\_mcnt\_mean\*/

/\*from zio.nhid\_gy3060\_antihis\_groupby\*/

/\*group by indi\_dscm\_no\*/

/\*order by indi\_dscm\_no; \*/

/\*quit; \*/

/\*proc sort data=zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*by indi\_dscm\_no descending last\_tot\_mcnt; \*/

/\*run; \*/

\* 현재 nhid\_gy3060\_AntiYesNoAll 에는 anti 처방, anti 아닌 처방이 여러개 섞여 있음;

\* drop dup 하기 전에 복사;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_AntiYesNoAll;

**run**;

\* 사람별 첫 처방일자, 마지막 처방일자 열 생성;

\* 여기는 anti = 0, 1이 모두 섞여있으므로, anti = 1인 건수만을 대상으로 첫 처방일자를 찾아야 함;

**data** tmp;

set zio.nhid\_gy3060\_antihis\_allprescrip;

if anti = **1**;

**run**;

**proc** **sql**;

create table tmp as

select indi\_dscm\_no, min(mdcare\_strt\_dt) as first\_MEDI\_date

from tmp

group by indi\_dscm\_no

order by indi\_dscm\_no;

**quit**;

**proc** **sql**;

create table zio.nhid\_gy3060\_antihis\_allprescrip as

select p.\*, q.\*

from zio.nhid\_gy3060\_antihis\_allprescrip as p left join tmp as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

\* change to datetime;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_antihis\_allprescrip;

first\_MEDI\_date\_new = input(first\_MEDI\_date, yymmdd10.);

format first\_MEDI\_date\_new yymmdd10.;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy3060\_antihis\_allprescrip

where first\_medi\_date\_new ^= **.** ;

**quit**; \* 3,634,084 명 확인;

\* 여기까지 backup 하나 만들어놓자;

**data** zio.nhid\_gy3060\_antihis\_allpres\_bu;

set zio.nhid\_gy3060\_antihis\_allprescrip;

**run**;

/\*\* 사람별 drop duplicates; \*/

/\*\* (기존) 사람별, 약제별 drop duplicates; \*/

/\*proc sort data=zio.nhid\_gy3060\_antihis\_groupby nodupkey;\*/

/\*by indi\_dscm\_no;\*/

/\*run;\*/

/\*\*/

/\*\* 복용기간에 따른 15, 30, 60, 90, 180days, 1year, 3year, 5year, 7year 변수 생성; \*/

/\*data zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*set zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*if tot\_mcnt\_final >= 15 then R06\_15days = 1; \*/

/\*if tot\_mcnt\_final >= 30 then R06\_30days = 1; \*/

/\*if tot\_mcnt\_final >= 60 then R06\_60days = 1; \*/

/\*if tot\_mcnt\_final >= 90 then R06\_90days = 1; \*/

/\*if tot\_mcnt\_final >= 180 then R06\_180days = 1; \*/

/\*if tot\_mcnt\_final >= 365 then R06\_1year = 1; \*/

/\*if tot\_mcnt\_final >= 1095 then R06\_3year = 1; \*/

/\*if tot\_mcnt\_final >= 1825 then R06\_5year = 1; \*/

/\*if tot\_mcnt\_final >= 2555 then R06\_7year = 1; \*/

/\*run; \*/

/\*\*/

/\*\* null -> 0으로 채우기; \*/

/\*data zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*set zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*if R06\_15days = . then R06\_15days = 0; \*/

/\*if R06\_30days = . then R06\_30days = 0; \*/

/\*if R06\_60days = . then R06\_60days = 0; \*/

/\*if R06\_90days = . then R06\_90days = 0; \*/

/\*if R06\_180days = . then R06\_180days = 0; \*/

/\*if R06\_1year = . then R06\_1year = 0; \*/

/\*if R06\_3year = . then R06\_3year = 0; \*/

/\*if R06\_5year = . then R06\_5year = 0; \*/

/\*if R06\_7year = . then R06\_7year = 0; \*/

/\*run; \*/

/\*\*/

/\*\* intensity 변수 생성; \*/

/\*\* 이 때 intensity가 1인 사람은 1회 처방에 N일 받아갔고, 그 이후는 처방받지 않은 사람들임; \*/

/\*data zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*set zio.nhid\_gy3060\_antihis\_groupby; \*/

/\*intensity = tot\_mcnt\_final / last\_first\_gap; \*/

/\*run; \*/

\* 1년 연속으로 사용 vs 3일에 하루 3년 복용 변수 생성해야 함 -> 이게 되나? ;

\* (23.02.23) (최종) 0일 복용자 포함하여 397만명으로 작업;

\* (최종) 2002~2004에 30일 이상 먹은 사람들 잘라서 총원 179만명 가량으로 줄었음;

\* (1차 수정) 여기까지 zio.nhid\_gy3060\_antihis\_groupby 는 사람별 총복용일수와 첫 복용일자가 들어있는 table, 총원 179만명 가량이므로 행도 사람수만큼 존재함;

\* (기존) 여기까지 zio.nhid\_gy3060\_antihis\_groupby 는 사람별, MEDI 별 총복용일수와 첫 복용일자가 들어있는 table이므로, 한 사람당 여러 개의 MEDI가 존재할 수 있음;

/\* ANTIHISTAMINE end. ---------------------------------------------------------------------------------------------------------------------------------------------------------- \*/

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

/\* 1. T20, T40 에서 NPD 진단자 및 진단일자 뽑기 \*/

\* join key, 상병코드, 진단일자;

\* 주상병 + 부상병1로 좁혀서 확인;

\* T20 - NPD 진단 case만 뽑기 (공단 == 김진엽 선생님이 주신 코드 일치 (한국표준질병사인분류코드(KCD) 기준) 하는 거 확인, 일단 이 코드 돌려서 n수 체크하기);

\* 주상병이나 부상병 중 한번이라도 NPD 진단 받은 적 있으면; \* 완료;

\* (23.01.31) 정신과 진단만 포함 (form\_cd in 09, 10, 11);

**data** zio.nhid\_gy20\_npd\_2MajorSymp;

set zio.nhid\_gy20;

if (sick\_sym1 in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",

"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448") OR

sick\_sym2 in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",

"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448"))

and form\_cd in ("09", "10", "11")

;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy20\_npd\_2MajorSymp;

**quit**; \* 정신과 특정하니 32,468 명;

\* (23.01.31) T40은 정신과를 특정할 수 없어서 사용 x;

/\*\* T40 - NPD 진단 case만 뽑기 (공단 == 김진엽 선생님이 주신 코드 일치 (한국표준질병사인분류코드(KCD) 기준) 하는 거 확인, 일단 이 코드 돌려서 n수 체크하기);\*/

/\*\* 완료; \*/

/\*data zio.nhid\_gy40\_npd; \*/

/\*set zio.nhid\_gy40; \*/

/\*if mcex\_sick\_sym in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",\*/

/\*"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448"); \*/

/\*run; \*/

/\*\*/

/\*\* T20 에서 id, mdcare\_strt\_dt 가져와서 T40에 붙이기 - 완료 ; \*/

/\*proc sql; \*/

/\*CREATE table zio.nhid\_gy40\_npd as \*/

/\*SELECT p.\*, q.indi\_dscm\_no, q.cmn\_key, q.mdcare\_strt\_dt\*/

/\*FROM zio.nhid\_gy40\_npd as p LEFT JOIN zio.nhid\_gy20 as q\*/

/\*ON p.cmn\_key = q.cmn\_key; \*/

/\*quit; \*/

/\*\*/

/\*PROC FREQ DATA=ZIO.nhid\_gy40\_npd;\*/

/\*TABLE mcex\_sick\_sym;\*/

/\*RUN;\*/

/\*\*/

/\*proc sql; \*/

/\*title 'count'; \*/

/\*select count(distinct indi\_dscm\_no)\*/

/\*from zio.nhid\_gy40\_npd; \*/

/\*quit; \* 1,865,958명; \*/

\* T20 - MAIN\_SICK, SUB\_SICK 한 열로 합쳐서 질환명 살려놓기 & NPD 진단여부 열 (진단자만 뽑았으니 전부 1) 생성; \* 완료;

**data** zio.nhid\_gy20\_npd\_2MajorSymp;

set zio.nhid\_gy20\_npd\_2MajorSymp;

if sick\_sym1 in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",

"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448")

THEN NPD\_name = sick\_sym1; \* 주상병으로 진단받았으면 NPD\_name = 주상병;

if sick\_sym1 not in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",

"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448")

AND sick\_sym2 in ("F05", "F03", "F067", "R418", "G308", "F320", "F321", "F330", "F331", "F338", "F339", "F340", "F348", "F349", "F380", "F381", "F388", "F39", "F400","F410",

"F411","F412","F413","F418","F419", "F511", "G478", "F639", "F819", "G25", "R451", "F900", "R454", "R458", "R440", "R441", "R442", "R443", "R448")

THEN NPD\_name = sick\_sym2; \* 부상병2로 진단받았으면 NPD\_name = 부상병2;

NPD = **1**;

**run**;

**proc** **freq** data=zio.nhid\_gy20\_npd\_2MajorSymp;

table NPD\_name;

**run**;

/\*\* T40 - SICK\_SYM 열 이름 NPD\_name으로 바꾸기 (T20과 통일) & NPD 진단여부 열 (진단자만 뽑았으니 전부 1) 생성; \*/

/\*data zio.nhid\_gy40\_npd; \*/

/\*set zio.nhid\_gy40\_npd; \*/

/\*rename mcex\_sick\_sym = NPD\_name; \*/

/\*NPD = 1; \*/

/\*run; \*/

\* T20, T40 에서 필요한 변수만 뽑기;

**data** nhid\_gy20\_npd\_new; \* dump folder;

set zio.nhid\_gy20\_npd\_2MajorSymp;

keep indi\_dscm\_no cmn\_key mdcare\_strt\_dt NPD\_name NPD;

**run**;

/\*data nhid\_gy40\_npd\_new; \* dump folder; \*/

/\*set zio.nhid\_gy40\_npd; \*/

/\*keep indi\_dscm\_no cmn\_key mdcare\_strt\_dt NPD\_name NPD; \*/

/\*run; \*/

/\*\* 뽑아서 합치기; \*/

/\*data zio.nhid\_gy2040\_npd\_2MajorSymp; \*/

/\*set nhid\_gy20\_npd\_new nhid\_gy40\_npd\_new; \*/

/\*run; \*/

**data** zio.nhid\_gy20\_npd\_2MajorSymp;

set nhid\_gy20\_npd\_new;

**run**;

\* 최종 인원수 확인 - 32,468 명;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy20\_npd\_2MajorSymp;

**quit**;

\* (01.31) 정신과 진단만으로 특정; \* nhid\_gy20\_npd\_2MajorSymp;

\* (01.20) zio.nhid\_gy2040\_npd\_2MajorSymp 로 파일 수정;

\* 여기까지 zio.nhid\_gy2040\_npd 는 사람별, 진단일자별 NPD 진단여부이므로 한 사람당 여러 개의 처방이 존재할 수 있음;

\* 만약 A가 2002년 2월에 진단받고, 또 2002년 5월에 진단받았다면 2개의 행이 생길 것;

/\* NPD end. ---------------------------------------------------------------------------------------------------------------------------------------------------------- \*/

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

/\* 1. 02-08, 09-17 변수 통일

2. 05-10만 다시 뽑기 \*/

\* 02-17 합치기;

**data** zio.nhid\_gj\_0217;

set zio.nhid\_gj\_0208 zio.nhid\_gj\_0917;

**run**;

\* count;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0217;

**run**;

\* descriptive for numerical variables;

**proc** **means** data=zio.nhid\_gj\_0217\_new n nmiss mean median min max std;

**run**;

\* 필요한 변수만 뽑기;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217;

keep exmd\_bz\_yyyy indi\_dscm\_no hme\_dt

g1e\_hght g1e\_wght g1e\_bmi

g1e\_bp\_sys g1e\_bp\_dia g1e\_urn\_prot g1e\_hgb g1e\_fbs g1e\_tot\_chol g1e\_sgot g1e\_sgpt g1e\_ggt g1e\_gfr

g1e\_phx\_yn g1e\_phx\_lvdz g1e\_phx\_htn g1e\_phx\_stk g1e\_phx\_htdz g1e\_phx\_dm g1e\_phx\_cc g1e\_phx\_etc /\* 과거병력 \*/

/\* q\_phx1\_dz\_v0001 q\_phx2\_dz\_v0001 q\_phx3\_dz\_v0001 /\* 과거병력, 질환 1 2 3, 질환종류 - 모든 값이 null \*/

q\_phx1\_dz\_v0208 q\_phx2\_dz\_v0208 q\_phx3\_dz\_v0208 /\* 과거병력, 질환 1 2 3, 질환종류 \*/

q\_fhx\_lvdz q\_fhx\_htn q\_fhx\_stk q\_fhx\_htdz q\_fhx\_dm q\_fhx\_cc q\_fhx\_etc /\* 가족력 \*/

q\_drk\_frq\_v0108 q\_drk\_amt\_v0108 /\* 음주 \*/

q\_smk\_yn q\_smk\_now\_amt\_v0108 q\_smk\_drt /\* 흡연 \*/

q\_pa\_frq /\* 운동 \*/

/\*q\_pa\_drt 운동 - 모든 값이 null\*/

q\_phx\_dx\_stk q\_phx\_dx\_htdz q\_phx\_dx\_htn q\_phx\_dx\_dm q\_phx\_dx\_dld q\_phx\_dx\_ptb q\_phx\_dx\_etc /\* 과거력, 진단여부 \*/

q\_smk\_pre\_drt q\_smk\_pre\_drt q\_smk\_now\_drt q\_smk\_now\_amt\_v09n /\* 흡연 \*/

q\_drk\_frq\_v09n q\_drk\_amt\_v09n /\* 음주 \*/

q\_pa\_vd q\_pa\_md q\_pa\_walk /\* 신체활동 \*/ ;

**run**;

/\* 1. 변수 통일 \*/

\* 1-1. 과거병력;

\* 질환별 변수 새로 생성;

\* past\_lvdz: 과거력-간장질환, 0:없음, 1:있음;

\* past\_htn: 과거력-고혈압, 0:없음, 1:있음;

\* past\_stk: 과거력-뇌졸중, 0:없음, 1:있음;

\* past\_htdz: 과거력-심장병, 0:없음, 1:있음;

\* past\_dm: 과거력-당뇨병, 0:없음, 1:있음;

\* past\_cc: 과거력-암, 0:없음, 1:있음;

\* past\_etc: 과거력-기타, 0:없음, 1:있음;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

if g1e\_phx\_lvdz = **1** or q\_phx1\_dz\_v0208 = **3** or q\_phx2\_dz\_v0208 = **3** or q\_phx3\_dz\_v0208 = **3** then past\_lvdz = **1**; else past\_lvdz = **0**;

if g1e\_phx\_htn = **1** or q\_phx1\_dz\_v0208 = **4** or q\_phx2\_dz\_v0208 = **4** or q\_phx3\_dz\_v0208 = **4** or q\_phx\_dx\_htn = **1** then past\_htn = **1**; else past\_htn = **0**;

if g1e\_phx\_stk = **1** or q\_phx1\_dz\_v0208 = **6** or q\_phx2\_dz\_v0208 = **6** or q\_phx3\_dz\_v0208 = **6** or q\_phx\_dx\_stk = **1** then past\_stk = **1**; else past\_stk = **0**;

if g1e\_phx\_htdz = **1** or q\_phx1\_dz\_v0208 = **5** or q\_phx2\_dz\_v0208 = **5** or q\_phx3\_dz\_v0208 = **5** or q\_phx\_dx\_htdz = **1** then past\_htdz = **1**; else past\_htdz = **0**;

if g1e\_phx\_dm = **1** or q\_phx1\_dz\_v0208 = **7** or q\_phx2\_dz\_v0208 = **7** or q\_phx3\_dz\_v0208 = **7** or q\_phx\_dx\_dm = **1** then past\_dm = **1**; else past\_dm = **0**;

if g1e\_phx\_cc = **1** or q\_phx1\_dz\_v0208 = **8** or q\_phx2\_dz\_v0208 = **8** or q\_phx3\_dz\_v0208 = **8** then past\_cc = **1**; else past\_cc = **0**;

if g1e\_phx\_etc = **1** or q\_phx1\_dz\_v0208 = **1** or q\_phx2\_dz\_v0208 = **1** or q\_phx3\_dz\_v0208 = **1**

or q\_phx1\_dz\_v0208 = **2** or q\_phx2\_dz\_v0208 = **2** or q\_phx3\_dz\_v0208 = **2**

or q\_phx1\_dz\_v0208 = **9** or q\_phx2\_dz\_v0208 = **9** or q\_phx3\_dz\_v0208 = **9**

or q\_phx\_dx\_dld = **1** or q\_phx\_dx\_ptb = **1**

then past\_etc = **1**; else past\_etc = **0**;

**run**;

\* 1-2. 가족력;

\* 질환별 변수 새로 생성;

\* family\_lvdz: 가족력-간장질환, 0:없음, 1:있음;

\* family\_htn: 가족력-고혈압, 0:없음, 1:있음;

\* family\_stk: 가족력-뇌졸중, 0:없음, 1:있음;

\* family\_htdz: 가족력-심장병, 0:없음, 1:있음;

\* family\_dm: 가족력-당뇨병, 0:없음, 1:있음;

\* family\_cc: 가족력-암, 0:없음, 1:있음;

\* family\_etc: 가족력-기타, 0:없음, 1:있음;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

if q\_fhx\_lvdz = **2** then family\_lvdz = **1**; else family\_lvdz = **0**;

if q\_fhx\_htn = **2** then family\_htn = **1**; else family\_htn = **0**;

if q\_fhx\_stk = **2** then family\_stk = **1**; else family\_stk = **0**;

if q\_fhx\_htdz = **2** then family\_htdz = **1**; else family\_htdz = **0**;

if q\_fhx\_dm = **2** then family\_dm = **1**; else family\_dm = **0**;

if q\_fhx\_cc = **2** then family\_cc = **1**; else family\_cc = **0**;

if q\_fhx\_etc = **2** then family\_etc = **1**; else family\_etc = **0**;

**run**;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table family\_dm;

**run**;

\* 1-3. 음주;

\* q\_drk\_frq\_v0108 음주 빈도;

\* 1: 거의 마시지 않는다;

\* 2: 월 2~3회;

\* 3: 일주일 1~2회;

\* 4: 일주일 3~4회;

\* 5: 거의 매일;

\* q\_drk\_amt\_v0108 주간 섭취량;

\* 1: 소주 반병 이하;

\* 2: 소주 한병;

\* 3: 소주 1병반;

\* 4: 소주 2병 이상;

\* q\_drk\_frq\_v09n 주간음주일수;

\* 0~7일 / 주;

\* q\_drk\_amt\_v09n 일일음주량;

\* 잔;

\* 음주 빈도 / 일일음주량 두가지 변수 생성;

\* drk\_freq: 주간 음주 빈도, 0: 거의 마시지 않는다, 1: 일주일에 1~2회 마신다, 3: 일주일에 3회 이상 마신다;

\* drk\_amount: 일일음주량, 0: 소주 반병 미만, 1: 소주 반병 이상;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

if q\_drk\_frq\_v0108 = **.** or q\_drk\_frq\_v0108 = **1** or q\_drk\_frq\_v0108 = **2** or q\_drk\_frq\_v09n = **.** or q\_drk\_frq\_v09n = **0** then drk\_freq = **0**;

if q\_drk\_frq\_v0108 = **3** or q\_drk\_frq\_v09n = **1** or q\_drk\_frq\_v09n = **2** then drk\_freq = **1**;

if q\_drk\_frq\_v0108 >= **4** or q\_drk\_frq\_v09n >= **3** then drk\_freq = **2**;

if q\_drk\_amt\_v0108 = **1** or q\_drk\_amt\_v09n < **4** then drk\_amount = **0**;

if q\_drk\_amt\_v0108 > **1** or q\_drk\_amt\_v09n >= **4** then drk\_amount = **1**;

**run**;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table drk\_freq;

**run**;

\* 1-4. 흡연;

\* q\_smk\_pre\_drt;

\* 금연전흡연기간 (년);

\* q\_smk\_now\_drt;

\* 흡연기간;

\* q\_smk\_now\_amt\_v09n;

\* 흡연량;

\* q\_smk\_now\_amt\_v0108: 현재흡연량;

\* 1: 반갑 미만;

\* 2: 반갑 이상 ~ 한갑 미만;

\* 3: 한갑 이상 ~ 두갑 미만;

\* 4: 두갑 이상;

\* q\_smk\_drt: 현재흡연기간;

\* 1: 5년미만;

\* 2: 5~9년;

\* 3: 10~19년;

\* 4: 20~29년;

\* 5: 30년 이상;

\* 흡연기간, 흡연량의 두가지 변수로 합치기;

\* smk\_term: 과거 혹은 현재 흡연기간, 0: 아예 피지 않음, 1: 5년 미만, 2: 5년~9년, 3: 10년~19년, 4: 20년~29년, 5: 30년 이상;

\* smk\_amount: 과거 혹은 현재 흡연량, 0: 반갑 미만, 1: 반갑 이상 ~ 한갑 미만, 2: 한갑 이상 ~ 두갑 미만, 3: 두갑 이상;

\* 이때, 한갑의 기준은 20개피임;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

if (q\_smk\_pre\_drt = **.** or q\_smk\_pre\_drt = **0**) and (q\_smk\_now\_drt = **.** or q\_smk\_now\_drt = **0**) then smk\_term = **0**;

if (q\_smk\_pre\_drt > **0** and q\_smk\_pre\_drt < **5**) or (q\_smk\_now\_drt > **0** and q\_smk\_now\_drt < **5**) or q\_smk\_drt = **1** then smk\_term = **1**;

if (q\_smk\_pre\_drt >= **5** and q\_smk\_pre\_drt < **10**) or (q\_smk\_now\_drt >= **5** and q\_smk\_now\_drt < **10**) or q\_smk\_drt = **2** then smk\_term = **2**;

if (q\_smk\_pre\_drt >= **10** and q\_smk\_pre\_drt < **20**) or (q\_smk\_now\_drt >= **10** and q\_smk\_now\_drt < **20**) or q\_smk\_drt = **3** then smk\_term = **3**;

if (q\_smk\_pre\_drt >= **20** and q\_smk\_pre\_drt < **30**) or (q\_smk\_now\_drt >= **20** and q\_smk\_now\_drt < **30**) or q\_smk\_drt = **4** then smk\_term = **4**;

if (q\_smk\_pre\_drt >= **30**) or (q\_smk\_now\_drt >= **30**) or q\_smk\_drt = **5** then smk\_term = **5**;

if q\_smk\_now\_amt\_v09n < **10** or q\_smk\_now\_amt\_v0108 = **1** then smk\_amount = **0**;

if (q\_smk\_now\_amt\_v09n >= **10** and q\_smk\_now\_amt\_v09n < **20**) or q\_smk\_now\_amt\_v0108 = **2** then smk\_amount = **1**;

if (q\_smk\_now\_amt\_v09n >= **20** and q\_smk\_now\_amt\_v09n < **40**) or q\_smk\_now\_amt\_v0108 = **3** then smk\_amount = **2**;

if (q\_smk\_now\_amt\_v09n >= **40**) or q\_smk\_now\_amt\_v0108 = **4** then smk\_amount = **3**;

**run**;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table q\_smk\_yn;

**run**;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table smk\_term smk\_amount;

**run**;

\* 1-5. 운동;

\* q\_pa\_vd: 신체활동/고강도 - 0~7일/주;

\* q\_pa\_md: 신체활동/중등도 - 0~7일/주;

\* q\_pa\_walk: 신체활동/걷기 - 0~7일/주;

\* 1: 안한다;

\* 2: 1~2회;

\* 3: 3~4회;

\* 4: 5~6회;

\* 5: 거의 매일;

\* q\_pa\_frq: 신체활동-빈도;

\* 한가지 변수로 통합;

\* exec\_freq: 주간운동횟수, 0: 안한다, 1: 주 1~2회, 2: 주 3회 이상;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table q\_pa\_frq;

**run**;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

if (q\_pa\_vd = **0** and q\_pa\_md = **0** and q\_pa\_walk = **0**) or q\_pa\_frq = **1** or q\_pa\_frq = **.** then exec\_freq = **0**;

if (q\_pa\_vd >= **1** and q\_pa\_vd <= **2**) or (q\_pa\_md >= **1** and q\_pa\_md <= **2**) or (q\_pa\_walk >= **1** and q\_pa\_walk <= **2**) or q\_pa\_frq = **2** then exec\_freq = **1**;

if (q\_pa\_vd >= **3**) or (q\_pa\_md >= **3**) or (q\_pa\_walk >= **3**) or q\_pa\_frq >= **3** then exec\_freq = **2**;

**run**;

**proc** **freq** data=zio.nhid\_gj\_0217\_new;

table exec\_freq;

**run**;

\* 2005-2010, 필요한 변수만 뽑기;

\* change to datetime;

**data** zio.nhid\_gj\_0217\_new;

set zio.nhid\_gj\_0217\_new;

hme\_dt\_new = INPUT(hme\_dt, yymmdd10.);

FORMAT hme\_dt\_new yymmdd10.;

**RUN**;

\* 2005-2010 뽑기;

\* (23.02.23) 2002-2010 으로 바꿔보자;

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select \*

from zio.nhid\_gj\_0217\_new

where exmd\_bz\_yyyy >= '2002' and exmd\_bz\_yyyy <= '2010';

**quit**;

\* 쓸 변수만 뽑기;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

keep exmd\_bz\_yyyy indi\_dscm\_no hme\_dt\_new

g1e\_hght g1e\_wght g1e\_bmi

g1e\_bp\_sys g1e\_bp\_dia g1e\_urn\_prot g1e\_hgb g1e\_fbs g1e\_tot\_chol g1e\_sgot g1e\_sgpt g1e\_ggt g1e\_gfr

past\_lvdz past\_htn past\_stk past\_htdz past\_dm past\_cc past\_etc

family\_lvdz family\_htn family\_stk family\_htdz family\_dm family\_cc family\_etc

drk\_freq drk\_amount

smk\_term smk\_amount

exec\_freq;

**run**;

\* g1e\_gfr 삭제, 결측이 너무 많음;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

drop g1e\_gfr;

**run**;

\* hght, wght 삭제 - BMI 사용 예정;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

drop g1e\_hght g1e\_wght;

**run**;

\* descriptive for numerical variables;

**proc** **means** data=zio.nhid\_gj\_0210 n nmiss mean median min max std;

**run**;

\* 자격 테이블에서 성별, AGE GROUP 가져와서 (ID, YEAR) 조합으로 검진에 붙이기;

\* 자격 테이블에서 필요한 변수만 뽑기 - sex;

**proc** **sql**;

create table nhid\_jk\_sex as

select indi\_dscm\_no, sex\_type

from zio.nhid\_jk

order by indi\_dscm\_no;

**quit**;

\* sex table은 drop duplicate 해야 함; \* 400만명;

**proc** **sort** data=nhid\_jk\_sex nodupkey;

by indi\_dscm\_no;

\* 자격 테이블에서 필요한 변수만 뽑기 - 출생연도;

**proc** **sql**;

create table nhid\_jk\_age\_group as

select indi\_dscm\_no, std\_yyyy, byear

from zio.nhid\_jk

order by indi\_dscm\_no;

**quit**;

\* 검진 테이블에 붙이기 - SEX (join key ID만 사용);

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select p.\*, q.indi\_dscm\_no, q.sex\_type

from zio.nhid\_gj\_0210 as p left join nhid\_jk\_sex as q

on p.indi\_dscm\_no = q.indi\_dscm\_no

order by indi\_dscm\_no, exmd\_bz\_yyyy;

**quit**;

\* 검진 테이블에 붙이기 - 출생연도 (join key ID+YEAR 사용);

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select p.\*, q.indi\_dscm\_no, q.byear

from zio.nhid\_gj\_0210 as p left join nhid\_jk\_age\_group as q

on p.indi\_dscm\_no = q.indi\_dscm\_no and p.exmd\_bz\_yyyy = q.std\_yyyy

order by indi\_dscm\_no, exmd\_bz\_yyyy;

**quit**;

\* 사람별 첫 검진일자 정의;

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select \*, min(hme\_dt\_new) as first\_exam\_date format=yymmdd10.

from zio.nhid\_gj\_0210

group by indi\_dscm\_no

order by indi\_dscm\_no;

**run**;

\* age 정의;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

age = exmd\_bz\_yyyy - byear;

**run**;

\* 검진 인원수 확인 (이상치 처리, na drop 전) - 2,298,631 명;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210;

**quit**;

\* 이상치 처리;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

if **13** > g1e\_bmi or **50** < g1e\_bmi then g1e\_bmi=**.** ;

if **5** > g1e\_tot\_chol or **740** < g1e\_tot\_chol then g1e\_tot\_chol=**.** ;

if (g1e\_bp\_sys > **250** or g1e\_bp\_sys <**60**) then g1e\_bp\_sys =**.**; \* SBP;

if (g1e\_bp\_dia >**150** or g1e\_bp\_dia < **40** ) then g1e\_bp\_dia =**.**; \* DBP;

if **1** > g1e\_ggt or **999** <= g1e\_ggt then g1e\_ggt = **.** ;

if **1** > g1e\_sgot or **999** <= g1e\_sgot then g1e\_sgot=**.** ; \* 기존 2900 -> 999로 변경;

if **1** > g1e\_sgpt or **999** <= g1e\_sgpt then g1e\_sgpt=**.** ; \* 기존 2900 -> 999로 변경;

if **1** > g1e\_fbs or **800**< g1e\_fbs then g1e\_fbs=**.** ;

**run**;

\* 쓸 변수에서 null 모두 drop;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

if hme\_dt\_new = **.** then delete;

if g1e\_bmi = **.** then delete;

if g1e\_hgb = **.** then delete;

if g1e\_bp\_sys = **.** then delete;

if g1e\_bp\_dia = **.** then delete;

if g1e\_urn\_prot = **.** then delete;

if g1e\_fbs = **.** then delete;

if g1e\_tot\_chol = **.** then delete;

if g1e\_sgot = **.** then delete;

if g1e\_sgpt = **.** then delete;

if g1e\_ggt = **.** then delete;

if age = **.** then delete;

if smk\_term = **.** then delete;

**run**;

\* 검진 인원수 확인 (이상치 처리, na drop 후) - 2,292,945 명;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210;

**quit**;

**proc** **freq** data=zio.nhid\_gj\_0210; \* urn\_prot check;

table g1e\_urn\_prot;

**run**;

\* descriptive for numerical variables;

**proc** **means** data=zio.nhid\_gj\_0210 n nmiss mean median min max std;

**run**;

\* 검진도 여기까지 돌려놓은 것 backup;

**data** zio.nhid\_gj\_0210\_bu;

set zio.nhid\_gj\_0210;

**run**;

\* backup본 copy가 필요할 때;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210\_bu;

**run**;

/\* 검진 end. ---------------------------------------------------------------------------------------------------------------------------------------------------------- \*/

/\* antihistamine, NPD, 검진 1차 preprocess 완.

antihistamine 최종 table : zio.nhid\_gy3060\_antihis\_allprescrip - 397만명

NPD 최종 table : zio.nhid\_gy20\_npd\_2MajorSymp - 3만 2천명

검진 최종 table : zio.nhid\_gj\_0210 - 229만명

\*/

\* NPD table 사람별 하나의 첫 진단일자, 진단여부만 나오도록 하기;

\* table 복사;

**proc** **sql**;

create table zio.nhid\_gy20\_npd\_2MSymp2 as

select \*

from zio.nhid\_gy20\_npd\_2MajorSymp;

**quit**;

\* change to datetime;

**data** zio.nhid\_gy20\_npd\_2MSymp2;

set zio.nhid\_gy20\_npd\_2MSymp2;

mdcare\_strt\_dt\_new = INPUT(mdcare\_strt\_dt, yymmdd10.);

FORMAT mdcare\_strt\_dt\_new yymmdd10.;

**RUN**;

\* 진단일자 사람별 오름차순 정렬;

**proc** **sort** data=zio.nhid\_gy20\_npd\_2MSymp2;

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

**run**;

\* 진단일자 2002년 이상으로 끊기;

\* 제외하고 받았으나, 혹시 모르니 한번더;

**data** zio.nhid\_gy20\_npd\_2MSymp2;

set zio.nhid\_gy20\_npd\_2MSymp2;

if mdcare\_strt\_dt >= '20020101';

**run**;

\* 사람별 맨 위 행만 남기기;

**proc** **sort** data=zio.nhid\_gy20\_npd\_2MSymp2 NODUPKEY;

by indi\_dscm\_no;

**run**;

\* 32,468명;

\* 검진 table 이랑 left join 해서, NPD 열이 null 인 사람은 0으로 넣어주면 됨;

\* (23.03.20) NPD washout 추가, 진단일자 2002~2004인 사람은 삭제;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy20\_npd\_2MSymp2

where npd\_date >= '20020101' and npd\_date <= '20041231';

**quit**; \* 16명;

**data** zio.nhid\_gy20\_npd\_2MSymp2;

set zio.nhid\_gy20\_npd\_2MSymp2;

if npd\_date >= '20020101' and npd\_date <= '20041231' then delete;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy20\_npd\_2MSymp2;

**quit**; \* 32,452명;

\* NPD 2차 preprocess 완. ;

/\* antihistamine, NPD, 검진 2차 preprocess 완.

antihistamine 최종 table with all prescriptions : zio.nhid\_gy3060\_antihis\_allprescrip

NPD 최종 table : zio.nhid\_gy20\_npd\_2MSymp2

검진 최종 table : zio.nhid\_gj\_0210

\*/

\* (23.03.20) NPD\_final4도 total rate check 에만 사용함;

\* (23.02.23) 이하 프로세스에서, NPD\_final4는 사용하지 않음;

\* NPD의 정의는 정신과에서 진단된 모든 건으로;

\*\*\* (23.01.23) NPD 정의 수정 (NPD\_final4) \*\*\*;

\* 만약 검진기준으로 검진+약물+npd를 붙여버리면 id당 검진 n회x약물 n회xnpd 1회 이렇게 됨, 행이 너무 많아져서 제대로 볼수없다;

\* 일단 검진에 약물복용여부와 npd 진단여부만 0/1로 붙여두자;

\* 그러면 약물xnpd를 left join으로 먼저 했을 때 잘리는 npd 진단자도 검진에서는 진단되었다고 잘 표시되어 있음;

\* npd 재정의는 약물 x npd table에서 진행하고 검진에 붙이자;

\* 검진 총원 check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210;

**quit**; \*2,292,945;

**proc** **sort** data=zio.nhid\_gj\_0210;

by indi\_dscm\_no exmd\_bz\_yyyy;

**run**; \* 인당 여러 회 검진 모두 살아있음;

\* 검진에 약물복용여부 (1일이상) 와 npd 진단여부만 붙여놓기;

\* 기타 npd, medi 관련 변수 다 붙여놓기;

/\*data zio.nhid\_gy3060\_antihis\_allprescrip; \*/

/\*set zio.nhid\_gy3060\_antihis\_allprescrip; \*/

/\*antihistamine = 1; \*/

/\*run; \*/

**proc** **sql**;

create table tmp as

select distinct(indi\_dscm\_no), anti, first\_medi\_date\_new

from zio.nhid\_gy3060\_antihis\_allprescrip;

**quit**;

\* 내림차순 정렬 -> drop dup in id -> 먹은사람은 1, 안먹은 사람은 0 남음;

**proc** **sort** data=tmp;

by indi\_dscm\_no descending anti;

**run**;

**proc** **sort** data=tmp nodupkey;

by indi\_dscm\_no;

**run**; \*3978548명;

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select p.\*, q.\*

from zio.nhid\_gj\_0210 as p left join tmp as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

**proc** **sql**;

create table zio.nhid\_gj\_0210 as

select p.\*, q.indi\_dscm\_no, q.NPD, q.npd\_date\_new

from zio.nhid\_gj\_0210 as p left join zio.nhid\_gy20\_npd\_2MSymp2 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

\* 총원 check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0510;

**quit**; \*2,160,153;

\* 총원 check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210;

**quit**; \*2,292,945;

\* 검진에 약물을 붙였는데 아예 anti = . 으로 안 붙은 사람은, 검진은 받았으나 약물 복용 이력이 아예 없어서 약물 복용 테이블에서 안 잡히는 사람;

\* 이런 사람들은 anti = 0 처리해 줘도 됨;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

if anti = **.** then anti = **0**;

**run**;

\* (23.03.09) 약물 첫 처방 이전에 발생한 NPD 진단자는? ;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210

where npd = **1** and anti = **1** and npd\_date\_new < first\_medi\_date\_new;

**quit**; \*919명; \* 2002~2004 npd washout 적용시 918명으로 변경; \* 여기까지 돌리고 5개 질병군 table 생성하러 감;

\* (23.03.09) 이 밑에서 anti = 1인 사람만 뽑아서 약물 정의 작업 시작하므로,

여기서 약물 첫 처방 이전에 발생한 NPD 처리;

\* 삭제하지 말고, 약물 첫 처방 이전에 발생했을 경우 anti = 0, NPD = 1로 변경;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

if npd = **1** and anti = **1** and npd\_date\_new < first\_medi\_date\_new then npd\_before\_medi = **1**;

by indi\_dscm\_no;

**run**;

**data** zio.nhid\_gj\_0210;

set zio.nhid\_gj\_0210;

if npd\_before\_medi = **1** then anti = **0**;

**run**;

\* npd 인원 check ;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0510

where npd = **1**;

**quit**; \* 7,134;

\* npd 인원 check ;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210

where npd = **1**;

**quit**; \* 8,772; \* npd washout 적용 후 8,770;

\* 검진에 약물, npd 여부 붙여놓기 완;

\* anti = 1, npd = 1;

\* 23.03.09 여기까지 돌려놓고 감;

\* 약물 전처리 시작;

\* 이제는 anti = 1인 사람만 뽑아서;

\* 복사;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_AntiYesNoAll;

**run**;

\* anti = 1만 뽑기;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_antihis\_allprescrip;

if anti = **1**;

**run**;

\* 약물 table with all prescription (only anti = 1);

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy3060\_antihis\_allprescrip;

**quit**; \*3,634,084;

\* 첫 진단만 남긴 npd table check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gy20\_npd\_2MSymp2;

**quit**; \* 32,468; \* npd washout 적용 후 32,452;

\* 진단일자 변수명 겹치니까 rename;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_antihis\_allprescrip;

rename mdcare\_strt\_dt = medi\_date;

**run**;

**data** zio.nhid\_gy20\_npd\_2MSymp2;

set zio.nhid\_gy20\_npd\_2MSymp2;

rename mdcare\_strt\_dt = npd\_date;

rename mdcare\_strt\_dt\_new = npd\_date\_new;

**run**;

\* 사람별 첫 처방일자, 마지막 처방일자 열 생성;

**proc** **sql**;

create table zio.nhid\_gy3060\_antihis\_allprescrip as

select \*, min(medi\_date) as first\_MEDI\_date

from zio.nhid\_gy3060\_antihis\_allprescrip

group by indi\_dscm\_no

order by indi\_dscm\_no;

**quit**;

\* change to datetime;

**data** zio.nhid\_gy3060\_antihis\_allprescrip;

set zio.nhid\_gy3060\_antihis\_allprescrip;

first\_MEDI\_date\_new = input(first\_MEDI\_date, yymmdd10.);

format first\_MEDI\_date\_new yymmdd10.;

**run**;

\* 약물 x npd table left join;

**proc** **sql**;

create table zio.nhid\_antihis\_npd as

select p.indi\_dscm\_no, p.cmn\_key, p.medi\_date, p.first\_medi\_date\_new, p.tot\_mcnt, p.anti, q.\*

from zio.nhid\_gy3060\_antihis\_allprescrip as p left join zio.nhid\_gy20\_npd\_2MSymp2 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

**data** tmp;

set zio.nhid\_antihis\_npd;

if indi\_dscm\_no = **187857**;

**run**;

\* npd n check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd

where npd = **1**;

**quit**; \* 29,983; \* npd washout 적용 후 29,969;

**data** tmp;

set zio.nhid\_antihis\_npd;

if npd = **1** and npd\_date\_new < first\_medi\_date\_new;

**run**;

\* 약물 첫 처방 이전에 발생한 npd 삭제;

\* (23.03.09) 이 사람들은 삭제하지 말고, anti = 0, NPD = 1로 변경하기로 함;

\* 검진 테이블에 붙여놨으므로, anti = 1만 뽑은 여기서는 삭제해도 됨;

**data** zio.nhid\_antihis\_npd;

set zio.nhid\_antihis\_npd;

if npd = **1** and anti = **1** and npd\_date\_new < first\_medi\_date\_new then delete;

by indi\_dscm\_no;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd;

if indi\_dscm\_no = **187857**;

**run**;

\* npd n check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd

where npd = **1**;

**quit**; \* 24,625; \* npd washout 적용 후 24,622;

**proc** **sort** data=zio.nhid\_antihis\_npd;

by indi\_dscm\_no medi\_date;

**run**;

**data** zio.nhid\_antihis\_npd;

set zio.nhid\_antihis\_npd;

medi\_date\_new = input(medi\_date, yymmdd10.);

format medi\_date\_new yymmdd10.;

**run**;

\* 처방일자 + 처방일수 생성;

**data** zio.nhid\_antihis\_npd;

set zio.nhid\_antihis\_npd;

medi\_date\_plus\_tot\_mcnt = intnx('day', medi\_date\_new, tot\_mcnt);

format medi\_date\_plus\_tot\_mcnt yymmdd10.;

**run**;

\* calculate gap between two prescriptions;

\* lag 변수 활용;

**data** tmp;

set zio.nhid\_antihis\_npd;

keep indi\_dscm\_no cmn\_key medi\_date\_new medi\_date\_plus\_tot\_mcnt;

**run**;

**proc** **sort** data=tmp;

by indi\_dscm\_no medi\_date\_plus\_tot\_mcnt; **run**;

**data** tmp; set tmp;

lag\_medi\_date = lag(medi\_date\_plus\_tot\_mcnt);

format lag\_medi\_date yymmdd10.;

**run**;

**data** tmp; set tmp;

by indi\_dscm\_no;

if first.indi\_dscm\_no then lag\_medi\_date = **.**;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd2 as

select p.\*, q.cmn\_key, q.lag\_medi\_date

from zio.nhid\_antihis\_npd as p left join tmp as q

on p.cmn\_key = q.cmn\_key;

**quit**;

**proc** **sort** data=zio.nhid\_antihis\_npd2;

by indi\_dscm\_no medi\_date\_plus\_tot\_mcnt; **run**;

/\*data zio.nhid\_antihis\_npd2; \*/

/\*set zio.nhid\_antihis\_npd2; \*/

/\*prescrip\_interval = intck('day', lag\_medi\_date, medi\_date\_plus\_tot\_mcnt); \*/

/\*put prescrip\_interval = ; \*/

/\*run; \* 30분 이상 돌렸는데 안됨; \*/

**data** zio.nhid\_antihis\_npd2;

set zio.nhid\_antihis\_npd2;

prescrip\_interval = medi\_date\_plus\_tot\_mcnt - lag\_medi\_date;

**run**; \* 그냥 이렇게 해보자;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd2

where prescrip\_interval ^= **.**;

**run**; \*3,368,016;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd2

where npd = **1** and prescrip\_interval ^= **.**;

**run**; \* 23,675;

\* 2주 이내의 처방 다 합치기;

**data** zio.nhid\_antihis\_npd3;

set zio.nhid\_antihis\_npd2;

by indi\_dscm\_no;

retain group\_ **1**;

if first.indi\_dscm\_no then group\_ = **1**;

if prescrip\_interval >= **14** then group\_ = group\_ + **1**;

**run**;

\* chunk 별 약물 복용량 합, 해당 chunk 의 medi\_date 구하기, min(medi\_date) 로;

\* 해당 chunk의 약물 복용이 끝난 날짜는, max(medi\_date) + 마지막 처방일수;

**proc** **sql**;

create table zio.nhid\_antihis\_npd4 as

select \*, sum(tot\_mcnt) as medi\_taken, min(medi\_date\_new) as medi\_date\_final, max(medi\_date\_new) as medi\_date\_final\_end

from zio.nhid\_antihis\_npd3

group by indi\_dscm\_no, group\_

order by indi\_dscm\_no, medi\_date\_new;

**quit**;

\*medi\_date\_final format 수정;

**data** zio.nhid\_antihis\_npd4;

set zio.nhid\_antihis\_npd4;

format medi\_date\_final yymmdd10.;

format medi\_date\_final\_end yymmdd10.;

**run**;

**proc** **sort** data=zio.nhid\_antihis\_npd4;

by indi\_dscm\_no group\_ medi\_date\_new;

**run**;

\* medi\_date\_final\_end에 마지막 처방일수 더해주기;

\* chunk에 처방이 하나밖에 없으면, 해당 처방일수가 더해짐;

**data** zio.nhid\_antihis\_npd4;

set zio.nhid\_antihis\_npd4;

by indi\_dscm\_no group\_;

if last.group\_ then medi\_date\_final\_end2 = intnx("day", medi\_date\_final\_end, tot\_mcnt);

format medi\_date\_final\_end2 yymmdd10.;

**run**;

\* id별, group별 마지막 항목에만 생성된 medi\_date\_final\_end2를 맨 위로 올리기;

**proc** **sort** data=zio.nhid\_antihis\_npd4;

by indi\_dscm\_no group\_ descending medi\_date\_final\_end2;

**run**;

\* 이제 id별, group별 drop dup을 하면 chunk를 하나의 처방처럼 생각할 수 있음;

**proc** **sort** data=zio.nhid\_antihis\_npd4 nodupkey;

by indi\_dscm\_no group\_;

**run**;

\* 이제 chunk별 첫 약물 복용일자는 medi\_date\_final;

\* chunk별 약물 복용이 끝난 일자는 medi\_date\_final\_end2;

\* medi\_date\_final과 medi\_date\_final\_end2 사이에 npd\_date가 존재하면, NPD\_final4 = 1;

\* 이때,

if npd = 1 and npd\_final4 = 1 then medi\_taken\_final = 그 chunk의 medi\_taken (발생했고, 약물 복용기간중 발생)

if npd = 0 and npd\_final4 = 0 then medi\_taken\_final = max(medi\_taken) from all evaluation period (발생 아예 안함)

if npd = 1 and npd\_final4 = 0 then medi\_taken\_final = max(medi\_taken) until npd occurs (발생했지만, 약물 복용기간 중에 발생하지는 않음);

\* 최종 재정의된 npd 진단변수는 npd\_final4, 약물복용변수는 medi\_taken\_final;

\* (23.02.24) 최종 outcome 변수를 NPD로 사용하기로 함;

\* 약물 복용 변수는 발생 전까지 전체복용기간의 sum 혹은 max(medi\_taken) until npd occurs 두가지 다 확인하기로 함;

\* 이때,

if npd = 1 then medi\_taken\_final\_max = max(medi\_taken) until npd occurs

if npd = 1 then medi\_taken\_final\_sum = sum(medi\_taken) until npd occurs

if npd = 0 then medi\_taken\_final\_max = max(medi\_taken) from all evaluation period

if npd = 0 then medi\_taken\_final\_sum = sum(medi\_taken) from all evaluation period;

\* 약물복용변수 정의;

\* npd = 1 then medi\_taken\_final\_max = max(medi\_taken) until npd occurs;

\* npd = 1 then medi\_taken\_final\_sum = sum(medi\_taken) until npd occurs;

\* tmp\_group = 1인 데까지가 사람별 npd 발생 전까지의 데이터;

\* npd\_final4 정의;

**proc** **sql**;

create table zio.nhid\_antihis\_npd5 as

select \*, case when (medi\_date\_final < npd\_date\_new <= medi\_date\_final\_end2) then **1** end as NPD\_final4

from zio.nhid\_antihis\_npd4

group by indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**data** zio.nhid\_antihis\_npd5;

set zio.nhid\_antihis\_npd5;

if NPD\_final4 = **.** then NPD\_final4 = **0**;

**run**;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd5

where npd\_final4 = **1**;

**quit**; \*15,662명; \* 956명;

\* npd\_final4에 대한 약물복용변수 정의;

**data** zio.nhid\_antihis\_npd5;

set zio.nhid\_antihis\_npd5;

if npd = **1** and npd\_final4 = **1** then medi\_taken\_final = medi\_taken;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd6 as

select \*, case when npd = **.** and npd\_final4 = **0** then max(medi\_taken) end as medi\_taken\_max

from zio.nhid\_antihis\_npd5

group by indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**data** zio.nhid\_antihis\_npd6;

set zio.nhid\_antihis\_npd6;

if npd = **.** and npd\_final4 = **0** then medi\_taken\_final = medi\_taken\_max;

**run**;

**data** zio.nhid\_antihis\_npd6;

set zio.nhid\_antihis\_npd6;

drop medi\_taken\_max;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd7 as

select \*, case when npd = **1** and npd\_final4 = **0** and medi\_date\_final\_end2 <= npd\_date\_new then **1** end as tmp\_group

from zio.nhid\_antihis\_npd6

group by indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd8 as

select \*, case when npd = **1** and npd\_final4 = **0** and tmp\_group = **1** then max(medi\_taken) end as medi\_taken\_max2

from zio.nhid\_antihis\_npd7

group by indi\_dscm\_no, tmp\_group

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

\* 이제 medi\_date\_final = . 이면 모두 medi\_taken\_max2로 채우면 됨;

**data** tmp;

set zio.nhid\_antihis\_npd8;

keep indi\_dscm\_no medi\_taken\_max2;

**run**;

**proc** **sort** data=tmp nodupkey;

by indi\_dscm\_no descending medi\_taken\_max2;

**run**;

**proc** **sort** data=tmp nodupkey;

by indi\_dscm\_no;

**run**;

**data** zio.nhid\_antihis\_npd8;

set zio.nhid\_antihis\_npd8;

drop medi\_taken\_max2;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd9 as

select p.\*, q.\*

from zio.nhid\_antihis\_npd8 as p left join tmp as q

on p.indi\_dscm\_no = q.indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**data** zio.nhid\_antihis\_npd9;

set zio.nhid\_antihis\_npd9;

if medi\_taken\_final = **.** then medi\_taken\_final = medi\_taken\_max2;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd9;

table medi\_taken\_final;

**run**;

**data** tmp2;

set zio.nhid\_antihis\_npd9;

where medi\_taken\_final = **.**;

**run**;

**data** tmp2;

set zio.nhid\_antihis\_npd9;

where indi\_dscm\_no = **157557**;

**run**;

\* npd = . 이면 0 넣어주기;

**data** zio.nhid\_antihis\_npd9;

set zio.nhid\_antihis\_npd9;

if npd = **.** then npd = **0**;

**run**;

\* sum, max 변수 정의;

**data** zio.nhid\_antihis\_npd5;

set zio.nhid\_antihis\_npd9;

**run**;

**data** zio.nhid\_antihis\_npd5;

set zio.nhid\_antihis\_npd5;

drop tmp\_group medi\_taken\_max2;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd5 as

select \*, case when npd = **1** and medi\_date\_final <= npd\_date\_new then **1** end as tmp\_group

from zio.nhid\_antihis\_npd5

group by indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd6 as

select \*, case when npd = **1** and tmp\_group = **1** then max(medi\_taken) end as medi\_taken\_final\_max,

case when npd = **1** and tmp\_group = **1** then sum(medi\_taken) end as medi\_taken\_final\_sum

from zio.nhid\_antihis\_npd5

group by indi\_dscm\_no, tmp\_group

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

\* npd = 0 then medi\_taken\_final\_max = max(medi\_taken) from all evaluation period

\* npd = 0 then medi\_taken\_final\_sum = sum(medi\_taken) from all evaluation period;

**proc** **sql**;

create table zio.nhid\_antihis\_npd7 as

select \*, case when npd = **.** then max(medi\_taken) end as medi\_max\_tmp,

case when npd = **.** then sum(medi\_taken) end as medi\_sum\_tmp

from zio.nhid\_antihis\_npd6

group by indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

**data** zio.nhid\_antihis\_npd7;

set zio.nhid\_antihis\_npd7;

if medi\_taken\_final\_max = **.** then medi\_taken\_final\_max = medi\_max\_tmp;

if medi\_taken\_final\_sum = **.** then medi\_taken\_final\_sum = medi\_sum\_tmp;

drop medi\_max\_tmp medi\_sum\_tmp;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd7;

keep indi\_dscm\_no medi\_taken\_final\_max medi\_taken\_final\_sum;

**run**;

**proc** **sort** data=tmp nodupkey;

by indi\_dscm\_no descending medi\_taken\_final\_max descending medi\_taken\_final\_sum;

**run**;

**proc** **sort** data=tmp nodupkey;

by indi\_dscm\_no;

**run**;

**data** zio.nhid\_antihis\_npd7;

set zio.nhid\_antihis\_npd7;

drop medi\_taken\_final\_max medi\_taken\_final\_sum;

**run**;

**proc** **sql**;

create table zio.nhid\_antihis\_npd8 as

select p.\*, q.\*

from zio.nhid\_antihis\_npd7 as p left join tmp as q

on p.indi\_dscm\_no = q.indi\_dscm\_no

order by indi\_dscm\_no, medi\_date\_final;

**quit**;

/\*data zio.nhid\_antihis\_npd9; \*/

/\*set zio.nhid\_antihis\_npd9; \*/

/\*if medi\_taken\_final = . then medi\_taken\_final = medi\_taken\_max2; \*/

/\*run; \*/

/\*proc freq data=zio.nhid\_antihis\_npd9; \*/

/\*table medi\_taken\_final; \*/

/\*run;\*/

**data** tmp2;

set zio.nhid\_antihis\_npd8;

where medi\_taken\_final\_max = **.**;

**run**; \* 없음 확인;

\* npd = . 이면 0 넣어주기;

**data** zio.nhid\_antihis\_npd8;

set zio.nhid\_antihis\_npd8;

if npd = **.** then npd = **0**;

**run**;

\* 이제 dropdup 가능;

**data** zio.nhid\_antihis\_npd9;

set zio.nhid\_antihis\_npd8;

**run**;

**proc** **sort** data=zio.nhid\_antihis\_npd9 nodupkey;

by indi\_dscm\_no;

**run**; \* 3,628,737;

\* 3.10 여기까지 돌림;

**data** tmp;

set zio.nhid\_antihis\_npd9;

if medi\_taken\_final\_max = **0**;

**run**; \* id = 49918305 , 처방은 받았으나 tot\_mcnt = 0, drop;

**data** zio.nhid\_antihis\_npd9;

set zio.nhid\_antihis\_npd9;

if medi\_taken\_final\_max = **0** then delete;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd9;

table npd\_final4;

**run**;

**data** zio.nhid\_antihis\_npd10;

set zio.nhid\_antihis\_npd9;

**run**;

\*npd4를 위해, 여기서 target 군만 뽑아보자;

\* 실험군 (5개 질병) 만 남긴 table 만들기;

/\*proc sql; \*/

/\*create table zio.nhid\_antihis\_npd10 as\*/

/\*select p.\*, q.\*\*/

/\*from zio.nhid\_antihis\_npd10 as p inner join target.target\_disease as q\*/

/\*on p.indi\_dscm\_no = q.indi\_dscm\_no; \*/

/\*quit; \*/

/\*proc freq data=zio.nhid\_antihis\_npd10; \*/

/\*table npd\_final4; \*/

/\*run; \* 191만 중 7명; \*/

\* npd4 total rate는 그냥 362만을 대상으로 여기서 확인하자;

\* medi\_taken\_final;

\* npd\_final4 변수 위해 그대로 range 유지;

\* (04.07) 180days + 를 마지막으로 range 변경, 1y+는 삭제함;

**data** zio.nhid\_antihis\_npd10;

set zio.nhid\_antihis\_npd10;

if medi\_taken\_final = **.** or medi\_taken\_final = **0** then medi\_final\_category = **0**;

if medi\_taken\_final >= **1** and medi\_taken\_final < **7** then medi\_final\_category = **1**;

if medi\_taken\_final >= **7** and medi\_taken\_final < **30** then medi\_final\_category = **2**;

if medi\_taken\_final >= **30** and medi\_taken\_final < **90** then medi\_final\_category = **3**;

if medi\_taken\_final >= **90** and medi\_taken\_final < **180** then medi\_final\_category = **4**;

if medi\_taken\_final >= **180** then medi\_final\_category = **5**;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd10;

table medi\_final\_category;

**run**;

\* 0~30;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **0** or medi\_final\_category = **1** or medi\_final\_category = **2**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **0** or medi\_final\_category = **1** or medi\_final\_category = **2**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **0** or medi\_final\_category = **1** or medi\_final\_category = **2**;

**quit**;

\* 90+;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **4** or medi\_final\_category = **5** or medi\_final\_category = **6**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **4** or medi\_final\_category = **5** or medi\_final\_category = **6**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **4** or medi\_final\_category = **5** or medi\_final\_category = **6**;

**quit**;

\* 180+;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **5** or medi\_final\_category = **6**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **5** or medi\_final\_category = **6**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **5** or medi\_final\_category = **6**;

**quit**;

\* 1~7;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **1**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **1**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **1**;

**quit**;

\* 7~30;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **2**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **2**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **2**;

**quit**;

\* 30~90;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **3**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **3**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **3**;

**quit**;

\* 90~180;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **4**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10;

if (medi\_final\_category = **4**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10

where medi\_final\_category = **4**;

**quit**;

\* chi sq test;

**data** zio.nhid\_antihis\_npd10;

set zio.nhid\_antihis\_npd10;

if medi\_final\_category = **0** then delete;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd10;

tables npd\_final4\*medi\_final\_category / chisq;

**run**;

\* 같은 작업 60세 이상에서도 진행;

\* 아.. 검진이 안붙어 있네;

\* age만 붙이기;

\* 자격 테이블에서 필요한 변수만 뽑기 - 출생연도;

**proc** **sql**;

create table nhid\_jk\_age\_group as

select indi\_dscm\_no, std\_yyyy, byear

from zio.nhid\_jk

order by indi\_dscm\_no;

**quit**;

\* nhid\_antihis\_npd10에 약 복용년도 column 생성;

**data** zio.nhid\_antihis\_npd10;

set zio.nhid\_antihis\_npd10;

medi\_year = substr(medi\_date, **1**, **4**);

**run**;

\* nhid\_antihis\_npd10에 붙이기 - 출생연도 (join key ID+YEAR 사용);

**proc** **sql**;

create table zio.nhid\_antihis\_npd10\_over60 as

select p.\*, q.indi\_dscm\_no, q.byear

from zio.nhid\_antihis\_npd10 as p left join nhid\_jk\_age\_group as q

on p.indi\_dscm\_no = q.indi\_dscm\_no and p.medi\_year = q.std\_yyyy

order by indi\_dscm\_no, medi\_year;

**quit**;

\* age 생성;

**data** zio.nhid\_antihis\_npd10\_over60;

set zio.nhid\_antihis\_npd10\_over60;

age = medi\_year - byear;

**run**;

**data** zio.nhid\_antihis\_npd10\_over60;

set zio.nhid\_antihis\_npd10\_over60;

if age >= **60**;

**run**; \* 459,332명;

**proc** **freq** data=zio.nhid\_antihis\_npd10\_over60;

table npd\_final4\*medi\_final\_category;

**run**;

\* 1~7;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **1**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **1**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10\_over60

where medi\_final\_category = **1**;

**quit**;

\* 7~30;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **2**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **2**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10\_over60

where medi\_final\_category = **2**;

**quit**;

\* 30~90;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **3**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **3**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10\_over60

where medi\_final\_category = **3**;

**quit**;

\* 90~180;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **4**) and npd\_final4 = **1**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd10\_over60;

if (medi\_final\_category = **4**);

**run**;

**proc** **sql**;

title 'total rate';

select (sum(npd\_final4)/sum(medi\_taken\_final)) \* **100000**

from zio.nhid\_antihis\_npd10\_over60

where medi\_final\_category = **4**;

**quit**;

\* chi sq test;

**data** zio.nhid\_antihis\_npd10\_over60;

set zio.nhid\_antihis\_npd10\_over60;

if medi\_final\_category = **0** then delete;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd10\_over60;

tables npd\_final4\*medi\_final\_category / chisq;

**run**;

\* 03.21 여기까지 돌리고 감;

\* 이제 npd4는 신경 안쓰고, 밑에서는 max, sum 변수에 대해서만 하면 됨;

\* 질병 5개 세분화 하는 것 잊지말기;

\*\*\*\*\* (23.02.01) 검진 붙이기 \*\*\*\*\*\*\*

\* 검진 총원 check;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_gj\_0210;

**quit**; \*2,292,945;

\* 첫 검진만 남기기;

**proc** **sort** data=zio.nhid\_gj\_0210;

by indi\_dscm\_no exmd\_bz\_yyyy;

**run**;

**data** zio.nhid\_gj\_0210\_1;

set zio.nhid\_gj\_0210;

**run**;

**proc** **sort** data=zio.nhid\_gj\_0210\_1 nodupkey;

by indi\_dscm\_no;

**run**;

**data** zio.nhid\_gj\_0210\_1;

set zio.nhid\_gj\_0210\_1;

if npd = **.** then npd = **0**;

**run**;

**data** zio.nhid\_gj\_0210\_1;

set zio.nhid\_gj\_0210\_1;

rename npd = npd\_all;

**run**;

**data** tmp;

set zio.nhid\_gj\_0210\_1;

if anti = **.**;

**run**; \* 없음 확인;

\*final table;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam as

select p.\*, q.npd\_final4, q.medi\_taken\_final, q.medi\_taken\_final\_max, q.medi\_taken\_final\_sum

from zio.nhid\_gj\_0210\_1 as p left join zio.nhid\_antihis\_npd9 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

**data** tmp;

set zio.nhid\_antihis\_npd\_exam;

if anti = **1** and medi\_taken\_final\_max = **.**;

**run**; \* 이 사람들이 있으면 안됨; \* 49918305 1명 있음.;

**data** zio.nhid\_antihis\_npd\_exam;

set zio.nhid\_antihis\_npd\_exam;

if anti = **1** and medi\_taken\_final\_max = **.** then delete;

**run**; \* id = 49918305 , 처방은 받았으나 tot\_mcnt = 0, drop;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where anti = **0**;

**quit**; \*147,596;

**proc** **sql**;

title 'count';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where npd\_final4 = **1**;

**quit**; \*5;

\* 복용일수 변수 생성;

\* 분포 확인;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam n nmiss mean median min max std;

**run**;

\* medi\_taken\_final\_sum;

\* mean 87.4;

\* 0-30, 30-60, 60-90, 90-120, 120-180, 180-365, 1y-2y, 2y-3y, 3y-4y, 4y+;

\* (23.03.21 수정) 0, 1-7, 7-30, 30-90, 90-180, 180-1y, 1y+ ;

\* (23.04.07 수정) 0, 1-7, 7-30, 30-90, 90-180, 180+;

**data** zio.nhid\_antihis\_npd\_exam;

set zio.nhid\_antihis\_npd\_exam;

if anti = **0** then medi\_sum\_category = **0**;

if medi\_taken\_final\_sum >= **1** and medi\_taken\_final\_sum < **7** then medi\_sum\_category = **1**;

if medi\_taken\_final\_sum >= **7** and medi\_taken\_final\_sum < **30** then medi\_sum\_category = **2**;

if medi\_taken\_final\_sum >= **30** and medi\_taken\_final\_sum < **90** then medi\_sum\_category = **3**;

if medi\_taken\_final\_sum >= **90** and medi\_taken\_final\_sum < **180** then medi\_sum\_category = **4**;

if medi\_taken\_final\_sum >= **180** then medi\_sum\_category = **5**;

**run**;

/\*\* medi\_taken\_final\_max; \*/

/\*\* mean 13.26; \*/

/\*\* 0-5, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-60, 60-90, 90-120, 120-180, 180-365, 1y+; \*/

/\*data zio.nhid\_antihis\_npd\_exam; \*/

/\*set zio.nhid\_antihis\_npd\_exam; \*/

/\*if medi\_taken\_final\_max >= 0 and medi\_taken\_final\_max < 5 then medi\_max\_0\_5 = 1; else medi\_max\_0\_5 = 0; \*/

/\*if medi\_taken\_final\_max >= 5 and medi\_taken\_final\_max < 10 then medi\_max\_5\_10 = 1; else medi\_max\_5\_10 = 0; \*/

/\*if medi\_taken\_final\_max >= 10 and medi\_taken\_final\_max < 15 then medi\_max\_10\_15 = 1; else medi\_max\_10\_15 = 0; \*/

/\*if medi\_taken\_final\_max >= 15 and medi\_taken\_final\_max < 20 then medi\_max\_15\_20 = 1; else medi\_max\_15\_20 = 0; \*/

/\*if medi\_taken\_final\_max >= 20 and medi\_taken\_final\_max < 30 then medi\_max\_20\_30 = 1; else medi\_max\_20\_30 = 0; \*/

/\*if medi\_taken\_final\_max >= 30 and medi\_taken\_final\_max < 40 then medi\_max\_30\_40 = 1; else medi\_max\_30\_40 = 0; \*/

/\*if medi\_taken\_final\_max >= 40 and medi\_taken\_final\_max < 50 then medi\_max\_40\_50 = 1; else medi\_max\_40\_50 = 0; \*/

/\*if medi\_taken\_final\_max >= 50 and medi\_taken\_final\_max < 60 then medi\_max\_50\_60 = 1; else medi\_max\_50\_60 = 0; \*/

/\*if medi\_taken\_final\_max >= 60 and medi\_taken\_final\_max < 90 then medi\_max\_60\_90 = 1; else medi\_max\_60\_90 = 0; \*/

/\*if medi\_taken\_final\_max >= 90 and medi\_taken\_final\_max < 120 then medi\_max\_90\_120 = 1; else medi\_max\_90\_120 = 0; \*/

/\*if medi\_taken\_final\_max >= 120 and medi\_taken\_final\_max < 180 then medi\_max\_120\_180 = 1; else medi\_max\_120\_180 = 0; \*/

/\*if medi\_taken\_final\_max >= 180 and medi\_taken\_final\_max < 365 then medi\_max\_180\_1y = 1; else medi\_max\_180\_1y = 0; \*/

/\*if medi\_taken\_final\_max >= 365 then medi\_max\_1y\_more = 1; else medi\_max\_1y\_more = 0; \*/

/\*run; \*/

\* medi\_taken\_final\_max;

\* mean 13.26;

\* 0-5, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-60, 60-90, 90-120, 120-180, 180-365, 1y+;

\* (23.03.21 수정) 0, 1-7, 7-30, 30-90, 90-180, 180-1y, 1y+ ;

\* (23.04.07 수정) 0, 1-7, 7-30, 30-90, 90-180, 180+;

**data** zio.nhid\_antihis\_npd\_exam;

set zio.nhid\_antihis\_npd\_exam;

if anti = **0** then medi\_max\_category = **0**;

if medi\_taken\_final\_max >= **1** and medi\_taken\_final\_max < **7** then medi\_max\_category = **1**;

if medi\_taken\_final\_max >= **7** and medi\_taken\_final\_max < **30** then medi\_max\_category = **2**;

if medi\_taken\_final\_max >= **30** and medi\_taken\_final\_max < **90** then medi\_max\_category = **3**;

if medi\_taken\_final\_max >= **90** and medi\_taken\_final\_max < **180** then medi\_max\_category = **4**;

if medi\_taken\_final\_max >= **180** then medi\_max\_category = **5**;

**run**;

**data** tmp;

set zio.nhid\_antihis\_npd\_exam;

if medi\_taken\_final = **.** ;

**run**; \* 147,600;

/\*\* medi\_taken\_final; \*/

/\*\* npd\_final4 변수 위해 그대로 range 유지; \*/

/\*data zio.nhid\_antihis\_npd\_exam; \*/

/\*set zio.nhid\_antihis\_npd\_exam; \*/

/\*if medi\_taken\_final = . or medi\_taken\_final = 0 then medi\_final\_category = 0; \*/

/\*if medi\_taken\_final >= 1 and medi\_taken\_final < 7 then medi\_final\_category = 1; \*/

/\*if medi\_taken\_final >= 7 and medi\_taken\_final < 30 then medi\_final\_category = 2; \*/

/\*if medi\_taken\_final >= 30 and medi\_taken\_final < 90 then medi\_final\_category = 3; \*/

/\*if medi\_taken\_final >= 90 and medi\_taken\_final < 180 then medi\_final\_category = 4; \*/

/\*if medi\_taken\_final >= 180 and medi\_taken\_final < 365 then medi\_final\_category = 5; \*/

/\*if medi\_taken\_final >= 365 then medi\_final\_category = 6; \*/

/\*run; \*/

\* table;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam;

table npd\_all \* medi\_max\_category;

**run**;

/\*proc freq data=zio.nhid\_antihis\_npd\_exam; \*/

/\*table npd\_final4 \* medi\_final\_category; \*/

/\*run; \*/

/\*proc freq data=zio.nhid\_antihis\_npd\_exam; \*/

/\*table npd\_final4; \*/

/\*run; \*/

\* npd\_final4 관련변수 삭제;

**data** zio.nhid\_antihis\_npd\_exam;

set zio.nhid\_antihis\_npd\_exam;

drop NPD\_final4 medi\_taken\_final;

**run**;

\* 질병별 table 생성;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_j30 as

select p.\*, q.indi\_dscm\_no, q.sick

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease\_j30 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**; \* 1,480,337;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_h10 as

select p.\*, q.indi\_dscm\_no, q.sick

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease\_h10 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**; \* 1,222,034;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_j45 as

select p.\*, q.indi\_dscm\_no, q.sick

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease\_j45 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**; \* 814,414;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_l23 as

select p.\*, q.indi\_dscm\_no, q.sick

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease\_l23 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**; \* 1,191,918;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_l50 as

select p.\*, q.indi\_dscm\_no, q.sick

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease\_l50 as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**; \* 977,724;

\* 질병별 table에서 crosstable;

\* table;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_j30;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_j30;

table npd\_all \* medi\_max\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_h10;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_h10;

table npd\_all \* medi\_max\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_j45;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_j45;

table npd\_all \* medi\_max\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_l23;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_l23;

table npd\_all \* medi\_max\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_l50;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_l50;

table npd\_all \* medi\_max\_category;

**run**;

\* sum 0~30;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_0\_30)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_0\_30 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_0\_30 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_0\_30 = **1** and npd\_all = **1**;

**quit**;

\* sum 30~60;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_30\_60)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_30\_60 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_30\_60 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_30\_60 = **1** and npd\_all = **1**;

**quit**;

\* sum 60~90;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_60\_90)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_60\_90 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_60\_90 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_60\_90 = **1** and npd\_all = **1**;

**quit**;

\* sum 90~120;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_90\_120)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_90\_120 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_90\_120 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_90\_120 = **1** and npd\_all = **1**;

**quit**;

\* sum 120~180;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_120\_180)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_120\_180 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_120\_180 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_120\_180 = **1** and npd\_all = **1**;

**quit**;

\* sum 180~1y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_180\_1y)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_180\_1y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_180\_1y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_180\_1y = **1** and npd\_all = **1**;

**quit**;

\* sum 1y~2y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_1y\_2y)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_1y\_2y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_1y\_2y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_1y\_2y = **1** and npd\_all = **1**;

**quit**;

\* sum 2y~3y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_2y\_3y)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1** and npd\_all = **1**;

**quit**;

\* sum 2y~3y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_2y\_3y)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_2y\_3y = **1** and npd\_all = **1**;

**quit**;

\* sum 3y~4y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_3y\_4y)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_3y\_4y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_3y\_4y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_3y\_4y = **1** and npd\_all = **1**;

**quit**;

\* sum 4y+;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_sum\_4y\_more)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_4y\_more = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_4y\_more = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_sum\_4y\_more = **1** and npd\_all = **1**;

**quit**;

\* max 0~5;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_0\_5)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_0\_5 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_0\_5 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_0\_5 = **1** and npd\_all = **1**;

**quit**;

\* max 5~10;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_5\_10)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_5\_10 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_5\_10 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_5\_10 = **1** and npd\_all = **1**;

**quit**;

\* max 10~15;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_10\_15)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_10\_15 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_10\_15 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_10\_15 = **1** and npd\_all = **1**;

**quit**;

\* max 15~20;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_15\_20)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_15\_20 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_15\_20 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_15\_20 = **1** and npd\_all = **1**;

**quit**;

\* max 20~30;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_20\_30)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_20\_30 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_20\_30 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_20\_30 = **1** and npd\_all = **1**;

**quit**;

\* max 30~40;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_30\_40)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_30\_40 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_30\_40 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_30\_40 = **1** and npd\_all = **1**;

**quit**;

\* max 40~50;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_40\_50)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_40\_50 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_40\_50 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_40\_50 = **1** and npd\_all = **1**;

**quit**;

\* max 50~60;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_50\_60)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_50\_60 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_50\_60 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_50\_60 = **1** and npd\_all = **1**;

**quit**;

\* max 60~90;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_60\_90)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_60\_90 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_60\_90 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_60\_90 = **1** and npd\_all = **1**;

**quit**;

\* max 90~120;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_90\_120)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_90\_120 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_90\_120 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_90\_120 = **1** and npd\_all = **1**;

**quit**;

\* max 120~180;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_120\_180)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_120\_180 = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_120\_180 = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_120\_180 = **1** and npd\_all = **1**;

**quit**;

\* max 180\_1y;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_180\_1y)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_180\_1y = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_180\_1y = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_180\_1y = **1** and npd\_all = **1**;

**quit**;

\* max 1y+;

**proc** **sql**;

title 'total rate';

select sum(npd\_all)/sum(medi\_max\_1y\_more)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_1y\_more = **1**;

**quit**;

**proc** **sql**;

title 'medi';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_1y\_more = **1**;

**quit**;

**proc** **sql**;

title 'npd';

select count(distinct indi\_dscm\_no)

from zio.nhid\_antihis\_npd\_exam

where medi\_max\_1y\_more = **1** and npd\_all = **1**;

**quit**;

**data** tmp;

set zio.nhid\_antihis\_npd\_exam;

if anti = **0**;

**run**;

libname target "/userdata08/room553/data\_source/user\_data/0119";

\* 실험군 (5개 질병) 만 남긴 table 만들기;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_case as

select p.\*, q.\*

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_disease as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_case;

table npd\_all \* medi\_sum\_category;

**run**;

**proc** **freq** data=zio.nhid\_antihis\_npd\_exam\_case;

table npd\_all \* medi\_max\_category;

**run**;

\* 대조군 (5개 질병 X) 만 남긴 table 만들기;

**proc** **sql**;

create table zio.nhid\_antihis\_npd\_exam\_noncase as

select p.\*, q.\*

from zio.nhid\_antihis\_npd\_exam as p inner join target.target\_non\_disease as q

on p.indi\_dscm\_no = q.indi\_dscm\_no;

**quit**;

\* 이후 모델 돌릴 때는, 각 약물복용변수 vs. 0일 복용자만 빼서 복용변수별 table 만들어야 함. ;

\* descriptive;

**data** zio.nhid\_antihis\_npd\_exam\_case;

set zio.nhid\_antihis\_npd\_exam\_case;

drop NPD\_final4 medi\_taken\_final;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_case n nmiss mean median min max std;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_h10 n nmiss mean median min max std;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_j30 n nmiss mean median min max std;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_j45 n nmiss mean median min max std;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_l23 n nmiss mean median min max std;

**run**;

**proc** **means** data=zio.nhid\_antihis\_npd\_exam\_l50 n nmiss mean median min max std;

**run**;

\* descriptive by medi group;

\* medi sum category 1 (1~7days) vs else;

**data** tmp;

set zio.nhid\_antihis\_npd\_exam\_case;

if medi\_max\_category = **5**;

**run**;

**proc** **means** data=tmp n nmiss mean median min max std;

**run**;

\* End. Move to 2-1.PSM\_R06\_1year\_1vs1;