

〈건강검진코호트DB 결과 정리〉

- 9월 23일 Version
- Study Population 정의 관련

Goal) Study population 정의하는 방안 구상 / N수 파악

[Result]

1) 데이터 탐색 관련

① Raw data의 unique PERSON_ID 수 : 총 1,323,412명

- 2002년 ~ 2015년도 검진DB : 514,866명 (PERSON_ID가 결측인 개체 제외)
- 2016년 ~ 2017년도 검진DB : 346,694명
- 2018년 ~ 2019년도 검진DB : 347,205명
- 2007년 ~ 2015년 생애전환기검진DB : 91,662명
- 2016년 ~ 2017년 생애전환기검진DB : 22,985명

② Raw Data에서 사망일자(DTH_MDY)의 범위

- 2002년 ~ 2015년 자격DB : ~2015.12.31.일까지 있음
- 2016년 ~ 2019년 자격DB : ~2019.12.31.일까지 있음

2) Study population 포함 조건 적용

: 첫 번째 포함조건 “검진 DB에 BMI, WAIST, 공복 혈당 기록이 적어도 한 건 이상 존재해야 함. + 진료기록, 처방 기록 적어도 한 건 이상 존재해야 함.”을 적용한 후 unique PERSON_ID 수 파악

〈Coding 방법〉

① “검진 DB에 BMI, WAIST, 공복 혈당 기록이 적어도 한 건 이상 존재해야 함.” 조건 적용

: 검진DB에서 BMI, WAIST, 공복혈당이 한 건이라도 있는 unique PERSON_ID 추출

(이때, 연도에 따라 DB 분할 필요 - 2002년 ~ 2008년 / 2009년 ~ 2017년 / 2018년 ~ 2019년)

→ 생애전환기검진DB에서도 동일한 조건을 가지는 unique PERSON_ID 추출

(연도에 따라 DB 분할되어 있음 - 2007년 ~ 2015년 / 2016년 ~ 2017년)

→ 두 PERSON_ID tbl 합침. --- “include1_total” tbl (unique PERSON_ID만 있음.)

② ①번 조건 + “진료기록 적어도 한 건 이상 존재해야 함.”

: 진료DB import (진료DB도 2002년 ~ 2015년 / 2016년 ~ 2019년으로 분할되어 있음.)

→ 각 진료DB에서 진료기록이 한 건이라도 있는(MAIN_SICK, SUB_SICK 변수 이용) PERSON_ID, KEY_SEQ, MAIN_SICK, SUB_SICK, RECU_FR_DT 가져옴.

* unique PERSON_ID에 대해서 KEY_SEQ가 여러 건 존재함을 확인.

* PERSON_ID는 검진DB와의 joint key, KEY_SEQ는 T60와의 joint key이다.

→ 위 조건 적용한 PERSON_ID로 구성된 tbl과 “include1_total” tbl과 joint 시행

--- “T20_screening” tbl

- 포함한 변수 : PERSON_ID, KEY_SEQ, MAIN_SICK, SUB_SICK, RECU_FR_DT

③ ②번 조건 + “처방 기록 적어도 한 건 이상 존재해야 함.”

: 처방DB(T60) import (T60도 2002년 ~ 2015년 / 2016년 ~ 2019년으로 분할되어 있음.)

→ 각 처방DB에서 진료기록이 한 건이라도 있는(DIV_CD 변수 이용) KEY_SEQ, DIV_CD, RECU_FR_DT 가 저음.

→ 위 조건 적용한 tbl과 “T20_screening” tbl과 joint 시행 (joint key는 “KEY_SEQ”)

--- “T60_T20_screening” tbl

- 포함한 변수 : KEY_SEQ, DIV_CD, RECU_FR_DT

〈N수 파악 결과〉

① “검진 DB와 생애전환기검진DB에 BMI, WAIST, 공복 혈당 기록이 적어도 한 건 이상 존재해야 함.”을 만족하는 개체 수 : 총 514,862명

- 2002년 ~ 2008년 검진DB에 WAIST, BMI, 공복 혈당 기록이 적어도 한 건 이상인 개체 수 : 총 514,840명
- 2009년 ~ 2017년 검진DB에 WAIST, BMI, 공복 혈당 기록이 적어도 한 건 이상인 개체 수 : 총 456,254명
- 2018년 ~ 2019년 검진DB에 WAIST, BMI, 공복 혈당 기록이 적어도 한 건 이상인 개체 수 : 총 347,099명
- 2007년 ~ 2015년 생애전환기검진DB에 WAIST, BMI, 공복 혈당 기록이 적어도 한 건 이상인 개체 수 : 총 91,660명
- 2016년 ~ 2017년 생애전환기검진DB에 WAIST, BMI, 공복 혈당 기록이 적어도 한 건 이상인 개체 수 : 총 22,985명

② ①번 조건 + “진료기록 적어도 한 건 이상 존재해야 함.”을 만족하는 개체 수 : 총 514,671명

③ ②번 조건 + “처방 기록 적어도 한 건 이상 존재해야 함.”을 만족하는 개체 수 : 총 514,096명

[질문 & 논의사항]

1) 진료DB 관련

: 진료DB가 “의과_보건기관(T1)”과 “치과_한방(T2)”으로 나누어져 있는데 의과_보건기관(T1)의 진료, 처방기록만 이용해도 괜찮은지.

2) 처방DB 관련

약제성분 관련 변수는 다음과 같이 두 가지가 있다.

DIV_CD	GNL_NM_CD
- 수가 코드	
- 구성 : 기본 code 5자리 + 산정 코드 3자리)	- 약품의 일반(성분) code

예시)

DIV_CD	GNL_NM_CD
A04203581	234601ACH
A04203581	234601ACH
A06102491	259700ATB
A07600231	157901ATB
A43850611	101405ATR
A21402211	308200ATB
F03300021	268000ATB
A43850611	101405ATR
A07600231	157901ATB

문제점 1) GNL_NM_CD 변수가 2016년부터는 기록에 포함되지 않는다. -- DIV_CD 변수를 이용하면 되지 않을까?

RECU_FR_DT	DIV_TYPE_CD	DIV_CD	GNL_NM_CD
20151230	3	644904160	297400ATB
20151208	3	659900150	299100AGN
20151209	3	654100320	178405CSI
20151202	3	643506770	334700CSI
20151211	3	653602330	112705ATB



RECU_FR_DT	DIV_TYPE_CD	DIV_CD
20171226	3	142301ATB
20171226	3	234401ATB
20171226	3	186101ATB
20171226	3	189001ACH

문제점 2) DIV_CD 변수가 2010년부터 코드 구성(분류 코드)가 변경됨.

RECU_FR_DT	DIV_CD	RECU_FR_DT	DIV_CD
20021209	A60650201	20171226	142301ATB
20021209	A60650151	20171226	234401ATB
20021209	A06904711	20171226	186101ATB

--> Medication 관련 covariates, Outcome 정의할 때 연도 기준으로 ATC code 따로 파악할 필요 있음.
(2010년부터의 DIV_CD는 기존 ATC code 형태가 아닌 듯.)