

# 〈건강검진코호트DB 결과 정리〉

- 10월 6일 Version

- Study population 포함조건 모두 적용

## [Share]

1. 검진DB Cleaning Plan 작성 완료 (회의록 지속적으로 follow up 함.)

--- 공유 필요 여부 확인

2. coding 과정 기록한 file 새로 생성 (추후의 필요성 & 과정 공유 위해)

## [Result]

1. Study population 포함조건 만족하는 unique PERSON\_ID의 수 : 총 177,570명

--- 결과 table을 살펴보니, Disease\_onset이 검진 때 많이 파악된 것을 확인함.

〈포함 조건 충족하는 code 적용 후 얻은 tbl 구성〉

: “KEY\_SEQ”가 검진DB와 진료&처방DB 구분해주는 역할 (검진DB는 이 변수 값이 결측)

PERSON_ID	KEY_SEQ	Disease_onset

## 2. Cohort entry date coding 계획

0. “Disease\_onset\_date” tbl의 PERSON\_ID column만 있는 tbl 생성

--- “include\_ID” tbl

① max\_date(Disease onset date, 최초 명세서 청구 날짜, 최초 처방 명세서 날짜) 구하기

①-1. “include\_ID” tbl과 “T60\_T20\_screening” tbl joint

--> PERSON\_ID, RECU\_FR\_DT 기준으로 정렬 / 가장 과거의 기록 가져오기 (최초 명세서, 처방 날짜)

--- “first\_RECU” tbl (이때, “RECU\_FR\_DT” 변수명 “Date”로 변경 필요)

①-2. “Disease\_onset\_date” tbl과 “first\_RECU” tbl row binding

--> PERSON\_ID, Date 기준 정렬 / last.PERSON\_ID 가져오기

--- “cohort1” tbl

② min\_date(Disease onset date 이후 검진일자 & BMI와 BLDS 기록 존재) 구하기

②-1. “screening\_total” tbl과 “Disease\_onset\_date” tbl joint (joint key는 ‘PERSON\_ID’)

--> WHERE) BLDS & BMI가 결측이 아님 AND Disease\_onset\_date <= HME\_DT

--> KEEP PERSON\_ID, HME\_DT

--- “cohort2” tbl (Disease onset date 이후의 모든 검진기록)

③ Cohort entry date 구하기

③-1. “cohort1” tbl과 “cohort2” tbl column binding

--> IF) cohort1.Date <= cohort2.Date / PERSON\_ID, HME\_DT만 가져오기

--- “cohort\_entry” tbl

③-2. “cohort\_entry” tbl PERSON\_ID, HME\_DT 기준으로 정렬

--> 가장 과거의 기록만 가져오기

--> unique PERSON\_ID 가져오기 --- “cohort\_entry\_date” tbl