

〈건강검진코호트DB Coding 결과 정리〉

- Study population 정의 관련
- 9월 16일 Update Version

Goal) 모집단 정의하는 코드 구성

현재 내가 할 수 있는 part)

- : study population 포함 조건 중 “첫 번째 조건” (검진DB에 BMI, WAIST 기록이 적어도 한 건 이상 존재) 만족하는 객체(PERSON_ID) 파악.
- : 검진DB & 생애전환기검진DB에서 “두 번째 조건”(공복혈당수치가 126 이상 OR 당뇨병 과거력 존재) 만족하는 객체(PERSON_ID) 파악.
- : Raw Data row 개수 파악
- : 검진DB & 생애전환기검진DB에 기록이 적어도 하나 이상 존재하는 객체 수 파악.

[Result]

1) 데이터 탐색 관련

- : 검진항목 / 문진항목에 결측값이 꽤 존재.
- : 2008년 이전에도 WAIST가 관측된 기록이 있음.
- : 2002년 ~ 2015년 검진DB에 PERSON_ID(개인일련번호)가 기록되지 않은 관측치 존재.
- : 사망일자(DTH_MDY) 값이 날짜 혹은 빈 칸으로 구성되므로 결측치 파악 통해 사망 여부 추측 가능.
- : 자격DB에 소득분위변수(CTRB_PT_TYPE_CD)가 존재. / 의료급여수급권자는 “0분위”
- : 검진DB에도 정확한 검진일자가 존재하는 것을 확인 - “HME_DT” 변수
- : 60T에는 “RECU_FR_DT” 변수가 존재 - 이는 최초내원일자, 조제투여일자를 의미.

2) 파악 결과

① Raw data 수 : 총 3,731,518(2018년도 ~ 2019년도 기록 제외)

- 2002년 ~ 2015년도 검진DB : 3,203,488개
- 2016년 ~ 2017년도 검진DB : 413,382개
- 2007년 ~ 2015년도 생애전환기검진DB : 91,663개
- 2016년도 ~ 2017년도 생애전환기검진DB : 22,985개

② 검진DB & 생애전환기검진DB에 기록이 적어도 하나 이상 존재하는 객체 수 : 총 820,496명

- 2002년 ~ 2015년도 검진DB : 514,741명
- 2016년도 ~ 2017년도 검진DB : 202,978명
- 2007년 ~ 2017년도 생애전환기검진DB : 102,777명

3) 코딩 과정

[검진DB에서 가져와야 하는 변수]

- 공통변수 : HCHK_YEAR(검진 연도), HCHK_MDT(검진 일자), PERSON_ID(개인일련번호), BLDS(공복 혈당), BMI, WAIST
- ~2008년 기록 : HCHK_PMH_CD1, HCHK_PMH_CD2, HCHK_PMH_CD3 (과거병력문진항목)
- 2009년 기록 ~ : HCHK_DIABML_PMH_YN (과거병력문진항목)

[생애전환기검진DB에서 가져와야 하는 변수]

- : PERSON_ID(개인일련번호), HME_DT(검진 일자), BMI, WAIST, BLDS(공복 혈당), HCHK_DIABML_PMH_YN (과거병력문진항목)

(나의 Plan)

