

〈건강검진코호트DB 결과 정리〉

-- 11월 4일 Version
-- Covariate 정의 part

〈TO DO LIST〉

1) Demographic covariate 정의

: “age”, “sex”, “income”

2) Family History covariate 정의

: “FMLY_Heart_disease”, “FMLY_Diabetes”, “FMLY_Cancer”

〈Shared thing〉

0) 전처리한 tbl 중 후에 필요할 것 같다고 예상되는 tbl은 따로 저장

: 연도별로 split / binding한 검진DB, 생애전환기검진DB, 자격DB,
최종 study population에 포함되는 개체들의 PERSON_ID, cohort entry date로 구성된 tbl

1) Demographic covariate 정의할 때 자격DB에서 가져온 관측연도(STND_Y)

: cohort entry date의 연도와 동일한 연도의 관측값 가져옴

(이유 1). 성별은 변할 가능성이 드물기 때문에 어떤 년도의 값을 가져와도 상관없다고 판단.

(이유 2). 연령은 cohort entry date 시점 당시의 연령이 필요하기 때문에 자격DB의 연도가 cohort entry date의 연도와 같아야 한다고 생각함.

(이유 3). 사회경제적 정보가 연도에 따라 많이 변하지 않을 것 같다고 예상, 소득분위에 해당하는 변수값이 1년 차이에 따라 많이 변하지 않을 것이라고 판단.

2) coding 결과, CTRB_PT_TYPE_CD 변수(소득분위)가 결측인 개체는 없는 것을 확인

3) Family history 변수는 결측인 경우가 다수 존재

: 암 이력의 경우 ~2008년까지의 문진 항목만 이용해 확인 / 나머지 연도에 대해서는 결측으로 처리
: 총 141,212명의 Study population 중
- 세 개의 가족력 중 하나라도 missing 인 obs = 2490명
- 세 개의 가족력 모두 missing인 obs = 680명