

## 〈건강검진코호트DB 결과 정리〉

- 11월 16일 Version
- Covariate define part

### 〈TO DO LIST〉

- 1) 음주습관(DRNK\_FREQUENCY\_HABIT) covariate 정의
- 2) 운동습관(EXERCI\_FREQUENCY\_HABIT) / 운동지속시간(EXERCI\_METS\_minutes) covariate 정의

### 〈Share〉

- 1) 음주습관 관련 변수 재정의할 때 이전에 정한 방안대로 값 단위를 “daily”로 통일
  - 2) 음주습관 관련 covariate 값 중 결측치 존재  
: “DRNK\_FREQUENCY\_HABIT” 변수값이 결측인 객체 수 --> 5706명(약 4.04%)
  - 3) 운동습관(EXERCI\_HABIT) covariate 생성 결과 결측치가 존재  
: “EXERCI\_HABIT” 변수값이 결측인 객체 수 --> 2547명(약 1.8%)
- 
- 4) 운동지속시간(MET\_minutes) 변수를 정의할 때 관측시점 기준으로 운동습관 변수 정의한 tbl split함.  
(~08년 / 09년~15년 / 16~17년)
  - 5) 운동습관(EXERCI\_HABIT) covariate 관측 시점이 2008년 이전인 사람에 대한 “운동지속시간” 변수 정의  
--> 94,268명의 개체가 대상
- 5)-① : 2007년 ~ 2008년 생애전환기검진DB에는 운동지속시간(MET\_minutes) 변수를 정의하는데 이용하는 “MOV20\_WEK\_FREQ\_ID” 변수 , “MOV30\_WEK\_FREQ\_ID” 변수, “WLK30\_WEK\_FREQ\_ID” 변수가 존재함을 확인  
--- 따라서, 운동습관(“EXERCI\_HABIT” 변수)이 관측된 tbl과 ~08년도까지의 생애전환기검진DB와 먼저 joint 시행한 뒤, 운동지속시간(MET\_minutes) 변수 생성  
--> 이 과정을 통해 운동지속시간(MET\_minutes) 변수가 정의되는 객체 수는 6390명
- 5)-② : 5)-① 과정을 통해 운동지속시간(MET\_minutes)가 정의되는 개체 외 나머지 개체들을 가져와 09년 ~15년 검진DB / 생애전환기검진DB와 joint  
--> 가장 과거의 기록 가져와 운동지속시간(MET\_minutes) 변수 생성  
--> 이 과정을 통해 운동지속시간(MET\_minutes) 변수가 정의되는 객체 수는 5)-①에서 정의된 객체 제외 나머지 87,877명 중 75891명  
--- 11987명은 운동지속시간(MET\_minutes) 변수값이 결측

6) 운동습관 변수 관측시점이 16~17년인 경우 16~17년도에 해당하는 검진DB / 생애전환기검진DB와 joint 한 후 운동지속시간(MET\_minutes) 변수 정의하려고 시도

: 이때, 기록이 생애전환기검진DB, 검진DB 두 군데 중복되어 기록되어 있는 객체가 존재 (275명), 중복된 기록 값이 같기 때문에 unique value만 가져와 tbl 생성

7) 운동지속시간(MET\_minutes) 변수값이 결측인 객체 존재

--> 총 13,175명(약 9.33%)