

〈건강검진코호트DB - Meeting 자료〉

- 5월 30일 Version

〈What TO DO〉

- Covariate 정의
- Outcome 정의
- Table1 생성

〈Share & Result〉

1) Covariate 정의

- : Covariate 측정 timing은 “cohort entry date에 가장 근접한 과거 시점” 이다.
- : Outcome 별, 모집단 포함 관계 이용해 **Mortality outcome에 대응하는 모집단의 covariate를 먼저 정의**

■ Demographic factor : AGE / SEX / CTRB_PT_TYPE_CD(보험료 10분위 - 소득분위)

- 이전과 동일한 방식으로 진행
- **결측이 있는 obs 없음.**

■ 가족력 : Heart disease / Diabetes / Cancer family history

- 모든 연도에 걸쳐 한 번이라도 positive이면, 가족력 있는 것으로 간주 → **결측이 있는 obs 없음.**

	가족력이 존재하는 객체 수	% (Percent)
Heart Disease	8,277	14.36%
Diabetes	19,188	33.28%
Cancer	12,793	22.19%

■ Biomarker(수정된 부분)

: TOT_CHOLE / BLDS / SGOT_AST / SGPT_ALT / GAMMA_GTP / BP_HIGH / BP_LSWT / HMG

- **결측이 있는 obs가 존재함**을 확인.

	결측이 있는 객체 수	% (Percent)
BP_HIGH	81	0.14%
BP_LWST	81	0.14%
HMG	79	0.14%
TOT_CHOLE	4,129	7.16%
BLDS	73	0.13%
SGOT_AST	83	0.14%
SGPT_ALT	85	0.15%
GAMMA_GTP	73	0.13%

■ 음주 습관 : DRNK_HABIT_RSPS_CD

- 이전과 동일한 방식으로 진행, 결측 대체 과정 진행하지 않음.

	결측이 있는 객체 수	% (Percent)
DRINK_FREQ	7,424	12.88%

■ 운동 습관 : MOV20_WEK_FREQ_ID / MOV30_WEK_FREQ_ID / WLK30_WEK_FREQ_ID

- 이전과 동일하게 진행, 결측 대체 과정은 진행하지 않음.

■ 운동 지속 시간 : MOV20_WEK_FREQ_ID / MOV30_WEK_FREQ_ID / WLK30_WEK_FREQ_ID

- 이전과 동일한 방식으로 진행

- 먼저, 09~16년도로 MET minutes를 정의한 뒤, MET minutes가 정의되지 않는 객체 유형 파악

i) 09~16년도 사이 검진 기록이 **있으나**, but cohort entry date 전 검진 기록이 없는 객체 - **15,799명**

ii) 09~16년도 사이 검진 기록도 **없고**, cohort entry date 전 검진 기록도 없는 객체 - **920명**

→ 두 유형 모두 cohort entry date가 04~08년도

- i)의 유형에 대해서는 위 표준화 방안 이용해 MET minutes 정보 가져옴

- 최종적으로 생산한 tbl을 살펴보니, “운동 빈도” = 0인 obs 대상, 두 가지 유형의 obs가 존재함을 확인

iii) 운동 빈도 = 0 AND MET minutes is not missing

iv) 운동 빈도 = 0 AND MET minutes is missing

→ iii)이 발생한 이유는 위 i) 유형의 MET minutes를 09~16년도 기록을 이용해, 대체하였기 때문이다.

	결측이 있는 객체 수	% (Percent)
EXERCI_FREQ	7,375	12.79%
MET_MINUTE	8,301	14.4%

■ 흡연 습관 / 흡연 기간 / 흡연량 / Pack year

- 이전과 동일한 방식으로 진행, 결측 대체 과정은 진행하지 않음

- 흡연 습관 = “never smoker” 이면서 흡연량 및 흡연 기간이 결측인 관측치에 대해서는 모두 “0”으로 대체 진행함.

	결측이 있는 객체 수	% (Percent)
흡연 습관	1,287	2.23%
흡연 기간	2,592	4.5%
흡연량	4,303	7.47%
Pack year	4,311	7.48%

* “흡연 습관”이 missing이면, “흡연 기간”과 “흡연량” 모두 결측

* “**과거 흡연자**”의 경우, 흡연량 정보가 결측인 case가 대부분이다.

* “**현재 흡연자**”는 흡연 기간 정보가 결측인 case가 대부분이다.

■ 개인 과거력(Hypertension / MI / Heart failure / TIA or Stroke / Cancer)

- 개인 과거력은 문진 + T20 이용해 정의

- 이전과 다르게, 한 번이라도 과거력이 있으면 “과거력 존재”로 판단

- 결측은 없음

■ Comorbidities / Medication history

- 수정한 대로, 각 항목 수정해 comorbidities / 약물 처방 이력 모두 정의

- 결측은 없음

■ 각 Secondary outcome 별 covariate table 생성

- 먼저, mortality outcome 기준 생성한 covariate table 모두 MERGE 후

- outcome 별, 모집단에 속하는 객체의 정보만 GET + outcome에 대응하는 개인 과거력 변수 제외
→ 각 Outcome 별, 모집단에 속하는 객체의 모든 정보가 담겨 있음을 확인

2) Outcome 정의

: 모집단 별, Outcome 정의

- Primary outcome인 mortality는 cohort entry date 이후 자격 table 이용,
“사망 일자” 변수 결측 여부 이용
- Secondary outcome은 cohort entry date 이후의 T20 기록 이용, ICD-10 code 이용

	발생 N수(%)
Mortality	4,511(7.82%)
MI	997(1.75%)
Heart failure	3,346(6.63%)
TIA or Stroke	5,625(11.11%)
Cancer	5,331(9.75%)