

〈건강검진코호트DB - Meeting 자료〉

- 2월 22일 Version

〈What TO DO〉

- Covariate 정의

〈Share & Result〉

0) 새로 생성한 code 참고해 제외되는 객체 수가 늘어난 이유 파악

- “Cohort entry date와 Disease onset date와의 격차가 2년 초과인 경우 제외”
 - Cohort entry date가 정의되는 객체 수가 늘어남으로써 자연스럽게 따라오는 부분(8명 추가 제외됨)
- “특정 과거 질병 이력이 존재하는 객체 제외”
 - 이전 version(01.25) code에서 과거력 여부 tagging 시, main_sick(주상병)만 사용하였음
→ 현재 version(02.08) code에서는 sub_sick(제 1-부상병)까지 고려함.

1) Covariate 정의

: Covariate 측정 timing은 “cohort entry date에 가장 근접한 과거 시점” 이다.

- Demographic factor : AGE / SEX / CTRB_PT_TYPE_CD(보험료 10분위 - 소득분위)
 - 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	4,954	13.69%
Diabetes	12,416	34.31%
Cancer	8,025	22.18%

- Biomarker : TOT_CHOLE(총 콜레스테롤) / BLDS / SGOT_AST / SGPT_ALT / GAMMA_GTP
 - 결측이 있는 obs가 존재함을 확인.

	결측이 있는 객체 수	% (Percent)
TOT_CHOLE	10	0.03%
BLDS	0	0%
SGOT_AST	9	0.02%
SGPT_ALT	11	0.03%
GAMMA_GTP	0	0%

- 음주 습관 : DRNK_HABIT_RSPS_CD

(표준화 방안)

연도	표준화 방안
02~08년도 검진+생애전환기검진DB	value=1 then 0 value=2 then 2.5/30.5 value=3 then 1.5/7 value=5 then 1
09~16년도 검진+생애전환기검진DB	변수 값 / 7

- 음주 습관 관련 변수가 결측이면 “0”으로 대체 → 결측이 있는 obs는 없음.

- 운동 습관 : MOV20_WEK_FREQ_ID / MOV30_WEK_FREQ_ID / WLK30_WEK_FREQ_ID
(표준화 방안)

연도	표준화 방안
02~08년도 검진DB	변수 값 / 7
07~08년도 생애전환기검진DB + 09~16년도 검진+생애전환기검진DB	{(MOV20_WEK_FREQ_ID+MOV30_WEK_FREQ_ID+WLK30_WEK_FREQ_ID)/3}/7

- 운동 관련 변수들이 결측이면 미리 “0”으로 변환 → **결측이 있는 obs는 없음.**

- 운동 지속 시간 : MOV20_WEK_FREQ_ID / MOV30_WEK_FREQ_ID / WLK30_WEK_FREQ_ID
(표준화 방안)

연도	표준화 방안
Cohort entry date가 02~08년도 + 09~16년도 사이 검진 기록 부재	09~16년도 기록 중 가장 과거의 기록 이용
07~08년도 생애전환기검진DB + 09~16년도 검진+생애전환기검진DB	{(MOV20_WEK_FREQ_ID+MOV30_WEK_FREQ_ID+WLK30_WEK_FREQ_ID)/3}/7

- 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 - **13,272명**
 - ii) 09~16년도 사이 검진 기록도 **없고**, cohort entry date 전 검진 기록도 없는 객체 - **714명**
→ 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 = 300
 - iv) 운동 빈도 = 0 AND MET minutes is missing
→ iii)이 발생한 이유는 위 i) 유형의 MET minutes를 표준화 방안 기준, 대체하였기 때문이다.
∴ iii), iv)의 유형에 해당하는 obs에 대해, MET minutes를 모두 **“0”으로 대체**
- 결측 대체 과정을 거쳤음에도, 여전히 결측이 존재하는 obs가 존재 (더 이상 대체 방안이 없는 유형)

	결측이 있는 객체 수	% (Percent)
MET_minutes	10	0.03%