

〈건강검진코호트DB - Meeting 자료〉

- 3월 7일 Version

〈What TO DO〉

- Cohort entry date 및 모집단 재 정의
- Covariate 재 정의

〈Share & Result〉

0) Cohort entry date 및 모집단 재 정의

- 배경 : Comorbidities 정의 과정 중, “Cohort entry date 전 의료 이용 내역이 없는 객체 존재”, 이는 cohort entry date 정의와 맞지 않는 부분으로, 수정이 필요하다고 여겨짐.
- Disease onset date 이후, 검진 기록이 있는 객체(단, BMI와 BLDS가 not null) -- **85,872명**
 - Disease onset date 이후 검진 기록이 있는 객체 중, 첫 의료 이용 일자 이후의 검진 기록이 있는 객체 — **85,837명**
 - cohort entry date가 2016년 이전인 객체 -- **69,880명** (이전 ver : 69,933명)
→ cohort entry date 이전, 모든 객체 대상 진료 내역이 존재함을 확인
- 제외 조건 적용
 - Disease onset date와 cohort entry date 간 차이가 2년 초과 — **21,707명**
 - Cohort entry date 전, 특정 과거력 존재 — **19,215명**
→ 총 33,997명이 제외됨
∴ 최종 모집단 = 35,883명 (이전 ver : 36,187명)

1) Exposure 재 정의

: 정의 방안은 이전과 동일

GRP	BMI 기준	N수
Underweighted	<18.5	546
Moderate	18.5-22.9	9,089
Overweight	23-24.9	9,764
Obesity	25.0-29.9	14,663
2th-obesity	30-34.9	1,684
3th-obesity	>=35	137

2) Covariate 정의

: 정의 방안은 이전과 동일하나, 따로 결측치 대체 과정은 수행하지 않음.

- 가족력 : Heart disease / Diabetes / Cancer family history

	가족력이 존재하는 객체 수	% (Percent)
Heart Disease	4,918	13.71%
Diabetes	12,327	34.35%
Cancer	7,969	22.21%

- Biomarker : TOT_CHOLE(총 콜레스테롤) / BLDS / SGOT_AST / SGPT_ALT / GAMMA_GTP

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

- 음주 습관(빈도) : 이전과 다르게, missing인 관측치를 0으로 대체하지 않음

	결측이 있는 객체 수	% (Percent)
음주 습관(빈도)	476	1.33%

- 운동 빈도

- Raw-data 확인 결과, 운동 빈도 관련 변수 값이 non-missing이면서, 0인 관측치는 없었음.
- 이 또한, missing인 관측치를 0으로 대체하지 않음.

- 운동 지속 시간(MET minutes)

	결측이 있는 객체 수	% (Percent)
운동 빈도	462	1.29%
운동 지속 시간	1,120	3.12%

- 흡연 습관 / 흡연 기간 / 흡연량 / Pack year

- 이전에 생성한 표준화 방안 이용해 변수 정의하였음.
- missing category는 따로 대체 과정 진행 안 함.
 - 흡연 습관 = “*never smoker*” 이면서 흡연량 및 흡연 기간이 결측인 관측치에 대해서는 모두 “0”으로 대체 진행함.

	결측이 있는 객체 수	% (Percent)
흡연 습관	1,054	2.94%
흡연 기간	1,055	2.94%
흡연량	2,500	6.97%
Pack year	2,501	6.97%

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

* “**과거 흡연자**” 대상, “흡연량”이 missing이면서 “흡연 기간”도 missing인 관측치는 없음

- 흡연량이 missing + 흡연 기간은 non-missing : 1,446명

- 흡연량이 non-missing + 흡연 기간은 missing : 1명

- 개인 과거력 / Comorbidites / 약물 처방 이력
 - 세 covariate에 대해, **결측이 있는 관측치는 없음.**