

〈건강검진코호트DB 질문 정리〉

- 10월 28일 Version

- Covariate 관련

Q1) 가족력도 cohort entry date 이전의 기록 중 가장 최근의 기록만 가지고 파악?

-- cohort entry date 이후에 가족력이 발생하는 경우는?

Q2) Missing 대체 - logical imputation remind 필요

Q3) 건강검진DB 관련 covariate

~2008년 : 흡연상태 / 흡연기간, 하루 흡연량(결측이 많음) / 음주습관, 운동빈도(결측이 많음)

-- pack year 계산?

2009년 ~ 2017년 : 과거 / 현재 흡연 기간(결측이 많음), 하루 흡연량(결측이 많음) / 과거 병력, 가족력(NA가 존재)

Q4) 소득분위 변수(CTRB_PT_TYPE_CD) : 2015년까지는 결측이 없으나, 2016년 ~ 2019년 사이 결측인 사람이 존재 --> 이전 값으로 대체?

--> Covariate 결측 처리 방법 구체적 제시 필요