

〈건강검진코호트DB Meeting 내용 압축 정리〉

- 3월 24일 ~ 4월 3일 Version

- 작성자 : 이은경

〈주요 안건〉

- ① Comorbidities(병력), Medication(약물 처방 이력) 변수 간 2x2 분할표 생성해 다중 공선성 의심되는 변수 조합 찾아보기.
- ② Coxph model에 사용되는 모든 공변량 간 상관계수 살펴보기.
- ③ ② 결과를 토대로 Coxph model 다시 적합해 보고 회귀계수 살펴보기.

① 관련 파일은 “Binary_Contingency_table.xlsx” file입니다.

[Sheet 설명]

- : 첫 번째 sheet는 Comorbidities 변수 간 분할표 생성한 결과
- : 두 번째 sheet는 Comorbidities 변수, Medication 변수 간 분할표 생성한 결과
- : 세 번째 sheet는 Medication 변수 간 분할표 결과

①-1). “Hypertension(고혈압) 발병” 과 “GI_Disorders” 발병

: Hypertension 병력이 있으면서 GI_Disorders 이력이 있는 사람이 85,752명 (50.94%)

①-2). “CNS(Central Nerve System)” 약물 처방 이력과 “Hypertension(고혈압)” 발병

: Hypertension 병력이 있으면서 CNS 약물 처방 이력이 있는 사람이 81,439명(48.83%)

①-3). “GI_Disorders” 발병과 “Non_opioid” 약물 처방 이력

: GI_Disorders 병력이 있으면서 Non_opioid 약물 처방 이력이 있는 사람은 90,203명 (53.58%)

①-4). “Hypertension” 발병과 “Non_opioid” 약물 처방 이력

: 고혈압 병력이 있으면서 Non_opioid 약물 처방 이력이 있는 사람은 114,314명 (67.91%)

①-5). “CNS(Central Nerve System)” 약물 처방 이력과 “Non_opioid” 약물 처방 이력

: CNS 약물 처방 이력이 있으면서 Non_opioid 약물 처방 이력이 있는 사람은 86,136명 (51.17%)

② 관련 파일은 “Correlation_table.xlsx” file입니다.

: Coxph model 적합 때 사용하는 모든 공변량 간 상관계수를 측정하여 정리한 파일입니다.

: 변수 간 상관계수 대부분이 0.1 미만의 매우 적은 값을 보였다.

(눈에 띄는 상관계수)

- ① “SEX” / “SMK_STAT_TYPE_RSPS_CD”(흡연상태) : -0.545 (“남성”이 현재 흡연자인 경우가 많다.)
- ② “SMK_STAT_TYPE_RSPS_CD”(흡연상태) / “PAST_SMK_TERM”(과거 흡연기간) : 0.969
- ③ “statin” / “Hyperlipidemia” : 0.70531
- ④ “SGOT_AST” / “SGPT_ALT” : 0.694

③ 관련 파일은 “Coxph_Version2.xlsx” file입니다.

: ② 결과를 반영하여 Coxph model에 사용하는 공변량의 투입 여부를 조절하였습니다.

③-1). “statin” 변수와 “Hyperlipidemia” 변수의 상관계수가 0.7 이상이므로, “statin” 약물 처방 이력 변수를 제외해 Coxph model 적합한 다음, “Hyperlipidemia” 변수의 회귀계수 다시 살펴보았음.

- “statin” 약물 처방 이력 변수를 제외해도, “Hyperlipidemia” 변수 회귀계수의 부호는 여전히 음수이다.

③-2). “SMK_STAT_TYPE_RSPS_CD”(흡연상태)가 “현재 흡연자”일 때의 HR이(Baseline은 “Never Smoker”) 1만 이상이었다.

↳ “SEX”, “PAST_SMK_TERM”, “PAST_PACK_YEAR”, “PAST_DSQTY”, “CUR_SMK_TERM”, “CUR_DSQTY”, “CUR_PACK_YEAR” 변수를 모두 제외하고 Coxph model 적합한 뒤 “흡연상태”의 회귀계수 다시 살펴봄.

- “Never Smoker”에 비해 “Past Smoker”의 HR이 1.388배 높다.
- “Never Smoker”에 비해 “Current Smoker”의 HR이 1.818배 높다.
- 위의 두 결과 모두 유의수준 0.05하에서 유의하다고 판단됨.

⇒ SEX 변수는 중요한 공변량 중 하나이므로, 설명변수로 사용하는 것은 불가피하다고 판단함.

↳ “PAST_SMK_TERM”, “PAST_PACK_YEAR”, “PAST_DSQTY”, “CUR_SMK_TERM”, “CUR_DSQTY”, “CUR_PACK_YEAR” 변수를 모두 제외하고 Coxph model 적합한 뒤 “흡연상태”의 회귀계수 다시 살펴봄. (⇔ “SEX” 변수 다시 input함)

- “Never Smoker”에 비해 “Past Smoker”의 HR이 1.108배 높다.
- “Never Smoker”에 비해 “Current Smoker”의 HR이 1.487배 높다.
- 위의 두 결과 모두 유의수준 0.05하에서 유의하다고 판단됨.

--- 해당 결과를 excel file에 정리하였습니다.