

# 20230620\_전체회의\_내용정리

👤 담당자	은경 이
☀ 상태	완료
📅 마감일	@2023년 6월 20일
🏷 태그	Meeting comment

## What TO DO)

: 6월 20일 전체 회의 Comment 정리 + TO DO LIST 정리

## Meeting Comment)

: 중간보고(7/17)을 위한 model 적합 + 고려하는 model 후보 설정 +  
Schedule 공유

### (공통사항)

- Study population을 5개년도 / 성별로 split
- 고려하는 Outcome은 폐암 / 백혈병 발병 - 폐암을 먼저 파악
- 가능한 Outcome 발생 경우
  - 근무 도중 암 발병
  - 휴직기(직전 회사 고용보험 상실일 ~ 다음 회사 고용보험 상실일)에 암 발병 : 휴직기도 하나의 row로 분석용 데이터에 input 필요
  - 마지막 회사 고용보험 상실일 이후 암 발병

### (Model 후보)

#### 1) Model1 (Baseline)

: 적합한 공변량이 오직 "entry" 변수 (Dummy variable로 time-invariant covariate - 95~99년도 입사가 Baseline factor)

(Data example)

INDI_ID	Episode	Start	Stop	Event	entry
1	1	직장_10	직장_11	0	95~99
1	2	직장_11	직장_20	0	95~99
1	3	직장_20	직장_21	0	95~99
1	4	직장_21	직장_30	0	95~99
1	5	직장_30	직장_31	0	95~99
1	6	직장_31	Event 발생일 혹은 Censoring date	0 or 1	95~99

→ (직장\_10, 직장\_11)은 (첫 번째 직장 고용보험 취득일, 첫 번째 직장 고용보험 상실일)을 의미하며, (직장\_11, 직장\_20)은 두 번째 직장 고용보험 취득일 이전 기간인 "휴직기간"을 의미한다.

— 나머지 row도 의미 동일함.

## 2) Model2

: 적합하는 공변량이 “entry” 변수 + 암 이력(폐암 또는 백혈병 발병 이전 다른 암 발생 여부를 뜻하며, time-varying covariate로 정의)

(Data example)

INDI_ID	Episode	Start	Stop	Event	entry	Cancer
1	1	직장_10	직장_11	0	95~99	0
1	2	직장_11	직장_20	0	95~99	0
1	3	직장_20	직장_21	0	95~99	0
1	4	직장_21	직장_30	0	95~99	0
1	4	직장_30	암_1	0	95~99	0
1	5	암_1	직장_31	0	95~99	1
1	6	직장_31	Event 발생일 혹은 Censoring date	0 or 1	95~99	1

→ 암 이력을 시간-가변 공변량으로 정의하므로, 암 이력이 존재하는 경우 [Start, Stop) 형태를 위 예시처럼 더 세분화할 필요가 있음.

## 3) Model3

: 적합하는 공변량이 “entry” 변수 + 사업장 / 직종 정보

→ Model1의 Upgrade version

## 4) Model4

: 적합하는 공변량이 “entry” 변수 + 암 이력 + 사업장 / 직종 정보

→ Model2의 Upgrade version

## 5) Another Model1 (가장 먼저 적합하는 Model!)

: 위의 1) ~ 4)는 시간 상 오래 걸리는 데이터 형태이므로 보다 더 시간적으로 효율적인 model 설립

: 적합하는 공변량은 “entry” 변수 + time-invariant 암 이력 변수(즉, at least once 암 발병하면 1 아니면 0의 값을 가지는 이변량 변수)

→ 해당 경우에는 2)에서의 Data Example 처럼 암 발병 시점으로 구분해 row를 나눌 필요 없음

(Data Example)

INDI_ID	Episode	Start	Stop	Event	entry	Cancer
1	1	직장_10	직장_11	0	95~99	1
1	2	직장_11	직장_20	0	95~99	1
1	3	직장_20	직장_21	0	95~99	1
1	4	직장_21	직장_30	0	95~99	1
1	5	직장_30	직장_31	0	95~99	1
1	6	직장_31	Event 발생일 혹은 Censoring date	0 or 1	95~99	1

## (결과 보고 방식)

: 기본적으로 Excel file에 정리

: 출생연도 / 성별로 split한 데이터 별로

- 공변량에 대응하는 fitted coefficients + S.E + P-value
- C-index 성능지표 값

: model 적합 때, train set / test set 따로 구분하지 않음. 즉, 실제 y값과 예측값 y 비교

: “prog phreg”에서 “concordance” option 추가하면 구할 수 있을 거라 예상함.

---

## TO DO LIST

- 6월 27일 까지 5) model 적합 + 결과 파일 생성
  - Demographic variable(성별, 출생연도, entry 변수) 정의한 뒤 바로 Outcome 정의
  - Outcome 정의(폐암 / 백혈병 발병 여부)한 후 암 이력 변수 정의