

〈산업 안전 보건 연구원 - Meeting 자료〉

- 6월 22일 Version

- 작성자 : 이은경

〈What TO DO〉

: 이전에 정의한 폐암 혹은 백혈병 발병 여부 + 진단 일자 / 사망 여부 + 사망 시점 이용해 counting process format data 생성 위한 데이터 재정비 — **현시점 고려하는 Outcome은 폐암**

〈Share & Result〉

1) 이전에 정의한 폐암 혹은 백혈병 발병 여부 + 진단 일자(lung_cancer / leukemia / lung_cancer_date / leukemia_date) 변수 이용해 폐암 대상, “Outcome_Category” 변수, “Outcome_date” 변수 새로 정의

1)-1. “Outcome_category” 변수

: 폐암 발병한 사람은 “lung_cancer”, 폐암 발병 안 함 + 사망한 사람은 “Death” + 어떤 사건도 발생하지 않은 사람은 “last_follow” 값을 가지는 범주형 변수 생성

1)-2. “Outcome_date” 변수

: “Outcome_category” 변수 기준으로 Outcome 발생 시점 뜻함.

(“last_follow”인 경우 ‘20181231’로 설정)

1)-3. “between_yes” 변수

: “Outcome_date”가 각 반복 측정치의 (취득일, 상실일) 사이에 속하는지의 여부 뜻함.

(생성한 후 확인한 데이터 일부)

연번	취득일	상실일	INDI_ID	성별	BYEAR	Entry	Cancer_History	Outcome_Category	Outcome_date	between_yes
5546155	20180102	20190321	1000008126694	여	1962	4	0	last_follow	20181231	1
1028146	20030215	20031201	1000008126695	여	1962	1	0	last_follow	20181231	0
979550	20031201	20040901	1000008126695	여	1962	1	0	last_follow	20181231	0
2855780	20040901	20050322	1000008126695	여	1962	1	0	last_follow	20181231	0
3401889	20090917	20091201	1000008126695	여	1962	1	0	last_follow	20181231	0
5372977	20100601	20100711	1000008126695	여	1962	1	0	last_follow	20181231	0
1299207	20100825	20151201	1000008126695	여	1962	1	0	last_follow	20181231	0
3953786	20051020	20051115	1000008126696	여	1962	2	0	lung_cancer	20140602	0
5557887	20100914	20101120	1000008126696	여	1962	2	0	lung_cancer	20140602	0
6430125	20151007	20151119	1000008126696	여	1962	2	0	lung_cancer	20140602	0
4140056	20110309	20120331	1000008126698	여	1962	3	0	last_follow	20181231	0
2850558	20120402	20181231	1000008126698	여	1962	3	0	last_follow	20181231	0

[이상 Case 정리]

① 마지막 반복 측정치의 상실일이 “2018.12.31.” 이후 + 폐암 혹은 사망 사건이 발생하지 않음(last_follow)

연번	취득일	상실일	INDI_ID	성별	BYEAR	Entry	Cancer_History	Outcome_Category	Outcome_date
5546155	20180102	20190321	1000008126694	여	1962	4	0	last_follow	20181231

↳ 2018년 이후의 상실일은 전부 2018.12.31.로 변경

② 같은 INDI_ID + 고용보험 취득일 동일 + 고용보험 상실일이 다른 반복 측정치 존재

1000000757060	20170101	20190101	last_follow	20181231
1000000757060	20170101	20200101	last_follow	20181231

↳ 상실일을 모두 2018.12.31.로 변경하고, 행 하나만 가지고 오기.

③ 사망일보다 상실일이 더 미래시점인 측정치 존재

: 행정 상의 문제로 의심됨.

1000000757235	19950701	19960913	Death	19960912
1000000757236	19950701	19971014	Death	19971013