

산보연 Data check 6월 20일 ver

작성자 : 이은경

1. 사업장 별 "통합 누적 발생률"과 대조군의 "통합 누적 발생률" 비교

[대조군 생성 계획]

: UP2 상관없이, SEX, CAL, ECNY_AGE, YEAR가 똑같은 집단에 대한 백혈병 발생 건수, 추적인년합계 변수 aggregate
-> 이를 UP2, SEX, CAL, ECNY_AGE, YEAR에 따라 grouping한 데이터와 연결

① grouping key ; UP2, SEX, CAL, ECNY-AGE, YEAR

UP2	SEX	CAL	ECNY-AGE	YEAR	통합누적발생률 (파란색)	Sum -FY-BZ	Sum -leukemia-BZ

② 대조군 ; grouping key는 ①의 grouping key에서 "UP2" 제외 전부!

SEX	CAL	ECNY-AGE	YEAR	Sum -FY-BZ-2	Sum -leukemia 2	Weighted 통합누적발생률 $= W_i \times \frac{\text{Sum-leukemia 2}}{\text{Sum-FY-BZ 2}}$ (이때 $W_i = \frac{\text{Sum-FY-BZ}}{\text{Sum-FY-BZ-2}}$)
						병합 후 생성하는 변수

⇒ ① tibble 라 ② tibble 을 SEX, CAL, ECNY-AGE, YEAR 변수 기준으로 병합, 그리고

"Weighted 통합누적발생률" 변수 생성

⇒ 위의 계획처럼, 대조군 생성 후에 통합 누적 발생률 계산. 각 사업장에 대한 통합 누적 발생률("origin_rate" 변수명)과 대조군의 통합 누적 발생률("weight_rate" 변수명) 겹쳐서 시각화

----- 다음 페이지로 -----

① y축의 scale을 조정하지 않은 경우

: 이때, 그래프의 순서는 2018년 기준 추적인년합계의 순위(오름차순 정렬)

: 빨간색이 사업장 별 통합누적발생률, 검은색이 대조군의 통합누적발생률 의미



⇒ 위 그래프에서 두 통합 누적 발생률의 차이가 두드러지는 부분은 빨간색 테두리가 있는 부분이다. 경향을 보면, 추적 인년 합계가 적은 사업장일수록 대조군과의 차이가 두드러져 보이는 것을 알 수 있다. 해당 사업장명은 다음과 같다.

UP2	사업장명
96	개인 사업
90	예술, 스포츠업, 건설업 등 각종 사업
37	하수, 폐기물 정화업 (하수, 폐수, 분뇨)
5	광업(석탄, 원유, 채굴업)
12	제조업(담배)
36	용수 공급업
2	임업
39	환경 정화업
6	광업(금속, 철, 비금속)
97	가구내 고용 활동
8	광업(광업, 자원, 원유)
98	가사 생산 활동

② y축의 scale을 조정한 경우

: 이때, 그래프의 순서는 2018년 기준 추적인년합계의 순위(오름차순 정렬)

: 빨간색이 사업장 별 통합누적발생률, 검은색이 대조군의 통합누적발생률 의미



⇒ 위 그래프에도 두 통합 누적 발생률의 차이가 두드러지는 부분을 빨간색으로 표시했다. 여기서도 경향을 보면, 추적 인년 합계가 적은 사업장일수록 대조군과의 차이가 두드러져 보이는 것을 알 수 있다. 해당 사업장명은 다음과 같다.

UP2	사업장명
96	개인 사업
73	수의업, 디자인업
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)
5	광업(석탄, 원유, 채굴업)
12	제조업(담배)
36	용수 공급업
2	임업
39	환경 정화업
6	광업(금속, 철, 비금속)
97	가구내 고용 활동
8	광업(광업, 자원, 원유)
98	가사 생산 활동

: 이때, UP2가 76, 34인 경우 “업종코드.csv” file에 명시되어 있지 않은 업종 코드였다.

2. 각 사업장 별 통합 누적 발생률과 대조군의 통합 누적 발생률 “차이”(Relative Risk)가 가장 큰 TOP10 사업장은?

UP2	사업장명	발생률 차이
8	광업(광업, 자원, 원유)	0.0000392
97	가구 내 고용활동	0.0000383
12	제조업(담배)	0.0000287
98	가사 생산 활동	0.0000265
3	어업	0.0000216
90	예술, 스포츠업, 건설업 등 각종 사업	0.0000167
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)	0.0000144
5	광업(석탄, 원유, 채굴업)	0.0000142
99	국제 및 외국기관	0.0000120
42	건설업(건물 구축 관련)	0.0000119

3. 각 사업장 별 통합 누적 발생률과 대조군의 통합 누적 발생률의 “비율”(간접 SIR)가 가장 큰 TOP10 사업장은?

UP2	사업장명	간접 SIR
97	가구 내 고용활동	2.07
90	예술, 스포츠업, 건설업 등 각종 사업	1.56
12	제조업(담배)	1.46
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)	1.32
76	(명시되어 있지 않은 업종 코드)	1.30
19	제조업(석유 정제품, 코크스 관련)	1.28
5	광업(석탄, 원유, 채굴업)	1.22
15	가죽 관련 제조업	1.19
7	광업(비금속 광물, 소금 등)	1.15
16	제조업(목재 관련)	1.15

4. 백혈병 통합 누적 발생률 TOP 10인 사업장은? (2018년 기준)

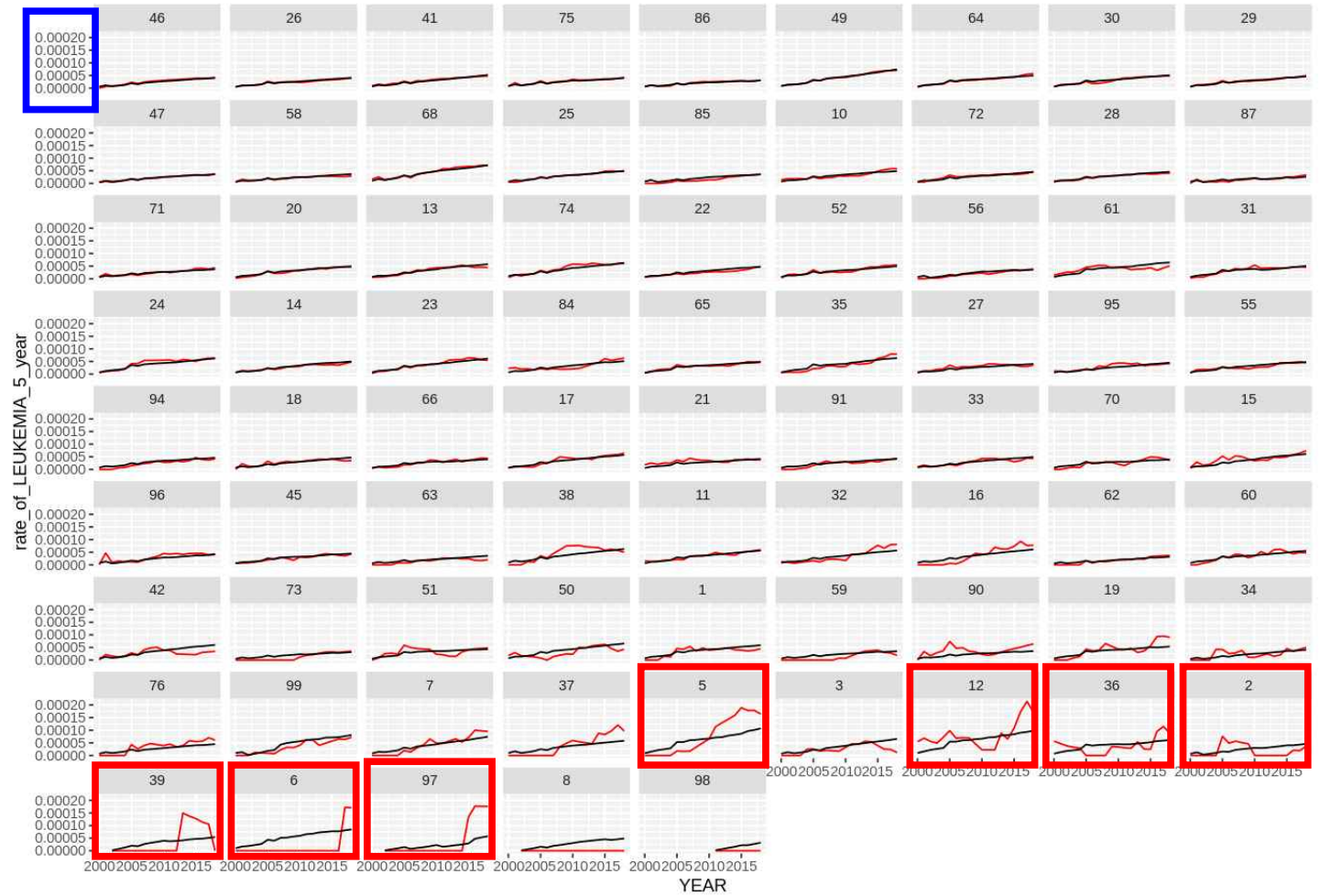
UP2	사업장명	통합 누적 발생률
12	제조업(담배)	0.0000915
5	광업(석탄, 원유, 채굴업)	0.0000786
97	가구나 고용활동	0.0000743
7	광업(비금속 광물, 소금 등)	0.0000593
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)	0.0000591
68	부동산업	0.0000557
38	하수, 폐기물 정화업(방사성 물질, 폐기물, 건설 폐기물)	0.0000531
74	사업 건물 관리, 조경 관리 서비스업	0.0000523
19	제조업(석유 정제품, 코크스 관련)	0.0000519
16	제조업(유리, 목재 등)	0.0000517

5. 각 사업장 별로, 대조군 별로 5년 단위 통합 누적 발생률 계산 후 비교

: 계산 후, 추적 인년 합계 기준으로 오름차순 정렬 뒤, 시각화(사업장 별 5년 단위 통합 누적 발생률과 대응하는 대조군의 5년 단위 통합 누적 발생률 겹쳐서 그림)

: **빨간색이 사업장 별** 5년 단위 통합 누적 발생률, **검은색이 대조군의** 5년 단위 통합 누적 발생률 의미

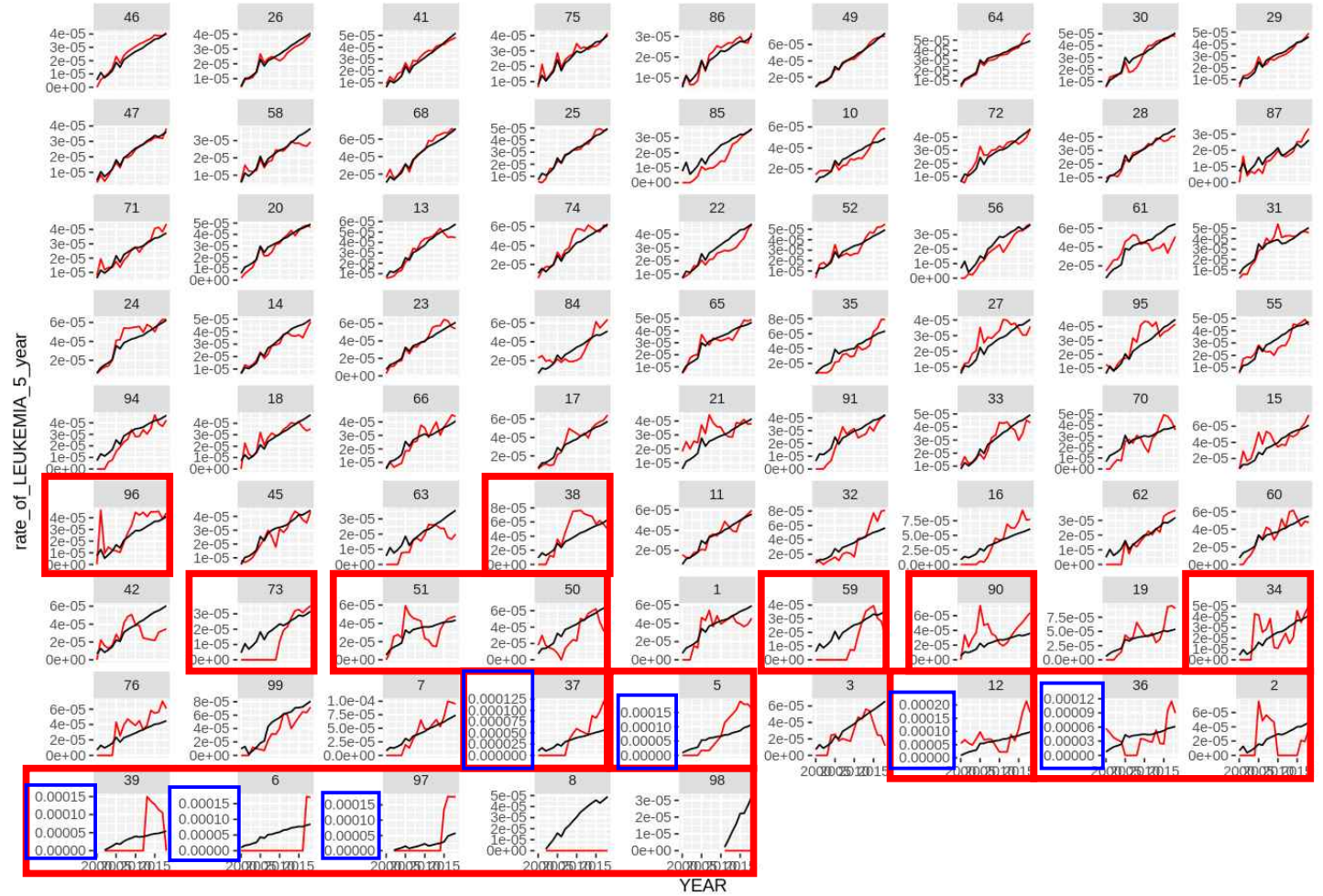
① y축의 scale을 조정하지 않은 경우 : y축의 scale이 1번 버전 그래프(year 단위 고려하지 않음)와 다름.



⇒ 위 그래프에서 두 통합 누적 발생률의 차이가 두드러져 보이는 그래프에 빨간색 사각형으로 표시하였다. 여기서도 경향을 보면, 추적 인년 합계가 적은 사업장일수록 대조군과의 차이가 두드러져 보이는 것을 알 수 있다. 해당 사업장명은 다음과 같다.

UP2	사업장명
5	광업(석탄, 원유, 채굴업)
12	제조업(담배)
36	용수 공급업
2	임업
39	환경 정화업
6	광업(금속, 철, 비금속)
97	가구내 고용 활동

② y축의 scale을 조정한 경우



⇒ 위 그래프에서 두 통합 누적 발생률의 차이가 두드러져 보이는 그래프에 빨간색 사각형으로 표시하였고, y축의 scale이 연도 단위를 고려하지 않았을 때와 비교해서 커진 경우에는 파란색 사각형으로 표시하였다. 여기서도 경향을 보면, 추적 인 년 합계가 적은 사업장일수록 대조군과의 차이가 두드러져 보이는 것을 알 수 있다. 해당 사업장명은 다음과 같다.

UP2	사업장명
96	개인 사업
38	하수, 폐기물 정화업(방사성 물질, 폐기물, 건설 폐기물)
73	수의업, 디자인업
51	항공 운수업
50	수상 운수업
59	영화, 비디오, 오디오물 출판, 상영업
90	예술, 스포츠업, 건설업 등 각종 사업
34	(명시되어 있지 않은 업종 코드)
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)
5	광업(석탄, 원유, 채굴업)
12	제조업(담배)
36	용수 공급업
2	임업
39	환경 정화업
6	광업(금속, 철, 비금속)
97	가구 내 고용 활동
8	광업(광업, 자원, 원유)
98	가사 생산 활동

6. 각 사업장 별 5년 단위 통합 누적 발생률과 대조군의 5년 단위 통합 누적 발생률의 “비율”(간접 SIR)가 가장 큰 TOP10 사업장은?

UP2	YEAR	사업장명	SIR
36	2000	용수 공급업	7.86
12	2000	제조업(담배)	5.63
5	2005	광업(석탄, 원유, 채굴업)	4.81
97	2015	가구 내 고용 활동	4.60
12	2001	제조업(담배)	4.26
34	2004	(명시되어 있지 않은 업종 코드)	4.00
97	2016	가구 내 고용 활동	3.70
36	2001	용수 공급업	3.66
39	2013	환경 정화업	3.60
96	2001	개인 사업	3.60

: 파란색으로 표시한 부분은 3번 결과(연도 별로 나누지 않고 계산한 간접 SIR)와 겹치는 사업장 의미

7. 백혈병 5년 단위 통합 누적 발생률 TOP 10인 사업장은? (연도 고려하지 않음.)

UP2	YEAR	사업장명	5년 단위 통합 누적 발생률
12	2017	제조업(담배)	0.000214
5	2015	광업(석탄, 원유, 채굴업)	0.000189
5	2017	광업(석탄, 원유, 채굴업)	0.000179
5	2016	광업(석탄, 원유, 채굴업)	0.000178
97	2016	가구 내 고용 활동	0.000177
97	2017	가구 내 고용 활동	0.000176
97	2018	가구 내 고용 활동	0.000176
12	2016	제조업(담배)	0.000173
6	2017	광업(금속, 철, 비금속)	0.000172
12	2018	제조업(담배)	0.000172

: 파란색으로 표시한 부분은 4번 결과(2018년 기준 통합 누적 발생률)와 겹치는 사업장 의미

8. 2018년 기준 백혈병 5년 단위 통합 누적 발생률 TOP10은?

UP2	사업장명	2018년 기준 통합 누적 발생률
97	가구 내 고용 활동	0.000176
12	제조업(담배)	0.000172
6	광업(금속, 철, 비금속)	0.000170
5	광업(석탄, 원유, 채굴업)	0.000164
37	하수, 폐기물 정화업(하수, 폐수, 분뇨)	0.0000968
7	광업(비금속 광물, 소금)	0.0000950
19	제조업(석유 정제품, 코크스 관련)	0.0000905
36	용수 공급업	0.0000892
32	제조업(금속 가구, 목재 가구)	0.0000809
34	전기, 가스, 증기 및 수도사업	0.0000790

: 파란색으로 표시한 부분은 4번 결과(2018년 기준 통합 누적 발생률)와 겹치는 사업장 의미