

## 근로자 집단 암 발병 예측 모형 시범 개발

**연구기간** 2022년 05월 ~ 2022년 12월

**핵심 단어** 발병률, 추계, 백혈병, 폐암, 사업장기반 암 감시자료

**Highlights** 사업장 기반 질환 감시 자료를 활용하여, 백혈병과 폐암을 대상으로 업종별 특성을 반영한 질환발병률 추계 통계모형을 시범적으로 설계 분석한 결과 일부 규모가 큰 업종에 대하여는 통계적인 기법 기반의 단기추계(3-5년)가 가능함을 확인함. 그러나 사업장 단위의 자료원의 한계점으로 암 발병의 가장 큰 위험요인인 나이가 반영되지 못한 결과라는 점에서 향후 사업장 기반 자료 구축 시 추계모형 설계 활용을 고려한 근로자 집단의 나이 정보 수집 방안을 제시하였음.

### 1

#### 연구배경

##### ○ 질환발병 추계 모형 시범 설계 및 분석

- 암 발생 추계 모형은 국가별 또는 대규모 집단별 추계 연구에서 간접적으로 수행되어 오거나 특정 암을 중심으로 제시되어 왔음. 각 층(예-성별 등)의 범주화된 나이와 시간(calendar time)의 모든 조합에서 암 발생 또는 사망 자료가 주어지는 경우 일반화 선형 모형 또는 age-period-cohort(APC) 모형을 이용하였으나 나이와 시간에 대한 정보가 다른 형태로 주어지는 경우 이러한 분석방법을 직접적으로 적용하기에는 한계가 있음.
- 2021년에 산업안전보건연구원에서 수행한 빅데이터를 활용한 직업성질환 코호트운영(1)을 통해 수집된 사업장 기반 질환 감시 자료를 활용하여, 시범적으로 백혈병과 폐암을 대상으로 업종별 특성을 반영한 질환발병률 추계 통계모형을 설계/제안하고자 함.

### 2

#### 주요연구내용

##### 연구결과

##### ○ 예측모형의 타당성 평가

- 이번 조사에서는 2015-2018년을 이용하여 정량적인 평가를 실시하는 방안을 고려하였으며, 타당성검토 지표로 훈련오류와 검증오류를 고려하여 평가함.
- 훈련오류 (training error, in-sample error)는 2015-2018년 발병자수를 전체(-2018년분)로부터 예측(적합)하였을 때, 예측값의 실제값대비 오차를 뜻하며, 검증오

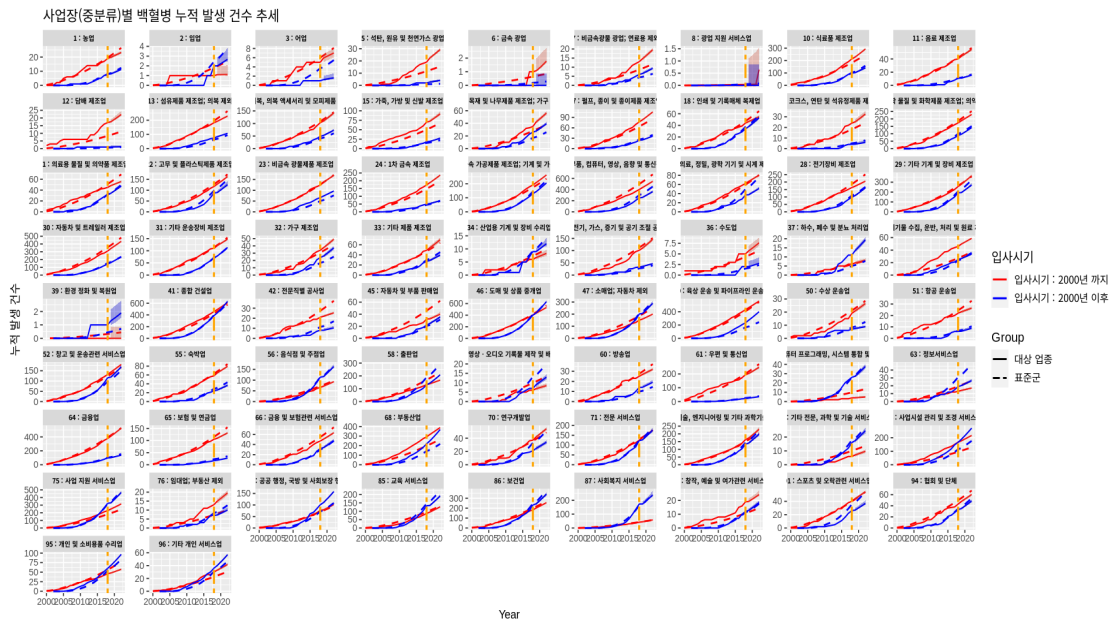
류 (validation error, out-sample error)는 2015-2018년 발병자수를 과거(-2014년분)로부터 예측(추계)하였을 때, 예측값의 실제값대비 오차를 의미함. 두 오류에 대하여 오차의 지표는 제곱근평균오차비율(root-mean-squared error ratio, RMSE ratio) 이용하였음.

$$\text{제곱근평균오차비율} = \frac{\sqrt{\frac{1}{4} \sum_{t=2015}^{2018} (\tilde{y}_t - y_t)^2}}{\sqrt{\frac{1}{4} \sum_{t=2015}^{2018} y_t^2}}$$

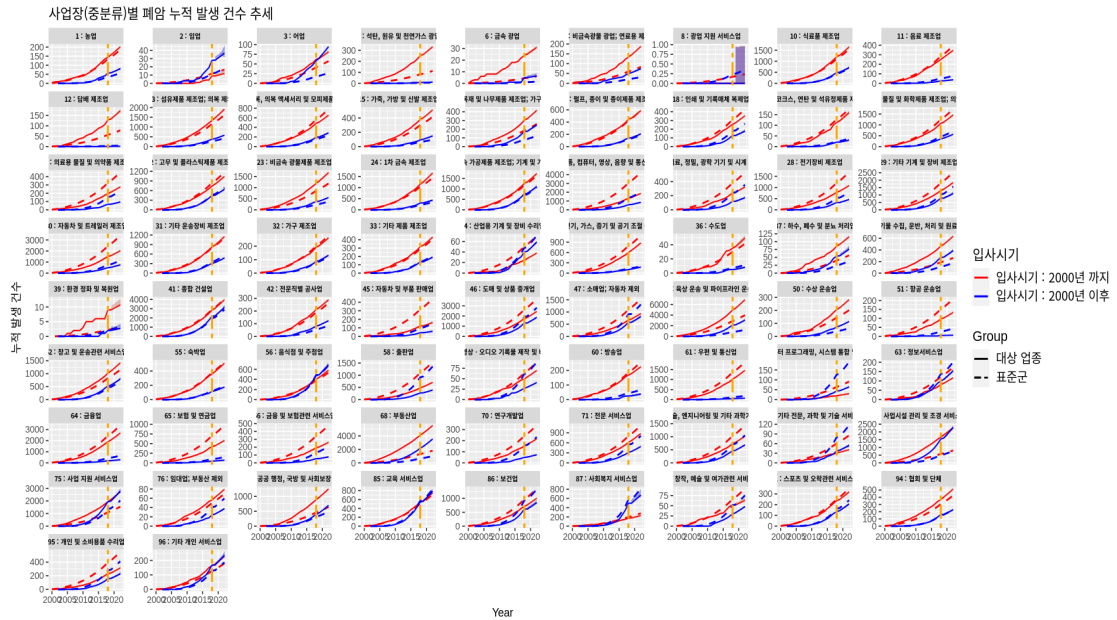
$y_t$ :  $t$ 년도에서의 누적 발병건수 관찰값

$\tilde{y}_t$ : 제안된 방법 (모형적합, 앙상블, 계층시계열 조정)을 거친  $y_t$ 의 예측값

- 타당성검토를 통해 사업장 기반 질환 감시 자료를 활용한 발생률 추계는 일정 수의 발생건수를 보인 업종에 대하여는 통계적인 시계열추계 기법 기반의 단기추계(3-5년)값은 안정적으로 얻을 수 있음을 확인하였음. 이에 2018년까지 수집된 사업장기반 감시자료를 활용하여 2022년의 추계 값 및 발생률이 높은 고위험 업종을 중분류 수준에서 시범 분석하였음. (그림1, 그림2)



〈그림 2〉 중분류 업종수준에 따른 백혈병 누적 발생 건수 추계



〈그림 3〉 중분류 업종수준에 따른 폐암 누적 발생 건수 추세

○ 2022년 추계값과 참조집단(전체업종 합계) 발병률과 비교하여 발병비율이 가장 높은 10개 중분류 업종을 위험집단으로 삼아 입사시기(2000년 및 그 이전, 2000년 이후)와 성별에 따라 제시하였음.

〈표 5〉 2022년 추계값 성별 표준화발병비 기준 상위 10개 백혈병 및 폐암의 고위험 업종

암	입사시기	업종 코드	업종명	1995-2018년 관찰값		2019-2022 예측값	
				발병자	발병비*	발병자	발병비*
백혈병	2000년 및 그 이전 입사자	12	담배 제조업	17	2.1	5	1.7
		76	임대업 (부동산제외)	13	1.9	7	2.3
		05	석탄, 원유 및 천연가스 광업	19	1.6	10	3.3
		07	비금속 광물 광업 (연료용제외)	13	1.4	7	2.3
		74	사업시설 관리 및 조정 서비스업	168	1.8	49	1.5
		90	창작, 예술 및 여가관련서비스업	19	1.7	5	1.7
		38	폐기물 수집, 운반, 처리 및 원료재생업	46	1.7	13	1.3
		16	목재 및 나무제품 제조업(가구제외)	35	1.3	21	2.1
		68	부동산업	292	1.5	97	1.4
		15	가죽, 가방 및 신발제조업	70	1.4	22	1.2
	2000년 이후 입사자	37	하수, 폐수 및 분뇨처리업	11	2.2	8	4.0
		68	부동산업	219	1.5	149	2.0
		49	육상 운송 및 파이프라인 운송업	234	1.4	166	2.1
		74	사업시설 관리 및 조정 서비스업	166	1.4	105	1.7
		16	목재 및 나무제품제조업 (가구제외)	26	1.5	13	1.6
		32	가구 제조업	24	1.4	13	1.6
		07	비금속 광물 광업 (연료용제외)	6	1.5	3	1.5
		84	공공행정, 국방 및 사회보장행정	101	1.4	57	1.5
		60	방송업	14	1.4	5	1.3
		23	비금속 광물제품 제조업	66	1.3	31	1.2

표(계속). 2022년 추계값 성별 표준화발병비 기준 상위 10개 백혈병 및 폐암의 고위험 업종

암	입사시기	업종 코드	업종명	1995-2018년 관찰값		2019-2022 예측값	
				발병자	발병비*	발병자	발병비*
폐 암	2000년 및 그 이전 입사자	06	금속 광업	24	6.0	7	3.5
		68	부동산업	4087	3.1	1450	2.9
		05	석탄, 원유 및 천연가스광업	250	3.0	82	2.6
		74	사업시설 관리 및 조경 서비스업	1660	2.9	616	2.8
		38	폐기물수집, 운반, 처리 및 원료재생업	466	2.5	171	2.3
		07	비금속광물광업(연료용제외)	133	2.2	55	2.3
		12	담배 제조업	128	2.2	45	2.1
		37	하수, 폐수 및 분뇨처리업	86	2.2	30	1.8
		84	공공행정, 국방 및 사회보장행정	889	1.9	359	2.0
		75	사업 지원 서비스업	1940	1.8	789	1.8
	2000년 이후 입사자	03	어업	58	3.1	37	4.6
		68	부동산업	2173	3.2	1366	4.3
		74	사업시설 관리 및 조경 서비스업	1536	3.0	763	3.0
		49	육상 운송 및 파이프라인 운송업	2138	2.5	905	2.4
		07	비금속광물광업(연료용제외)	51	2.3	24	2.4
		38	폐기물수집, 운반, 처리 및 원료재생업	272	2.3	115	2.1
		16	목재 및 나무제품제조업(가구제외)	172	2.0	77	1.9
		37	하수, 폐수 및 분뇨처리업	54	2.1	22	1.8
		2	임업	28	2.3	9	1.3
		50	수상 운송업	73	1.8	41	2.1

\* 성별 층화 전체업종 발생률 대비 해당업종의 발생률의 비 (발병자수 2명 이하(2018년 기준)의 중분류는 제외)

### ■ 시사점

- 1995-2014년의 자료를 이용하여 2015-2018년도 업종별 누적발병건수를 예측하고 정량적 타당성 검증을 수행한 결과 2000년도까지의 입사자 집단에 대하여는 대부분의 업종에서 모두 오차율 10% 미만의 정확한 추계가 가능하였음. 그러나 2000년 이후 입사자 집단에 대하여는 업종별 오차율의 중앙값이 20-25% 수준이었음.

## 3

### 연구활용방안

### ■ 개선방안

- 사업장기반 암 감시자료를 활용하여 안정적인 추세가 존재하는 업종에 대하여는 통계적인 기법 기반의 단기추계(3-5년)가 가능함을 확인하였음. 그러나 사업장 단위의 자료원을 활용한 이번 설계에서는 자료원의 한계로 암 발병의 가장 큰 위험요인인 나이가 반영되지는 못한 결과라는 한계점이 있음. 이에 추계모형 설계를 위한 자료 수집 시 매 시점 나이를 입력할 수 있는 방안을 제시함.

- 암발병 추계의 궁극적인 목적이 알려진 위험요인의 기여도 평가와 업종별 집단 감시라면, 위험기여도 평가 모형과 감시모형은 추계모형과 독립적으로 운용될 필요가 있다. 산업안전보건연구원이 확보 가능한 빅데이터로부터 위험기여도 평가모형과 감시 모형을 개발하기 위한 추가적인 연구와 검토가 요구됨.

✉ **연락처**

- 연구책임자: 숙명여자대학교 최영근교수
- 연구상대역: 산업안전보건연구원 직업건강연구실 중부권역학조사팀 이경은
- 연락처: 032-510-0753
- e-mail: kyeong85@kosha.or.kr