

건강보험 빅데이터를 통한 전체 근로자 및 공무원 근로자의 암 발생률 분석

백성욱^{1,2,3} · 이완형⁴ · 유기봉⁵ · 이우리⁵ · 이원태^{1,3,6} · 김민석^{1,3,6} · 임성실^{1,3,7} · 김지현^{6,8} · 최준혁⁹ · 이경은⁹ · 윤진하^{3,10*}

¹세브란스병원 직업환경의학과, ²연세대학교 일반대학원 의학과, ³연세대학교 의과대학 산업보건연구소,
⁴가천대학교 의과대학 길병원 직업환경의학과, ⁵연세대학교 보건과학대학 보건행정학과,
⁶연세대학교 일반대학원 보건학과, ⁷인천 근로자건강센터, ⁸수원 KMI 한국의학연구소,
⁹안전보건공단 산업안전보건연구원, ¹⁰연세대학교 의과대학 예방의학교실

Analyzing Cancer Incidence among Korean Workers and Public Officials Using Big Data from National Health Insurance Service

Seong-Uk Baek^{1,2,3} · Wanhyung Lee⁴ · Woo-Ri Lee⁵ · Ki-Bong Yoo⁵ · Won-Tae Lee^{1,3,6} ·
Min-Seok Kim^{1,3,6} · Sung-Shil Lim^{1,3,7} · Jihyun Kim^{6,8} · Jun-Hyeok Choi⁹ · Kyung-Eun Lee⁹ · Jin-Ha Yoon^{3,10*}

¹Department of Occupational and Environmental Medicine, Severance Hospital

²Graduate School, Yonsei University College of Medicine

³Institute for Occupational Health, Yonsei University College of Medicine

⁴Department of Occupational and Environmental Medicine, Gil Medical Center,
College of Medicine, Gachon University

⁵Department of Health Administration, Yonsei University Graduate School

⁶Department of Public Health, Graduate School, Yonsei University

⁷Incheon Workers Health Center

⁸Korea Medical Institute, Suwon Center

⁹Occupational Safety and Health Research Institute, Korea Occupational Safety and Health Agency

¹⁰Department of Preventive Medicine, Yonsei University College of Medicine

ABSTRACT

Objectives: This study aimed to establish a control group based on the big data from National Health Insurance Service. We also presented the number of incidences for each cancer, and analyzed the cancer incidence rate among Korean workers.

Methods: The cohort definition was separated by 'baseline cohort', 'dynamic cohort', and 'fixed-industry cohort' according to the definition. Cancer incidence was calculated based on the Korean Standard Classification of Disease code. Incidence rate was calculated among the group of all workers and public officials. Based on the study subjects and each cohort definition, the number of observations, incidences, and

*Corresponding author: Jin-Ha Yoon, Tel: 02-2228-1579, E-mail: flyinyou@gmail.com

Department of Preventive Medicine, Yonsei University College of Medicine, 50-1, Yonsei-ro, Seodaemun-gu, Seoul, Korea, 03722

Received: August 25, 2022, Revised: September 11, 2022, Accepted: September 22, 2022

 Seong-Uk Baek <https://orcid.org/0000-0002-0882-8326>

 Woo-Ri Lee <https://orcid.org/0000-0002-2700-6221>

 Min-Seok Kim <https://orcid.org/0000-0002-6975-5677>

 Jihyun Kim <https://orcid.org/0000-0002-1825-0097>

 Kyung-Eun Lee <https://orcid.org/0000-0001-5112-7747>

 Wanhyung Lee <https://orcid.org/0000-0001-6408-7668>

 Won-Tae Lee <https://orcid.org/0000-0001-7438-3153>

 Sung-Shil Lim <https://orcid.org/0000-0001-6733-7933>

 Jun-Hyeok Choi <https://orcid.org/0000-0002-4938-1348>

 Jin-Ha Yoon <http://orcid.org/0000-0003-4198-2955>

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the incidence rate according to sex and age groups was calculated. The incidence rate was estimated based on the incidence per 100,000 person-year, and 95% confidence intervals calculated according to the Poisson distribution.

Results: The result shows that the number of cancer cases in the all-worker group decreases after the age of 55, but the incidence rate tends to increase, which is attributed to the retirement of workers over 55 years old. Despite the specific characteristics of the workers, the trend and figures of cancer incidence revealed in this study are similar to those reported in previous studies of the overall South Korean population. When comparing the incidence rates of all workers and the control group of public officials, the incidence rate of public officials is generally observed to be higher in the age group under the age of 55. On the other hand, for workers aged 60 or older, the incidence rates were 1,065.4 per 100,000 person-year for all workers and 1,023.7 per 100,000 person-year for civil servants.

Conclusions: This study analyzed through health insurance data including all workers in Korea, and analyzed the incidence of cancer of workers by sex and age. In addition, further in-depth researches are needed to determine the incidence of cancer by industry.

Key words: National Health Insurance Service, big data, worker, cancer, incidence

I. 서 론

암은 현대 사회에서 주요한 사망원인 중 하나이다. 한국에서는 매년 약 240,000명이 암으로 진단받고 약 79,000명이 암으로 사망한다고 보고된다(Kang et al., 2022). 암을 예방하기 위한 관점에서 직업적 원인에 대한 암 발생을 규명하는 것은 중요한데, 한 선행연구에 따르면 모든 암 발생 중 약 2-8% 정도가 직업적 요인으로 발생하고 있다(Purdue et al., 2015). 사회적 보상의 측면에서도 의학적·법적 합리적인 기준에 따라 개별 근로자의 직업성 암 발생 원인을 추론하는 것은 첨예한 관심사이다(Kim & Kim, 2020).

산업보건 분야에서 근로자 암 발생 위험과 질병 발생을 분석하려는 노력이 지속되어 왔으며, 최근 빅데이터에 관한 관심이 증가함에 따라 이를 활용한 건강 위험 분석이 꾸준히 이어져 왔다. 예를 들어, Lee et al.(2019)은 고용보험 자료를 활용하여 직종별 표준화 암 발생률을 비교 분석한 바가 있다. 개인 혹은 집단에서 발생한 질병이 직업적 노출에 기인하였다고 추론하기 위해서는 유해 요인과 질병 사이의 합리적 인과성이 증명됨과 동시에 노출 집단에서 질병 발생이 유의하게 증가하여야 한다(ILO, 2009). 역학 연구를 통해 확인되는 집단 수준의 질병 이환율 증가는 업무 관련성 판단을 위한 핵심적 증거로 사용된다. 대표적인 예로, 세계보건 기구 산하의 국제 암 연구기관(International Agency for Research on Cancer)는 인간 대상의 역학적 연구 결과를 바탕으로 발암물질과 공정을 분류하고 있다(Loomis et al., 2018). 노출 집단의 질병 발생

이 실제 대조군에 비해 유의미하게 증가하여 있는지 확인하기 위해서는 연령, 성별 등 교란변수를 보정한 통계적 지표를 산출해야 하며, 연관성을 확인하는 데 활용되는 대표적 척도는 표준화 발생률, 표준화 사망률, 위험비 등이 있다. 하지만, 위와 같은 세부 노출 별 연관성 지표를 산출해내기 전에, 우리나라 근로자 전체 집단의 암 종별 발생 건수와 발생률을 먼저 개괄적으로 분석하여 일반 인구 대비 근로자 집단의 건강 수준과 암 발생 수준을 전체적으로 짚어볼 필요가 있을 것이다. 국립암센터는 매년 한국 암등록통계자료를 바탕으로 하여 전체 국민의 암 발생 추이와 발생률을 계산하여 발표하고 있지만, 아직 우리나라 전체 근로자 집단의 암 발생의 현황은 보고된 바 없다.

대한직업환경의학회는 국민 건강보험 공단과 양해각서를 2017년에 체결하고 직업성 질환 코호트를 구축한 바 있다. 해당 자료는 전 국민을 아우르는 방대한 자료로서, 직업 연관성 변수가 추가로 포함되어 활용할 수 있다. 건강보험공단의 청구자료를 바탕으로 한 코호트는 특정 사업장이나 지역으로 한정되었던 국내의 기존 코호트들의 단점을 극복하여 발생률이 낮은 암과 같은 질환에 대해 종적 연관을 확인할 수 있는 핵심적인 자료로 활용될 수 있다.

더불어, 최근 국내 학계 및 연구소를 통해 집단 수준의 직업성 질병 발생위험도에 대한 역학 연구들이 광범위하게 진행되고 있으나 이러한 역학 연구들은 현재 표준화된 대조집단을 공유하고 있지 않기 때문에 상호 비교 분석이 불가능하다는 한계가 산업보건 학계에서 꾸준히 지적되어온 바가 있다. 표준 대조집단이 구축되지

아니함으로써 역학 연구들 사이 상이한 결과들이 도출될 수 있고 이는 역학 연구의 결과의 일반화 과정에서 한계점으로 작용하며, 정책 개발 과정에서 의견 충돌과 갈등을 유발할 수 있는 요소이다. 본 연구는 이와 같은 학술적·사회적 요구에 발맞추어 한국 근로자 집단에 대해 대표성을 가지면서 충분한 인구수가 포함된 대조 집단을 구축하고 분석함으로써 우리나라 근로자 전체 집단의 전체적인 암 종별 발생 현황을 분석하여 제시하는 첫 번째 연구가 될 것이다.

정리하면, 본 연구의 목적은 첫째, 건강보험 빅데이터를 바탕으로 근로자 표준 대조집단을 구축하고 이를 위해 다양한 코호트 정의를 탐색하고 비교 분석하는 것이며, 둘째, 각 코호트 정의에 따라 성별, 연령별 암 발생률을 분석하여 제시하는 것이다. 본 연구는 건강보험 데이터를 바탕으로 암 관련 대조집단 구축 방법론을 다양하게 검토하고, 각 시나리오에 따라 한국 근로자의 세부 암종별 발생 현황과 특징을 제시하고자 한다.

II. 대상 및 방법

1. 연구 대상

본 연구는 국민건강보험 청구 데이터를 사용하여 수행되었다(NHISS, 2022). 국민건강보험은 대한민국 국민을 의무 가입 대상으로 하는 사회보험이며, 우리나라에 거주하는 98%의 인구를 포함하고 있다(Seong et al., 2017). 건강보험공단은 기준 년도인 2022년 약 5291만명에 대해서 적용되고 있으며, 이중 직장가입자는 약 3744만명으로 약 70.7%를 포함하고 있다(NHIS, 2022). 국민건강보험공단의 청구자료는 의료급여권자와 건강보험 대상자 전체에 대해서 상병 내역, 외래 및 입원 진료 내역, 처방 내역 등 의료 자료를 기록하고 있다. 본 연구의 분석에 사용된 직업코호트는 직장가입자를 대상으로 성, 연령, 사망, 지역, 보험 종류 등 인구사회학적 변수와 업종과 같은 직업 관련 변수를 바탕으로 구축되었다. 업종을 분류하는 과정은 표준산업분류 코드와 건보공단이 제시한 코드를 매칭하여 215개의 소분류로 분류하였다. 근로자 대조군을 정확하게 정의하는 것 역시 성별 및 연령에 따른 질병 발생률을 파악하는데 핵심적인 요소이다. 데이터 정제 과정에서 전체 근로자 대조군은 직장가입자만을 포함하였으며, 공무원 대조군의 경우 공무원 직역으로 포함되어있으면서 직종이 일반직과 교육직인 경우로 정의하였다.

2. 연구 방법

본 연구의 분석을 위한 코호트 구축은 코호트 정의별 기준에 따라 시작시점 코호트(baseline cohort), 다이나믹 코호트(dynamic cohort), 고정업종 코호트(fixed-industry cohort)로 나누어 분석하였다. 시작시점 코호트는 코호트의 시작 시점을 2009년을 기준으로 하여 해당 년도에 포함된 근로자에 대해서 추적 관찰을 진행한 코호트이며, 이 때 관찰 시작 1년 전 해당 질병이 발생하였거나 2009년 이후에 유입되는 관찰 대상은 배제되었다. 다이나믹 코호트의 정의는 코호트 추적 기간 내 한 번 이상 포함된 근로자를 대조군으로 정의하였으며, 이때 관찰 시작 1년에 질병이 발생한 경우 배제되었다. 또한 최소 관찰 기간 2년을 충족하지 못하는 2013년 이후 유입된 관찰 대상 역시 대조군에서 배제되었다. 마지막으로 고정업종 코호트의 정의는 관찰 기준 년도인 2009년을 기준으로 3년 동안 업종이 바뀌지 않은 대상자만을 포함하였으며, 2007-2009년 이나 2008-2010년, 혹은 2009-2011년도에 업종이 일정하게 유지된 대상자만을 포함하였고 따라서 업종 관찰년도 3년 동안 업종이 변한 경우는 배제되었다. 이때 관찰 시작 1년에 질병이 발생한 경우나 업종 관찰년도(3년) 동안 질병이 발생한 경우, 2013년 이후에 유입된 대상자는 제외하였다. 워시아웃(wash-out)의 경우 분석 관심 대상이 되는 질병 암종에 따라 해당 암종이 관찰 시작 1년 전에 발생한 경우(시작시점 코호트, 다이나믹 코호트, 고정업종 코호트), 관찰 시작 1년에 발생한 경우(다이나믹 코호트), 혹은 업종 관찰년도 3년 동안 발생한 경우 제외하였다.

분석 대상이 되는 전체 및 공무원 근로자 수는 코호트 정의 별로 달라지지만, 시작시점 및 다이나믹 코호트의 경우 약 1070만명-1100만명의 전체 근로자가 포함되었으며, 고정 업종 코호트는 약 832만명의 근로자가 포함되었다. 공무원 근로자의 경우 시작시점 및 다이나믹 코호트에서는 약 68만명 포함되었으며, 고정업종 코호트의 경우 61만명 포함되었다(Table 1).

암 질환은 세계보건기구의 제8차 한국질병분류(Korean Standard Classification of Disease, KCD-8)을 기준으로 분류되었다. 악성 신생물(C00-C97) 및 제자리암종(D00-D97)이 포함되었으며, 양성 신생물(D10-D36) 및 행동 양식 불명 또는 미상의 신생물(D37-D48)의 발생률 분석은 부가 자료에서만 제시되었다. 성별과 연령 집단에 따른 코호트별 대상자 수 및 발생자 수는 Table 1에 제시

Table 1. Baseline characteristics of study samples among total workers and public officials according to each cohort design

		Total workers		
		Baseline cohort N (%)	Dynamic cohort N (%)	Fixed-industry cohort N (%)
Total		10,768,989 (100.0)	11,050,398 (100.0)	8,317,710 (100.0)
Sex	Male	7,167,927 (66.5)	7,362,615 (66.6)	5,740,042 (69.0)
	Female	3,618,073 (33.5)	3,687,783 (33.4)	2,577,668 (31.0)
Age	25-29	1,772,555 (16.4)	1,853,391 (16.8)	1,406,466 (16.9)
	30-34	1,854,080 (17.2)	1,907,448 (17.3)	1,459,580 (17.5)
	35-39	1,890,273 (17.5)	1,925,593 (17.4)	1,525,911 (18.3)
	40-44	1,670,085 (15.5)	1,697,347 (15.4)	1,298,749 (15.6)
	45-49	1,466,671 (13.6)	1,494,043 (13.5)	1,178,732 (14.2)
	50-54	1,126,037 (10.4)	1,147,382 (10.4)	814,843 (9.8)
	55-59	643,775 (6.0)	657,070 (5.9)	421,468 (5.1)
	60-64	362,524 (3.4)	368,124 (3.3)	211,961 (2.5)
Year	2007			4,999,016 (60.1)
	2008			6,050,237 (72.7)
	2009	10,786,000 (100.0)	10,786,000 (97.6)	8,317,710 (100.0)
	2010	10,717,997 (99.4)	10,801,025 (97.7)	8,314,709 (100.0)
	2011	10,667,341 (98.9)	10,835,097 (98.1)	8,293,307 (99.7)
	2012	10,605,077 (98.3)	10,828,096 (98.0)	8,250,626 (99.2)
	2013	10,541,774 (97.7)	10,734,444 (97.1)	8,207,232 (98.7)
	2014	10,511,101 (97.5)	10,686,858 (96.7)	8,186,613 (98.4)
	2015	10,491,048 (97.3)	10,655,973 (96.4)	8,170,683 (98.2)
		Public officials		
		Baseline cohort N (%)	Dynamic cohort N (%)	Fixed-industry cohort N (%)
Total		677,485 (100.0)	677,667 (100.0)	611,833 (100.0)
Sex	Male	338,652 (50.0)	338,793 (50.0)	300,080 (49.0)
	Female	338,833 (50.0)	338,874 (50.0)	311,753 (51.0)
Age groups	25-29	71,960 (10.6)	71,982 (10.6)	66,409 (10.9)
	30-34	110,036 (16.2)	110,061 (16.2)	101,535 (16.6)
	35-39	111,524 (16.5)	111,571 (16.5)	104,072 (17.0)
	40-44	111,381 (16.4)	111,415 (16.4)	104,121 (17.0)
	45-49	113,637 (16.8)	113,657 (16.8)	105,520 (17.2)
	50-54	90,020 (13.3)	90,042 (13.3)	82,259 (13.4)
	55-59	53,141 (7.8)	53,151 (7.8)	41,496 (6.8)
	60-64	15,786 (2.3)	15,788 (2.3)	6,421 (1.0)
Year	2007			5,744 (0.9)
	2008			9,253 (1.5)
	2009	677,485 (100.0)	677,485 (100.0)	611,833 (100.0)
	2010	676,882 (99.9)	676,963 (99.9)	611,832 (100.0)
	2011	676,160 (99.8)	696,292 (99.8)	611,823 (100.0)
	2012	675,336 (99.7)	675,494 (99.7)	611,221 (99.9)
	2013	674,555 (99.6)	674,696 (99.6)	610,582 (99.8)
	2014	673,734 (99.4)	673,866 (99.4)	609,908 (99.7)
	2015	672,919 (99.3)	673,045 (99.3)	609,207 (99.6)

되었다. 주 분석에서 전체 근로자 대상의 모든 신생물에 대한 종합 발생률이 계산되어 제시되었으며(Table 2), 공무원 집단에서의 발생률 역시 제시되었다(Table 3). 또한 Figure 1과 2에서는 주요 암종인 위암, 폐암, 대장암, 백혈병에 대한 전체 근로자 및 공무원 집단의 연령별 발생률

이 제시되었다. 마지막으로 Supplementary table 1-6에서는 전체 근로자 및 교육직 공무원의 세부 암종별 발생률이 계산되었다.

설정된 연구 대상 및 질환 정의를 바탕으로 본 연구는 성별 및 연령군에 따른 관찰인년, 발생자 수, 발생률을

Table 2. Number of cases, observations, and incidence rate for all neoplasms combined (ICD code: C00–C97, D00–D09) among total workers in each cohort design [CI: confidence interval]

Age groups	Baseline cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	13,564	121	112.1 [110.2–114.0]	3,329	60.9	54.6 [52.8–56.5]	10,235	60.1	170.3 [167.1–173.7]
30–34	19,614	127	154.4 [152.3–156.6]	7,303	82.2	88.9 [86.9–90.9]	12,311	44.8	274.6 [269.8–279.5]
35–39	27,796	129.4	214.8 [212.3–217.3]	12,488	90.8	137.5 [135.1–139.9]	15,308	38.6	396.8 [390.5–403.1]
40–44	33,740	114.5	294.7 [291.5–297.8]	16,315	79.4	205.4 [202.3–208.6]	17,425	35.1	496.7 [489.3–504.1]
45–49	40,741	100.2	406.4 [402.5–410.4]	23,202	70.1	331.1 [326.8–335.4]	17,539	30.2	579.8 [571.3–588.4]
50–54	42,382	76.4	554.6 [549.3–559.9]	29,291	55	532.6 [526.5–538.7]	13,091	21.3	613.6 [603.2–624.2]
55–59	32,955	43.2	763.4 [755.2–771.7]	25,691	31.7	811.3 [801.4–821.3]	7,264	11.5	631.7 [617.2–646.3]
60–64	25,480	23.9	1,065.4 [1,052.3–1,078.5]	21,321	18.1	1,179.0 [1,163.3–1,195.0]	4,159	5.8	713.0 [691.5–735.0]
Age groups	Dynamic cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	13,671	124.2	110.0 [108.2–111.9]	3,377	63.3	53.3 [51.5–55.2]	10,294	60.8	169.2 [166.0–172.5]
30–34	19,730	129.2	152.7 [150.6–154.9]	7,361	83.9	87.7 [85.7–89.7]	12,369	45.2	273.3 [268.6–278.2]
35–39	27,896	130.8	213.4 [210.9–215.9]	12,528	91.9	136.3 [133.9–138.7]	15,368	38.9	394.9 [388.7–401.2]
40–44	33,943	115.6	293.7 [290.6–296.8]	16,389	80.2	204.4 [201.3–207.6]	17,554	35.4	495.6 [488.3–503.0]
45–49	41,031	101.3	404.9 [401.0–408.8]	23,353	70.8	330.1 [325.9–334.3]	17,678	30.6	578.0 [569.5–586.6]
50–54	42,644	77.2	552.0 [546.8–557.3]	29,459	55.6	530.0 [524.0–536.1]	13,185	21.7	608.5 [598.2–619.0]
55–59	33,181	43.7	759.9 [751.7–768.1]	25,864	32.1	806.2 [796.4–816.0]	7,317	11.6	631.7 [617.3–646.3]
60–64	25,592	24.1	1,062.6 [1,049.7–1,075.7]	21,411	18.2	1,178.6 [1,162.9–1,194.5]	4,181	5.9	706.6 [685.4–728.4]
Age groups	Fixed-industry cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	11,300	114.8	98.5 [96.7–100.3]	3,339	62.8	53.2 [51.4–55.0]	7,961	52	153.1 [149.8–156.5]
30–34	15,814	120.4	131.3 [129.3–133.4]	6,900	83.1	83.0 [81.1–85.0]	8,914	37.3	238.8 [233.8–243.8]
35–39	22,002	125.8	174.9 [172.5–177.2]	11,139	92.2	120.9 [118.6–123.1]	10,863	33.7	322.7 [316.6–328.8]
40–44	25,359	106.5	238.1 [235.2–241.1]	13,740	76.7	179.2 [176.2–182.2]	11,619	29.8	389.5 [382.4–396.6]
45–49	31,638	96.2	328.7 [325.1–332.3]	20,275	69.9	290.0 [286.0–294.0]	11,363	26.2	432.9 [425.0–440.9]
50–54	29,403	65.4	449.5 [444.3–454.6]	21,894	48.8	448.3 [442.4–454.3]	7,509	16.6	452.8 [442.6–463.2]
55–59	20,852	33.5	622.4 [614.0–631.0]	16,986	25.4	668.3 [658.3–678.4]	3,866	8.2	473.4 [458.6–488.6]
60–64	14,701	16.8	873.3 [859.3–887.6]	12,782	13.2	964.7 [948.0–981.6]	1,919	3.7	523.4 [500.2–547.3]

*Incidence rate per 100,000 person-years

Table 3. Number of cases, observations, and incidence rate for all neoplasms combined (ICD code: C00–C97, D00–D09) among public officials in each cohort design [CI: confidence interval]

Age groups	Baseline cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	687	5	137.4 [127.3–148.1]	61	1.1	56.3 [43.1–72.3]	626	3.9	159.8 [147.6–172.9]
30–34	1,676	7.7	218.6 [208.3–229.3]	257	2.6	99.5 [87.7–112.4]	1,419	5	283.8 [269.2–299.0]
35–39	2,296	7.7	299.5 [287.4–312.0]	400	3.2	123.1 [111.3–135.7]	1,896	4.4	429.3 [410.2–449.1]
40–44	3,000	7.7	391.3 [377.4–405.6]	838	3.9	214.0 [199.7–228.9]	2,162	3.8	576.5 [552.5–601.4]
45–49	3,723	7.8	480.4 [465.1–496.1]	1,634	4.6	356.5 [339.4–374.2]	2,089	3.2	659.7 [631.7–688.6]
50–54	3,656	6.2	592.9 [573.8–612.4]	2,411	4.4	545.9 [524.3–568.1]	1,245	1.8	711.4 [672.5–752.1]
55–59	2,867	3.6	800.1 [771.1–829.9]	2,160	2.7	810.0 [776.2–844.9]	707	0.9	771.3 [715.5–830.3]
60–64	1,109	1.1	1,023.7 [964.3–1,085.8]	890	0.8	1,068.0 [999.0–1,140.5]	219	0.2	876.0 [763.8–1,000.0]
Age groups	Dynamic cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	687	5	137.4 [127.3–148.1]	61	1.1	56.3 [43.1–72.3]	626	3.9	159.8 [147.6–172.9]
30–34	1,676	7.7	218.6 [208.3–229.3]	257	2.6	99.5 [87.7–112.4]	1,419	5	283.8 [269.2–299.0]
35–39	2,296	7.8	296.3 [284.3–308.6]	400	3.2	123.1 [111.3–135.7]	1,896	4.4	429.3 [410.2–449.1]
40–44	3,000	7.7	391.3 [377.4–405.6]	838	3.9	214.0 [199.7–228.9]	2,162	3.8	576.5 [552.5–601.4]
45–49	3,723	7.8	480.4 [465.1–496.1]	1,634	4.6	356.5 [339.4–374.2]	2,089	3.2	659.7 [631.7–688.6]
50–54	3,657	6.2	593.0 [574.0–612.6]	2,412	4.4	546.1 [524.5–568.4]	1,245	1.8	711.4 [672.5–752.1]
55–59	2,867	3.6	800.1 [771.1–829.9]	2,160	2.7	810.0 [776.2–844.9]	707	0.9	771.3 [715.5–830.3]
60–64	1,109	1.1	1,023.7 [964.3–1,085.8]	890	0.8	1,068.0 [999.0–1,140.5]	219	0.2	876.0 [763.8–1,000.0]
Age groups	Fixed-industry cohort								
	Total			Male workers			Female workers		
	Cases	Person-years (in 100,000)	Incidence rate* [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]	Cases	Person-years (in 100,000)	Incidence rate [95% CI]
25–29	482	4.7	103.3 [94.3–112.9]	45	1	45.0 [32.8–60.2]	437	3.7	119.2 [108.3–130.9]
30–34	1,112	7.1	157.0 [147.9–166.5]	160	2.4	66.2 [56.3–77.3]	952	4.7	204.0 [191.2–217.4]
35–39	1,508	7.2	210.4 [199.9–221.3]	265	3.1	85.9 [75.9–96.9]	1,243	4.2	298.3 [282.0–315.4]
40–44	1,929	7.2	269.2 [257.3–281.4]	554	3.7	151.1 [138.8–164.2]	1,375	3.5	392.9 [372.4–414.2]
45–49	2,334	7.2	321.9 [309.0–335.3]	1,072	4.2	252.2 [237.4–267.8]	1,262	2.9	432.7 [409.1–457.2]
50–54	2,286	5.6	409.4 [392.8–426.6]	1,550	4.1	379.6 [360.9–399.0]	736	1.6	464.8 [431.9–499.7]
55–59	1,491	2.8	526.2 [499.9–553.6]	1,102	2	551.0 [518.9–584.5]	389	0.8	518.7 [468.4–572.9]
60–64	298	0.4	715.2 [636.3–801.2]	231	0.3	693.0 [606.5–788.4]	67	0.1	804.0 [623.1–1,021.1]

*Incidence rate per 100,000 person-years

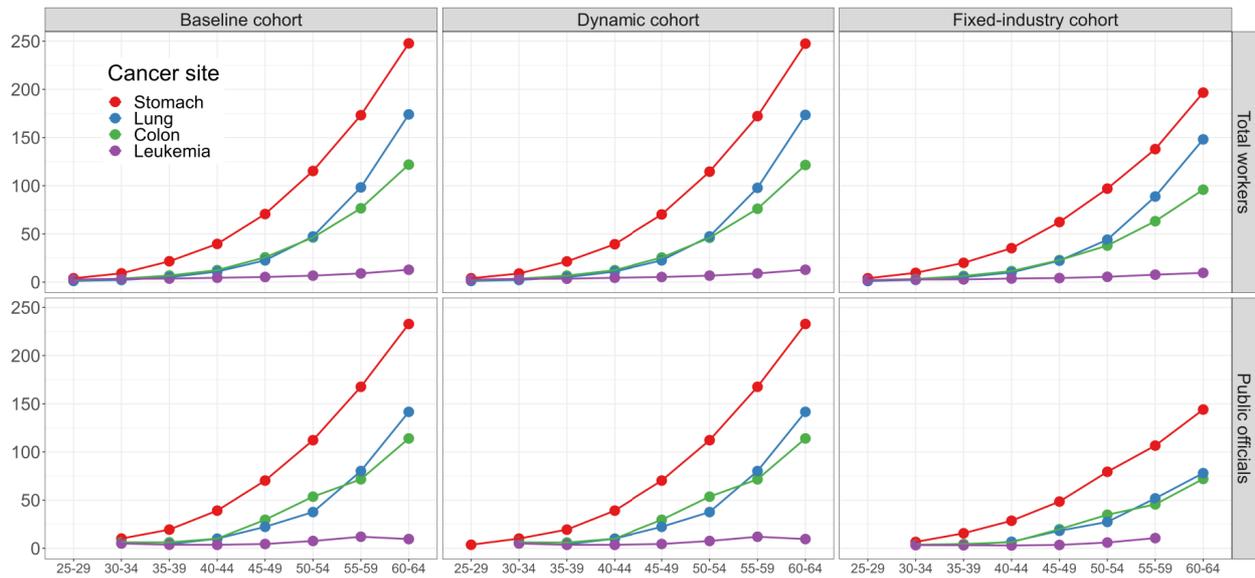


Figure 1. Age-specific incidence rates of common cancers among male Korean workers (Incidence rate per 100,000 person-year)

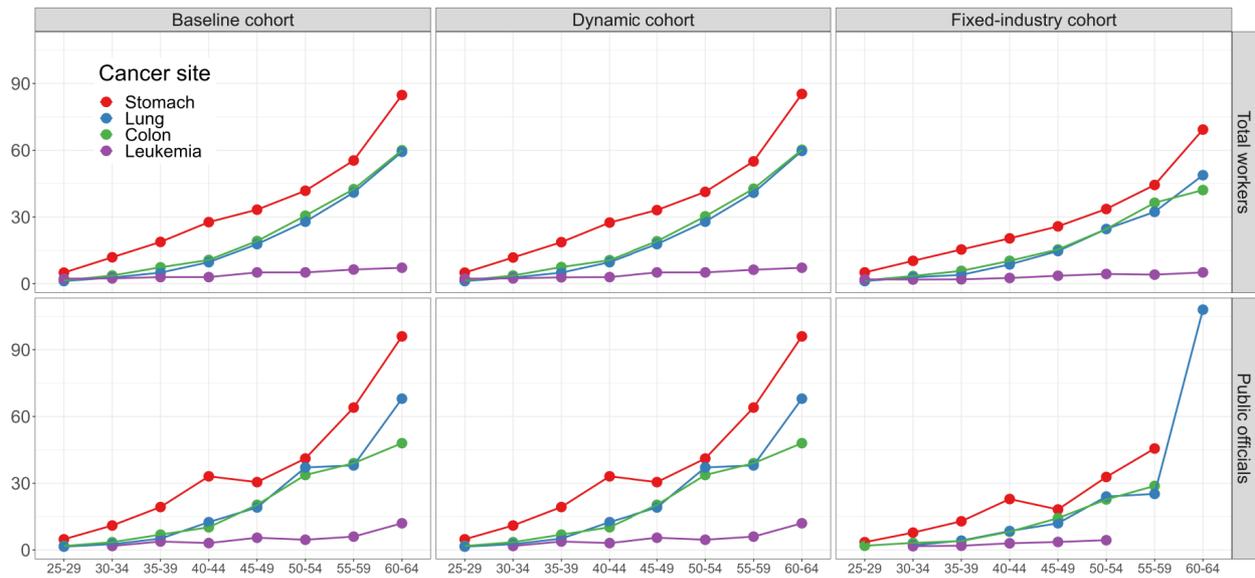


Figure 2. Age-specific incidence rates of common cancers among female Korean workers (Incidence rate per 100,000 person-year)

계산하였다. 발생률은 100,000인년 당 발생을 기준으로 계산하였으며, 포아송 분포에 따른 95% 신뢰구간이 계산되었다.

통계 분석과 시각화에 통계 프로그램 SAS(Version 9.4. SAS Institute, Cary, North Carolina)와 R (Version 4.2.0; R Foundation for Statistical Computing, Vienna, Austria)를 사용하였다. 본 연구

는 연세대학교 생명윤리심의위원회(Institutional Review Board, IRB)의 승인을 받았다(Y-2017-0100).

III. 결 과

Table 1에서 전체 및 공무원 대조집단의 성과 연령에 따른 총 인원 및 전체 악성신생물 발생자가 제시되

어 있다. 전체 근로자의 경우 시작시점 코호트에서 남성 근로자의 수는 7,167,927명으로 3,618,073의 여성 근로자에 비해 약 2배 포함되어 있다. 반면 공무원 대조집단의 경우 남성과 여성 공무원 수가 모두 약 33만 8000명으로 1:1의 비율을 보인다.

Table 2에서 전체 악성신생물에 대한 발생자, 관찰인년 수, 발생률이 성과 연령별로 층화되어 제시되었다. Table 2에서 나타난 것처럼, 전체 근로자와 공무원 대조집단에서 55세를 기점으로 발생자 수는 감소하지만, 발생률은 증가한다. 또한 성과 연령으로 층화하여 볼 때, 양 집단 모두에서 55세 미만 신생물 발생률의 경우 여성이 남성에 비해 높으나, 55세 이상에서 남성의 발생률이 여성의 발생률을 역전하는 것으로 나타난다. 예를 들어, 시작시점 코호트의 남성 및 여성 근로자 대조집단의 경우 45-49세의 악성신생물 발생률은 각각 100,000인년 당 331.1 건과 579.8 건인데 비해 60세 이상 근로자의 경우 남성이 1179.0건, 여성이 713.0건이다. 또한 60세 이상 근로자의 발생률이 30세 미만 근로자에 비해 남성의 경우 약 20배 이상, 여성의 경우 4-5배 높은 것으로 분석된다. 마지막으로 전체 근로자와 공무원 대조집단의 발생률을 비교할 때, 55세 미만의 연령 집단에서 공무원의 발생률이 대체로 높게 관찰된다. 반면 60세 이상의 근로자에서는 전체 근로자는 100,000인년 당 1,065.4건, 공무원은 100,000인년당 1,023.7명으로 나타난다.

Figure 1과 2에서는 한국의 주요 암종으로 볼 수 있는 위암, 폐암, 대장암, 그리고 혈액암인 백혈병에 대해서 나이별 발생률을 성별에 따라 제시하였다. 전체적으로 나이가 증가할수록 모든 암의 발생률이 증가하는 것으로 관찰되었으며, 고정업종 코호트 정의를 바탕으로 한 발생률은 시작시점 코호트나 다이내믹 코호트 정의를 사용한 분석보다 낮은 발생률이 관찰되었다.

Supplementary table 1-6에서는 전체 근로자와 공무원 대조집단 각각에 대해서 세부 암종별 발생률을 확인할 수 있도록 나누어 제시하였다. 위암, 대장암, 간암, 폐암, 유방암과 같은 발생률이 비교적 높고 잘 알려진 암뿐만 아니라, 중추신경계 암, 피부암, 림프종, 혈액암 등 희귀암에 대한 발생률도 제시되었다.

IV. 고 찰

본 연구는 건강보험공단의 직업코호트를 사용하여 근

로자의 암 발생에 대한 대조집단을 구축하고 전체 근로자의 세부 암종별 발생 현황을 파악하는 것을 목표로 하였다. 본 코호트 집단은 2009년부터 현재까지 건강보험에 등록된 모든 근로자를 포함하고 있으므로, 대한민국 전체를 대표한다는 점에서 의의가 있다. 본 연구는 의료인에 의해 공식적으로 등록된 암 발생에 대한 건강정보를 담고 있으며, 근로자의 세부 업종에 대한 자료 역시 포함되어 있어 업종별 암 발생을 파악하기 위한 핵심 자료로 사용될 수 있다.

다양한 나라들의 연구진들은 국가 규모 수준의 데이터를 활용하여 근로자의 암 발생을 분석한 바 있다. 예를 들어, 미국의 한 연구는 27개 주의 사망자료를 바탕으로 미국 여성의 직업성 폐암 사망을 분석하였으며 (Robinson et al., 2011), 일본의 한 연구진 역시 약 2700만명 이상의 남성 근로자의 폐, 위, 대장암 분석을 수행한 바 있다(Eguchi et al., 2017). 국내에서도 고용보험 자료를 활용하여 직종별 암 발생 위험도를 분석한 바 있다(Lee et al., 2020). 또한 국내 건강보험공단 자료를 활용하여 업종별 암 위험도를 분석한 선행연구들이 있다. 대표적으로 항공 운수업 종사자의 혈액암 발생 (Lee et al., 2019), 육상 운수업 종사자의 암 발생(Lee et al., 2020a), 외부 공기 매개체 발암물질 노출과 소화기 암 발생(Lee et al., 2020b) 등이 분석된 바 있다.

본 연구는 표준 대조집단 구축을 위해 세 가지의 서로 다른 코호트 정의를 설정하였다. 본 연구에서 시작시점 코호트로 정의되었던 코호트 정의 방식은, 기존 역학연구에서 닫힌 코호트(closed cohort)로 불려왔으며, 관찰 시점에 포함된 고정된 참여자들을 대상으로 추적 관찰하는 기법을 뜻한다(Wang & Kattan, 2020). 시작시점 코호트의 경우 많은 선행연구에서 사용되는 전통적인 코호트 정의 방식이나, 시작시점에 포함되지 않은 근로자들이 배제되는 문제가 있다. 이 문제를 보완하기 위해 다이내믹 코호트 정의를 추가하여 더 많은 근로자를 포함하는 방식을 취했다. 다이내믹 코호트는 코호트 추적기간 동안 한번 이상 등록된 근로자를 포함하기 때문에 시작시점 코호트 보다 더 많은 실험 대상을 모집할 수 있다는 장점이 있다(Williamson et al., 2001). 다만 위 두 가지 코호트 정의에서는 업종이 빈번하게 바뀌는 근로자들에 대해서 업종과 특정 암 질환 발생 사이의 연관을 파악하기 힘들다는 단점이 있어, 3년 이상 동일 업종에서 일한 근로자만을 대상으로 한 고정업종 코호트를 고안하여 적용하였다. 고정업종

코호트 분석 방식은 기존 역학 문헌들에서 자주 사용되는 분석 방식은 아니지만, 빅데이터를 활용하여 직업성 질환을 분석한 선행연구에서 동일한 코호트 정의 방식이 활용된 바 있다(Lee et al., 2018). 하지만, 고정업종 코호트에서는 포함된 근로자 수와 암 발생자 수가 감소할 뿐만 아니라, 암 발생률 역시 감소하는 것으로 보이며, 이는 보수적인 대상자 선정 기준으로 인해 상대적으로 암 발생에 취약한 근로자들이 배제되어 손실 바이어스(attrition bias)가 일어난 것으로 해석될 수 있다(Nunan et al., 2018). 추후 연구에서는 본 연구의 표준 대조집단 구축에서 활용된 3가지 코호트 정의 방식의 장·단점을 검토하고 업종별 암 발생 위험을 적절하게 파악할 수 있는 방법론을 구축하는 과정이 필요할 것으로 보인다.

분석 결과 근로자 대조집단에서 55세 이후 암 발생자 수가 감소하지만, 발생률은 증가하는 추세를 나타냈고, 이는 55세 이상의 근로자들이 은퇴함에 따라 코호트에서 탈락하기 때문으로 사료된다. 이러한 근로자 집단의 특수성에도 불구하고, 본 연구에서 드러난 암 발생률 추세와 수치는 한국 전체 인구를 대상으로 한 선행연구에서 보고된 수치와 유사하다(Kweon, 2018).

또한 분석 결과를 성별로 층화하여 볼 때, 젊은 세대에서는 암 발생이 여성이 상대적으로 높지만, 55세를 기점으로 남성의 암 발생률이 여성을 초과하는 것으로 드러난다. 성별에 따른 연령별 세부 암 발생률을 계산한 분석에서도, 대한민국 국민의 전체 암 발생률 수치와 비슷한 수치와 추세를 보이고 있으나(Kang et al., 2022), 고정 업종 코호트에서는 발생률이 전체적으로 낮아지는데, 이는 본 연구의 업종 선택 기간인 3년 동안 암 발생자의 위험이 높은 근로자들이 탈락하면서 발생하는 현상으로 해석된다.

건강근로자 효과는 대조집단 구축에 있어 고려해야 할 핵심적인 사안이다. 건강근로자 효과는 노출 집단에 해당 되는 근로자의 건강 상태가 기저 수준에서 이미 우수하기 때문에 발생하는 바이어스로, 직업성 질환의 역학 연구에 있어서 유해인자의 효과가 과소 추정되는 주요 원인으로 지목된다(Li & Sung, 1999). 과거에는 건강근로자 효과는 암에 대해서는 영향이 적다는 주장도 있었으나(Choi, 1992), 최근 수십년간 직업성 암 발생에서도 해당 교란 효과가 업종별 암 발생 효과를 과소 추정할 가능성이 있다는 점이 지적되어 왔다(Kirkeleit et al., 2013). 해당 바이어스를 극복하기

위해 대조집단을 근로자 중 유해인자 노출 빈도가 적고 안정된 환경에서 근무하는 공무원으로 설정하는 방식이 사용될 수 있다. 특히, 암 관련 분석에 있어 일반직 및 교육직 공무원은 직업적 발암물질 노출로부터 안전한 환경에서 근무하므로, 대조집단으로 설정될 당위성이 있다. 대표적으로, 업종별 암 발생 위험도를 분석하였던 국내 선행연구들에서도 공무원 집단을 최적의 근로환경을 가진 대조집단으로 설정하고 특정 업종 종사자와 비교한 바 있으며, 이 경우 전체 근로자를 대조집단으로 설정하였을 때에 비해 위험도의 상승이 공통적으로 관찰된 바 있다(Lee et al., 2019; Lee et al., 2020a). 이에 본 연구는 건강보험공단 청구 자료를 바탕으로 전체 근로자뿐만 아니라 공무원 대조집단을 추가로 구축하고 해당 집단의 암 발생률 역시 분석하여 비교하였다.

공무원과 전체 근로자를 비교한 연구에서 전체 공무원의 암 발생률이 전체 근로자에 비해 유사하거나 다소 높게 나타났다. 이는 공무원 집단의 높은 교육 수준과 의료 이용률 때문에 암의 조기 진단이 이루어지는 것으로 설명할 수 있다. 예를 들어 의료 이용 접근성에 대한 선행연구들은 높은 교육수준을 가질수록 의료 이용 빈도가 유의하게 높아지는 것을 보고하고 있다(Gruber & Kiesel, 2010). 이를 비추어 볼 때, 암 병기와 조기 암, 치료 성적을 고려한 추후 분석이 공무원에서 보고되는 높은 암 발생률을 자세히 설명해줄 수 있을 것으로 보인다.

본 코호트의 광범위한 데이터 범위에도 불구하고, 다음과 같은 단점이 있다. 먼저, 자영업자, 일용직 근로자 등 취약 근로자들이 본 대조집단 구축에서 제외되었기 때문에, 본 결과의 해석에 주의가 필요하다. 본 데이터는 근로자의 업종에 대한 분류를 제공하고 있지만, 근로시간, 직종 등 근로환경에 대한 세부정보를 포함하고 있지 않아 성·연령 표준화 발생률 산출 이외에 콕스 비례위험 분석 등 교란변수 보정이 필요한 정밀한 위험도 분석에는 제한이 있다. 또한, 본 연구의 고정업종 코호트 정의는 3년 이상 동일한 업종에 종사한 대상자만 포함하였는데, 이 경우 발생할 수 있는 중도 탈락 및 건강 근로자 효과를 보정하기 위한 추가적인 연구 설계가 필요할 수 있다. 그럼에도 불구하고 본 코호트는 대한민국 근로자의 건강정보를 분석하여 성, 연령 별 세부 암종에 대한 암 발생률을 산출하는데 이르렀다는 점에서 의의가 있다.

V. 결 론

본 연구는 대한민국 전체 근로자 집단을 포함하는 건강보험 자료를 통해 빅데이터를 구축하고, 이를 바탕으로 성, 연령별 근로자의 암 발생률을 분석하였다. 다만 업종의 정의에 따라 코호트의 정의가 달라질 수 있으며, 정확한 코호트를 설계하기 위한 추후 연구가 필요하다. 또한 본 데이터를 바탕으로 업종별 암 발생률을 분석할 수 있으며, 업종과 직업성 암의 관련성 파악에 활용할 수 있다.

감사의 글

본 연구의 데이터는 대한직업환경의학회와 건강보험공단의 MOU의 일환으로 제공 받았습니다. 본 연구는 국민건강보험공단의 빅데이터를 활용하여 분석하였으며 자료를 제공해주신 건강보험공단의 관계자에게 깊은 감사를 드립니다. 본 연구는 산업안전보건연구원 학술용역에 의거하여 수행된 연구입니다. 산업안전보건연구원 "근로자 표준 대조집단 구축 및 활용 건강위험 선별 프로그램 개발" 연구의 일환으로 수행되었으며, 펀드 제공자는 연구의 방법론과 방향 어느 것에도 영향을 미치지 않았습니다.

References

- Choi BC. Definition, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect. *J Occup Med* 1992;34(10):979-88
- Eguchi H, Wada K, Prieto-Merino D, Smith DR. Lung, gastric and colorectal cancer mortality by occupation and industry among working-aged men in Japan. *Sci Rep* 2017 Feb 23;7:43204 (<https://doi.org/10.1038/srep43204>)
- Gruber S, Kiesel M. Inequality in health care utilization in Germany? Theoretical and empirical evidence for specialist consultation. *J public health* 2010;18(4):351-65 (<https://doi.org/10.1007/s10389-010-0321-2>)
- ILO. Identification and recognition of occupational diseases: criteria for incorporating diseases in the ILO list of occupational diseases. Geneva; ILO.; 2009. p. 7
- Kang MJ, Won YJ, Lee JJ, Jung KW, Kim HJ et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2019. *Cancer Res Treat* 2022;54(2):330-344 (<https://doi.org/10.4143/crt.2022.128>)
- Kim K, Kim J. A work-relatedness assessment in epidemiological case investigation of occupational cancers: I. Principles. *Ann Occup Environ Med* 2020;32(1):e30 (<https://doi.org/10.35371/aoem.2020.32.e30>)
- Kirkeleit J, Riise T, Bjørge T, Christiani DC. The healthy worker effect in cancer incidence studies. *Am J Epidemiol* 2013;177(11):1218-24 (<https://doi.org/10.1093/aje/kws373>)
- Kweon SS. Updates on cancer epidemiology in Korea, 2018. *Chonnam Med J* 2018;54(2):90-100
- Lee HE, Zaitso M, Kim EA, Kawachi I. Cancer incidence by occupation in Korea: Longitudinal analysis of a nationwide cohort *Saf Health Work* 2020;11(1):41-49 (<https://doi.org/110.1016/j.shaw.2019.12.004>)
- Lee SG, Kim EA, Lee KE, Um HS, Shin MY et al. Construction of Occupational Disease Cohort Using Big Data. *Occupational Safety and Health Research Institute*. 2018;p10. Available from: URL:<https://kosha.or.kr/oshri/professionalBusiness/occupationalDiseaseReferenceRoom.do>
- Lee W, Kang MY, Kim J, Lim SS, Yoon JH. Cancer risk in road transportation workers: a national representative cohort study with 600,000 person-years of follow-up. *Sci Rep* 2020 Jul 9;10(1):11331 (<https://doi.org/10.1038/s41598-020-68242-5>)
- Lee W, Kang MY, Yoon JH. Cancer incidence among air transportation industry workers using the national cohort study of Korea. *Int J Environ Res Public Health* 2019 Aug 14;16(16):2906 (<https://doi.org/10.3390/ijerph16162906>)
- Lee W, Kim J, Lim SS, Kim Y, Ahn YS et al. External airborne-agent exposure increase risk of digestive tract Cancer. *Sci Rep* 2020 May 25;10(1):8617 (<https://doi.org/10.1038/s41598-020-65312-6>)
- Li CY, Sung FC. A review of the healthy worker effect in occupational epidemiology. *Occup Med (Lond)* 1999;49(4):225-9 (<https://doi.org/10.4068/cmj.2018.54.2.90>)
- Loomis D, Guha N, Hall AL, Straif K. Identifying occupational carcinogens: an update from the IARC Monographs. *Occup Environ Med* 2018;75(8):593-603 (<https://doi.org/110.1016/10.1136/oemed-2017-104944>)
- National Health Insurance Service(NHIS). Population coverage. 2020. Available from: URL:<https://www.nhis.go.kr/>

nhis.or.kr/english/wbheaa02400m01.do
National Health Insurance Sharing Service(NHISS).
Sample research DB: details of DB and cost. 2022.
Available from: URL:https://nhiss.nhis.or.kr/bd/
ab/bdaba 022eng.do
Nunan D, Aronson J, Bankhead C. Catalogue of bias:
attrition bias. *BMJ Evid Based Med* 2018 Feb;
23(1):21-22. (https://doi.org/10.1136/ebmed-2017-
110883)
Purdue MP, Hutchings SJ, Rushton L, Silverman DT. The
proportion of cancer attributable to occupational
exposures. *Ann Epidemiol* 2015;25(3):188-192
(https://doi.org/10.1016/j.annepidem.2014.11.009)
Robinson CF, Sullivan PA, Li J, Walker JT. Occupational
lung cancer in US women, 1984-1998. *Am J Ind Med*
2011 Feb;54(2):102-17 (https://doi.org/10.1002/
ajim.20905)
Seong SC, Kim YY, Khang YH, Park JH, Kang HJ et al. Data

resource profile: the national health information
database of the National Health Insurance Service in
South Korea. *Int J Epidemiol* 2017 Jun 1;46(3):
799-800 (https://doi.org/10.1093/ije/dyw253)
Wang X, Kattan MW. Cohort studies: design, analysis,
and reporting. *Chest* 2020 Jul;158(1S):S72-S78
(https://doi.org/10.1016/j.chest.2020.03.014)
Williamson JM, Satten GA, Hanson JA, Weinstock H,
Datta S. Analysis of dynamic cohort data. *Am J
Epidemiol* 2001 Aug 15;154(4):366-72 (https://
doi.org/10.1093/aje/154.4.366)

<저자정보>

백성욱(전공의), 이완형(교수), 유기봉(교수), 이우리
(박사과정), 이원태(전공의), 김민석(전공의), 임성실
(전문의), 김지현(전문의), 최준혁(과장), 이경은(선임
연구위원), 윤진하(교수)