A Two-Decade Analysis of Participation Rates in Lung-Damaged Industries and Lung Cancer Mortality Across Canadian Provinces

December 8, 2023

1 Introduction

Health inequities have always been a challenge facing all human societies. This unfairness exists in all aspects, such as income gap, living environment and geographical location, policies, food, etc. Different industries have different environments and have different impacts on health. For example, blue-collar workers working in dusty construction sites and white-collar workers sitting in clean offices face different risks of lung disease, which will create health inequities. This article will use data on work rates, lung cancer rates, and number of cigarettes smoked per day in different industries in different provinces in Canada over the past 20 years to study the impact of different industries on lung cancer rates.

Several noteworthy studies have contributed to our understanding of these patterns. The research conducted by William J. Blot and Joseph F. Fraumeni Jr., titled "Geographic Patterns of Lung Cancer: Industrial Correlations," in the United States during 1950-1969, identified elevated lung cancer mortality rates among males in counties with paper, chemical, petroleum, and transportation industries. Notably, this correlation persisted even after adjusting for demographic influences, emphasizing the impact of specific industries on lung health. Additionally, A.B. Miller's study, "Recent Trends in Lung Cancer Mortality in Canada," highlights the escalating trend in lung cancer mortality, particularly in the last decade. The analysis reveals a consistent increase in mortality rates across age groups and birth cohorts, underscoring the urgency of addressing the factors contributing to this concerning trend. Furthermore, a cohort study conducted by M.M. Finkelstein and D.K. Verma focused on mortality among Ontario pipe trades workers, revealing significant increases in lung cancer mortality rates, particularly among plumbers, pipefitters, and sprinkler fitters. The findings suggest occupational exposure, notably to asbestos, as a significant contributor to the observed patterns.

The existing articles above tell us that different industries do have different impacts on different aspects of health. Although the first two articles showed very well that different industries do have different associations with lung cancer rates, they did not take smoking into account in their models. The model of this article will introduce the new variable smoking, and control the influence of cigarette consumption and demographic variables. We will run regressions on the employment-to-population ratio and lung cancer rates in different industries. Examine which industry in Canada best explains the lung cancer rate data, excluding the effects of smoking and population.

2 Data Source

The analysis in this essay draws upon a comprehensive dataset to unravel the intricate connections between industrial activities, lung cancer mortality, and the role of cigarette consumption in Canada over the past two decades. The University of Waterloo's dataset on Average Daily Cigarette Consumption by Province, spanning from 1999 to 2020, provides a nuanced perspective on tobacco use trends across the country, allowing for a thorough exploration of its potential influence on lung cancer rates. Complementing this information, Statistics Canada's datasets on Employment by Industry (annual), Population by Province, and Number and Rates of New Cases of Primary Cancer (by cancer type, age group, and sex) offer a comprehensive view of the economic landscape, workforce dynamics, and lung cancer incidence across provinces over the same period. This multifaceted dataset forms the foundation for a rigorous examination of the interplay between industrial factors, cigarette consumption, and lung cancer outcomes, fostering a more holistic understanding of the complexities surrounding health inequalities in Canada.

3 Methodology

The methodology employed in this study seeks to elucidate the relationship between industries with potential lung damage and lung cancer rates while mitigating the confounding impact of cigarette consumption. We will try to find the regression model of the industry that best correlates the variation of lung cancer rate. The theory is that if there are no health inequities problems between industries, then there will be no industry that has a regression model with industry employment rate that explains the lung cancer rate well. If there is such an industry, this industry has the potential to contribute to the lung cancer rate. The regression analyses are performed separately for distinct industry categories. So the dependent variable data which is the lung cancer rate is the same across different industries when we run regressions, we will switch the corresponding data of industry employment rate data when switching the industry.

To discern patterns across Canadian provinces from 2000 to 2020, we initially calculated the ratio of Employment by dividing the total number of people in that industry by the total population and times 100,000 of that province in the corresponding year. providing a proportional measure of industry-specific employment. This ratio allows us to assess the relative contribution of different industries to the workforce in each province over the specified period.

$$\text{employment ratio in this industry(per 100,000)} = \text{number of people in that industry} \div \text{total population} \times 100,000$$

Subsequently, we conduct multiple linear regressions, treating lung cancer rate as the dependent variable and industry-specific employment ratios as independent variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where Y denotes the lung cancer rate (cases per 100,000 population) for a given province and year, the independent variables include X_1(per 100,000 population), the average employment ratio in a specific industry for the corresponding province and year, X_2 the average daily cigarette consumption for the corresponding province and year, X_3 the population (in 100,000) of the province in that year. The list of industries considered in this analysis encompasses a diverse range, such as Accommodation and Food Services, Agriculture, Business, Building, and Other Support Services, Construction, Educational Services, Finance, Insurance, Real Estate, Rental, and Leasing, among others.

To assess the relationship between each industry's employment rate X_1 and the lung cancer rate Y while controlling for daily cigarette consumption X_2. The coefficients (β_1, β_2, and β_3) provide insights into the strength and direction of these relationships. The inclusion of the average daily cigarette consumption and population as control variables ensures that the observed associations are specific to industrial factors, excluding the confounding influence of smoking habits and population size.

The statistical indicators, including R^2 and p-values are crucial for evaluating the goodness of fit and the significance of the regression models. These metrics aid in determining the explanatory power of the model.

4 Results

4.1 Scatter Plot

NAICS is the abbreviation of the North American Industry Classification System. In Figures 1 and 2, the scatter plots visualize the lung cancer rate (unit: cases per 100,000 population) against the ratio of employment (unit: per 100,000 population) for various industries. Each point corresponds to data for a Canadian province in one year. It can be seen from the figure that there is no obvious relationship between the employment ratio of most industries and the lung cancer rate. This is actually not that surprising, because an industry such as the energy industry is also divided into many types, such as thermal power plants. More workers may have to deal with coal mines and there is a potential risk of lung damage. However, clean energy such as wind power, those working in Those who are probably more likely to be engineers, porters, etc. are not at such a high risk of lung injury. Interestingly, it can be seen from the figure that industries such as Public Administration and Health Care and social assistance seem to have a positive correlation with lung cancer rates which means that the greater the proportion of the population employed in these industries associated with a greater lung cancer rate.

Figure 1: Lung Cancer Rates versus Industry-Specific Employment Ratios, Stratified by Industry - A Visual Exploration Across Canadian Provinces (2000-2020)

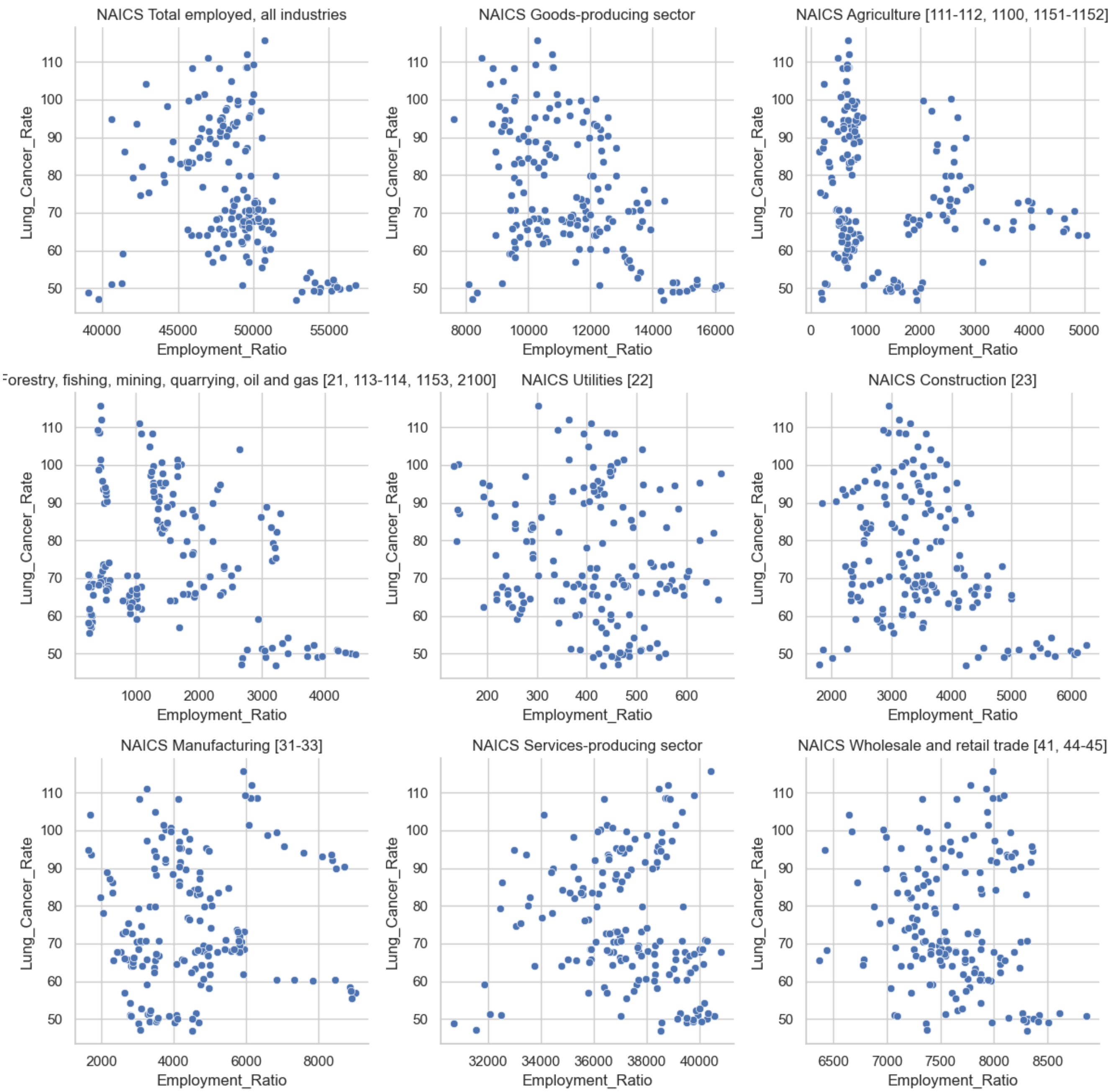
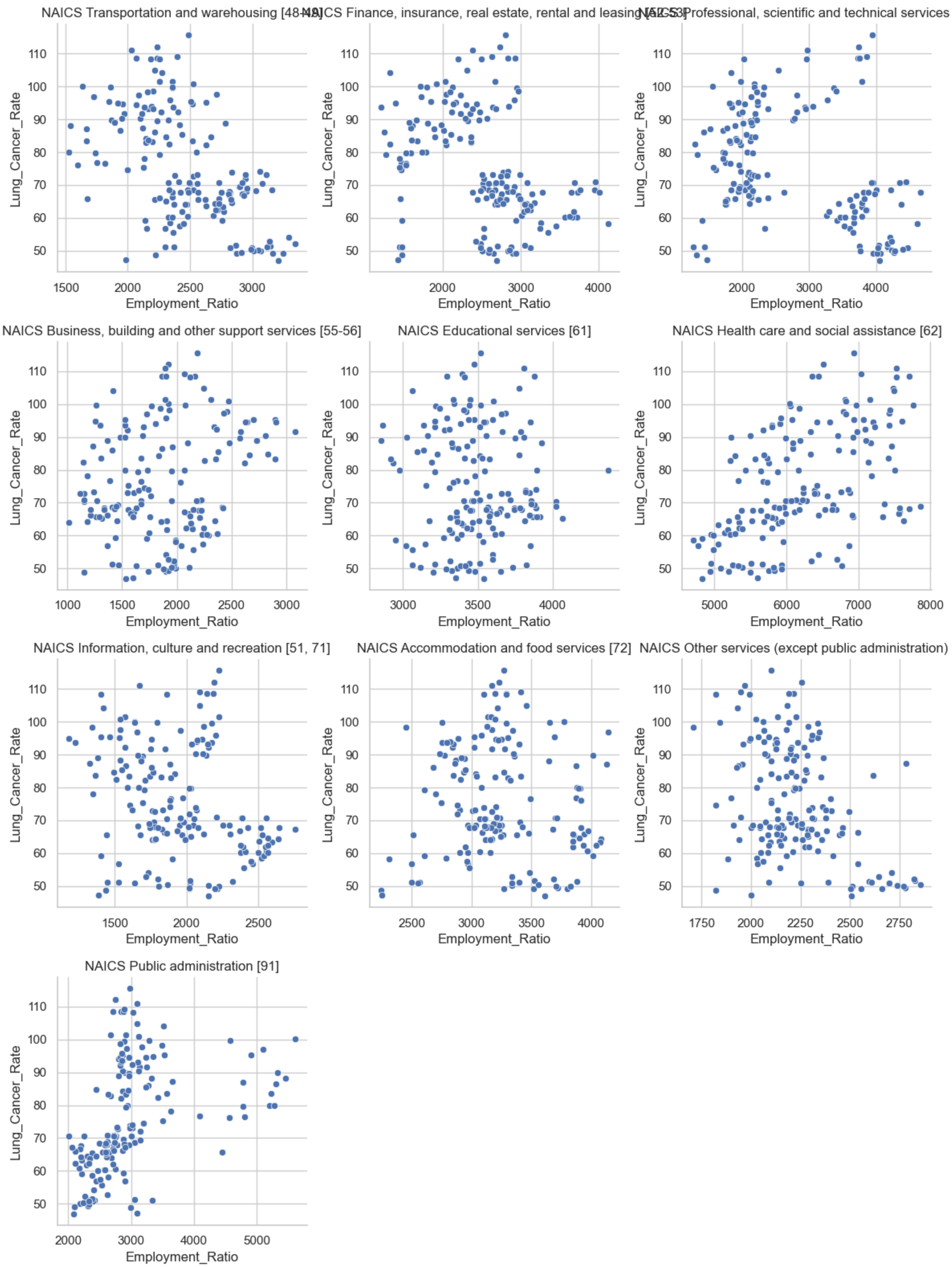


Figure 2: Lung Cancer Rates versus Industry-Specific Employment Ratios, Stratified by Industry - A Visual Exploration Across Canadian Provinces (2000-2020)



4.2 Regression Result

We use the following index to indicate the industry. For example number 5 is the construction industry.

	NAICS
0	Total employed, all industries
1	Goods-producing sector
2	Agriculture [111-112, 1100, 1151-1152]
3	Forestry, fishing, mining, quarrying, oil and gas [21, 113-114, 1153, 2100]
4	Utilities [22]
5	Construction [23]
6	Manufacturing [31-33]
7	Services-producing sector
8	Wholesale and retail trade [41, 44-45]
9	Transportation and warehousing [48-49]
10	Finance, insurance, real estate, rental and leasing [52-53]
11	Professional, scientific and technical services [54]
12	Business, building and other support services [55-56]
13	Educational services [61]
14	Health care and social assistance [62]
15	Information, culture and recreation [51, 71]
16	Accommodation and food services [72]
17	Other services (except public administration) [81]
18	Public administration [91]

Table 1: Regression Results - Examining the Impact of Industry-Specific Employment Ratios on Lung Cancer Rates, Controlling for Cigarette Consumption and Population Across Canadian Provinces (2000-2020)

	0	1	2	3	4	5	6
const	139.7872***	104.7895***	71.1319***	76.9219***	67.3453***	101.0975***	61.0601***
	(25.2542)	(13.1088)	(12.2081)	(11.2076)	(13.2520)	(14.7699)	(11.9687)
Employment_Ratio	-0.0016***	-0.0041***	-0.0044***	-0.0079***	-0.0209*	-0.0067***	0.0033***
	(0.0004)	(0.0006)	(0.0011)	(0.0012)	(0.0108)	(0.0014)	(0.0011)
Smoke_per_day	0.8724	1.3150*	1.0046	1.1532	1.3246	0.0337	0.3824
	(0.7929)	(0.7242)	(0.7841)	(0.7208)	(0.8134)	(0.8160)	(0.8655)
Population	-0.0327	-0.0608**	-0.1199***	-0.1747***	-0.0785**	-0.0784**	-0.1631***
	(0.0335)	(0.0293)	(0.0336)	(0.0328)	(0.0326)	(0.0308)	(0.0445)
R-squared	0.1244	0.2461	0.1305	0.2547	0.0733	0.1640	0.0977
R-squared Adj.	0.1081	0.2320	0.1143	0.2408	0.0559	0.1484	0.0809

Table 2: Regression Results - Examining the Impact of Industry-Specific Employment Ratios on Lung Cancer Rates, Controlling for Cigarette Consumption and Population Across Canadian Provinces (2000-2020)

	7	8	9	10	11	12
const	55.5360*	86.5221***	131.0246***	100.8941***	84.1514***	47.0910***
	(32.3083)	(23.4215)	(14.2166)	(14.9020)	(13.6860)	(12.1781)
Employment_Ratio	0.0001	-0.0043	-0.0221***	-0.0141***	-0.0075***	0.0103***
	(0.0007)	(0.0030)	(0.0029)	(0.0031)	(0.0020)	(0.0030)
Smoke_per_day	1.4048*	1.6194*	0.0279	0.4522	0.6878	0.8675
	(0.8435)	(0.8230)	(0.7188)	(0.7917)	(0.8014)	(0.7982)
Population	-0.0751**	-0.0612*	-0.0497*	0.1009**	0.0559	-0.1051***
	(0.0369)	(0.0338)	(0.0283)	(0.0492)	(0.0468)	(0.0329)
R-squared	0.0511	0.0631	0.3029	0.1595	0.1270	0.1166
R-squared Adj.	0.0334	0.0457	0.2899	0.1438	0.1107	0.1002

Table 3: Regression Results - Examining the Impact of Industry-Specific Employment Ratios on Lung Cancer Rates, Controlling for Cigarette Consumption and Population Across Canadian Provinces (2000-2020)

	13	14	15	16	17	18
const	74.4460***	-90.1086***	80.3293***	69.4024***	137.6891***	33.9077***
	(26.4722)	(17.6064)	(13.6388)	(16.4376)	(17.1726)	(12.0856)
Employment_Ratio	-0.0039	0.0159***	-0.0159***	-0.0034	-0.0330***	0.0094***
	(0.0054)	(0.0016)	(0.0048)	(0.0032)	(0.0054)	(0.0018)
Smoke_per_day	1.1798	4.3520***	1.7414**	1.3442*	0.9746	0.9455
	(0.8622)	(0.7032)	(0.7930)	(0.8106)	(0.7359)	(0.7525)
Population	-0.0793**	0.0692**	0.0165	-0.0775**	-0.0890***	-0.0055
	(0.0336)	(0.0295)	(0.0418)	(0.0329)	(0.0297)	(0.0328)
R-squared	0.0541	0.4151	0.1121	0.0577	0.2280	0.1955
R-squared Adj.	0.0364	0.4042	0.0955	0.0401	0.2136	0.1805

The multiple regression analysis results, summarized in Table 1, Table 2, and Table 3, provide valuable insights into the relationships between industry-specific factors and lung cancer rates while controlling for average daily cigarette consumption and population size. The standard error is in the parentheses. The statistically significant level is *: p-value less than 0.1, **: p-value less than 0.05, and ***: p-value less than 0.01. The dependent variable is the lung cancer rate (cases per 100,000 population) and the number indicates the industry category.

The const is the average lung cancer rate (per 100,000 population) in each province in each year in Canada when no people are working and smoking. For example, in category 0 which is all industries, the average lung cancer rate in each province each year is 139.7872 cases per 100,000 people. The coefficient for the Employment Ratio indicates the expected change in the average lung cancer rate(per 100,000 population) for a one-unit change in the employment ratio in this industry(per 100,000 people) in each province per year while controlling the effect of cigarette consumption per day and population. For example, in category 0 the value is -0.0016 which means on average in each province each year that if 1 of 100,000 people working in this category increases, it is associated with 0.0016 of 100,000 people of lung cancer decreases. The coefficient for Smoke per day indicates the expected change in the average lung cancer rate(per 100,000 population) for a one-unit change in cigarette consumption per day in each province per year while controlling other variables. For example, in category 0, the value is 0.8724, which means on average in each province each year that if the daily cigarette consumption increases by 1, it is associated with a unit 0.8724 of 100,000 people of lung cancer increases. The coefficient for Population indicates the expected change in the average lung cancer rate(per 100,000 population) for a one-unit change in population(in 100,000 people) while controlling other variables. In category 0, the value is -0.0327, which means on average in each province each year that if a 100,000 population increases, it is associated with a unit 0.0327 of 100,000 people of lung cancer decreases for that industry.

Because there are too many industries, the regression results have too many coefficients. Figures 3(a) and 3(b) visualize the coefficients of Employment Ratio and Smoke per day with bar charts. Figures 4(a) and 4(b) visualize the coefficients of population and R square of each regression model with bar charts that provide the direction and size of each coefficient. We will focus on the four industries with the highest R-squared in Figure 4b. They are Industry 1(Goods-producing sector), 3(Forestry, fishing, mining, quarrying, oil and gas), 9(Transportation and warehousing), and 14(Health care and social assistance). They have corresponding R-square values: 0.2461, 0.2547, 0.3029, and 0.4151. These four industries have better data that explain the variation of the lung cancer rate better.

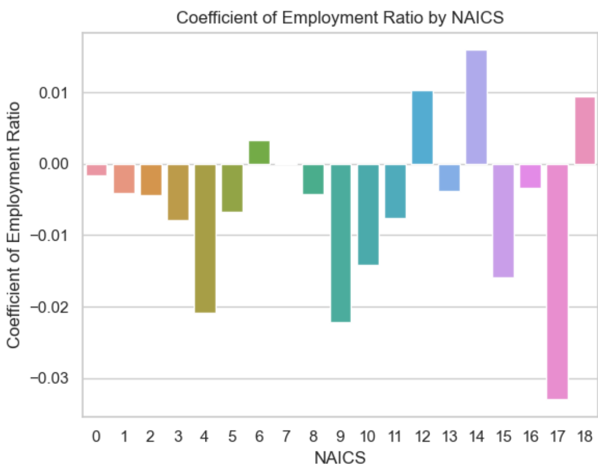
Industry 1 (Goods-producing sector) although explains almost 25% of the variation of lung cancer rate data. From Figures 3(a), 3(b), and 4(a), we can see these coefficients are small which means in this industry the changes in average employment rate across provinces have a very small negative relationship with lung cancer rate while controlling cigarette consumption per day and population.

In Industry 3 (Forestry, Fishing, Mining, Quarrying, Oil and Gas), higher employment ratios are associated with lower lung cancer rates, suggesting a potential protective effect. The impact of daily cigarette consumption on lung cancer rates is not statistically significant in this industry. The negative coefficient for the population variable implies that higher population sizes are associated with lower lung cancer rates in Industry 3.

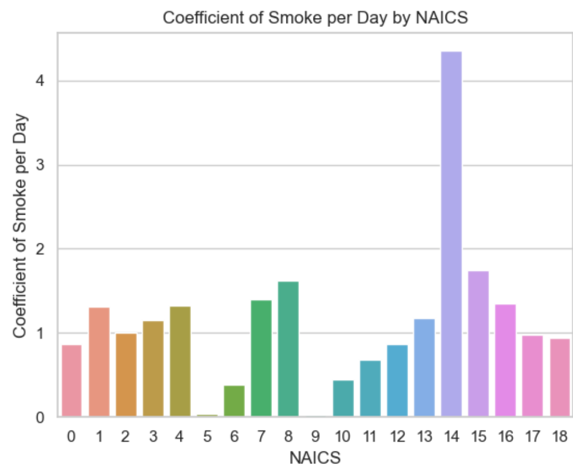
In Industry 9 (Transportation and warehousing), higher employment ratios are associated with lower lung cancer rates, suggesting a potential protective effect. The impact of daily cigarette consumption on lung cancer rates is not statistically significant in this industry. The negative coefficient for the population variable implies that higher population sizes are associated with lower lung cancer rates in Industry 9.

In Industry 14(Health care and social assistance), higher employment ratios are associated with higher lung cancer rates, suggesting a positive association. The positive coefficient for the population variable implies that higher population sizes are associated with higher lung cancer rates in Industry 14. While the positive correlation between employment and lung cancer rates is observed in Industry 14, the high coefficient for cigarette consumption suggests a need for cautious interpretation, so there is a potential influence of omitted variables, particularly smoking habits.

Figure 3

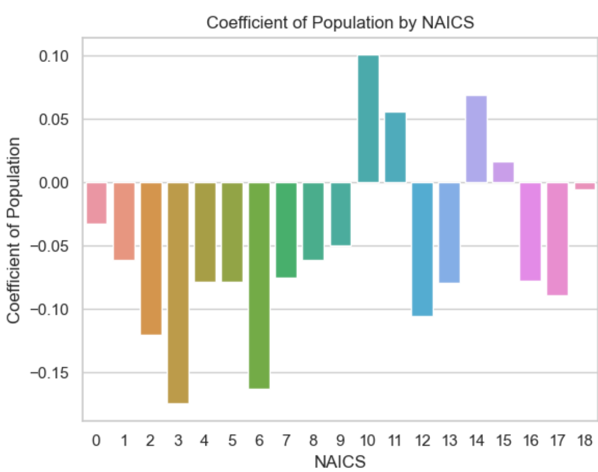


(a) Bar Graph: Coefficients of Employment Ratio across industries

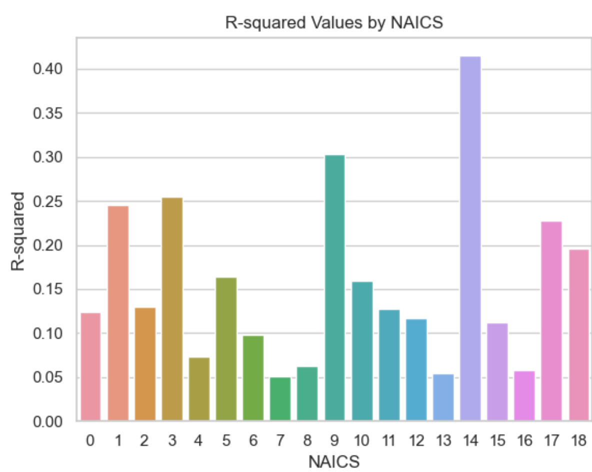


(b) Bar Graph: Coefficients of Cigarette Consumption per day across industries

Figure 4



(a) Bar Graph: Coefficients of Population across industries



(b) Bar Graph: R square for regression results across industries

5 Conclusion

In conclusion, this two-decade analysis investigating the relationship between participation rates in lung-damaged industries, cigarette consumption, and lung cancer mortality across Canadian provinces reveals nuanced and diverse associations. While the regression model for the health care and social assistance industry explained the highest proportion of variation in lung cancer rates (41%), caution is warranted due to the observed positive correlation with cigarette consumption. Industries such as the goods-producing sector, forestry, fishing, mining, quarrying, oil and gas, and transportation and warehousing exhibited intriguing patterns, with higher employment ratios in the latter two associated with lower lung cancer rates.

While the chosen methodology provides a robust framework for investigating the relationship between working industries and lung cancer rates, it is important to acknowledge certain limitations inherent in the study design. Utilizing provincial-level data introduces the possibility of regional biases and undetected variations within provinces, potentially overlooking localized patterns that may exist at a more representative level. Additionally, the time lag between working in a particular industry and the development of lung cancer introduces a temporal challenge. Lung cancer, often associated with prolonged exposures, may not manifest immediately, leading to delayed detection of its association with occupational factors. This time discrepancy poses a challenge in establishing a direct and immediate link between current industrial exposures and contemporaneous lung cancer rates. This model can be improved when using small-scope data such as individual data.

References

1. Government of Canada, Statistics Canada. (2021, January 8). Employment by industry, annual, provinces and economic regions, inactive. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410009201#archived

2. Government of Canada, Statistics Canada. (2023a, May 16). Number and rates of new cases of primary cancer, by cancer type, age group and sex. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310011101&pickMembers%5B0%5D=2.1&pickMembers%5B1%5D=3.1&pickMembers%5B2%5D=4.22&cubeTimeFrame.startYear=2016&cubeTimeFrame.endYear=2020&referencePeriods=20160101%2C20200101

3. Government of Canada, Statistics Canada. (2023b, September 27). Population Estimates, quarterly. Population estimates, quarterly. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=1999&cubeTimeFrame.endMonth=10&cubeTimeFrame.endYear=2022&referencePeriods=19990101%2C20221001

4. Smoking in the provinces. Tobacco Use in Canada. (2023, October 6). https://uwaterloo.ca/tobacco-use-canada/adult-tobacco-use/smoking-provinces

5. BLOT, W. J., & FRAUMENI, J. F. (1976). GEOGRAPHIC PATTERNS OF LUNG CANCER: INDUSTRIAL CORRELATIONS. American Journal of Epidemiology, 103(6), 539–550. https://doi.org/10.1093/oxfordjournals.aje.a112258

6. Miller, A. B. (1977). Recent trends in lung cancer mortality in Canada. Canadian Medical Association Journal, 116(1), 28a–230.

7. Finkelstein, M. M., & Verma, D. K. (2004). A cohort study of mortality among Ontario pipe trades. Occupational and Environmental Medicine (London, England), 61(9), 736-.