

## Article

# Analysis of environmental factors caused by exposure to air pollution in Yeosu industrial complex: Analyzing Google Trend Trends in Cancer Generation through Big Data

Kil Yong Choi<sup>1</sup>, BuSoon Son<sup>2</sup>, WonHo Yang<sup>3\*</sup>

<sup>1</sup> Department of Environmental Energy Engineering, Anyang University, Anyang-si 14028, Republic of Korea; bestchoi94@anyang.ac.kr (K.Y.C);

<sup>2</sup> Department of Environmental Health Science, Soonchunhyang University, Asan-si 31538, Republic of Korea; sonbss@sch.ac.kr (B.S.S.)

<sup>3</sup> Department of Occupational Health, Catholic University of Daegu, Gyeongsan-si 38430, Republic of Korea; whyang@cu.ac.kr (W.H.Y.)

\* Correspondence: whyang@cu.ac.kr; Tel.: +82-53-850-3739, Fax : +82-53-850-3736

**Abstract:** Depending on the characteristics of the industrial area, toxicity evaluation of human body, risk assessment and health impact assessment may directly cause cancer due to air pollution. Environmental data collection is from August 2018 to January 31, 2019, and the average, minimum, and maximum values of air pollution data respectively. According to the global data on global trends using the Big Data, high blood pressure is confirmed at 33rd place in the world, and myocardial infarction among the environmental diseases is confirmed to be lower than Korea. Disease that occurred in Jeolla province industrial complex considering the characteristics of our country was identified as representative. Air pollutants are considered to be the causes of allergic diseases in Korea. PM10 was found to be higher than the control area (28.8804348 ( $\mu\text{g} / \text{m}^3$ ), 31.7065217 ( $\mu\text{g} / \text{m}^3$ ) and 32.8532609 ( $\mu\text{g} / \text{m}^3$ ). The mean concentrations of PM2.5 in the middle and high exposure areas were lower than those of the control areas, but the highest in the intermediate exposure areas was 16.5978261 ( $\mu\text{g} / \text{m}^3$ ), 16.1086957 ( $\mu\text{g} / \text{m}^3$ ) and 17.1847826 ( $\mu\text{g} / \text{m}^3$ ) respectively. The relationship between the major variables of environmental exposure in Yeosu was confirmed to be correlated with high blood pressure, chronic obstructive pulmonary disease (COPD), bronchitis, cerebrovascular, diabetes, thyroid disease, sinus infection, anemia and pneumonia.

**Keywords:** Google Trend; Particulate Matter; National Ambient Air Quality Monitoring Information System; Chronic obstructive pulmonary disease; Big Data

## 1. Introduction

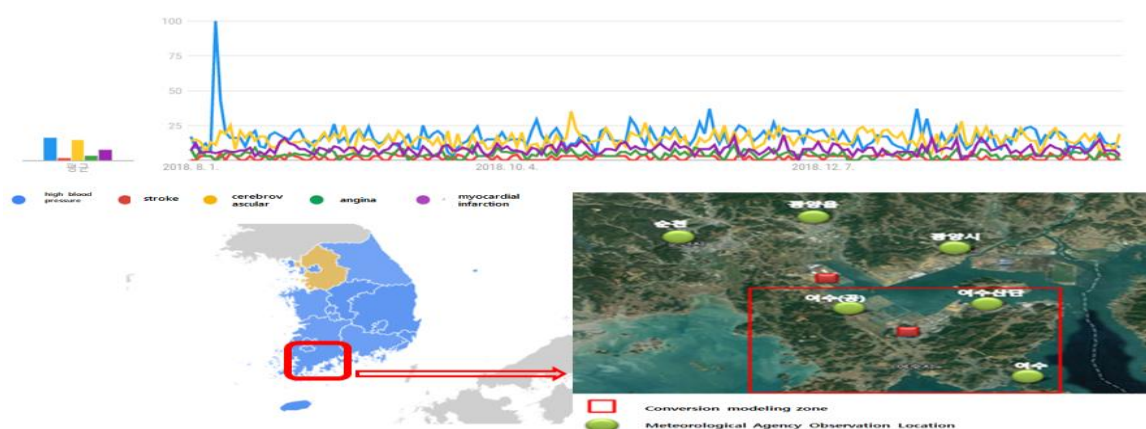
In Korea, diseases caused by the environment exposed to the occupants in the industrial complex area can occur. This has been found to affect exposure of secondary pollutants through diffusion of the atmosphere [1, 2]. Depending on the characteristics of the industrial area, toxicity evaluation of human body, risk assessment and health impact assessment may directly cause cancer due to air pollution [3, 4]. Previous studies of the past have shown that air pollution directly affects lung disease and respiratory diseases in the United States in 1948, a case of air pollution in Pennsylvania, USA [5, 6, 7]. In Korea, it has been confirmed that the application of industrial deterrence (Yeosu, Gwangyang, Ulsan, Pohang, Ansan) has increased as a factor of cancer occurrence and disease in recent years through the process of establishing industrial complexes in major economic

development since 1970 [8, 9] Among them, heavy metals and chemicals in Korea are identified as air pollutants, and fine dust due to fuel consumption is also confirmed [10, 11]. As a result, the emission factors of odor, air pollution, soil pollution, and heavy metals directly affect the residents of the area due to the composition of the industrial complex. Through these results, it becomes a socially very serious environmental impact. Based on this, epidemiological studies have been actively conducted in Korea, and the characteristics of this study are to provide a management method to reduce the possibility of exposure to environmental pollutants considering regional characteristics [12, 13]. It is necessary to investigate the health damage caused by exposure to environmental pollutants in the industrial complex area using BIG DATA, and to identify the cause of cancer in residents of exposed areas [14, 15]. Therefore, it is necessary to confirm whether the exposure from environmental pollution directly affects the health effect in the industrial complex area [16, 17, 18], and it is necessary to use the BIG DATA as the basis of various studies for the purpose of environmental health for evaluation. This study has been actively carried out in Korea recently to develop the major factors by grasping the magnitude and severity of health impacts and grasping pathways for exposure to environmental harmful factors [19]. I would like to identify the cause of long-term exposure by using direct treatment of cancer occurrence and hospital treatment by checking with Google trend through exposure data in the area.

## 2. Materials and methods

### 2.1. Search term (GT, Google Trend)

As an important indicator of this study, environmental data were collected by focusing on "Percentage Per Day in the Year of 2018". Based on previous research data, we searched based on multi-disciplinary research methodology with expert (e.g., Clinical specialists, epidemiologists, research methodologists and statisticians) knowledge. The search was done in a quantitative way to minimize reliability and risk. According to the global data on global trends using the Big Data, high blood pressure is confirmed at 33rd place in the world, and myocardial infarction among the environmental diseases is confirmed to be lower than Korea. Based on this search, we will examine the tendency of cancer diagnoses in KOREA (Figure 1).



**Fig. 1** Comparisons of global Google Trends (GT) and environmental concerns with countries

### 2.2. Air pollution measurement data

The air pollution concentration data of the measuring station were measured as the air pollution monitoring network data operated by the national and local governments. In this study, the measured values of fine dust (PM10, PM2.5) annual mean value (less than  $50 \mu\text{g} / \text{m}^3$ , less than  $25 \mu\text{g} / \text{m}^3$ ) and 24h average (less than  $100 \mu\text{g} / \text{m}^3$ , less than  $50 \mu\text{g} / \text{m}^3$  /  $\beta$ -Ray Absorption Method). This method captures suspended particulate matter suspended in the air on a filter paper for a

certain period of time, permeates the betaine, and continuously measures the weight concentration of the particulate matter.

The sampling flow rate should be constantly inhaled at a set flow rate of 16.7 liters per minute for the correct separation of particulate matter and the flow rate (15.0 liters per minute or less or 18.4 liters per minute or more) The sampling should be discontinued and the sampler checked. The concentration data measured at the measuring station were analyzed using the information gathered by the National Ambient Air Quality Monitoring Information System (NAMIS) based on the Internet.

2.3. Statistical analysis

Statistical analysis was done using the statistical program SAS 9.4 and the log value was taken for the normality test of the concentration distribution of air pollution in the search of Google Trends. After confirming the normal value the geometric mean (GM) and geometric standard deviation (GSD). We confirmed the correlation of environmental concentration values by checking the average search data of Google Trends for environmental diseases (e.g., high blood pressure, COPD: chronic obstructive pulmonary disease, asthma) on a daily basis. The t-test was conducted by applying environmental variables and continuous variables (disease of Google Trends).

3. Results

3.1. PM<sub>10</sub> concentration in air pollution in Korea

The disease (e.g., Major diseases in the region: high blood pressure, stroke, cerebrovascular, angina, myocardial infarction, diabetes, anemia, COPD: chronic obstructive pulmonary disease, bronchitis) that occurred in Jeolla province industrial complex considering the characteristics of our country was identified as representative. Air pollutants are considered to be the causes of allergic diseases in Korea. It was found that the concentration of airborne pollutants was slightly higher in the high and medium exposure areas than in the control areas of the substance. Figure 2 shows that the environmental data collection is from August 2018 to January 31, 2019, and the average, minimum, and maximum values of air pollution data respectively.

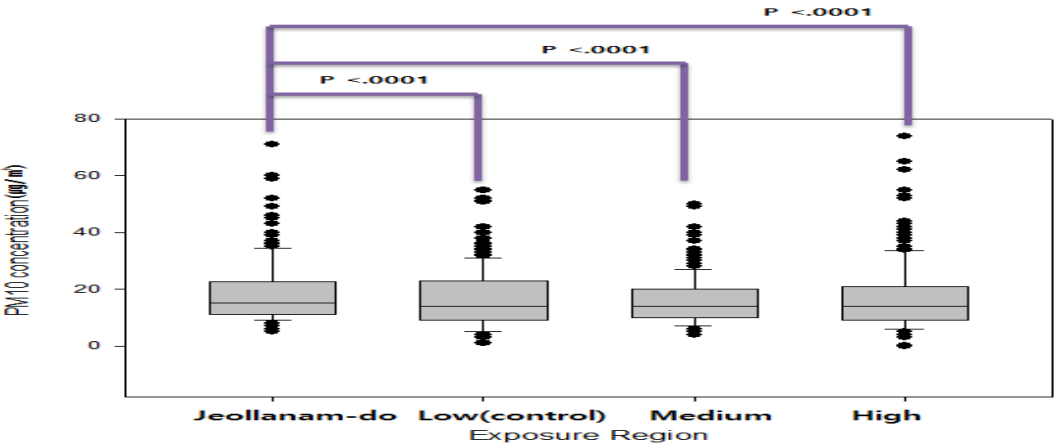


Fig. 2 Analysis of Correlation between Low, Medium and High Exposure (PM<sub>10</sub>) Areas in Jeollanam - d  
o using Air Korea Data

Table 1 shows the results of checking the exposure area of the environment through the indicators of environmental data. Compared with the control area (Yulchon-myeon), the mean values of the medium (e.g., Myongdong-dong, Dundek-dong, Huyang-myeon and Yeochun-dong) and high exposure areas (e.g., Samil-dong and Haesan-dong) were statistically significant. Compared to the control area, the PM<sub>10</sub> (Particulate Matter) was found to be higher than the

control area (28.8804348 ( $\mu\text{g} / \text{m}^3$ ), 31.7065217 ( $\mu\text{g} / \text{m}^3$ ) and 32.8532609 ( $\mu\text{g} / \text{m}^3$ ). The mean concentrations of PM2.5 in the middle and high exposure areas were lower than those of the control areas, but the highest in the intermediate exposure areas was 16.5978261 ( $\mu\text{g} / \text{m}^3$ ), 16.1086957 ( $\mu\text{g} / \text{m}^3$ ) and 17.1847826 ( $\mu\text{g} / \text{m}^3$ ) respectively.

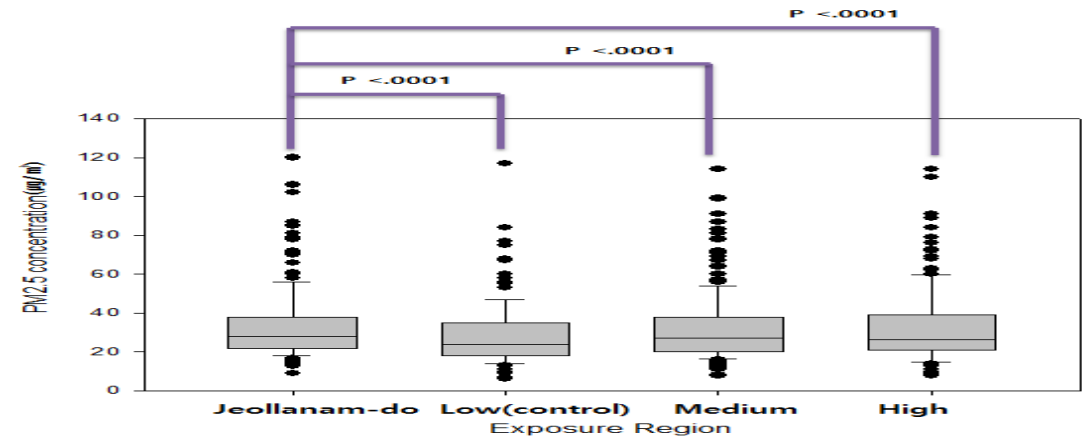
**Table 1** Measurement of outdoor air pollution concentration using Air Korea data

	N	Minimum	Maximum	Average
PM <sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )*	184	28.8804348	6	117
PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )*	184	16.5978261	1	55
PM <sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )**	184	31.7065217	8	114
PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )**	184	16.1086957	4	50
PM <sub>10</sub> ( $\mu\text{g}/\text{m}^3$ ***	184	32.8532609	8	114
PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ ***	184	17.1847826	0	74

\* Low(control) exposure region, \*\* Medium exposure region, \*\*\* High exposure region

3.2. PM<sub>2.5</sub> concentration in air pollution in Korea

The results of analyzing air pollution according to the characteristics of the industrial complex in Korea were confirmed. This confirmed the risk of air pollution in Jeollanam-do due to the high exposure area after the fine dust was designated as a special law by the National Institute of Environmental Science. In addition, it is necessary to confirm the cancer diagnosis data of the disease and to discuss ways to manage it postoperatively. This will require management according to the characteristics of Korea’s industrial complex, and the overall concentration of Jeolla-do was statistically significant in low, medium, and high concentrations. Considering these points in the future, the high concentration of Jeolla Province should be managed. This is thought to be the cause of cancer in the long term, and may affect even chronic diseases through management.



**Fig. 3** Analysis of Correlation between Low, Medium and High Exposure (PM<sub>2.5</sub>) Areas in Jeollanam - do using Air Korea Data.with countries

Table 2 shows the results of correlation analysis between air pollution data (atmospheric environment) and diseases are as follows. In PM10, medium (1.15(1.07-1.23)) and high (1.99(1.97-2.01)) exposure were statistically significant, respectively, compared to low exposure. Even in PM2.5, medium (1.20(1.11-1.30)) and high (1.93(1.88-1.98)) exposure were statistically significant, respectively, compared to low exposure. High exposure areas were higher than low exposure and medium exposure.

**Table 2** Analysis of Correlation between Air Pollution Data (atmospheric environment) and Disease

control region	Mean	95% CL Mean		Pr >  t
Medium exposure area****	1.1486	1.0657	1.2316	<.0001***
High exposure area****	1.9911	1.9734	2.0088	
Medium exposure area*****	1.2027	1.1089	1.2965	<.0001***
High exposure area*****	1.9286	1.8801	1.977	

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$

\*\*\*\* control region, Medium exposure area, High exposure area

\*\*\*\*\* control region, Medium exposure area, High exposure area

### 3.3. Correlation between Air Pollution and Disease

As shown in Table 3, the relationship between PM<sub>10</sub>, which is fine dust in environmental data, and disease was confirmed as follows. As a result, high blood pressure (Higher Region Mean =20.3, 95% CL Mean = 18.6-22.1), COPD (Higher Region Mean =24.4, 95% CL Mean = 20.5-28.3) and pneumonia (Higher Region Mean =40.4, 95% CL Mean = 37.2-43.6) were found to be statistically correlated in the control area ( $P = 0.0286, 0.0437, 0.044$ ). And high blood pressure (Higher Region Mean =20.4, 95% CL Mean = 18.7-22.1), bronchitis (Higher Region Mean =19.6, 95% CL Mean = 15.6-23.6), and pneumonia (Higher Region Mean =40.3, 95% CL Mean = 37.3-43.3) were found to be statistically correlated in the intermediate exposure area ( $P = 0.0227, 0.0487, 0.0176$ ). Finally, in high exposure areas, cerebrovascular (Higher Region Mean =21.7, 95% CL Mean = 18.5-25.0), diabetes (Higher Region Mean =48.9, 95% CL Mean = 46.7-51.0), thyroid disease (Higher Region Mean =33.6, 95% CL Mean = 30.7-36.5), bronchitis (Higher Region Mean =19.2, 95% CL Mean = 15.2-23.1), and pneumonia (Higher Region Mean =40.1, 95% CL Mean = 37.0-43.2) were found to be statistically correlated with each other ( $P = 0.0531, 0.001, 0.0121, 0.033, 0.0109$ ).

**Table 3** Analysis of Correlation between Air Pollution Data (Pm<sub>10</sub>,  $\mu\text{g}/\text{m}^3$ ) and Disease

	Control Region		Medium Exposure area		High Exposure area	
	Mean(95% CL)	P-value	Mean(95% CL)	P-value	Mean(95% CL)	P-value
<sup>1</sup> Disease Lower(average)	24.0(20.9-27.1)	0.0286	24.3(20.9-27.8)	0.0227	23.1(20.7-25.5)	0.1938
Higher(average)	20.3(18.6-22.1)		20.4(18.7-22.1)		21.0(18.9-23.2)	
<sup>2</sup> Disease Lower(average)	13.5(8.8-18.1)	0.9653	15.0(9.8-20.2)	0.4496	14.2(9.2-19.3)	0.6643
Higher(average)	13.3(9.4-17.2)		12.5(8.9-16.2)		12.9(9.1-16.6)	
<sup>3</sup> Disease Lower(average)	26.7(21.9-31.6)	0.1137	27.1(21.8-32.3)	0.1198	27.6(22.1-33.0)	0.0531
Higher(average)	21.9(18.4-25.5)		22.1(18.7-25.5)		21.7(18.5-25.0)	
<sup>4</sup> Disease Lower(average)	27.6(23.7-31.5)	0.8695	28.0(23.9-32.1)	0.9294	28.5(24.3-32.6)	0.7215
Higher(average)	28.0(24.5-31.6)		27.8(24.4-31.2)		27.5(24.1-30.9)	
<sup>5</sup> Disease Lower(average)	37.8(33.9-41.7)	0.5385	38.1(33.6-42.6)	0.7228	39.3(35.1-43.5)	0.7373
Higher(average)	39.3(36.3-42.4)		39.0(36.2-41.9)		38.4(35.4-41.4)	
<sup>6</sup> Disease Lower(average)	53.1(49.2-57.1)	0.1833	52.8(48.6-57.1)	0.326	55.9(51.7-60.2)	0.001
Higher(average)	50.3(48.0-52.5)		50.7(48.4-52.9)		48.9(46.7-51.0)	
<sup>7</sup> Disease Lower(average)	29.7(25.3-34.0)	0.2249	30.8(26.2-35.5)	0.5977	29.0(24.5-33.5)	0.1294



	Higher(average)	33.3(29.3-37.4)		32.4(28.6-36.3)		33.5(29.6-37.4)	
<sup>8</sup> Disease	Lower(average)	38.4(33.8-43.1)	0.1291	38.9(33.7-44.1)	0.1097	40.3(35.5-45.2)	0.0121
	Higher(average)	34.4(31.5-37.4)		34.5(31.7-37.3)		33.6(30.7-36.5)	
<sup>9</sup> Disease	Lower(average)	31.0(25.8-36.2)	0.0437	27.9(22.7-33.2)	0.6656	29.3(23.7-34.8)	0.2993
	Higher(average)	24.4(20.5-28.3)		26.5(22.6-30.5)		25.7(21.9-29.6)	
<sup>10</sup> Disease	Lower(average)	25.7(20.1-31.3)	0.0874	26.7(20.5-33.0)	0.0487	27.2(20.9-33.4)	0.033
	Higher(average)	19.6(15.4-23.9)		19.6(15.6-23.6)		19.2(15.2-23.1)	
<sup>11</sup> Disease	Lower(average)	17.4(11.6-23.3)	0.954	16.5(10.2-22.8)	0.7436	17.1(10.9-23.2)	0.9215
	Higher(average)	17.2(13.0-21.4)		17.7(13.6-21.8)		17.4(13.3-21.5)	
<sup>12</sup> Disease	Lower(average)	14.4(9.8-19.0)	0.8344	12.4(7.8-16.9)	0.4652	15.3(10.5-20.1)	0.507
	Higher(average)	13.7(9.5-17.9)		14.8(10.7-18.9)		13.2(9.2-17.2)	
<sup>13</sup> Disease	Lower(average)	28.5(23.0-33.9)	0.5047	30.4(24.2-36.6)	0.1364	28.7(23.2-34.2)	0.4629
	Higher(average)	26.2(22.2-30.1)		25.3(21.7-29.0)		26.2(22.2-30.1)	
<sup>14</sup> Disease	Lower(average)	4.4(1.2-7.6)	0.3043	4.3(0.9-7.8)	0.3731	4.8(0.7-8.9)	0.5202
	Higher(average)	6.8(3.4-10.3)		6.7(3.4-9.9)		6.5(3.4-9.5)	
<sup>15</sup> Disease	Lower(average)	16.8(11.9-21.8)	0.2543	16.7(11.3-22.2)	0.3354	16.9(11.3-22.5)	0.2998
	Higher(average)	13.3(9.6-17.0)		13.6(10.1-17.1)		13.5(10.1-16.9)	
<sup>16</sup> Disease	Lower(average)	20.5(15.6-25.4)	0.9112	22.7(17.4-28.1)	0.2639	21.5(16.4-26.5)	0.5726
	Higher(average)	20.1(16.0-24.3)		19.0(15.1-22.9)		19.6(15.6-23.6)	
<sup>17</sup> Disease	Lower(average)	8.8(4.3-13.3)	0.697	6.1(2.0-10.2)	0.3004	8.0(3.6-12.5)	0.9444
	Higher(average)	7.7(4.0-11.3)		9.2(5.5-12.9)		8.2(4.6-11.8)	
<sup>18</sup> Disease	Lower(average)	8.8(4.3-13.3)	0.697	6.1(2.0-10.2)	0.3004	8.0(3.6-12.5)	0.9444
	Higher(average)	7.7(4.0-11.3)		9.2(5.5-12.9)		8.2(4.6-11.8)	
<sup>19</sup> Disease	Lower(average)	45.8(41.6-50.1)	0.044	46.9(42.3-51.5)	0.0176	47.0(42.6-51.4)	0.0109
	Higher(average)	40.4(37.2-43.6)		40.3(37.3-43.3)		40.1(37.0-43.2)	
<sup>20</sup> Disease	Lower(average)	2.7(-1.1-6.5)	0.6063	3.1(-1.3-7.5)	0.4482	3.0(-1.2-7.2)	0.5379
	Higher(average)	1.6(-0.6-3.9)		1.5(-0.6-3.6)		1.5(-0.6-3.7)	
<sup>21</sup> Disease	Lower(average)	14.6(8.6-20.7)	0.5671	12.5(6.4-18.5)	0.7172	14.0(8.0-20.0)	0.7952
	Higher(average)	12.5(8.5-16.6)		13.8(9.7-18.0)		13.0(8.9-17.2)	

<sup>1</sup>High blood pressure, <sup>2</sup>Stroke, <sup>3</sup>Cerebrovascular, <sup>4</sup>Angina, <sup>5</sup>Myocardial infarction, <sup>6</sup>Diabetes, <sup>7</sup>Anemia, <sup>8</sup>Thyroid disease, <sup>9</sup>COPD, chronic obstructive pulmonary disease, <sup>10</sup>Bronchitis, <sup>11</sup>Pulmonary emphysema, <sup>12</sup>Bronchiectasis, <sup>13</sup>Asthma, <sup>14</sup>Sinus infection, <sup>15</sup>Allergic coryza, <sup>16</sup>Allergy desmatitis, <sup>17</sup>Bronchial asthma, <sup>18</sup>Acute bronchitis, <sup>19</sup>Pneumonia, <sup>20</sup>Acute pneumonia, <sup>21</sup>Skin Disease

Table 4 shows the relationship between PM2.5, which is fine dust in environmental data, and disease. In contrast, high blood pressure (Higher Region Mean =20.1, 95% CL Mean = 18.5-21.8), COPD (Higher Region Mean =23.8, 95% CL Mean = 20.1-27.5), sinus infection (Higher Region Mean =7.9, 95% CL Mean = 4.2-11.5), and pneumonia (Higher Region Mean =40.3, 95% CL Mean = 37.1-43.4) were found to be statistically correlated with each other (P = 0.0157, 0.019, 0.0518, 0.035). High-blood pressure, cerebrovascular, anemia, and pneumonia were found to be statistically correlated in the middle-exposed areas. High blood pressure, cerebrovascular, thyroid disease, COPD and pneumonia were found to be statistically correlated in the high exposure area.

**Table 4** Analysis of Correlation between Air Pollution Data (Pm2.5, µg/m³) and Disease

	Control Region		Medium Exposure area		High Exposure area	
	Mean(95% CL)	P-value	Mean(95% CL)	P-value	Mean(95% CL)	P-value
<sup>1</sup> Disease Lower(average)	24.2(21.0-27.3)	0.0157	23.7(20.6-26.8)	0.0546	24.4(21.1-27.6)	0.0148
Higher(average)	20.1(18.5-21.8)		20.5(18.7-22.3)		20.2(18.5-21.9)	
<sup>2</sup> Disease Lower(average)	12.6(8.6-16.6)	0.6754	14.0(9.8-18.1)	0.7465	13.3(9.1-17.6)	0.981
Higher(average)	13.9(9.7-18.1)		13.0(8.8-17.1)		13.4(9.4-17.4)	
<sup>3</sup> Disease Lower(average)	26.9(22.1-31.8)	0.0813	27.5(22.3-32.6)	0.0445	27.8(22.6-33.0)	0.0449
Higher(average)	21.7(18.2-25.2)		21.5(18.2-24.8)		21.5(18.1-24.9)	
<sup>4</sup> Disease Lower(average)	27.3(23.4-31.2)	0.7181	27.2(23.3-31.1)	0.6928	26.7(22.6-30.9)	0.5183
Higher(average)	28.2(24.7-31.8)		28.3(24.7-31.8)		28.5(25.1-31.9)	
<sup>5</sup> Disease Lower(average)	39.2(35.5-42.9)	0.7481	39.5(35.7-43.4)	0.5957	39.7(35.5-43.9)	0.5395
Higher(average)	38.4(35.2-41.6)		38.2(35.1-41.3)		38.1(35.2-41.1)	
<sup>6</sup> Disease Lower(average)	53.3(49.3-57.2)	0.1434	53.9(49.9-57.8)	0.0604	52.9(48.8-57.0)	0.279
Higher(average)	50.1(48.0-52.3)		49.8(47.6-52.1)		50.5(48.3-52.8)	
<sup>7</sup> Disease Lower(average)	29.0(24.8-33.2)	0.0971	28.2(23.8-32.6)	0.0458	30.7(26.5-35.0)	0.5571
Higher(average)	33.9(29.8-38.0)		34.2(30.3-38.2)		32.6(28.5-36.6)	
<sup>8</sup> Disease Lower(average)	37.9(33.4-42.4)	0.2472	38.4(33.9-42.9)	0.1495	39.5(34.6-44.3)	0.0405
Higher(average)	34.7(31.6-37.8)		34.5(31.4-37.6)		34.0(31.1-36.9)	
<sup>9</sup> Disease Lower(average)	31.6(26.2-37.0)	0.019	29.3(24.0-34.5)	0.2647	31.6(26.1-37.0)	0.0311
Higher(average)	23.8(20.1-27.5)		25.6(21.6-29.5)		24.3(20.5-28.1)	
<sup>10</sup> Disease Lower(average)	25.1(19.4-30.7)	0.1523	25.5(19.8-31.3)	0.1124	25.8(19.9-31.8)	0.101
Higher(average)	20.0(15.7-24.2)		19.8(15.6-24.0)		19.8(15.7-23.9)	
<sup>11</sup> Disease Lower(average)	17.0(11.3-22.7)	0.8955	16.8(11.0-22.6)	0.8352	17.9(11.8-24.0)	0.7953
Higher(average)	17.5(13.2-21.7)		17.6(13.3-21.8)		16.9(12.8-21.0)	
<sup>12</sup> Disease Lower(average)	14.9(10.3-19.5)	0.6075	14.4(9.9-18.9)	0.8098	15.3(10.6-20.0)	0.4902
Higher(average)	13.3(9.1-17.5)		13.7(9.4-17.9)		13.2(9.1-17.3)	
<sup>13</sup> Disease Lower(average)	28.8(23.8-33.8)	0.3884	26.2(21.0-31.4)	0.664	26.9(21.7-32.2)	0.9455
Higher(average)	25.9(21.8-30.1)		27.7(23.6-31.7)		27.2(23.1-31.2)	
<sup>14</sup> Disease Lower(average)	3.0(0.4-5.6)	0.0518	3.8(0.3-7.4)	0.1743	4.7(0.7-8.7)	0.4656
Higher(average)	7.8(4.2-11.5)		7.2(3.9-10.4)		6.6(3.4-9.7)	
<sup>15</sup> Disease Lower(average)	15.6(10.8-20.5)	0.6064	15.8(10.7-20.9)	0.5501	17.0(11.8-22.3)	0.2409
Higher(average)	14.0(10.3-17.8)		13.9(10.3-17.5)		13.3(9.8-16.8)	
<sup>16</sup> Disease Lower(average)	21.4(16.5-26.2)	0.569	21.8(16.7-26.8)	0.4554	22.3(17.5-27.2)	0.3096
Higher(average)	19.5(15.4-23.7)		19.3(15.3-23.4)		19.1(14.9-23.2)	
<sup>17</sup> Disease Lower(average)	8.0(3.7-12.2)	0.9269	6.0(2.3-9.8)	0.2382	8.7(4.1-13.3)	0.7648
Higher(average)	8.2(4.5-12.0)		9.5(5.6-13.4)		7.8(4.2-11.4)	
<sup>18</sup> Disease Lower(average)	8.0(3.7-12.2)	0.9269	6.0(2.3-9.8)	0.2382	8.7(4.1-13.3)	0.7648
Higher(average)	8.2(4.5-12.0)		9.5(5.6-13.4)		7.8(4.2-11.4)	
<sup>19</sup> Disease Lower(average)	45.9(41.7-50.1)	0.035	46.1(41.9-50.4)	0.0302	46.0(41.6-50.4)	0.0486

Higher(average)	40.3(37.1-43.4)		40.3(37.1-43.4)		40.6(37.5-43.7)	
<sup>20</sup> Disease Lower(average)	2.6(-1.1-6.3)	0.6389	2.7(-1.1-6.6)	0.614	2.9(-1.2-7.0)	0.5261
Higher(average)	1.6(-0.7-4.0)		1.6(-0.6-3.9)		1.6(-0.6-3.7)	
<sup>21</sup> Disease Lower(average)	16.0(9.8-22.3)	0.2003	13.5(7.6-19.4)	0.9569	15.1(8.4-21.7)	0.4458
Higher(average)	11.5(7.7-15.4)		13.3(9.1-17.4)		12.4(8.6-16.1)	

<sup>1</sup>High blood pressure, <sup>2</sup>Stroke, <sup>3</sup>Cerebrovascular, <sup>4</sup>Angina, <sup>5</sup>Myocardial infarction, <sup>6</sup>Diabetes, <sup>7</sup>Anemia, <sup>8</sup>Thyroid disease, <sup>9</sup>COPD, chronic obstructive pulmonary disease, <sup>10</sup>Bronchitis, <sup>11</sup>Pulmonary emphysema, <sup>12</sup>Bronchiectasis, <sup>13</sup>Asthma, <sup>14</sup>Sinus infection, <sup>15</sup>Allergic coryza, <sup>16</sup>Allergy desmatitis, <sup>17</sup>Bronchial asthma, <sup>18</sup>Acute bronchitis, <sup>19</sup>Pneumonia, <sup>20</sup>Acute pneumonia, <sup>21</sup>Skin Disease

4. Discussion

This study was carried out based on BIG DATA for the evaluation of exposure due to environmental pollution in industrial area in recent 1 years. In the case of residential areas for high-exposure areas, middle-exposure areas and low-exposure areas (comparison) of Yeosu in the background of industrial complexes under study, the risk of cancer development including death in medium and high Was confirmed to be high. As a result, consistent disease according to the characteristics of Yeosu area according to all types of cancer will occur and follow up of the disease will be necessary. According to the recent data of Yeosu, the rate of housing distribution has been confirmed as 99.3%. As a result, considering the special features of oil refinery and petrochemical formation in Yeosu National Industrial Complex, It can be confirmed that the diffusion due to pollution is damaged in some areas in the natural environment [20, 21]. In this study, it was also confirmed that the external factors related to the disease are affecting climate change and global warming. The aim of this study was to evaluate the effects of exposure to air pollution on 21 outbreaks of environmental diseases including "High blood pressure, Stroke, Cerebrovascular, Angina, Myocardial infarction, Diabetes, Anemia, Thyroid disease, Chronic obstructive pulmonary disease: COPD, Bronchitis, Pulmonary emphysema, Acute bronchitis, Acute pneumonia, and Skin Disease "were identified in patients with Alzheimer 's disease, Asthma, Sinus infection, Allergic coryza, Allergy desmatitis, Bronchial asthma, Acute bronchitis. This disease was mainly searched and the GT (Google Trends) was used in Korea and international case study, and the systematic review of the search was conducted without any problems [22, 23]. The second was that the disease was exposed to the Korean environment and it was thought that the value as BIG DATA would be increased. Exposure to environmentally hazardous substances is a major variable and should be identified as a major factor in future disease management. As a result of confirming the air quality of the industrial complex in Cholla province where the representative pollutants of Korea are generated, the results of the preliminary analysis of the health index of the target area and the affinity of the health of the urban environment suggests a way to construct the health city of Yeosu , But the analysis of the indicators using BIG data seems to be very poor. In this study, daily mean values were calculated using trend data of PM10 and PM2.5 measured in the air pollution monitoring network in Korea. The data were applied to the requirements of the analysis according to the weather conditions using the generalized linear model, And concentration characteristics of industrial complex in Jeolla province were analyzed. Although there is no limit to the limit of Google Trends, the quality of air collected by the National Institute of Environmental Research (AIR KOREA) has been analyzed since August, 2018, so it is possible to use data for about one year recently. And the relationship between exposure to air and chronic diseases and some diseases. Compared with the control area, the high tendency of air pollution was found to be correlated with some diseases in the medium exposure area and the high exposure area. According to the results of previous studies, the correlation between air pollutants, environmental diseases such as atopic dermatitis and asthma disease has been reported. Based on this, it will be necessary to use BIG DATA for further verification. As a result of completing the primary screening check, it is necessary



to confirm the relationship between exposure to environmental pollution and environmental pollution. The relationship between PM10, which is a fine dust in the environmental data, and disease was confirmed to be correlated with high blood pressure and COPD and Pneumonia in the control area. High blood pressure, bronchitis, and pneumonia were found to be correlated in the intermediate exposure area. Cerebrovascular, Diabetes, Thyroid disease, Bronchitis and Pneumonia were found to be correlated in high exposed areas. The relationship between PM2.5, which is the fine dust in the environmental data, and the disease, was confirmed to be correlated with high blood pressure, COPD, sinus infection, and pneumonia in the control area. High blood pressure, Cerebrovascular, Anemia, and Pneumonia were found to be correlated in the intermediate exposure area. In high exposed areas, high blood pressure, cerebrovascular, thyroid disease, and chronic obstructive pulmonary disease (COPD) and pneumonia were found to be correlated. This study was conducted to confirm the correlation between air pollutants and exposure of the disease and to reduce the time by controlling, matching, and correcting disturbance variables through previous studies. Because of the selective bias, we were able to confirm the importance of this study.

Conclusion, the results show that there is a limit to the information based on some data that affect life style or health that can be obtained based on 21 major variables of exposure of environment designated by NESI (National Environmental Science Institute) , Analyzing the regional impacts and personal factors of women has been limited. There are very few researches to produce these research results in Korea, and it is necessary to use them as appropriate indicators in consideration of the regional characteristics of the infrastructure. The relationship between the major variables of environmental exposure in Yeosu was confirmed to be correlated with high blood pressure, chronic obstructive pulmonary disease (COPD), bronchitis, cerebrovascular, diabetes, thyroid disease, sinus infection, anemia and pneumonia. Therefore, the results obtained through the analysis of this study are meaningful in that they can present directions for future research rather than their own interpretation. It is expected that it will be used as a basic data when conducting studies to evaluate the health effects of exposure to industrial pollution using secondary data. Therefore, in order to investigate further precise epidemiological studies, we conducted a linkage study on environmental pollution data and health insurance data for major diseases that have been found to be significantly related to environmental exposures in industrial complexes or industrial complexes that are related to industrial pollution. You need to do it.

**Acknowledgments:** The study was approved by Soonchunhyang University institutional review board "Title : Monitoring of Environmental Pollutants Exposure and Health Effects in Industrial Residents(Gwangyang-Yeosu, 3rd Phase, 1st Year)" (IRB no. 1040875-201807-BR-028).

**List of abbreviations:** COPD: chronic obstructive pulmonary disease; GT: Google Trend; PM: Particulate Matter; GM: geometric mean; GSD: geometric standard deviation; NAMIS: National Ambient Air Quality Monitoring Information System; NESI: National Environmental Science Institute; CI confidence interval.

**Availability of data and materials:** The datasets analyzed in the current study are available from the corresponding author upon reasonable request.

## References

1. Nam, D. H., LEE, D. P., 2006. Monitoring for Pb and Cd pollution using feral pigeons in rural, urban, and industrial environments of Korea. *Science of the Total Environment* 357(1-3), 288-295.
2. Park, H. j., Lee, K. H., Moon, C. S., Woo, K. S., Kang, T. S., Chung, E. K., Son, B. S., 2015. Simultaneous exposure to heavy metals among residents in the industrial complex: Korean national cohort study. *International journal of environmental research and public health* 12(6), 5905-5917.
3. HERTWICH, E. G., PEASE, W. S., MCKONE, T. E., 1998. *Environmental Policy Analysis: Evaluating Toxic*

- Impact Assessment Methods: What Works Best. *Environmental science & technology* 32(5), 138A-144A.
4. LANDRIGAN, P. J., Kimmel, C. A., Correa, A., Eskenazi, B., 2004. Children's health and the environment: public health issues and challenges for risk assessment. *Environmental health perspectives* 112(2), 257-265.
  5. Gilliland, F. D., McConnell, R., Peters, J., Jr, H. G., 1999. A theoretical basis for investigating ambient air pollution and children's respiratory health. *Environmental health perspectives* 107(3), 403-407.
  6. Joel, S., 1994. What are people dying of on high air pollution days. *Environmental research* 64(1), 26-35.
  7. DOCKERY, D. W., SCHWARTZ, J., SPENGLER, J. D., 1992. Air pollution and daily mortality: associations with particulates and acid aerosols. *Environmental research* 59(2), 362-373.
  8. CHO, S. H., KANG, D. H., KO, K. S., KWON, H. J., KIM, D. H., HA, M., HAN, S. H., JU, Y. S., 1997. Estimates of occupational cancer in Korea. *Journal of occupational health* 39(3), 192-196.
  9. Cho, J. Y., Son, J. G., Park, B. J., Chung, B. Y., 2009. Distribution and pollution sources of polycyclic aromatic hydrocarbons (PAHs) in reclaimed tidelands and tidelands of the western sea coast of South Korea. *Environmental monitoring and assessment* 149(1-4), 385-393.
  10. Kim, K. H., Choi, G. H., Kang, C. H., Lee, J. H., Kim, J. Y., Youn, Y. H., Lee, S. R., 2003. The chemical composition of fine and coarse particles in relation with the Asian Dust events. *Atmospheric Environment* 37(6), 753-765.
  11. Choung, S. W., Oh, J. S., Han, W. S., Chon, C. M., Kwon, Y. S., Kim, D. Y., Shin, W. S., 2016. Comparison of physicochemical properties between fine (PM<sub>2.5</sub>) and coarse airborne particles at cold season in Korea. *Science of the Total Environment* 541, 1132-1138.
  12. Cole, M. A., Elliott, R. J. R., Shimamoto, K., 2005. Industrial characteristics, environmental regulations and air pollution: an analysis of the UK manufacturing sector. *Journal of environmental economics and management* 50(1), 121-143.
  13. Bell, M. L., Cifuentes, L. A., Davis, D. L., Cushing, E., Telle, A. G., Gouvei, N., 2011. Environmental health indicators and a case study of air pollution in Latin American cities. *Environmental research* 111(1), 57-66.
  14. Chen, X., Shao, S., Tian, Z., Xie, Z., Yin, P., 2017. Impacts of air pollution and its spatial spillover effect on public health based on China's big data sample. *Journal of cleaner production* 142, 915-925.
  15. Chen, M., Yang, J., Hu, L., Hossain, M. S., Muhammad, G., 2018. Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *IEEE Communications Magazine* 56(11), 14-20.
  16. Bergstr, A. D., Brunekreef, B., Burdorf, A., 2018. The effect of industry-related air pollution on lung function and respiratory symptoms in school children. *Environmental Health* 17(1), 30.
  17. Torres, P., Ferreira, J., Monteiro, A., Costa, S., Pereira, M. C., Madureira, J., Mendes, A., Teixeira, J. P., 2018. Air pollution: A public health approach for Portugal. *Science of the total environment* 643, 1041-1053.
  18. Chang, T. Y., Zivin, J. G., Gross, T., Neidell, M., 2019. The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China. *American Economic Journal: Applied Economics* 11(1), 151-72.
  19. Wang, Y., Kung, L. A., Wang, W. Y. C., Cegielski, C. G., 2018. An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management* 55(1), 64-79.
  20. Won, Y. J., Jung, K. W., Oh, C. M., Park, E. H., Kong, H. J., Lee, D. H., Lee, K. H., 2018. Geographical Variations and Trends in Major Cancer Incidences throughout Korea during 1999-2013. *Cancer research and treatment: official journal of Korean Cancer Association* 50(4), 1281.
  21. Hong, E. J., Lee, S. W., Kim, G. B., Kim, T. J., Kim, H. W., Lee, K. H., Son, B. S., 2018. Effects of Environmental Air Pollution on Pulmonary Function Level of Residents in Korean Industrial Complexes. *International journal of environmental research and public health* 15(5), 834.
  22. Faridi, S., hamsipour, M., Krzyzanowski, M., Künzli, N., Amini, H., Azimi, F., Malkawi, M., Momeniha, F., Gholampour, A., Hassanvand, M. S., Naddafi, K., 2018. Long-term trends and health impact of PM<sub>2.5</sub> and O<sub>3</sub> in Tehran, Iran, 2006-2015. *Environment international* 114, 37-49.
  23. Morsy, S., Dang, T. N., Kamel, M. G., Zayan, A. H., Makram, O. M., Elhady, M., Hirayama, K., Huy, N. T., 2018. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiology & Infection* 146(13), 1625-1627.