Big Data Analysis to Expand Support for Cervical Cancer Vaccine

Ji-Hye Noh[1], Yu-Ra Choi[2], Yun-Seon Choi[3], Seong-Taek Park[4*]

[1]Department of Information and Statistics, Chungnam National University, Daejeon 34134, South Korea
[2]Department of Mechanical Engineering, Chungnam National University, Daejeon 34134, South Korea
[3]Department of Mathematics, Chungnam National University, Daejeon 34134, South Korea
[4]Strategy Planning Headquarters, CISTEP, Cheonan, 31035, South Korea

Abstract[1]

Background/Objectives: This study is a basic stage of research to expand support for cervical cancer vaccines. Methods/Statistical analysis: Data was collected and refined and analyzed using R and Excel, which are statistical tools for big data analysis. A survey was conducted on college students in their 20s and tried to analyze it. Findings: In addition, as a result of conducting a correlation analysis with the factors affecting cervical cancer through R, it was found that they were related to each other. Improvements/Applications: The results of this study suggested the need for cervical cancer vaccine support from the government and suggested ways to expand it.

Index Terms

Big data, Big data analysis, Cervical cancer, Cervical cancer vaccine, Vaccine support are expanded

Corresponding author : Seong-Taek, Park
solpherd@cistep.re.kr

I. INTRODUCTION

Due to the development of technology, our lives are facing many positive changes[1]. In particular, in the era of the Fourth Industrial Revolution, many benefits are enjoyed. However, due to the spread of COVID-19 in 2020, it is facing a pandemic worldwide[2-4].

Corona has not only a negative impact, but also a positive effect. In particular, studies have shown that infectious diseases have decreased by 46% compared to the previous year as face-to-face contact has decreased due to social distancing and refraining from going out.

Many infectious diseases have appeared even before COVID-19, and various vaccines and treatments have been developed and are currently being used.

Cervical cancer is the most common cancer in Korea among cancers occurring in female genitals and accounts for about 9% of all female cancers. According to data from the Central Cancer Registration Headquarters released in 2017, cervical cancer accounts for about 1.7% of the total cancer incidence in Korea, and ranked seventh among women's cancers. In addition, cervical cancer accounts for about 15% of female cancers worldwide, and ranks second among cancers occurring in women.

The incidence of cancer is slightly decreasing compared to the past, but among all cervical cancer patients, the number of early cervical cancer patients in the first stage is gradually increasing. The prognosis of early diagnosed cervical cancer is good, but if advanced, not only is it difficult to cure, but many patients suffer from side effects from treatment. Therefore, it is important to prevent cervical cancer in advance or diagnose and treat it early [5].

The cause of cervical cancer is HPV (human papilloma virus), which is known to be infected through sexual contact, and 70-80% of women living a normal sex life are infected once in their lifetime. As the age of sexual experience accelerates, the incidence of cervical cancer is increasing rapidly in the younger generation, and according to the Health Insurance Review and Assessment Service, the number of patients in their 20s and 30s increased by about 25% between 2010 and 2014. There is HPV vaccination as a way to prevent cervical cancer, and the cost of vaccination is much higher than that of other vaccines.

Currently, the government has provided free vaccination support for 12-year-old girls since 2016, but the age group for vaccination support is narrower than in developed countries.

In addition, cervical cancer is a disease that women develop, but HPV, the cause of the disease, can be infected in men, so it is necessary to expand vaccination support to men.

II. RELATED WORK

A. *Definition of human papillomavirus (HPV)*

HPV is known as an important cause of cervical cancer and is a dual helical DNA virus belonging to the Papillomavirus family. Of the 100 species of human papilloma viruses known to date, 40 species are found in reproductive organs and are known to cause pathological changes in the cervical epithelium.

Among them, the high-risk group, carcinogenic papilloma virus, is known to be highly associated with cervical cancer, and 16 and 18 of the carcinogenic papilloma viruses are the most important and are found in more than 70% of cervical cancer worldwide.

In addition, 31, 33, 35, 39, 45, 51, 52, 56, 68, 69, and 73 belong to the high-risk virus. Low risk group viruses found in warts or gonus (sexual warts), which are benign pathological changes in the cervix, are 6, 11, 34, 40, 42, 3, 44, 54, 61, 70, 72, 81 etc. [6].

B. *HPV vaccine*

In Korea, the Korean Society of Women's Oncology (KSGOC) has set guidelines for vaccination against human papilloma virus, and basic vaccinations are recommended for women aged 15 to 17 and women aged 18 to 26 to be vaccinated before HPV infection occurs through sexual contact.

The most important thing is to get vaccinated before being infected with HPV, and if you have already had sexual experience or have not been infected with HPV even after the age of 26, you can get preventive effects through vaccination. Women aged 9-14 and 27-45 are recommended to be vaccinated by medical judgment, so they can decide after consulting with an expert.

In addition, in Korea, as in the case of the United States, HPV prevention tetravalent vaccine has been approved for men to be vaccinated against Gardas in Korea. For men, the basic vaccination is at the right age between the ages of 11 and 12, before the sexual experience begins. If you have not been vaccinated before, it is recommended to be vaccinated at the age of 13 to 21, and men aged 22 to 26 can also be vaccinated shown the Table 1[5].

Table 1. CURRENT INCIDENCE OF CERVICAL CANCER (WORLD)

2012	Mexico	USA	Switzerland	Australia	Finland	Korea	South Africa
The number of cases per 100,000 women	23,300	6,600	3,600	5,500	4,300	9,500	31,700

C. HPV vaccine type and inoculation method

HPV vaccines are divided into divalent, tetravalent, and 9-valent vaccines. The divalent vaccine can prevent diseases caused by HPV 16, 18, and tetravalent vaccines for human papilloma viruses 6, 11, 16, and 18, while the divalent vaccine can prevent diseases caused by HPV 16, 18, 6, 11, 31, 33, 45, 52, and 58. There are differences depending on the type of HPV vaccine, but at the age of 9 to 13 (14), vaccinations are possible twice every 6 months at the first inoculation.

If the vaccine is first inoculated after the age at which two vaccinations are permitted, three vaccinations are required, and the 4-ga, 9-ga vaccine is administered every 0, 2, and 6 months for the 2-ga vaccine every 0, 1, and 6 months [6].

D. Comparison of vaccine support types and cervical cancer incidence in developed countries

1. Status of HPV Vaccine Support in Korea

The HPV vaccine can be inoculated from girls aged 9 or older, and has been approved for inoculation from 25 to 26.

The HPV vaccine vaccination recommendation guidelines vary from country to country, and in Korea, it is recommended to be vaccinated at the age of 12 in consideration of the age of onset of bacterial experience, immunogenicity, vaccination cost-effectiveness, and ease of vaccination.

HPV vaccine has been introduced as a national vaccination since 2016 and has been providing free vaccinations (two- and four-valent vaccines/six-month intervals) as a "Healthy Women's First Step Clinic Project.

Cancer prevention effects have not been proven in women after the age of vaccine approval, but cancer prevention effects can be expected theoretically for women over the age of 26 who have not started their sex life or have fewer exposure opportunities to HPV.

2. Status of HPV Vaccine Support in Developed Countries

Countries such as Australia, Denmark, the United States, and France that introduced HPV vaccines to national vaccination projects are already experiencing significant reductions in HPV-related diseases such as intrauterine tumors and genital warts.

Vaccines are provided free of charge by including 9 a.m. in the U.S. and Australia until the age of 15 in the national vaccination.

3. Current incidence of cervical cancer in various countries

The incidence rate is mainly high in underdeveloped countries, 1.44 times higher in Korea than in the United States, which has a wide range of vaccine support, and 1.73 times higher than in Australia.

III. DATA ANALYSIS OF FACTORS ASSOCIATED WITH CERVICAL CANCER

A. Relationship between income level and cervical cancer incidence

In order to find out the relationship between income level and incidence of cervical cancer, income data and incidence data from cities and provinces across the country were secured and analyzed using R shown the Table 2 [7-9].

Table 2. DATA BY CITY AND PROVINCE

		Income	Cancer
1	Seoul	17139.6	14.1
2	Busan	14511.8	18.0
3	Daegu	13911.2	16.0
4	Incheon	13494.4	15.6
5	Gwangju	13975.0	15.6
6	Daejeon	14647.2	16.2
7	Ulsan	17960.2	12.6
8	Gyeonggi	14149.8	13.7
9	Gangwon	12817.2	17.0
10	Chungbuk	13453.4	16.0
11	Chungnam	13600.4	16.3
12	Jeonbuk	13350.6	15.5
13	Jeonnam	12654.0	15.5
14	Gyeongbuk	13378.2	15.9
15	Gyeongnam	13844.6	14.6
16	Jeju	14472.4	16.0

The order of analysis is as follows. First, income data and incidence data of metropolitan cities in Korea were collected, and data analysis was performed through a data preprocessing process. Data analysis was performed using R, a big data statistical analysis program. The analysis results were visualized and applied[10-12].

```
Call:
lm(formula = cancer ~ income, data = income_cancer)

Residuals:
    Min      1Q   Median      3Q      Max
-1.87079 -0.60048 -0.01166  0.47096  2.62939

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.3951822  2.7297018   8.571 6.09e-07 ***
income      -0.0005530  0.0001912  -2.892   0.0118 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.056 on 14 degrees of freedom
Multiple R-squared:  0.374,     Adjusted R-squared:  0.3293
F-statistic: 8.365 on 1 and 14 DF,  p-value: 0.01182
```

The P-value of 0.01182 shows that income level and incidence are significantly related at 95% confidence level [7].

B. HIV Vaccine Survey for People in their 20s

A survey was conducted on 100 college students on perceptions of HPV virus vaccines and national support policies in their 20s. The questionnaire includes interest in vaccine information, the presence or absence of vaccination and the reason for non-vaccination, awareness of vaccination costs, and awareness of state support.

About 15% of respondents said they had been vaccinated against HPV virus, of which 19% were women and 5% were men. The inoculation age is 37% for those aged 16-20 and 53% for those aged 21-26, with more than 90% of those aged 16 or older getting inoculated. Through this, most people in their 20s did not get vaccinated at the recommended age, and this is expected to be because there was a lack of information on vaccines at the time of the recommended age and the cost burden was higher than now shown the Fig. 1.
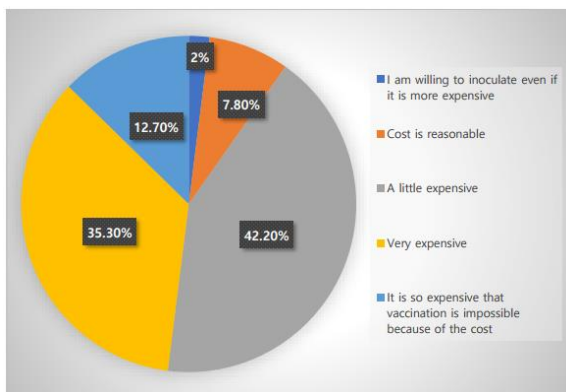

Fig. 1. Survey result I

As for the cost of vaccination, 90% of the respondents said they thought the inoculation cost

was high, as a result of a survey, saying, "The current 2-valent/4 vaccine, which is classified as a national vaccination, is 100,000 won/120,000 won per inoculation." Of these, more than 50% of respondents said they were willing to be vaccinated if they were less than 40% of the current cost shown the Fig. 2.
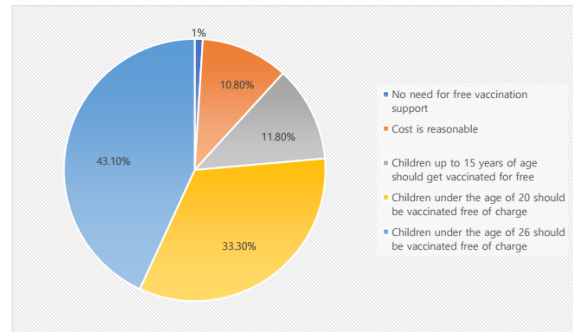

Fig. 2. Survey result II

According to a survey on the age of free vaccination, 90% of respondents in their 20s said the age of free vaccination support should be expanded.

As a result of the survey on the scope of free vaccinations, more than 90% of the respondents answered that "males should also be included" in the vaccination support list shown the Fig. 3.
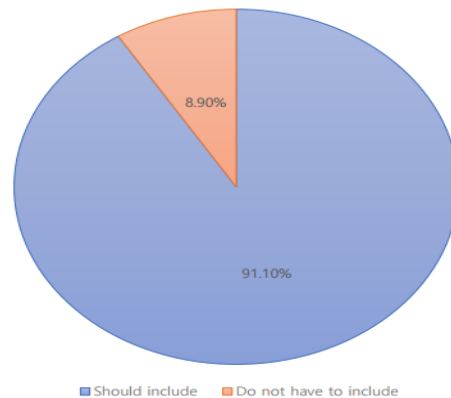

Fig. 3. Survey result III

Logit analysis was conducted on the difference in perception of men in women's view in response to whether they knew information on the existence and prevention rate of cervical cancer vaccines.

As a result, it can be seen that men's awareness of the virus decreases by 0.29 times at a fixed level compared to women.

13

```
Call:
vglm(formula = cbind(y1, y2, y3, y4, y5) ~ sex, family = cumulative(parallel = TRUE),
    data = HPVknowing)


Pearson residuals:
  logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3]) logit(P[Y<=4])
1        -0.6878         0.05529         0.2066         0.1616
2         0.3483        -0.06571        -0.1849        -0.1708

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -0.88689    0.26654  -3.327 0.000877 ***
(Intercept):2  0.07875    0.24560   0.321 0.748490
(Intercept):3  0.64441    0.25459   2.531 0.011369 *
(Intercept):4  1.48803    0.28964   5.138 2.78e-07 ***
sexmale       -1.21294    0.38361  -3.162 0.001567 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of linear predictors:  4

Names of linear predictors:
logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3]), logit(P[Y<=4])

Residual deviance: 0.7592 on 3 degrees of freedom

Log-likelihood: -14.8196 on 3 degrees of freedom

Number of iterations: 4
```

Major related diseases.	cervical cancer	AIDS.
The main route of infection.	Sex contact.	Sex contact.
Gender ratio between men and women.	100% female.	90% of men
Number of new cases.	3,582	1,062
The death toll.	967	104
National Support Expenses (2016)	Cervical cancer benefits of about 100 billion won + vaccine support of 18.4 billion won.	About 100 billion won for AIDS treatment.

Data analysis based on the results of the response to the relationship between vaccination cost and health insurance premium adjustment shows that the more people who want the P-value to lower the price from the 95% confidence level, the more willing they are to pay [8].

```
Call:
lm(formula = insur ~ lowercost, data = result)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0843 -0.6527  0.1623  0.6555  2.4089

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.09789    0.24278   8.641 1.07e-13 ***
lowercost   0.24660    0.09522   2.590   0.0111 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.007 on 98 degrees of freedom
Multiple R-squared:  0.06405,   Adjusted R-squared:  0.0545
F-statistic: 6.706 on 1 and 98 DF,  p-value: 0.01107
```

C. HPV and HIV comparison

Due to the increase in AIDS patients in Korea, the amount of treatment support is increasing by more than 10 billion won every year. AIDS is a disease that leads to death when treatment is stopped, and efforts should be made to prevent AIDS as it costs a lot, but the number of AIDS patients continues to increase, showing insignificant effects compared to input costs.

Table 2. HPV AND HIV COMPARISON

Types of viruses.	HPV(Human Papilloma Virus)	HIV(Human Immunodeficiency Virus)

On the other hand, cervical cancer is occurring in much more people than AIDS, and about 70% of them can be prevented with HPV vaccine. Relatively comparing the cost of state support, it can be seen that the cost of supporting AIDS treatment for five years is similar to the cost of HPV vaccination for all Korean women aged 9 to 24 (about 4.5 million).

IV. CONCLUSION

Since cervical cancer is a disease that can be prevented by about 70% with HPV vaccine, the state should strive to expand HPV vaccination. First of all, it is necessary to expand the national vaccination targets for HPV vaccines.

The age of HPV vaccination support should be extended from 12 to 18 years old, and men should be included in the application. In addition, the country is currently supporting divalent and tetravalent vaccines, and it should be changed to support divalent vaccines like developed countries.

In addition, in order to improve the HPV vaccination rate, the risk and vaccine effect of cervical cancer should be actively promoted, and the need for vaccination in men should be recognized by many people. In addition, the cost of vaccination, which is more expensive than other vaccines, should be lowered to ease the burden of people trying to vaccinate at their own expense in addition to those eligible for support.

REFERENCES

[1] Hwang, J. S., Liu, C., Li, G., & Park, S. T. (2020). A Study on Problem Solution of Culture Tourism Festival Through Big Data Analysis. International Journal of Emerging Multidisciplinary Research (IJEMR), 4(3), 17-21.

[2] Suh, K. D. (2021). A Study on deduction of important factors for new infectious diseases through big data analysis. Journal of Industrial Convergence, 19(3), 35-40.

[3] Suh, K. D., Choi, J. I., & Choi, P. A. (2020). Research on criminal policy measures for the prevention and management of infectious diseases: Focusing on Mers. Journal of Industrial Convergence, 18(6), 9-17.

[4] Suh, K. D., & Choi, P. A. (2020). A Study on Healthcare Policy Response to Risks of Future Infectious Diseases: Focused on Infectious Disease Surveillance Systems. Journal of Industrial Convergence, 18(3), 109-116.

[5] The National Health Information Portal, https://health.kdca.go.kr

[6] Seoul National University Hospital, http://www.snuh.org/health/nMedInfo/nView.do?category= DIS&medid=AA000277

[7] Noh, K. S, Kim, J. W, Park, S. T & Kim, B. S.(2018). Big data analysis using R. Wowpass.

[8] Chungnam National University WISET Project Group.(2018). Big Data Special Lecture Textbook. KASOM

[9] Park, S. T., & Kim, Y. K. (2019). A study on deriving an optimal route for foreign tourists through the analysis of big data. Journal of Convergence for Information Technology, 9(10), 56-63.

[10] Park, E. M., & Seo, J. H. (2019). A Study on Leadership Typology in Sports Leaders Based on Big Data Analysis. Journal of the Korea Convergence Society, 10(7), 191-198.

[11] Park, S. T., Lee, S. W., & Kang, T. G. (2018). A study on the trend of cloud service and security through text mining technique. International Journal of Engineering & Technology, 7(2.33), 127-132.

[12] Li, G., Dai, J. S., Park, E. M., & Park, S. T. (2017). A study on the service and trend of Fintech security based on text-mining: focused on the data of Korean online news. Journal of Computer Virology and Hacking Techniques, 13(4), 249-255.