JKMS

Editorial

Has the Incidence of Acute Pyelonephritis Increased in Korea? – Big Data as a Double-edged Sword

Jin-Hong Yoo [1,2]

[1]Division of Infectious Diseases, Department of Internal Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea
[2]Division of Infectious Diseases, Department of Internal Medicine, Bucheon St. Mary's Hospital, Bucheon, Korea

ORCID iDs
Jin-Hong Yoo
https://orcid.org/0000-0003-2611-3399

In the medical field, a vast amount of raw data is constantly being produced daily. Then, the raw data are gathered to build big data, which are waiting for medical researchers to use it. Since the government and public institutions have started to provide big data for the medical field, a lot of medical papers using them have been published. For example, in the Korean Health Insurance Review and Assessment Service (HIRA, Wonju, Korea), the National Health Insurance Service (NHIS, Wonju, Korea), or the Korean National Health and Nutrition Examination Survey (KNHANES, Cheongju, Korea), huge data are now available. This can be used to analyze the occurrence and trend of diseases. Especially, it is very useful for investigating a disease that cannot be analyzed by single institutional data alone. You can also analyze the risk factors of the disease, the nature of the treatment, and the costs associated with the disease. In addition, because of the enormous amount of data, significant conclusions can be drawn from statistical tests.[1,2]

However, these big data have a fundamental limitation in that they are not collected for the purpose of clinical research from the beginning. Moreover, it does not contain the private information of the patients and is limited to the 5-year deadline. Therefore, there is a problem with accuracy and validity in analyzing these vast amounts of data.[1] These data are inherently entered for insurance claim purposes and cannot guarantee the accuracy of the diagnosis.[2] That is, there is a high risk that the claimed diagnosis and the actual diagnosis could be different from each other, resulting in an unexpected result. Overestimation or underestimation of a disease can occur.

In this issue, Kim et al.[3] analyzed the status of acute pyelonephritis (APN) in Korea based on the HIRA data from 2010 to 2014. The authors' conclusion shows a very surprising result. The mean incidence for 5 years is 39.1/10,000. And the incidence, which was 35.6/10,000 in 2010, increased significantly to 43.8/10,000 in 2014. In the late 1990s, its incidence has been reported at 35.7 cases per 10,000 persons in Korea, and this figure has steadily maintained without any significant change until 2010, and has increased rapidly in just five years.[4]

If we accept this result as real, it means that the recent increase in the incidence of APN in Korea has emerged as a new serious problem. For what reason did this new problem arise?

The authors are suspicious of the increased antibiotic resistance of microorganisms causing APN. In fact, the antibiotic susceptibility of *Escherichia coli* isolated from patients with APN to ciprofloxacin and trimethoprim/sulfamethoxazole has gradually decreased.[5] The authors noted that the occurrence rate of Korean resistant strains of *E. coli* (ST131) increased from 19.7% in 2011 to 26.9% in 2016, assuming as a partially explainable basis for the increase in incidence of APN. Of course, this is not the evidence used directly in this study, it is only an estimate of the authors. Therefore, this assumption may be partially explained, but it cannot be a satisfactory explanation.

This is one of many blunders in data analysis using big data. As mentioned at the beginning, HIRA's big data consist of diagnostics entered for the purpose of claiming. However, the patient's personal information and various test result information are missing. The diagnosis of APN is also problematic in terms of accuracy. In practice, cystitis and asymptomatic bacteriuria are also likely to be included as APNs. Therefore, there is a limit to accurately and finely describing the fact that the incidence of APN has increased. To put it bluntly, it cannot be dismissed even if one concludes that the rate of claiming the APN, not the incidence of it, has increased.

While this paper is an excellent example of getting meaningful conclusions using big data, it also shows the loopholes of big data itself. In a nutshell, big data is not a panacea yet, but a double-edged sword.

Medical research using big data should continue to be encouraged. A big dataset is essentially an almost complete survey, and so you can obtain more truthful information than any previous study. However, medical research using big data in Korea is just at the beginning. In order to obtain valuable information by studying big data, researchers must discriminate between association and causality. Collaboration and coordination with statistical experts as well as the medical profession is needed in dealing with this huge amount of raw data in the future.

REFERENCES

1. Cho GJ Clinical research using medical big data. *Anesth Pain Med* 2017;12(1):9-14.
   CROSSREF

2. Ryu DR. Introduction to the medical research using national health insurance claims database. *Ewha Med J* 2017;40(2):66-70.
   CROSSREF

3. Kim B, Myung R, Kim J, Lee MJ, Pai H. Descriptive epidemiology of acute pyelonephritis in Korea, 2010–2014: population-based study. *J Korean Med Sci* 2018;33(49):e310.
   CROSSREF

4. Ki M, Park T, Choi B, Foxman B. The epidemiology of acute pyelonephritis in South Korea, 1997–1999. *Am J Epidemiol* 2004;160(10):985-93.
   PUBMED | **CROSSREF**

5. Kang CI, Kim J, Park DW, Kim BN, Ha US, Lee SJ, et al. Clinical practice guidelines for the antibiotic treatment of community-acquired urinary tract infections. *Infect Chemother* 2018;50(1):67-100.
   PUBMED | **CROSSREF**