Emerging Ethical Issues in Managing Real World Data

Kwang-Mo Yang

Department of Medical Humanities, Sungkyunkwan University School of Medicine, Seoul, Korea
Center for Health Promotion, Samsung Medical Center, Seoul, Korea

Data is important enough to be likened to crude oil in the fourth industrial revolution. Recently, the South Korean government announced its 'Bio-Health Innovation Strategy'. The most important element of that strategy is health care data. These days, it is also called 'Real World Data' because it shows what's happening in the real world. Health authorities have already declared that they will create a public data platform and have prepared a considerable budget [1].

In most countries around the world, including the Republic of Korea, it is legal to study health data after de-identifying personal information. However, some non-governmental organizations are concerned about such big data research. This is due to the view that it may be used commercially. Everyone acknowledges that health data includes highly sensitive personal information.

Most academic researchers conduct research for public interest purposes. But recently many researchers are conducting research with the aim of commercializing the outcomes. In South Korea, the quality of individual researchers is also important when looking at government-funded tasks, but sometimes (government) requires participation by companies that can commercialize the results.

As such, we cannot ignore or diminish the concerns of civic groups. Even if it is legal to use de-identified data, it is natural to suspect that there is a high probability of re-iden-

tification if it is exploited for commercial purposes. Indeed, many companies want to re-identify data to gain greater profit margins.

For example, suppose that a company called B researched patients with A disease. Naturally, the patients' information is de-identified. The study found out such characteristics as vulnerable age, gender, and race of the disease, and eventually company B made a new drug. If the drug is marketed to patients with A disease, B could make a bigger profit. Therefore, company B would wish to re-identify patient information with A disease and advertise directly at the same time by promoting it to the hospital where the patient might go [2].

There have been many similar cases in the past. Some pharmaceutical companies and some 'fake' data analytic companies that were working to help with their marketing activities had unethical transactions with patient information. Through this data, companies determined the market share of their drugs, and the information of hospitals that used their drugs was obtained. Of course, with that information, they could actively engage in sales to doctors.

Nowadays, such unethical activities have decreased and, on the contrary, almost all studies are in line with the public interest. The pharmaceutical industry is monitoring drug side effects almost in real time through 'Real World Data' study. And such evidence is helping to develop new drugs. Companies that use lots image data to learn artificial intelligence are expected to help radiologists. It is expected that the misdiagnosis rate will be lowered.

While the situation has changed, such as the research environment and the ethical consciousness of researchers,

the perception of non-governmental organizations has not changed dramatically. They say that individual patients' consent is necessary as far as possible, even if de-identified patient data is used for public interest purposes. Because the concept of de-identification is made to avoid consent and they believe that de-identified studies are increasing in prevalence, consent research will be reduced.

Furthermore, they also hold a negative view about linking personal health records from separate departments. They contend that 'the consent of the bereaved families' is needed for attempts to combine Electronic Medical Record data in each hospital with the death data of governments. If the agreement is a matter of cost, it is something that researchers or the government should bear. Governments and researchers are looking to reduce costs or increase efficiency by consolidating data. However, civic groups do not acknowledge this, saying it has not been proven.

The reason why non-governmental organizations are so conservative seems clear. It is unfair to demand that a potential patients, or people, citizen, give up their rights now in anticipation of future benefits that have not yet been realized. They are hence worried that the stance of government, industries, and researchers will reduce their rights.

A fundamental way to dispel these concerns would be to build solid trust between researchers and information providers, i.e. patients. If the perception that the researcher is a good manager is firm, there is no need for such controversy. By making the law, unethical researchers can be punished fairly. Unfortunately, it is now nearly impossible to suddenly do so.

What then are the practical solutions? When collecting data in the first place, it can also be considered to have 'broad consent'. The Common Rule of the United States allows broad consent if research is carried out on de-identified data and bio-samples or of an institutional review board is notified. General Data Protection Regulation of European Union also permits this in similar cases.

The broad consent and 'opt-out' can be said to be the basic approach of many developed countries, with an explanation of how the big data platform works. Of course, each time

a study is conducted, we can obtain consent by a portable smart device in an 'opt-in' manner. This is called 'Dynamic Consent' [3]. This type of consent has many advantages, but it also has the decisive drawback that it can be very inconvenient to study participants. But I believe that the problem can soon be resolved by technology.

What is the ideal solution for data scientists or researchers? It is proper for patients to be provided periodic and continuous data analytic results regarding how their data is currently being used. This is of course very difficult and inconvenient. But let's think about this from the patient's point of view. Many patients believe that periodic and continuous data analysis is more important and meaningful than a paper in a hard to understand medical journal.

It should be remembered that building trust with patients can be a fundamental solution. In order to do so, we need to make the information transparent and reveal what has been created as a result of using the personal health data.

ORCID

Kwang-Mo Yang (<http://orcid.org/0000-0002-7176-4935>)

References

1. Ministry of Health and Welfare. Investment of 4 trillion won in bio big data and research and development, fostered to world-class biohealth level [Internet]. Sejong, Korea: Ministry of Health and Welfare; 2019 [cited at 2019 Oct 15]. Available from: http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&CONT_SEQ=349512.
2. Tanner A. Our bodies, our data: how companies make billions selling our medical records. Boston (MA): Beacon Press; 2017.
3. Budin-Ljosne I, Teare HJ, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med Ethics* 2017;18(1):4.