

More Powerful Search Engine Invalidates Anonymity Guidelines for Case Reports

Shigeki Tani, PhD¹, Shinichi Higuchi¹, Goshi Fujimoto¹, Jun Nakaya, MD, PhD², Michio Kimura, MD, PhD¹

¹Department of Medical Informatics, Hamamatsu University, Hamamatsu; ²Department of Medical Informatics, Tokyo Medical and Dental University, Tokyo, Japan

In the United States when presenting case reports or research results, or when using related information for secondary purposes, the use of social security numbers of patients is restricted along with 18 items identified by the Health Insurance Portability and Accountability Act (HIPAA) Guidelines which must be kept anonymous [1]. In Japan, the Surgery Society Federation has established "Guidelines for protection of patient privacy in case reports, medical papers, and research meetings [2]" which prohibits even the use of patients' initials. Addresses must be recorded only by prefecture and city name and the event must be recorded only by the year and month, not by the date.

The following is an excerpt from an acceptable case report. "A male patient, in his 70s, visited our hospital in November 2007. His HbA1c test result was 7.4. He was diagnosed with diabetes and given voglibose. Two to three months later, he was retested and the HbA1c value was 5.9." (The details have been altered to protect the patient's privacy).

Hamamatsu University School of Medicine has a clinical information retrieval system D*D [3] which searches

10 years worth of records at high speed. The data include prescription/injection orders, laboratory test results, disease name registration, and basic patient information (385,815 persons, 73,709,298 total events as of June, 2008). This system searches the data according to time sequence or context. For example, a search for "Pravastatin (any titer, including generic drug) prescribed and an AST > 150 recorded within two weeks" took 112 seconds to compile a list of 83 patient hits.

A search using the information in the previously mentioned example case report resulted in identification of only ONE individual among 385,815 Hamamatsu University Hospital patients. We used this system and increased the degree of anonymity of information. We omitted the month of November and included test results with a range of "6.6-8.0" and "5.9-6.5" as defined by the Japan Diabetes Society to describe the seriousness of disease. The content of the search was as follows: "A male patient, in his 70's, visited our hospital in 2007. His HbA1c test result was (6.6-8.0). He was diagnosed with diabetes and given voglibose. Two to three months later, he was retested and the result was (5.9-6.5)." A search with this information identified 47 patients.

The national morbidity rate of diabetes is approximately 5%. Even for this rate, the content of search results must be ambiguous as exemplified above. For a case report of a very rare disease, e.g., one patient per tens of thousands of inhabitants, then the content of the search must be even more ambiguous to ensure the same degree of anonymity of information. As a result, the description becomes meaningless as a medical paper while conversely, the fact is that a highly identifiable case report for a rare disease is medically more meaningful.

One countermeasure to address this problem is to decrease the number of persons who have access to such papers; how-

Received for review: March 24, 2011

Accepted for publication: March 25, 2011

Corresponding Author

Michio Kimura, MD, PhD

Department of Medical Informatics, Hamamatsu University, 1-20-1 Handa, Hamamatsu 431-3192, Japan. Tel: +81-53-435-2770, Fax: +81-53-435-2769, E-mail: kimura@mi.hama-med.ac.jp

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2011 The Korean Society of Medical Informatics

ever, it is a historical fact that restricting information inhibits the advancement of science. Therefore, this countermeasure is impractical. Furthermore, in the present-day environment, it would be practically impossible to restrict access to internet published journals.

Therefore, to protect patient privacy, it is necessary to further ensure the confidentiality of information that may ultimately identify an individual, such as the names and addresses of patients. But healthcare ID numbers (the insurance number, patient hospital ID number, etc.) must be linked very carefully with identifiable information (a social security number, taxpayer ID number, etc.). The level of privacy is determined by the least protected ID information. For this reason, it is necessary to numerically evaluate the degree of anonymity of information, considering associated factors, for example, the morbidity rate of the disease.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. National Institute of Health. Protecting personal health information in research: understanding the HIPAA privacy rule. Bethesda: National Institute of Health; 2003 [cited at 2010 Mar 10]. Available from: http://privacyruleandresearch.nih.gov/pdf/HIPAA_Booklet_4-14-2003.pdf.
2. Japanese Society for Vascular Surgery. Guidelines for protection of patient privacy in case reports, medical papers and research meetings. Tokyo: Japanese Society for Vascular Surgery; 2005 [cited at Mar 11]. Available from: <http://www.jsvs.org/ja/info/040406.html>.
3. Kimura M, Tani S, Watanabe H, Naito Y, Sakusabe T, Watanabe H, Nakaya J, Sasaki F, Numano T, Furuta T, Furuta T. High speed clinical data retrieval system with event time sequence feature: with 10 years of clinical data of Hamamatsu University Hospital CPOE. *Methods Inf Med* 2008; 47: 560-568.