

Healthcare data research: The inception of the Taipei Veterans General Hospital Big Data Center

Ai Seon Kuan^{a,b}, Tzeng-Ji Chen^{b,c,d,*}

^aDivision of Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; ^bSchool of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC; ^cDepartment of Family Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; ^dBig Data Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

Major progress in data utilization has been made over the past decade in every area. Routinely collected healthcare data have become an increasingly important source of information for biomedical research. In Taiwan, the first official version of National Health Insurance Research Database (NHIRD), which contained both outpatient and inpatient data of nearly the entire Taiwanese population, was made available to researchers in 2000.^{1,2} NHIRD scientific community had flourished and published thousands of peer-reviewed articles since then. The *Journal of the Chinese Medical Association* was the pioneer and long-term supporter in publishing the results of NHIRD research.³⁻⁵ The NHIRD datasets, however, were not flawless. Information in the NHIRD was originally collected for claims purposes and subsequently recompiled for research use. One inherent problem with claims data was that the majority of diagnoses had not been validated against clinical criteria or objective data. The high risk of outcome misclassification affected the validity of research. Besides, the NHIRD lacked information on lifestyles, clinical measurements, and biomarkers, so that the types of feasible research were relatively limited. Furthermore, access to the NHIRD had been made difficult with the introduction of new regulations. Since 2016, data analysis can be performed only at the Ministry of Health and Welfare Data Science Center and its approved branches, with appointment queuing and strict user-unfriendly conditions that greatly hindered the availability.

Over the past few years, several large medical institutions in Taiwan have begun to set up their own data centers of electronic health records to promote clinical research, improve patient care, and develop personalized medicine.⁶ At the beginning of 2019, a new one was launched: the Taipei Veterans General Hospital (VGH) Big Data Center (BDC). Its initial missions included building a data platform for integration of healthcare data from all sources, creating system- or disease-specific datasets, and facilitating access by hospital personnel. As an early forerunner of medical informatics in Taiwan, the Taipei VGH has accumulated a massive amount of electronic data since 1978, ranging from medical records, pharmacy orders, clinical and

anthropometric measurements, laboratory results, examination reports, to radiological images, histopathological slides, and others. These important assets can be used to help clinical research that may transform clinical practices in the future. Because the methods of coding data continue to change over four decades, considerable effort of data preprocessing is required before data release for research use. In the first year of the Taipei VGH BDC, outpatient and inpatients claims data plus laboratory results and demographic information were open for application. A network with experts in the field was also provided to help research design and data analysis. All personnel of Taipei VGH are eligible to apply to the BDC for data after obtaining an approval from the institutional review board. Details can be found on the website of the BDC (<https://wd.vghtpe.gov.tw/bdc>).

The quality and comprehensiveness of data under the charge of the Taipei VGH BDC obviously outperform those of the NHIRD, so that more high-quality research can be expected. The Taipei VGH BDC will endeavor to broaden the extent and variety of its database in due course by incorporating unstructured data such as radiology reports and admission/progress/discharge notes into the database. Currently, one of the challenges to the BDC is to carry out appropriate data deidentification to protect the personal privacy without compromising the data usability. Besides, patients at Taipei VGH come from every corner of Taiwan and abroad. External data linkage with other hospitals and clinics is thus desirable for the accuracy and thoroughness of research. Concerted effort is being undertaken to link data among Taipei VGH, Taichung VGH, Kaohsiung VGH, and their branches within the system of Veterans Affairs Council. The Taipei VGH BDC will continue to overcome the challenges and strive for excellence.

REFERENCES

1. Chen YC, Wu JC, Chen TJ, Wetter T. Reduced access to database. A publicly available database accelerates academic production. *BMJ* 2011;342:d637.
2. Chen YC, Yeh HY, Wu JC, Haschler I, Chen TJ, Wetter T. Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics. *Scientometrics* 2011;86:365-80.
3. Su TP, Chen TJ, Hwang SJ, Chou LF, Fan AP, Chen YC. Utilization of psychotropic drugs in Taiwan: an overview of outpatient sector in 2000. *J Chin Med Assoc* 2002;65:378-91.
4. Chen TJ, Liu JY, Hwang SJ. Non-steroidal anti-inflammatory drug and antacid co-prescription in Taiwan: analysis of national insurance claims. *J Chin Med Assoc* 2002;65:588-93.
5. Lee KL, Chiu NC, Su CW, Tseng HS, Lee RC, Liu CA, et al. Less barium enema, more colonoscopy: a 12-year nationwide population-based study in Taiwan. *J Chin Med Assoc* 2019;82:312-7.
6. Tsai MS, Lin MH, Lee CP, Yang YH, Chen WC, Chang GH, et al. Chang gung research database: a multi-institutional database consisting of original medical records. *Biomed J* 2017;40:263-9.

*Address correspondence: Dr. Tzeng-Ji Chen, Department of Family Medicine, Taipei Veterans General Hospital, 201, Section 2, Shi-Pai Road, Taipei 112, Taiwan, ROC. E-mail address: tjchen@vghtpe.gov.tw (T.-J. Chen).

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2019) 82: 679.

Received June 13, 2019; accepted June 14, 2019.

doi: 10.1097/JCMA.000000000000144.

Copyright © 2019, the Chinese Medical Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).