

Invited Commentary

Nationwide Population Science Lessons From the Taiwan National Health Insurance Research Database

Ann W. Hsing, PhD; John P. A. Ioannidis, MD, DSc

Large population-based studies can inform us on the prevalence, incidence, natural history, treatment, correlates, and associations of disease, as well as the pattern of health care utilization. A special type of large population study encompasses



Related article page 1518

the population of an entire nation. Advantages include enormous sample size and lack of selection and participation bias. These advantages are enhanced further when the databases are rich in clinical, personal, and risk factor information and when different pieces of information are linked to permit joint analysis. Once the process for data accessing is established, vast amounts of information can be obtained at minimal cost, especially when additional collection and update of information is carried out routinely for purposes inherent in medical care and/or insurance coverage and reimbursement.

The Taiwan National Health Insurance (NHI) Research Database is one of the largest nationwide population databases in the world, covering approximately 23 million residents in Taiwan.¹ The NHI program was established in 1995 to deliver universal coverage provided by a government-run, single-payer compulsory insurance plan to centralize the disbursement of health care funding. The program covered 99.0% of the island population by 2004 and 99.5% of the population by 2010. Universal coverage reduces barriers to health care access for all citizens, regardless of socioeconomic background and/or residential location. In 1997, the National Health Research Institute was authorized to build the NHI Research Database to facilitate research on health care utilization, health economics, and biomedical and behavioral research, as well as to provide information guiding national health policy. Currently, the NHI Research Database includes more than 26 data files containing information on personal characteristics (sex, date of birth, place of residence, details of insurance, employment); family relationships; details of clinical information, including date, expenditures, and diagnosis related to both inpatient and outpatient procedures; prescription details; examinations; and operations. Data are updated biannually. Since 2003, the use of electronic medical records has been mandated and standardized across 321 hospitals, providing the foundation for digital health and data exchange.

By law, all Taiwan residents have a unique personal identification number that is used also for tax returns, bank accounts, and all transactions. Thus, NHI Research Database data are linkable to multiple national databases maintained by other departments, including registries of births, deaths, households, immunizations, cancer, reportable infectious diseases, and environmental exposures. This identification is also linked to the national immunization information system and

mass population-based screening for cancers of the oral cavity, colon, cervix, and breast, as well as myopia and urine screenings and regular examinations in school children, newborn screening, the rare disease and suicide registries, the Taiwan Pathogenic Microorganism Genome Database, and the blood donor registry. A health cloud is currently being built that will eventually further facilitate data linkage among various registries, databases, and departments. In addition, Academia Sinica is establishing a biobank of more than 200 000 residents; the data in the biobank will be linked with NHI Research Database data. With such extensive linkage, privacy is protected through laws that dictate encryption of the unique personal identification number (23 digits) and strict procedures for data access and human subject review. To date, no breach or leak is known to have occurred.

Since 1995, several large data centers have been established in academic universities in Taiwan to mine the NHI Research Database. More than 2016 articles had been published by mid-2014 (http://nhipapers.idv.tw/healthdata/advance_ch.php), including studies of drug adverse effects and risk of disease (n = 163), drug safety and efficacy (n = 131), drug prescription and utilization patterns (n = 169), clinical presentation (n = 117), prognosis and health outcomes research (n = 682), cost-effectiveness and health economics (n = 90), health services research (n = 81), health care utilization (n = 220), methodology (n = 46), ecologic correlation (n = 46), and risk factor analysis (n = 56), as well as descriptive studies (n = 215) and studies of other topics.²⁻⁴ Productivity is rapidly accelerating, with more than 430 studies published in 2013. The large and longitudinal nature of the data (nearly 20 years of follow-up data of the entire population) permits a wide range of study design, including cross-sectional, case-control, and retrospective and prospective cohort analyses, as well as family studies (family members can be linked through parents' identification numbers on birth certificates). The large sample size provides opportunities for the study of rare conditions. For example, in this issue of *JAMA Internal Medicine*, Kuo et al⁵ identified 18 283 patients with systemic lupus erythematosus (SLE) and their family members, including parents, siblings, and twins, to estimate SLE heritability, risk for specific family members, and relationships with other autoimmune diseases. Typically, SLE studies include up to a few hundred cases, and even large consortia, such as those conducting genome-wide association studies, usually include no more than several hundred to several thousand cases.

Some other countries (Japan, Korea, and Scandinavian countries [eg, Sweden and Denmark]) also have national databases with associated record linkage capacity. However, relative to Taiwan, the population size for Scandinavian

Table. Comparison of Some Nationwide Research Databases

Characteristic	Taiwan	Sweden	Korea
Year started	1995	1955	1989
Size, million	23	10	50
Single payer	Yes	Yes	Yes
National health insurance data			
Inpatient	Yes	Yes	Yes
Outpatient	Yes	Yes	Yes
Pharmacy data	Yes	Yes	Yes
Over-the-counter drugs	No	No	No
Chinese medicine	Yes	No	Yes
Clinician information	Yes	Yes	No
Hospital information	Yes	Yes	No
Linkage to other databases			
Registry			
Household	Yes	No	No
Birth certificate	Yes	Yes	Partial (date of birth, no information on parents or place of birth)
Death certificate	Yes	Yes	Yes
Cancer	Yes	Yes	Partial (through text mining)
Immunization record	Yes	No	Partial (through text mining)
Reportable infectious disease	Yes	No	No
Population-based mass screening program	Yes	No	Partial (if related to insurance)
Myopia registry	Yes	No	No
School children urine survey	Yes	No	No
Adult preventive care service ^a	Yes	No	No
Neonatal screening for congenital adrenal hyperplasia	No	Yes	No
Multigeneration registry	Possible ^b	Yes	No
Biobank	Being collected	Being designed	No

^a Biomarker data (eg, cholesterol levels) are available on 2 million individuals.

^b Registry data would be obtainable through linkage with unique personal identification numbers.

countries is small. For Japan and Korea, cross-linkage of registries is not as comprehensive. The **Table** shows some main features of the national databases in Taiwan, Sweden, and Korea. Biobank components from these national databases are limited. Several other countries, including China and those of the United Kingdom, have started large biobanks that store biological samples from more than half a million individuals.^{6,7}

In the United States, the Department of Veterans Affairs, Medicare, Medicaid, or health maintenance organizations (eg, Kaiser Permanente) also have large databases but are more limited in scope and underrepresent minority and underserved populations. Although several biobanks and biorepositories exist, there are many challenges for the United States in developing a nationwide database, including nonsharing multiple payer systems and health care professionals and electronic health record systems with firewalls. It is unclear whether the 2010 Affordable Care Act can help drive the process toward a national system of real world data.

Given the extensiveness of the Taiwan database and its ability to answer a wide variety of questions, should we abandon all other data collection efforts and await the definitive answers from this national treasure? No. Most obviously, the data reflect the genetics, health practices, disease patterns, and social and medical practices of the Taiwanese. Many factors can

vary among different countries and ethnicities. The prevalence of lactose intolerance, prevalence of hepatitis B, and provision of universal health insurance are just a few examples of things that we know are different in Taiwan and would limit generalizability of the findings. More broadly, as with all observational studies, nonexperimental results cannot prove causality. Nationwide observational studies assessing medications, procedures, and diet are particularly prone to bias by indication in estimating treatment effects. Only randomization can produce groups that are equivalent on measured and unmeasured factors. In this regard, nationwide databases may also offer opportunities to incorporate data in randomized trials. These trials would benefit from the extensive, routine collection of data and outcomes for trial participants in these linked national databases.

We look forward to continuing to learn from the Taiwan NHI Research Database and other national databases. Stringent quality control is critical to ensure data validity and minimize missing data. Centralized planning of the research agenda may be important to avoid fragmentation of analytic efforts or redundant or minimally informative findings. Ultimately, the value of large national databases will depend on the creativity and novel hypothesis, sound study design, and adherence to epidemiologic principles, as well as careful execution of the studies by researchers.

ARTICLE INFORMATION

Author Affiliations: Cancer Prevention Institute of California, Fremont (Hsing); Stanford Cancer Institute, School of Medicine, Stanford University, Stanford, California (Hsing, Ioannidis); Department of Health Research and Policy, School of Medicine, Stanford University, Stanford, California (Hsing, Ioannidis); Stanford Prevention Research Center, Department of Medicine, School of Medicine, Stanford University, Stanford, California (Ioannidis); Department of Statistics, School of Humanities and Sciences Stanford University, Stanford, California (Ioannidis).

Corresponding Author: John P. A. Ioannidis, MD, DSc, Stanford Prevention Research Center, Department of Medicine, School of Medicine, Stanford University, 1265 Welch Rd, Medical School Office Building, Room X306, Stanford, CA 94305 (jioannid@stanford.edu).

Published Online: July 20, 2015.
doi:10.1001/jamainternmed.2015.3540.

Conflict of Interest Disclosures: None reported.

Additional Contributions: San-Lin You, PhD, Fu-Jen University, Department of Public Health, Big Data Research Center, provided historical background and information on the Taiwan National Health Insurance Program and related legislations and research activities. There was no financial compensation.

REFERENCES

1. Cheng SH, Chiang TL. The effect of universal health insurance on health care utilization in Taiwan: results from a natural experiment. *JAMA*. 1997;278(2):89-93.
2. Wu CY, Chen YJ, Ho HJ, et al. Association between nucleoside analogues and risk of hepatitis B virus-related hepatocellular carcinoma recurrence following liver resection. *JAMA*. 2012; 308(18):1906-1914.
3. Tang ST, Wu SC, Hung YN, Chen JS, Huang EW, Liu TW. Determinants of aggressive end-of-life care for Taiwanese cancer decedents, 2001 to 2006. *J Clin Oncol*. 2009;27(27):4613-4618.
4. Lin HW, Tu YY, Lin SY, et al. Risk of ovarian cancer in women with pelvic inflammatory disease: a population-based study. *Lancet Oncol*. 2011;12(9):900-904.
5. Kuo C-F, Grainge MJ, Valdes AM, et al. Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune rheumatic diseases in affected families [published online July 20, 2015]. *JAMA Intern Med*. doi:10.1001/jamainternmed.2015.3528.
6. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: come and get it. *Sci Transl Med*. 2014; 6(224):224ed4. doi:10.1126/scitranslmed.3008601.
7. Chen Z, Chen J, Collins R, et al; China Kadoorie Biobank (CKB) collaborative group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40(6):1652-1666.