Scientometrics (2011) 86:365–380
DOI 10.1007/s11192-010-0289-2

Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics

Yu-Chun Chen · Hsiao-Yun Yeh · Jau-Ching Wu · Ingo Haschler · Tzeng-Ji Chen · Thomas Wetter

Received: 8 April 2010 / Published online: 9 September 2010

Abstract The trend to use administrative health care databases as research material is increasing but not well explored. Taiwan's National Health Insurance Research Database (NHIRD), one of the largest administrative health care databases around the world, has been used widely in academic studies. This study analyzed 383 NHIRD studies published between 2000 and 2009 to quantify the effects on overall growth, scholar response, and spread of the study fields. The NHIRD studies expanded rapidly in both quantity and quality since the first study was published in 2000. Researchers usually collaborated to share knowledge, which was crucial to process the NHIRD data. However, once the fundamental problem had been overcome, success to get published became more reproducible. NHIRD studies were also published diversely in a growing number of journals. Both general health and clinical science studies benefited from NHIRD. In conclusion, this new research material widely promotes scientific production in a greater magnitude.

Y.-C. Chen (✉) · I. Haschler · T. Wetter
Department of Medical Informatics, Heidelberg University, Im Neuenheimer Feld 305, Heidelberg 69120, Germany
e-mail: Yu-Chun.Chen@stud.uni-heidelberg.de; cbrain@self.twmail.cc
URL: http://tinyurl.com/Yu-ChunChen

H.-Y. Yeh · T.-J. Chen
Department of Family Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

J.-C. Wu
Department of Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

J.-C. Wu
Institute of Pharmacology, National Yang-Ming University, Taipei, Taiwan, ROC

J.-C. Wu
School of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC

T.-J. Chen
Institute of Hospital and Health Care Administration, National Yang-Ming University, Taipei, Taiwan, ROC

The experience of Taiwan's NHIRD should encourage national- or institutional-level data holders to consider re-using their administrative databases for academic purposes.

Keywords Administrative health care database · National Health Insurance Research Database · Bibliometric analysis · Knowledge growth

Introduction

Administrative health care databases have long been very important data sources for health studies. In recent decades, large-scale administrative health care databases featuring extremely large case numbers, long observation periods, and low access costs have been released to the public and brought quick gains of academic importance (Harpe 2009; Schneeweiss and Avorn 2005). Previous studies have posited the unique potentials of using large-scale administrative health care databases for health research (Hoffmann 2009; Hsiao et al. 2007; Tricco et al. 2008) so their dramatic growth is thus expected but rarely investigated.

Taiwan's National Health Insurance Research Database (NHIRD) is one of the most famous newly-released large-scale administrative health care databases in the world. Since its public release in 2000, it has been used widely in several studies. However, the real impact on the dynamics of scientific societies in response to the release of such a large-scale administrative health care database remains unclear.

This study aimed to analyze the growth of publication number, patterns of co-authorship, and spread among study fields of studies using the NHIRD as data source between 2000 and 2009. The results can quantify, visualize and help model the effects of the public release of a large-scale administrative healthcare database.

Related studies

Administrative health care databases

Administrative health care databases, also called claims-based databases or health care utilization databases, are made of routinely collected records originally intended for management purposes. In recent decades, some national data holders of administrative health care databases started to recompile these databases and made these available for academic research purposes (Schneeweiss and Avorn 2005). The most well-known administrative health care databases include the health maintenance organization (HMO) and Medicaid database in the US, General Practice Research Database (GPRD) in UK, Manitoba and Saskatchewan province database in Canada, and National Health Insurance Research Database (NHIRD) in Taiwan.

Although these databases have been widely used in health researches, especially in the fields of pharmaco-epidemiology and health policy (Harpe 2009), their influence and acceptance vary among scientific societies. Although the first research using administrative health care databases appeared as early as the 1980s, the vast majority of published studies is limited to epidemiology and health services research (Hoffmann 2009; Tricco et al. 2008). Only until the recent decade have administrative health care databases started to be used widely in the field of health therapeutic outcomes (Berger et al. 2009).

The dynamics of scientific societies can be thought of as a process of knowledge creation and information diffusion (Lambiotte and Panzarasa 2009). Bibliometric analyses

on literature using administrative health care databases can provide a better understanding of how a new idea generated by a local community influences the global scientific society.

Taiwan's National Health Insurance Research Database (NHIRD)

Taiwan, a country with more than 23 million people on a 35.8 thousand square kilometers island, launched its universal national health insurance (NHI) in the end of 1995. Under a single payer system operated by the Taiwanese government, national health insurance covered 99% of Taiwan's population by 2009. Several dozens of millions of health service claims are sent to the Bureau of National Health Insurance (BNHI) monthly where these undergo quality assurance checks and are approved for payment (342 million annual claims in 2007). After the billing process, the National Health Research Institutes (NHRI) compiles data from these insurance claims for research purposes. These data then becomes part of Taiwan's National Health Insurance Research Database (NHIRD).

In 2000, the first official version of NHIRD was made available publicly to researchers. Datasets comprehensively included Western medicine and Chinese medicine outpatient visits, dental visits, hospitalizations, medications prescribed, medications refilled,[1] laboratory and imaging examinations, and procedure codes. To protect privacy, individual and hospital identifiers were unique to the research database and not suited to be used either to trace individual patients or hospitals or linked to other census data, such as the cancer registry or household registry. Moreover, chart-level records such as physician notes, laboratory reports, and imaging studies were not available. As a result, the NHIRD was initially criticized by some researchers as a crippled database falling short of the requirements posed for clinical inferences (Hsiao et al. 2007; Harpe 2009).

The NHIRD is now publicly available to researchers in Taiwan and their collaborators. Although any citizen of Taiwan who fulfills the requirements of conducting research projects is eligible to apply for the NHIRD, its use is limited to research purposes only. The NHRI reviews research proposals and collects data use agreements prior to approval of data release. Only a small amount of data processing fee is charged, either NTD$ 500 (US$ 15) per compact disc or NTD$ 200 (US$ 6) per gigabyte data. No extra premium is required for any subsequent data use, hence the concern about grant support is less pressing.

The NHIRD has become an important research material. Due to its distinguishing large-scale and comprehensive data content, exceptional high coverage rate, and long observational period, it is very close to an ideal data source for epidemiologic research (Hsiao et al. 2007). Moreover, it has attracted more and more attention for its availability to researchers. Studying the prevailing use of NHIRD among Taiwan's researchers provides an extraordinary opportunity to observe the emergence of new knowledge and its influence.

Research questions

The main object of this study was to assess the influence of releasing an administrative health care database to the scientific community. As the NHIRD was a newly-released research database in Taiwan, studies using it as data source after its official release in 2000 were identified and analyzed. The results would indicate the overall effects brought by the emergence of NHIRD, its growing influence on authors, and its increasing impact on the scientific community.

[1] For patients in a chronic and stable disease condition in Taiwan, one is eligible to refill his or her medication in any contracted pharmacy without seeing a doctor again.

The following were the specific questions:

1. What is the publication-related effect resulting from the release of NHIRD? Does the number of studies increase and to what extent does the number of fields accumulate?
2. How do researchers deal with newly-released research material? How fast can a researcher take part in relevant fields and how do researchers collaborate?
3. Does the NHIRD affect only fields of health utilization or does it also have a significant influence on other fields?

Methods and procedures

Studies for analysis

Because PubMed extensively covered health-related studies, it was used to identify publications that might have drawn on the NHIRD as their primary data source. To capture as many studies as possible, a wide search query was built to include studies that mentioned "Taiwan" in their textual fields (title, abstract, MeSH index terms, and author's affiliation address) and met any of the following inclusion criteria: (1) the study was indexed with either medical subject heading (MeSH) "insurance, health" or "national health programs"; (2) any of the following terms appeared in any textual field of the study including title, abstract, and MeSH indexing terms. The search terms broadly included "health insurance", "national health", "national insurance", "claims data*", "claim data*", "insurance claim*", "insurance data*", "administrative data*", "nationwide data*", "national data*", "NHIRD", "NHI", "BNHI" and "population based"; and (3) either "population*" or "nationwide" was used in the study title.

The beginning of time span of the list was set to 1996, according to the year of launch of the NHI in Taiwan. To make the results comparable to other studies, only studies written in English or published with English abstracts were included.

Identifying studies using the NHIRD

The list of potential studies using NHIRD was so comprehensive that 1,517 studies were included. The first and second authors manually reviewed the abstracts of potential studies and classified them into one of following groups: "used NHIRD definitely", "not used NHIRD definitely" and "cannot determine by abstract". The reviewing process was performed independently and achieved a very high agreement ($k = 0.934$, Kappa test for inter-rater agreement).

For 62 studies either classified differently or classified as "cannot determine by abstract", the reviewers re-examined these by full text. As such, 383 studies were confirmed as having used the NHIRD as data source, was subsequently named "NHIRD studies", and been included in the analysis.

Authorship and collaboration network of NHIRD studies

Each author's productivity was measured using author names in last-name-plus-initial form as recorded in PubMed. Since the amount of commonly authored papers reflected collaboration among researchers, a social network reflected collaboration among researchers. A social network of co-authorship among NHIRD studies would visualize the dynamic of

information sharing among authors. Using 383 NHIRD studies between 2000 and 2009, a co-authorship network with nodes representing 163 proficient authors who published at least 3 studies was created. Pairs of nodes were interconnected for co-occurring authors. The width of connections between nodes, calculated by the count of co-authored studies, denoted the strength of collaboration between authors.

NLM indexed subject categories

Each study was designated with one or more subject categories according to the subject terms of the journal where the study appeared. The journals' subject terms were Medical Subject Headings (MeSH) obtained from the National Library of Medicine's online catalogue (National Library of Medicine 2010).[2] Each MeSH descriptor assigned to a journal was traced upwards in the MeSH tree and the resulting top level MeSH categories was assigned to all articles in that journal. To capture NHIRD studies relevant to "clinical science", studies published in journals indexed with any of the top-level MeSH categories A through G were subsequently defined as studies pertinent to "clinical science".

The broad concept of "clinical science" thus included "Anatomy", "Organisms", "Diseases", "Chemicals and Drugs", "Analytical, Diagnostic and Therapeutic Techniques and Equipment", "Psychiatry and Psychology", and "Biological Sciences". Studies published in journals indexed with any of the remaining 9 top-level MeSH categories were classified as "general health".

Overlay over global science map

Leydesdorff and Rafols (2009) developed a methodology to sketch science literature disciplinary structures into a global science map, which provided a robust skeleton for science mapping (Rafols and Meyer 2010). Based on the ISI Science Citation Index 2007 (www.isiknowledge.com), Leydesdorff and Rafols (2010) constructed the latest global science map, which covered 18 macro-disciplines and 221 subject categories, as framework to demonstrate dynamics among disciplines. Their approach of visualizing the evolution of study fields was used.

Data process and availability

Citations of NHIRD studies were downloaded from PubMed in the original XML format, stored in Microsoft SQL server 2008, and linked with the NLM Catalog and ISI Journal Citation Report 2008. Pajek was used for network analysis.

Findings

Growth of NHIRD studies

The number of NHIRD studies increased rapidly since the first study appeared as early as 2000, the year of NHIRD's public release (Table 1). The average annual growth rate of

[2] The NLM has assigned most of MEDLINE journals with one or more subject terms to describe journals' overall scopes. The subject terms are all valid MeSH terms and followed the MeSH tree structure ontology. The MeSH tree structure ontology is a hierarchical semantic tree and makes it possible to reorganize subject MeSH terms into higher conceptual levels.

Table 1 Annual number, cumulated numbers, average annual growth rates, and doubling time of NHIRD studies published between 2000 and 2009

Publish year	Number of NHIRD studies		No. of NHIRD studies indexed in JCR2008		Number of distinct authors publishing papers		Number of study fields		Number of journals	
	Annual count	Cumulated count	Annual count	Cumulated count	Annual count	Cumulated count	Annual count	Cumulated count	Annual count	Cumulated count
2000	1	1			8	8	1	1	1	1
2001	1	2	1	1	5	13	1	2	1	2
2002	7	9	5	6	23	36	6	7	5	7
2003	18	27	12	18	56	83	25	31	17	22
2004	32	59	28	46	93	154	41	61	25	41
2005	26	85	23	69	102	219	38	84	23	56
2006	55	140	55	124	149	310	59	119	43	88
2007	43	183	42	166	146	384	62	140	38	111
2008	93	276	85	251	226	510	128	202	76	159
2009	107	383	102	353	286	667	120	250	83	210
Average 5-year annual growth rate (%)[a]		45.8		51.1		34.2		33.0		39.0
Doubling time[b] (year)		2.65		2.49		3.69		3.66		3.15

[a] Annual growth rate = (no. of studies in current year − no. of studies in previous year)/no. of studies in previous year. The average 5-year annual growth rate is the average number of annual growth rate between 2005 and 2009

[b] Doubling time is estimated by fitted power growth model

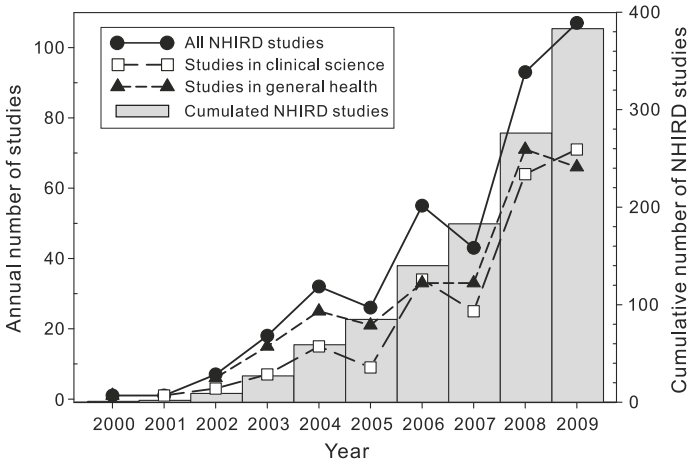

Fig. 1 Annual and cumulated numbers of NHIRD studies published between 2000 and 2009. All NHIRD studies are categorized into clinical science, general health, or both by their journals' coverage scopes

NHIRD studies was 45.8% over the last decade, which was considerably higher than all PubMed literature growth rate of 17% between 2002 and 2006 (DeShazo et al. 2009). In addition to its rapid growth, the NHIRD widely attracted 667 authors and it was used extensively in 383 studies published in 210 journals across 250 study fields at the end of 2009. The numbers of authors, journals and study fields doubled every 2 years.

The growth trend of high quality NHIRD studies was remarkable aside from their substantially rising number. Most (92.2%) of the NHIRD studies were indexed in the Science Citation Index 2008 at the end of 2009. Moreover, it took only 2.49 years to double the number of SCI indexed NHIRD studies. According to Fernandez-Cano et al. (2004), the number of scientific literature usually doubled every 10–20 years. They were not aware of any scientific discipline with this order of magnitude of duplication time.

The NHIRD studies achieved fast progress in both fields of clinical science and general health between 2000 and 2009. By the end of 2009, the cumulated number of general health studies had outweighed that of clinical science by 10.8% (70.6 vs. 59.8%). Nonetheless, the number of studies in clinical science grew at a faster average annual growth rate at 56.2% than general health studies (42.2%) (Fig. 1). Assuming all influential environmental factors remained the same, the cumulated number of studies in clinical science would exceed that of general health by 2012.

The above findings suggest that NHIRD studies have greatly expanded in both quantity and quality. The growing trend of NHIRD studies is similar to the proposed growth model of knowledge development (Szydlowski and Krawiec 2009). As the essential knowledge or technology required for conducting NHIRD studies accumulate, the NHIRD continuous to draw more and more scholars, and concomitantly broadens the fields of application. The NHIRD fosters strong growth of studies in either the field of clinical science or that of general health.

Authors' productivity patterns

A total of 667 authors had published 383 NHIRD studies by the end of 2009. Most (96.9%) NHIRD studies were co-authored, with the number of authors per study ranging from one

Table 2 Number of authors and NHIRD studies with their cumulated percentage to all, as grouped by the number of authored NHIRD studies between 2000 and 2009

No. of NHIRD studies per author	No. of authors	Cumulated percentage to all authors (%)	Cumulated no. of NHIRD studies	Cumulated percentage to all NHIRD studies (%)
≥20	8	1.2	175	45.7
10–19	18	3.9	231	60.3
2–9	245	40.6	365	95.3
1	396	100.0	383	100.0

to 13, with a median of four authors per study. Analyzing affiliations of first author, the majority were conducted in universities (69.5%, 266 out of 383 NHIRD studies) followed by hospitals (22.5%, 86 studies) and research institutes (6%, 23 studies). Moreover, the NHIRD had been used overseas. There were some NHIRD studies conducted by authors outside Taiwan (4%, 17 out of 383 NHIRD studies). The share of international work on NHIRD was lower than the previous reported share of clinical studies with international collaboration (4 vs. 13.6%) (Chen et al. 2007).

The advantage in conducting and publishing NHIRD studies is cumulative. Since the public release of NHIRD in 2000, the most proficient author (Lin HC) has published as many as 99 NHIRD studies whereas the majority (59.3%) of authors has been involved in only one study until the end of 2009. Moreover, the top eight most proficient authors have contributed nearly half (45.7%) of NHIRD studies, while 60.3% are produced by the 26 most proficient authors (3.9% of all authors) (Table 2). The data here fitted well with the general form of Lotka's Law[3] ($R(n) = C/n^a$), with Lotka's exponent (a) as small as 1.837 ($R^2 = 0.998$). The smaller Lotka's exponent of NHIRD studies than that of the humanities and natural sciences ($a = 2.5$ on average) and social science ($a = 3$ on average) suggests the existence of the phenomenon "success breeds success" and implies that success is more reproducible here than in other fields (Egghe 2005).

Researchers tended to collaborate on NHIRD studies in the beginning. At the end of 2005, nearly all (98.6%, 295 of 299) NHIRD studies were co-authored, with five authors per study as the median. Most researchers worked in groups and they were connected either directly or indirectly through co-written NHIRD studies between 2000 and 2005 (Fig. 2a). The pioneer research groups facilitated idea sharing and knowledge creation among researchers, including solving fundamental and technical problems in the emerging stage (Lambiotte and Panzarasa 2009; Larsen 2008).

By the end of 2009, new researchers continued to join pioneer research group members. The increase in number of co-written studies therefore enhanced the existing connection in terms of co-authorship (Fig. 2b). This was consistent with the model of knowledge growth (Gupta and Karisiddappa 2000; Larsen 2008; Szydlowski and Krawiec 2009), whereby once the essential conjectures had been proven and problems had been solved, more and more researchers would take part in research groups to share the information.

Harpe (2009) indicated that there were methodological obstacles in analyzing administrative health care databases. Researchers should familiarize themselves with such fundamental issues when conducting studies. Schneeweiss and Avorn (2005) suggested that researchers work closely with knowledgeable people about every detail of the database to

[3] General form of Lotka's law is $R(n) = C/n^a$, where $R(n)$ is the number of authors that produce n studies, C is a constant characteristic of the research field, and a is called "Lotka's exponent" ranges approximately from value of 2–4 (Hayes 2000).

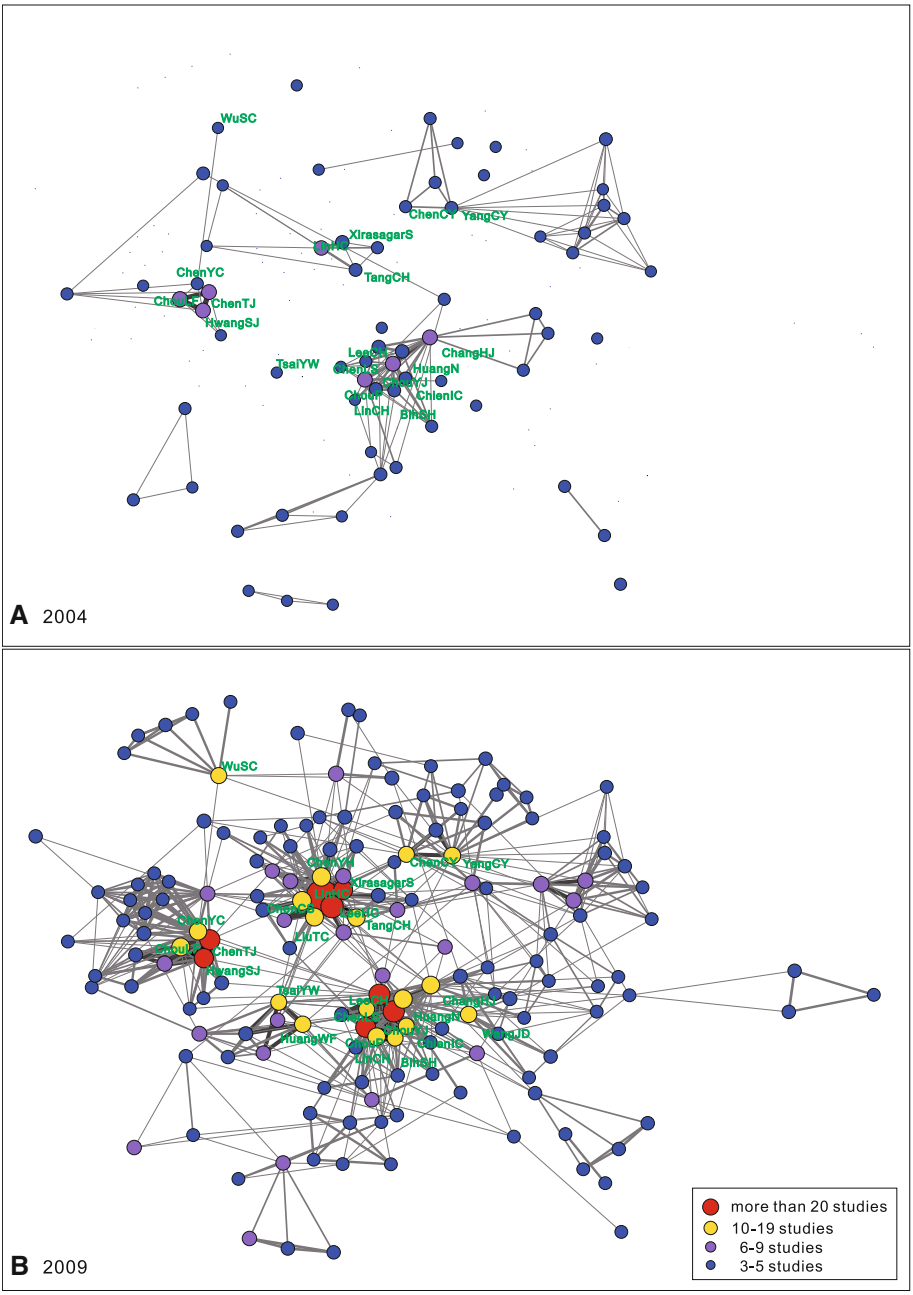

Fig. 2 Volumes of studies and co-authorship network among 163 proficient authors at the end of (**a**) 2004 and (**b**) 2009. Each node represents an author with its size proportional to the number of authored studies. Pairs of nodes are interconnected for co-written papers with the strength of collaboration denoted by numbers of co-authored studies expressed in gray gradient. The layout of the network is plotted using co-authorship data of NHIRD studies between 2000 and 2009 by the Pajek program

prevent important errors and resultant biases. This provided an excellent explanation to the evolved patterns of collaboration.

The findings here clearly demonstrate that researchers work closely on conducting NHIRD studies. The participation of subsequent researchers strengthens the existing collaboration network. Moreover, while the success in publishing NHIRD studies is reproducible, the "success breeds success" phenomena holds true for NHIRD studies.

Scattering of NHIRD studies

NHIRD studies spread across diverse journals. At the end of 2009, 383 NHIRD studies have been published in 210 journals, for an average of 1.8 studies per journal (Table 3). The majority (68%) of journals have published only one study whereas *Journal of the Formosan Medical Association*, the journal ranked first by number of published NHIRD studies, yielded 5% of NHIRD studies (Table 4). As the NHIRD studies expanded their coverage of journals, Bradford's "core publications" for NHIRD studies had not been identified yet.

The use of NHIRD for academic studies evolved rapidly. "Health occupation", followed by "Therapeutics", took the lead of the ten most used journal-level NLM indexed subject categories ranked by number of NHIRD studies between 2000 and 2009 (Table 5). By comparing the ranking and share of the ten most involved subject categories in the first 5 years (2000–2004) and last 5 years (2005–2009), "Health Occupations" remained the most used subject category although its share decreased by 21.4% because of the expanding number of new subject categories. Studies on "Therapeutics", "Health Care Quality, Access, and Evaluation", and "Health Services Administration" improved their ranks in terms of faster increase in numbers.

NHIRD had been applied in a growing number of study fields. In Fig. 3, 353 out of 383 NHIRD studies indexed in SCI & SSCI (2008) were overlaid on a global science map. The number of applied ISI study fields had doubled, from 30 fields in 2004 to 60 fields in 2009, with the average number of studies per field increasing from 3.8 (studies/field) to 5.9 (studies/field). While three macro disciplines—"Clinical Medicine" (34.8%, 123 out of 353 NHIRD studies indexed in SCI & SSCI 2008), "Health and Social Issues" (25.2%, 89 studies), and "Biomedical Sciences" (23.8%, 84 studies) continued developing, as three main streams of NHIRD studies between 2000 and 2009, interdisciplinary study fields such as "Economics" and "Computer Science, Information Systems" gradually emerged.

Analyzing indexed MeSH terms of each study, the NHIRD was used most frequently in cardiovascular disease studies (17.6%, 62 out of 353 NHIRD studies indexed in SCI & SSCI 2008), followed by mental disorders studies (14.4%, 51 studies), neurologic studies (13.9%, 49 studies), and psychological behavior studies (11.6%, 41 studies) (Table 6).

Table 3 Numbers of NHIRD studies and contribution journals grouped by published numbers of studies

No. of published studies per journal	No. of journals	Proportion to all journals	No. of yielded studies	Proportion to all NHIRD studies
≥4	19	9.0	128	33.4
3–2	48	22.9	112	29.2
1	143	68.1	143	37.3
Total	210	100.0	383	100.0

Table 4 Top 10 journals ranked by number of published NHIRD studies between 2000 and 2009

Journal name	Impact factor[a]	Rank in SCI subject category (total number of journals)[b]	No. of NHIRD studies	% to all NHIRD studies (%)
Journal of the Formosan Medical Association	0.640	83th in MEDICINE, GENERAL & INTERNAL (107)	19	5.0
Health policy	1.334	38th HEALTH CARE SCIENCES & SERVICES (62)	15	3.9
BMC health services research	1.680	29th in HEALTH CARE SCIENCES & SERVICES (62)	12	3.1
Pharmacoepidemiology and drug safety	2.611	94th in PHARMACOLOGY & PHARMACY (219)	8	2.1
BMC public health	2.029	46th in PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH (105)	7	1.8
Journal of affective disorders	3.271	34th in PSYCHIATRY (101), 35th in CLINICAL NEUROLOGY (156)	7	1.8
Schizophrenia research	4.174	18th in PSYCHIATRY (101)	6	1.6
Clinical therapeutics	0.350	209th in PHARMACOLOGY & PHARMACY (219)	6	1.6
Journal of the Chinese Medical Association	SCI-E[c]		5	1.3
Stroke	6.499	7th in PERIPHERAL VASCULAR DISEASE (56), 6th in CLINICAL NEUROLOGY (156)	5	1.3
Health policy and planning	1.953	23th in HEALTH CARE SCIENCES & SERVICES (62)	5	1.3
Vaccine	3.298	19th in MEDICINE, RESEARCH & EXPERIMENTAL (83), 42th in IMMUNOLOGY (121), 1st in VETERINARY SCIENCES (135)	5	1.3

[a] Based on the 2008 Journal Citation Reports—Science Edition, the Thomson Corporation, Philadelphia

[b] Based on the 2008 Journal Citation Reports—Science Edition. The numbers in parentheses were total number of journals in the given SCI subject category

[c] *Journal of the Chinese Medical Association* had been indexed only in the Science Citation Index Expanded (SCI-E) but not in the 2008 Journal Citation Reports. Hence, either the impact factor or rank in SCI subject category was not available

These findings illustrate the strength and extent of spread of NHIRD studies. NHIRD studies continue to expand their coverage in terms of number of journals and study fields. The study fields of clinical science and general health have substantially benefited from the release of NHIRD.

Limitations

This study has some limitations. First, to make the results comparable, NHIRD studies written and published in English were included in the analysis. This considerably underestimated the overall scientific output related to the NHIRD. If unpublished studies and those written in local languages were considered, the NHIRD-related output would be much higher. Second, an extensive search in PubMed was conducted because of its wide

Table 5 Top 10 NLM subject categories ranked by number of NHIRD studies between 2000 and 2009 compared to the numbers and rankings during the first (2000–2004) and last 5-years (2005–2009)

NLM subject categories[a]	2000–2004 ($n = 59$)			2005–2009 ($n = 324$)			2000–2009 ($n = 383$)		
	No. of studies	Proportion to all (%)	Rank	No. of studies	Proportion to all (%)	Rank[b]	No. of studies	Proportion to all (%)	Rank
Health Occupations [H02]	37	62.7	1	160	49.4	1	197	51.4	1
Therapeutics [E02]	4	6.8	8	43	13.3	2^-	47	12.3	2
Health Care Economics and Organizations [N03]	6	10.2	4	39	12.0		45	11.7	3
Health Care Quality, Access, and Evaluation [N05]	3	5.1	10	41	12.7	3^+	44	11.5	4
Health Care Facilities, Manpower, and Services [N02]	10	16.9	2	31	9.6	7^-	41	10.7	5
Natural Science Disciplines [H01]	6	10.2	4	35	10.8	5^-	41	10.7	5
Health Services Administration [N04]	5	8.5	7	34	10.5	6^+	39	10.2	7
Behavioral Disciplines and Activities [F04]	6	10.2	4	26	8.0	8^-	32	8.4	8
Social Sciences [I01]	7	11.9	3	24	7.4	10^-	31	8.1	9
Environment and Public Health [N06]	4	6.8	8	26	8.0	8	30	7.8	10

[a] NLM subject categories are indexed MeSH descriptors (MeSH code given in square bracket []) of each study's journal-level scope defined in the NLM catalog (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nlmcatalog)

[b] Superscripts denote change in rank compared with the first 5 year (2000–2004); (+ upward, − downward)

coverage of health related articles. However, as the study indicated, the NHIRD might not have its use limited only to health research. It might have greater influence on fields like social sciences or economics, which were not covered in PubMed. Third, a 10-year observation period might be insufficient to capture the whole dynamics of information spreading. As NHIRD is still in its infancy, the output is significant but its impact to society as a whole may require more time to validate.

Discussion and conclusions

The influence of claim-based studies may be increasing but not well explored. Taiwan's NHIRD is one of the large-scale, nationwide administrative health care databases around the world. Analyzing scientific output related to the NHIRD between 2000 and 2009, it can be concluded that the NHIRD has positively influenced the scientific society in Taiwan. Since its pioneering release to the public in 2000, the NHIRD has received more and more attention. It has expanded rapidly in quantity and quality of studies. Strong connections between researchers have been formed, which have enabled the "success-breeds-success"

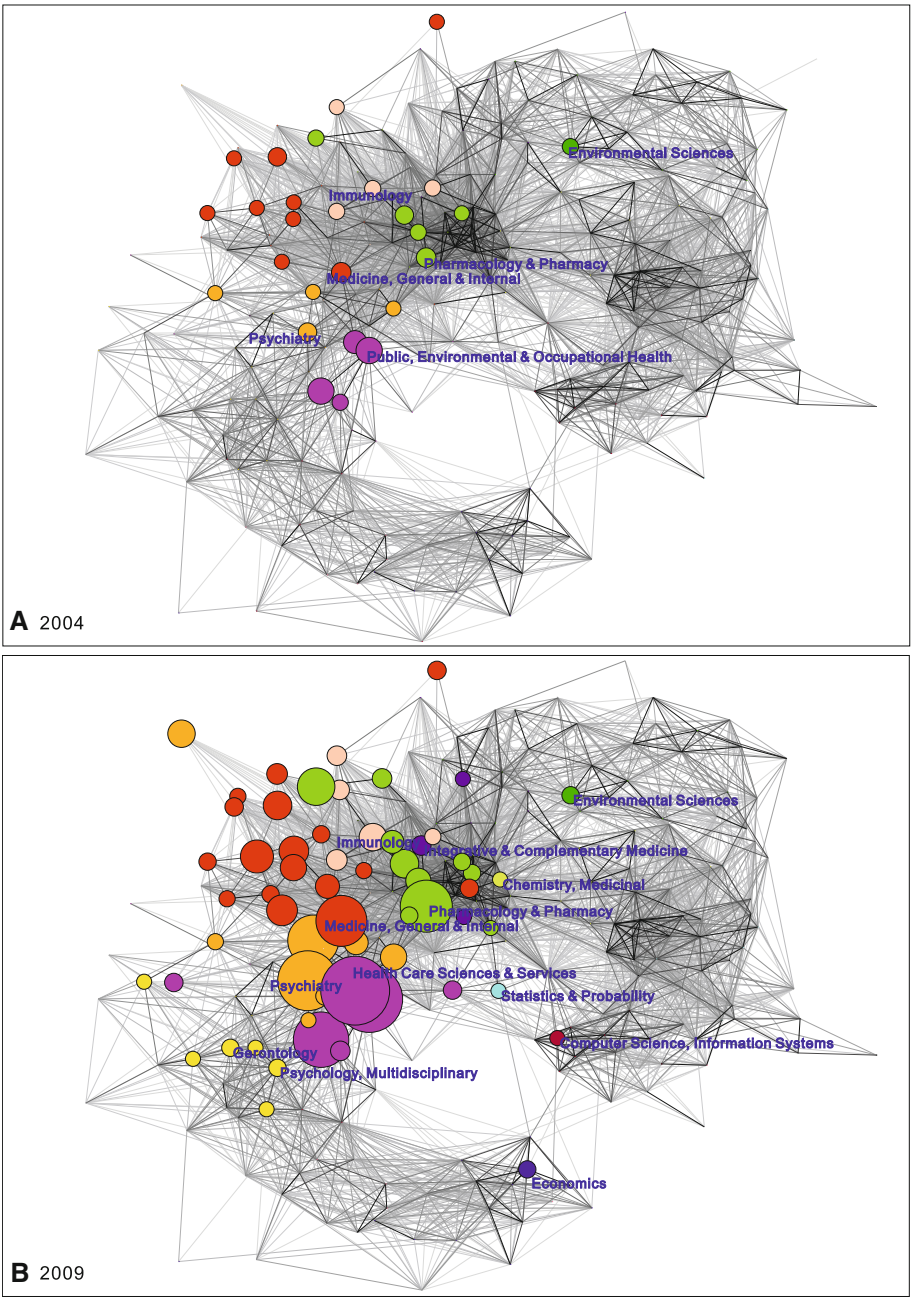

Fig. 3 Spread of NHIRD studies across 221 SCI and SSCI subject categorized from (**a**) 2004 to (**b**) 2009. Each node represents an SCI & SSCI subject category with its size proportional to the number of NHIRD studies. The nodes are colored in 18 macro disciplines and overlaid on a global science map constructed using SCI & SSCI (2007) citation data provided by Leydesdorff and Rafols (2010)

Table 6 Top 10 Medical Subject Headings (MeSH) and its leading 2 subheadings ranked by number of NHIRD studies indexed in JCR between 2000 and 2009

Covered Medical Subject Headings (MeSH)[a]	No. of NHIRD studies	% to NHIRD studies index in JCR2008 (%)	Average impact factor[b]
Cardiovascular diseases	62	17.6	3.29
Vascular diseases	51	14.4	3.42
Heart diseases	20	5.7	3.14
Mental disorders	51	14.4	3.26
Schizophrenia and disorders with psychotic features	20	5.7	3.52
Mental disorders, others	12	3.4	3.26
Nervous system diseases	49	13.9	3.10
Central nervous system diseases	32	9.1	3.66
Neurologic manifestations	6	1.7	2.39
Behavior and behavior mechanisms	41	11.6	3.34
Behavior	18	5.1	3.13
Psychology, social	17	4.8	4.18
Respiratory tract diseases	37	10.5	3.16
Lung diseases	23	6.5	3.25
Respiratory tract infections	22	6.2	3.49
Endocrine system diseases	31	8.8	3.62
Diabetes mellitus	30	8.5	3.61
Endocrine system diseases	1	0.3	4.12
Digestive system diseases	30	0.0 8.5	3.17
Gastrointestinal diseases	20	5.7	3.48
Liver diseases	7	2.0	2.92
Neoplasms	29	8.2	3.30
Neoplasms, others	14	4.0	3.59
Neoplasms by site	14	4.0	3.21
Nutritional and metabolic diseases	28	7.9	3.80
Metabolic diseases	27	7.6	3.80
Nutrition disorders	4	1.1	3.80
Female urogenital diseases and pregnancy complications	26	7.4	3.26
Pregnancy complications	17	4.8	3.41
Female urogenital diseases	10	2.8	2.78

[a] Based on the Medical Subject Headings (MeSH) version 2010 of U.S. National Library of Medicine. Two subheadings with highest number of studies within given MeSH were shown in italic

[b] Based on the 2008 Journal Citation Reports (JCR). The average impact factor was the average number of articles' journal impact factors in the given subject category

phenomena. Aside from general medicine and health science, the NHIRD has been continuously applied to a broad range of study fields.

This study clearly demonstrates the academic benefits of a publicly available, large-scale administrative health care database. The experience of Taiwan's NHIRD should encourage national- or institutional-level data holders to consider re-using their

administrative databases for academic purposes. For researchers, professional knowledge is still crucially required in handling database specific problems (Berger et al. 2009; Harpe 2009). Several information technology frameworks are proposed to facilitate collaborations among scholars and lower technical barriers (Diamond et al. 2009). If such aforementioned barriers are broken down, as this study shows, success can breed more success.

In conclusion, a public administrative health care database will significantly promote scientific production in many ways. As more and more large-scale data sources become available, there will be a need to track the use of similar databases in other countries and observe their influence on scientific communities worldwide.

Acknowledgments The authors thank two anonymous reviewers who generously contributed their valuable view points and helpful comments.

References

Berger, M. L., Mamdani, M., Atkins, D., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report-part I. *Value in Health, 12*(8), 1044–1052.

Chen, T. J., Chen, Y. C., Hwang, S. J., & Chou, L. F. (2007). International collaboration of clinical medicine research in Taiwan, 1990–2004: A bibliometric analysis. *Journal of the Chinese Medical Association, 70*(3), 110–116.

DeShazo, J. P., LaVallie, D. L., & Wolf, F. M. (2009). Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH. *BMC Medical Informatics and Decision Making, 9*, 7.

Diamond, C. C., Mostashari, F., & Shirky, C. (2009). Collecting and sharing data for population health: A new paradigm. *Health Affairs, 28*(2), 454–466.

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Oxford, UK: Elsevier.

Fernandez-Cano, A., Torralbo, M., & Vallejo, M. (2004). Reconsidering Price's model of scientific growth: An overview. *Scientometrics, 61*(3), 301–321.

Gupta, B. M., & Karisiddappa, C. R. (2000). Modeling the growth of literature in the area of theoretical population genetics. *Scientometrics, 49*(2), 321–355.

Harpe, S. E. (2009). Using secondary data sources for pharmaco-epidemiology and outcomes research. *Pharmacotherapy, 29*(2), 138–153.

Hayes, R. M. (2000). The mathematical foundations of bibliometrics. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield*. New Jersey, USA: Information Today.

Hoffmann, F. (2009). Review on use of German health insurance medication claims data for epidemiological research. *Pharmacoepidemiology and Drug Safety, 18*(5), 349–356.

Hsiao, F. Y., Yang, C. L., Huang, Y. T., & Huang, W. F. (2007). Using Taiwan's national health insurance research databases for pharmacoepidemiology research. *Journal of Food and Drug Analysis, 15*(2), 99–108.

Lambiotte, R., & Panzarasa, P. (2009). Communities, knowledge creation, and information diffusion. *Journal of Informetrics, 3*(3), 180–190.

Larsen, K. (2008). Knowledge network hubs and measures of research impact, science structure, and publication output in nanostructured solar cell research. *Scientometrics, 74*(1), 123–142.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348–362.

Leydesdorff, L., & Rafols, I. (2010). A user-friendly method for generating overlay maps. Retrieved Jan 15, 2010 from http://users.fmg.uva.nl/lleydesdorff/overlaytoolkit/.

National Library of Medicine (2010). NLM Catalog. Retrieved Jan 15, 2010 from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nlmcatalog.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of inter-disciplinarity: Case studies in bio-nanoscience. *Scientometrics, 82*(2), 263–287.

Schneeweiss, S., & Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology, 58*(4), 323–337.
SCI & SSCI (2007). Retrieved Jan 15, 2010 from http://isiknowledge.com/.
SCI & SSCI (2008). Retrieved Jan 15, 2010 from http://isiknowledge.com/.
Szydlowski, M., & Krawiec, A. (2009). Growth cycles of knowledge. *Scientometrics, 78*(1), 99–111.
Tricco, A. C., Pham, B., & Rawson, N. S. B. (2008). Manitoba and Saskatchewan administrative health care utilization databases are used differently to answer epidemiologic research questions. *Journal of Clinical Epidemiology, 61*(2), 192–197.