

REVIEW ARTICLE

A Review of Studies Using Japanese Nationwide Administrative Claims Databases

So Sato¹, Hideo Yasunaga¹

ABSTRACT

BACKGROUND

Administrative claims databases are increasingly being used worldwide for research purposes. We reviewed original published articles that used one of the four nationwide administrative claims databases in Japan: the National Database of Health Insurance Claims and Specific Health Checkups (NDB), NDB Open Data, the JMDC Claims Database, and the Diagnosis Procedure Combination (DPC) database.

METHODS

Studies published from January 2010 to July 2022 using the JMDC and DPC databases, and from January 2013 to July 2022 using the NDB and NDB Open Data were identified using PubMed. The number of original articles was divided into 19 fields. The annual growth rate of the number of studies was calculated using the four databases.

RESULTS

Overall, 1047 studies were included (95 for the NDB, 31 for the NDB Open Data, 222 for the JMDC database, and 699 for the DPC databases). Studies using one of these four databases increased from around 2010, and the average annual growth rate was approximately 41% from 2010 to 2021. DPC database studies had a higher proportion of articles on surgery (19.2%), urology (3.0%), neurosurgery (6.2%), anesthesiology (1.9%), and emergency medicine (14.0%), whereas the NDB and JMDC data had higher proportions of those regarding internal medicine.

CONCLUSIONS

Since 2010, these four databases have increasingly attracted attention, and the number of studies using them has grown rapidly. Our review suggests that each has unique features, and researchers should understand the database characteristics to operate their studies.

KEY WORDS

administrative claims database, nationwide database, NDB, JMDC, DPC

¹ Department of Clinical Epidemiology and Health Economics, Graduate School of Medicine, The University of Tokyo

Corresponding author: So Sato
Department of Clinical Epidemiology and Health Economics, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 1130033, Japan
E-mail: sato-so4k9@g.ecc.u-tokyo.ac.jp

Received: August 29, 2022

Accepted: December 16, 2022

J-STAGE Advance published date: January 28, 2023

No. 23008

© 2023 Society for Clinical Epidemiology

INTRODUCTION

The growing trend of recording data on all medical encounters in electronic format is increasing the popularity of large datasets in healthcare. This trend has prompted clinical epidemiologists to answer various research questions using database studies [1]. Administrative claims databases include routinely collected data on the primary purpose of healthcare billing. These include real-world data on diagnoses, procedures, and drug prescriptions, which can also be used for research purposes [2]. Compared to randomized controlled trials, studies using administrative claims databases have the following advantages: larger sample size, lower cost, and increased generalizability [3, 4].

There are several nationwide claims databases in Japan, including the National Database of Health Insurance Claims and Specific Health Checkups (NDB), NDB Open Data, JMDC Claims Database, and the Diagnosis Procedure Combination (DPC) database. These have been used in multiple fields, such as clinical epidemiology, pharmacoepidemiology, and health economics and policy. Japan also has other small-scale or regional health insurance databases; however, in this study we focused on these four databases as they are the only ones that have large nationwide data. To the best of our knowledge, there have been two previous reports on these databases [5, 6]. One report only focused on NDB and NDB Open Data [5], and the other study had only a short research period [6]. This study aimed to review studies using the NDB, NDB Open Data, JMDC Database, and DPC databases with over 10-year-research period to help researchers understand the current trends in studies using these databases, particularly focusing on difference in research area between the databases, and determine which can provide the best dataset for their research purposes.

METHODS

OVERVIEW OF THE FOUR DATABASES

In 1961, Japan established a universal healthcare coverage system [7]. Under this, the Ministry of Health, Labour, and Welfare (MHLW) launched the NDB in 2009 and began collecting anonymized electronic health insurance claims data for medical and dental services [8]. The NDB covers more than 126 million people and 1.9 billion electronic claims annually, with data from 99% of hospitals [9]. It can be used to understand the healthcare process for the Japanese population. NDB was publicly released

to researchers in 2011. To use NDB data, the study protocols must be approved by the advisory committee of the MHLW. It extracts data from the NDB and formats them into datasets depending on researchers' requests (i.e., special extraction, sampling dataset, and aggregated data in tabular form) [18]. The NDB contains information on patient age, sex, diagnoses, inpatient and outpatient medical data, dental service use, drug prescriptions, and health checkup data. To protect patient privacy, researchers are not allowed to link the NDB with other databases.

In 2016, the MHLW also began providing a free-access version of NDB (NDB Open Data), which anyone can access through its homepage. The NDB Open Data was created by aggregating a part of the NDB data without any confidential information [10]; therefore, researchers who use it cannot access patient- or facility-level information.

The JMDC Claims database contains commercially available data. The database has anonymized inpatient, outpatient, dispensing receipts, and medical examination data, collected from various health insurance associations [11–13]. As of September 2021, the total number of patients in the database was 13 million. The claims data include information from 2005 on patient enrollment, medical facilities, diagnoses, procedures, drugs and materials, annual health checkups, and the costs for each visit [14]. This database presents inherent limitations, including under-representation of the elderly, because data are collected from health insurance associations for employees and their dependents [15].

The DPC is a case-mix patient classification system, launched in 2002 by the MHLW, and is linked with a lump-sum per-diem payment system. This system has been adopted by more than 1700 large-to middle-sized hospitals in Japan. All 82 university hospitals have to participate in the database, whereas participation by community hospitals is voluntary [16]. DPC databases include administrative claims and discharge abstracts. The data items include unique hospital identifiers such as age, sex, main diagnoses, comorbidities at admission, and complications after admission (recorded with text data in Japanese and with codes from the International Classification of Diseases, 10th Revision), interventional/surgical procedures (indexed with Japanese original procedure codes), duration of anesthesia, length of stay, discharge status, and the total hospitalization cost. DPC databases also contain several clinical data points, including smoking status, body mass index, cancer stage, consciousness level, and activities of daily living. All patient data were recorded at discharge by the attending

physician. A previous validation study showed good sensitivity and specificity of diagnoses and procedure records in the database and high validity of cancer diagnoses [17].

SEARCH STRATEGY AND SELECTION CRITERIA

We searched PubMed for the JMDC and DPC (from January 2010 to July 2022) and for the NDB and NDB Open (from January 2013 to July 2022). The search terms used are described in **Appendix 1**. There are several kinds of DPC databases, but we excluded studies using the Medical Data Vision data, which is commercially available DPC data. The exclusion criteria were as follows: (1) non-English studies, (2) studies outside the period of interest, (3) non-original studies, (4) studies that did not use one of the four databases, and (5) studies with an inaccessible full-text link.

In the screening process, two reviewers screened the titles and abstracts to apply the exclusion criteria and then reviewed the study methods.

DATA EXTRACTION

We conducted a narrative review because our aim was not to compare or synthesize any specific statistical indicators among different studies. The components extracted from the reviewed studies were (1) used database and (2) research field. To define the research fields, we used the Japanese specialty board's categories for senior residents [19]. They were categorized into the following 19 fields: internal medicine, pediatrics, dermatology, psychiatry, surgery, orthopedics, obstetrics and gynecology, ophthalmology, otolaryngology, urology, neurosurgery, radiology, anesthesiology, pathology, clinical laboratory, emergency department, plastic surgery, rehabilitation, and general medicine. When a study was considered to be included in several fields, we selected one category. For example, a study titled "Ophthalmic Corticosteroids in Pregnant Women with Allergic Conjunctivitis and Adverse Neonatal Outcomes" [20] was assigned to ophthalmology. When a study was not included in any of the 19 fields, we selected the "other" category. We calculated the proportions of study fields for the four databases. The growth rate for the number of studies published in year i was calculated by dividing the difference between the number of studies published in year i and year $i-1$ by the number of studies published in year $i-1$. For example, if three studies were published in 2011 and six in 2012, the growth rate in 2012 was calculated as $(6 - 3)/3 = 100\%$.

RESULTS

INCLUSION AND EXCLUSION

The first screening retrieved 1652 studies (300 for NDB and NDB Open Data, 287 for JMDC database, and 1065 for DPC databases). Finally, 1047 studies were included in our review (95 for NDB, 31 for NDB Open Data, 222 for JMDC database, and 699 for DPC databases). **Fig. 1** describes the screening process, and the included studies are listed in **Supplemental Table 1–4**.

NUMBER OF STUDIES EACH YEAR

The numbers of published studies were 8, 18, 24, 27, 45, 68, 60, 53, 103, 105, 148, 221, and 167 in 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, and January to July of 2022, respectively. The average annual growth rate was approximately 41% from 2010 to 2021. The cumulative number of database studies published since 2010 is shown in **Fig. 2**.

RESEARCH FIELDS

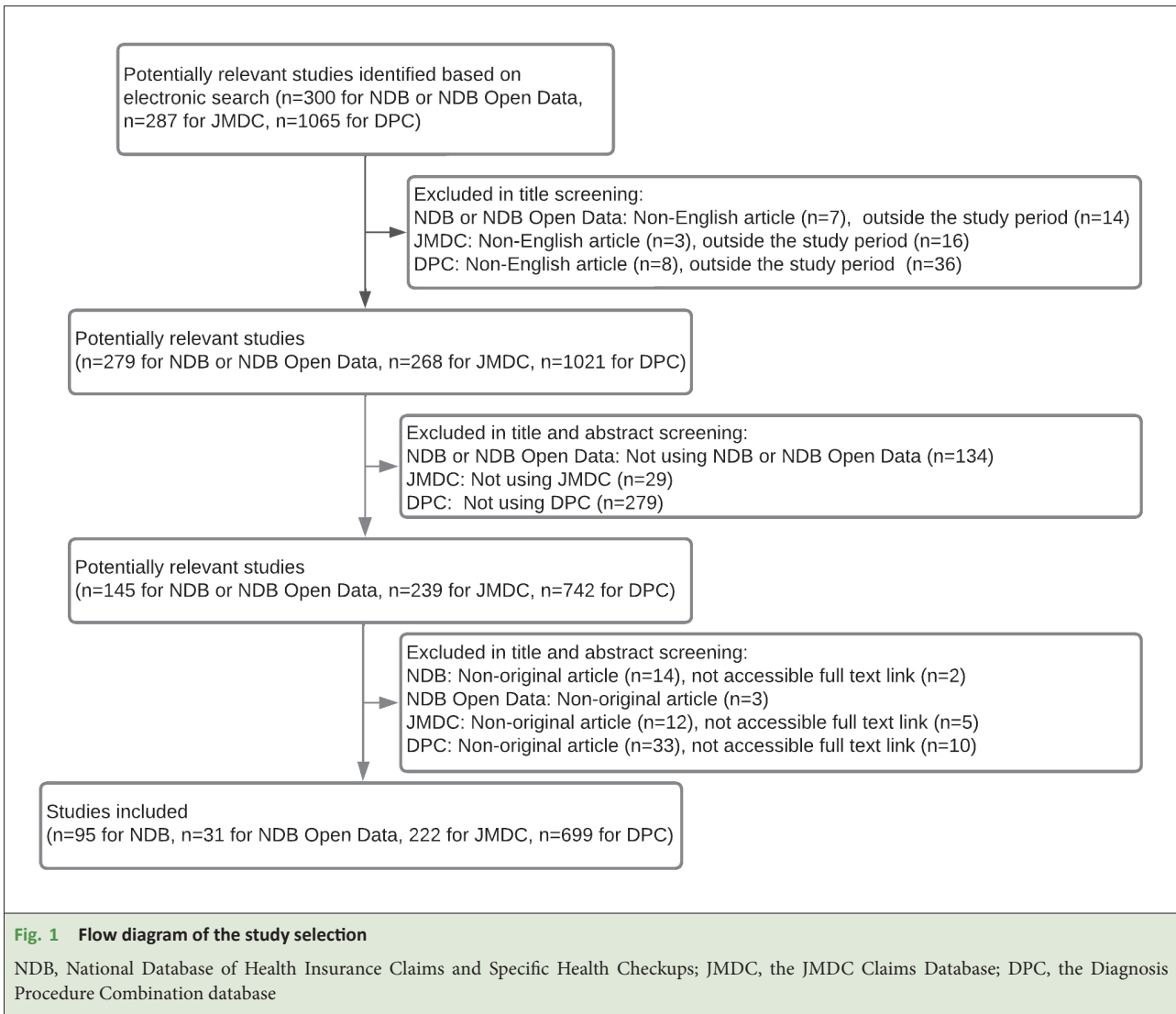
Regarding the research field, 400 studies were included in internal medicine, 138 in surgery, and 100 in emergency medicine. These three fields comprised approximately two-thirds of all studies. **Table 1** describes the number of studies that used the four databases divided by the study fields.

Studies using DPC databases accounted for higher proportions of surgery (19.2%), urology (3.0%), neurosurgery (6.2%), anesthesiology (1.9%), and emergency medicine (14.0%). More than half of the studies using the NDB and approximately two-thirds of the studies using the JMDC database were included in internal medicine.

DISCUSSION

Japanese nationwide administrative claims databases are widely used in academic research. This study investigated the trends in published studies using one of the four administrative claims databases. The number of studies has increased remarkably; the average annual growth rate of database studies was approximately 41% from 2010 to 2021.

The following are the reported number of studies using administrative claims databases in other countries: 1,427 in the United States (Medicare administrative claims databases) from 1979 to 2016 [21], 749 in the United Kingdom (General Practice Research Database) from 1995 to 2009 [22], 70 in Germany (German health insurance medication claims data) from 1998 to 2007 [23],



110 in France (French reimbursement databases) from 1988 to 2009 [24], 325 in Canada (Manitoba and Saskatchewan administrative health care utilization databases) from 1969 to 2004 [25], and 383 in Taiwan (National Health Insurance Research Database) from 2000 to 2009 [26]. We believe the number of Japanese studies was comparable to those in other nations.

Our study showed that many other studies used DPC databases. This may be because it contains more detailed patient data, including several clinical data (smoking status, body mass index, cancer stage, consciousness level, activities of daily living, and others), and these can be used to adjust for patient backgrounds.

The JMDC database is frequently used in research. This may be because of its uncomplicated accessibility. A previous study evaluated the accessibility of administrative healthcare databases in Asia-Pacific countries [27]. They scored database accessibility on a seven-point scale,

assigning a “level seven” score to the JMDC database, a “level three” score to the DPC database, and a “level two” score to the NDB.

Researchers often experience difficulties in gaining access to the NDB. To apply the NDB data, researchers must prepare a high-level security system for data management in their laboratory or use it only at onsite research centers. These security system requirements may limit access to the NDB. However, high accessibility does not necessarily benefit researchers and patients. To increase the NDB’s use in studies, the government must improve its accessibility.

The results suggest that DPC databases have a relatively high affinity for studies on surgery, anesthesiology, and emergency medicine, whereas the NDB and JMDC databases have been used in studies on internal medicine. DPC databases include data on inpatients admitted to acute care hospitals, and DPC database studies are likely

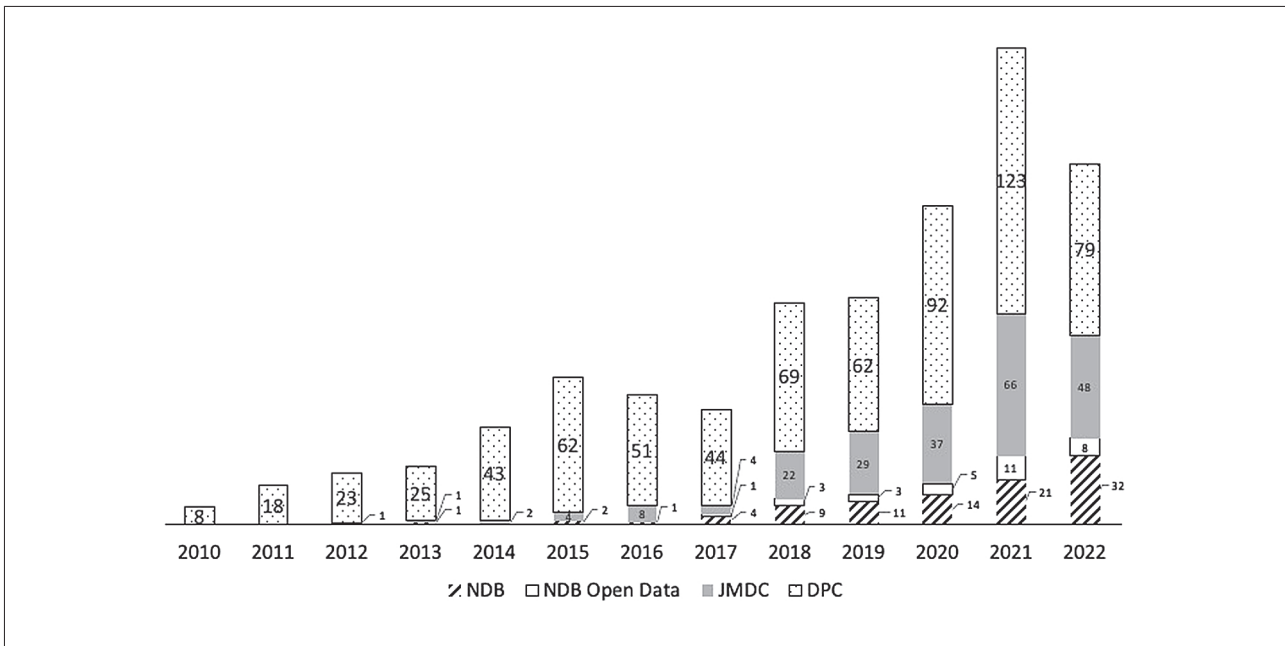


Fig. 2 The number of published articles in the four databases between 2010 and 2021
 NDB, National Database of Health Insurance Claims and Specific Health Checkups; JMDC, the JMDC Claims Database; DPC, the Diagnosis Procedure Combination database

	NDB, n (%)	NDB Open Data, n (%)	JMDC, n (%)	DPC, n (%)	All, n (%)
All	95	31	222	699	1047
Internal medicine	49 (51.6)	12 (38.7)	144 (64.9)	195 (27.9)	400 (38.2)
Pediatrics	3 (3.2)	1 (3.2)	12 (5.4)	43 (6.2)	59 (5.6)
Dermatology	1 (1.1)	1 (3.2)	4 (1.8)	3 (0.4)	9 (0.9)
Psychiatry	12 (12.6)	1 (3.2)	17 (7.7)	12 (1.7)	42 (4.0)
Surgery	3 (3.2)	1 (3.2)	0 0.0	134 (19.2)	138 (13.2)
Orthopedics	7 (7.4)	5 (16.1)	6 (2.7)	56 (8.0)	74 (7.1)
Obstetrics and gynecology	5 (5.3)	2 (6.5)	7 (3.2)	14 (2.0)	28 (2.7)
Ophthalmology	3 (3.2)	1 (3.2)	14 (6.3)	4 (0.6)	22 (2.1)
Otolaryngology	0 0.0	0 0.0	3 (1.4)	13 (1.9)	16 (1.5)
Urology	0 0.0	1 (3.2)	2 (0.9)	21 (3.0)	24 (2.3)
Neurosurgery	1 (1.1)	0 0.0	2 (0.9)	43 (6.2)	46 (4.4)
Radiology	0 0.0	1 (3.2)	0 0.0	3 (0.4)	4 (0.4)
Anesthesiology	0 0.0	0 0.0	0 0.0	13 (1.9)	13 (1.2)
Pathology	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Clinical laboratory	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Emergency medicine	0 0.0	0 0.0	2 (0.9)	98 (14.0)	100 (9.6)
Plastic surgery	0 0.0	0 0.0	0 0.0	3 (0.4)	3 (0.3)
Rehabilitation	4 (4.2)	2 (6.5)	2 (0.9)	24 (3.4)	32 (3.1)
General medicine	4 (4.2)	0 0.0	2 (0.9)	13 (1.9)	19 (1.8)
Other	3 (3.2)	3 (9.7)	5 (2.3)	7 (1.0)	18 (1.7)

NDB, National Database of Health Insurance Claims and Specific Health Checkups; JMDC, the JMDC Claims Database; DPC, the Diagnosis Procedure Combination database

to use short-term outcomes, including postoperative complications, length of stay, and in-hospital mortality. The NDB and JMDC databases include data on health checkups and outpatient data; therefore, studies using these are likely to investigate lifestyle diseases (including diabetes, hypertension, and dyslipidemia) or other chronic diseases.

Similar to a previous study [5], several limitations of the database studies were also reported in the reviewed studies (e.g., missing important data, unchecked validation of the data coding, and unclear causal relationship). To address the problem of validation of the data coding, several validation studies were conducted, mainly in DPC databases [17, 28–30], and they reported high validity for information in some specific areas. However, the number of validation studies in Japanese databases is less than in Western countries [31]. There is a need for more validation studies using Japanese databases.

This study has some limitations. First, although we reviewed only English articles, a previous study reported that the number of studies written in Japanese was not small [5]. Second, the search terms we used might not be perfect to find all of the relevant studies in the four data-

bases. Third, there are other Japanese health insurance databases, and we were unable to review all studies using administrative claims databases in Japan.

CONCLUSION

Since 2010, NDB, NDB Open Data, JMDC database, and DPC databases have increasingly attracted attention, and the number of studies using them has grown rapidly. Our review revealed that since each database has unique features, researchers should understand them to conduct their studies.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Health, Labour and Welfare, Japan (21AA2007 and 22AA2003), and the Ministry of Education, Culture, Sports, Science and Technology, Japan (20H03907). The supporting agencies only provided grant support. They were not involved in this research.

REFERENCES

- Schneeweiss S, Avorn J. A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics. *J Clin Epidemiol* 2005;58:323–37.
- Suissa S, Garbe E. Primer: Administrative Health Databases in Observational Studies of Drug Effects—Advantages and Disadvantages. *Nat Clin Pract Rheumatol* 2007;3:725–32.
- Harpe SE. Using Secondary Data Sources for Pharmacoepidemiology and Outcomes Research. *Pharmacotherapy* 2009;29:138–53.
- Gandhi SK, Salmon W, Kong SX, Zhao SZ. Administrative databases and outcomes assessment: an overview of issues and potential utility. *J Manag Care Pharm* 1999;5:215–22.
- Hirose N, Ishimaro M, Morito K, Yasunaga H. A review of studies using the Japanese national database of health insurance claims and specific health checkups. *Ann Clin Epidemiol* 2020;24:13–26.
- Fujinaga J, Fukuoka T. A Review of Research Studies Using Data from the Administrative Claims Databases in Japan. *Drugs - Real World Outcomes* 2022.
- Reich MR, Ikegami N, Shibuya K, Takemi K. 50 Years of Pursuing a Healthy Society in Japan. *Lancet* 2011;378:1051–3.
- Toyokawa S, Maeda E, Kobayashi Y. Estimation of the Number of Children with Cerebral Palsy Using Nationwide Health Insurance Claims Data in Japan. *Dev Med Child Neurol* 2017;59:317–21.
- Okumura Y, Sakata N, Takahashi K, Nishi D, Tachimori H. Epidemiology of Overdose Episodes from the Period Prior to Hospitalization for Drug Poisoning Until Discharge in Japan: An Exploratory Descriptive Study Using a Nationwide Claims Database. *J Epidemiol* 2017;27:373–80.
- Ministry of Health, Labour and Welfare. NDB Open Data [in Japanese]. Available from: <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177182.html>, Accessed 2022 Aug 4.
- Yasunaga H. Real world data in Japan: chapter I. *Ann Clin Epidemiol* 2019;1:28–30.
- Tanaka S, Seto K, Kawakami K. Pharmacoepidemiology in Japan: medical databases and research achievements. *J Pharm Health Care Sci* 2015;1:16.
- Nagai K, Tanaka T, Kodaira N, Kimura S, Takahashi Y, Nakayama T. Data resource profile: JMDC claims databases sourced from Medical Institutions. *J Gen Fam Med* 2020;21:211–8.
- Kimura S, Sato T, Ikeda S, Noda M, Nakayama T. Development of a database of health insurance claims: standardization of disease classifications and anonymous record linkage. *J Epidemiol* 2010;20:413–9.
- Laurent T, Simeone J, Kuwatsuru R, Hirano T, Graham S, Wakabayashi R, et al. Context and Considerations for Use of Two Japanese Real-World Databases in Japan: Medical Data Vision and Japanese Medical Data Center. *Drugs - Real World Outcomes* 2022;2:175–87.
- Yasunaga H, Horiguchi H, Kuwabara K, Matsuda S, Fushimi K, Hashimoto H, et al. Outcomes after laparoscopic or open distal gastrectomy for early-stage gastric cancer: a propensity-matched analysis. *Ann Surg* 2013;257:640–6.
- Yamana H, Moriwaki M, Horiguchi H, Kodan M, Fushimi K, Yasunaga H. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. *J Epidemiol* 2017;27:476–82.
- Ministry of Health, Labour and Welfare. The tenth expert committee of providing anonymous medical information [in Japanese]. Available from: https://www.mhlw.go.jp/stf/index_0002-0.html, Accessed 2022 Aug 4.
- Japanese Specialty Board. For the senior residents [in Japanese]. Available from: <https://jmsb.or.jp/senkoi/#an02>, Accessed 2022 Aug 4.
- Hashimoto Y, Michihata N, Yamana H, Shigemi D, Morita K, Matsui H, et al. Ophthalmic Corticosteroids in Pregnant Women with Allergic Conjunctivitis and Adverse Neonatal Outcomes: Propensity Score Analyses. *Am J Ophthalmol* 2020;220:91–101.
- Mues KE, Liede A, Liu J, Wetmore JB, Zaha R, Bradbury BD, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled

- populations in the US. *Clin Epidemiol* 2017; 9:267–77.
22. Chen YC, Wu JC, Haschler I, Majeed A, Chen TJ, Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. *PLoS One* 2011;6:e21404.
23. Hoffmann F. Review on use of German health insurance medication claims data for epidemiological research. *Pharmacoepidemiol Drug Saf* 2009;18:349–56.
24. Martin-Latry K, Bégaud B. Pharmacoepidemiological research using French reimbursement databases: yes we can! *Pharmacoepidemiol Drug Saf* 2010;19:256–65.
25. Tricco AC, Pham B, Rawson NS. Manitoba and Saskatchewan. Administrative health care utilization databases are used differently to answer epidemiologic research questions. *J Clin Epidemiol* 2008;61:192–7.
26. 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.
27. Milea D, Azmi S, Reginald P, Verpillat P, Francois C. A Review of Accessibility of Administrative Healthcare Databases in the Asia-Pacific Region. *J Mark Access Heal Policy* 2015;3:28076.
28. Shigemi D, Morishima T, Yamana H, Yasunaga H, Miyashiro I. Validity of initial cancer diagnoses in the Diagnosis Procedure Combination data in Japan. *Cancer Epidemiol* 2021;74:102016.
29. Ono S, Ishimaru M, Ida Y, Yamana H, Ono Y, Hoshi K, Yasunaga H. Validity of diagnoses and procedures in Japanese dental claims data. *BMC Health Serv Res* 2021; 21:1116.
30. Konishi T, Yoshimoto T, Fujiogi M, Yamana H, Tanabe M, Seto Y, Yasunaga H. Validity of operative information in Japanese administrative data: a chart review-based analysis of 1221 cases at a single institution. *Surg Today* 2022.
31. Ooba N, Setoguchi S, Ando T, Sato T, Yamaguchi T, Mochizuki M, et al. Claims-based Definition of Death in Japanese Claims Database: Validity and Implications. *PLoS One* 2013;8:1–7.