A Review of Research Studies Using Data from the Administrative Claims Databases in Japan

Jun Fujinaga1 · Toshio Fukuoka1

Accepted: 28 August 2022 / Published online: 15 September 2022
© The Author(s) 2022

Abstract
Background In Japan, several research studies have used administrative claims databases. However, no study has compared the studies conducted in each database.

Objective We assessed the type, design, and research area of each research study using the administrative databases in Japan to illustrate the suitability of the database used for the types of research studies.

Methods We reviewed studies from four administrative claims databases (the Japanese National Database of Health Insurance Claims and Specific Health Checkups [NDB]; the DPC Study Group database [DPC]; Japan Medical Data Center [JMDC]; and Medical Data Vision [MDV]). The PubMed database was searched from January 2015 to October 2020. We assessed the type, design, and research area of the abstract or full text of each research study.

Results Overall, 643 studies were included. The number of studies increased from 59 in 2015 to 171 in 2020. Descriptive studies accounted for 62.7% (42 studies) of the studies from the NDB, while the treatment effectiveness studies accounted for 81.7% (211 studies) of the studies from the DPC database and were the most common in the JMDC and MDV. Cohort studies accounted for only 17.9% of the studies on NDB but 45.5% on DPC. The most common research area was medicine, general, and internal medicine (53 studies, 8.4%).

Conclusions The type and design of the studies conducted differed in each database and were influenced by the characteristics of the database used. In the future, it may be necessary to integrate various real-world data databases to increase their comprehensiveness.

Key Points
Numerous real-world databases are available in Japan. We reviewed 643 studies conducted in administrative claims databases in Japan.

Studies using databases have shown a rapid increase, which may reflect changes in laws related to database research.

The type and design of the studies were influenced by the characteristics of the database used.

1 Introduction
The development of large-scale medical databases using real-world data (RWD) in Japan has led to the development of real-world evidence, and numerous studies have been published [1]. The US Food and Drug Administration defines RWD as patient health status and/or the delivery of healthcare data collected routinely from electronic health records, claims and billing activities, products, and disease registries, excluding the context of conventional randomized controlled trials [2]. The Japanese Society for Pharmacoepidemiology’s “Task Force on Pharmacoepidemiology and Databases” surveyed the characteristics of medical databases in Japan that can be applied to clinical epidemiology and pharmacoepidemiology [3]. The large medical databases that are listed here include registries such as the National Clinical Database, administrative claims databases such as the Diagnostic Procedure Combination (DPC) Study Group database, Japan National Database of Health Insurance Claims and Specific Checkup (NDB), Japan Medical Data Center (JMDC) claims database, and
Medical Data Vision (MDV) database. The registries are created for specific purposes; for example, conducting research, documenting rare diseases, monitoring provider metrics, and improving the quality of care [4]. However, the administrative claims databases contain medical information collected on a routine basis for reimbursement purposes [4]. Although these data are not intended for research, the administrative claims databases are a valuable source of information for outcomes research and have been used increasingly in clinical research [5].

However, researchers need to be aware of certain points when using administrative claims data. First, there are inherent limitations, including incompleteness and variation in the coding accuracy. It has been noted that even when the same codes are used, the accuracy varies for different populations in different settings, which may affect the validity and generalizability of the results [5]. Researchers also should select a database that is appropriate for the purpose of their studies. For case-control and cohort studies, insurer-based databases are suitable because they track the participants, while hospital-based data are suitable for studies that require the outcomes observed during hospitalization. A propensity score analysis, which is used to adjust for patient backgrounds and control for indication bias, requires information on the underlying disease or disease severity of the participant [6]. In addition, the characteristics of the population selected for this study may reduce the generalizability of the study results. Therefore, it is necessary to understand the characteristics of each database to select the most appropriate database for each research study. The characteristics of each database and aspects to consider when conducting research have already been reported [3, 7], but it is not clear whether the research conducted in each database corresponds to the features already pointed out. The only study that has examined the types of studies conducted in the Japanese administrative claims databases was for the NDB [8], and there have been no studies that have compared the types and designs of studies conducted in each database.

2 Objectives

Accordingly, we aimed to assess the research types and designs that have been conducted thus far using large-scale medical databases, especially the administrative claims databases in Japan, and to illustrate the suitability of the database used for the types of research studies that use each one.

3 Methods

This research study was based on a survey conducted as part of the research group project (Grant no. 19IA2024) of the Ministry of Health, Labour and Welfare (MHLW) during the period 1 April, 2021 to 31 March, 2022. We conducted the literature search on 1 August, 2021. We considered the time lag until the studies were listed in PubMed and the databases’ homepages, covering the literature search until 31 December, 2020. We reviewed the studies using the administrative claims databases, focusing particularly on studies conducted using the administrative claims databases in Japan. We included studies in English using any of the following databases: NDB, DPC, JMDC claims, and MDV. The NDB began providing data to third parties in 2013, the latest of the abovementioned databases. Therefore, we included studies published from January 2015.

In Japan, universal health coverage was achieved in 1961, approximately 40 years after social insurance was established in 1922. Subsequently, employee insurance and community health insurance were established, and today there are approximately 3500 insurance policies [9]. The NDB is a database of health insurance claims and specific health check-ups for the preparation, implementation, and evaluation of medical cost optimization plan built by the MHLW (https://www.mhlw.go.jp/english/) based on the “Act on Assurance of Medical Care for Elderly People” [3, 10]. The specific health check-up program was launched in 2008 for adults aged 40–74 years. This is a nationwide health check-up focusing on the metabolic syndrome to prevent lifestyle-related diseases. The health check-up includes a body mass index measurement, blood pressure measurement, blood test, and urine test. The NDB covers almost the entire Japanese population, with almost everyone’s medical care being insured [11]. Furthermore, because the data are collected from insurers, traceability is ensured even if patients visit different institutions. In 2013, the MHLW began providing NDB data to third parties [10]. The DPC Study Group database is provided to researchers by the DPC Research Institute (http://dpceri.or.jp) [10]. The DPC Study Group database is the most widely used DPC data for research. The database has approximately seven million registered patients per year, more than 1000 participating hospitals, with a coverage of more than 50% of all the acute inpatients in Japan [12]. In addition, the MHLW began providing DPC data in 2019. Registries such as the Japan Registry of All Cardiovascular Diseases also use DPC data. The DPC is a diagnostic group classification unique to Japan and is a method of calculating medical costs based on the classifications defined by patients, diagnoses, and procedures. The DPC-based per-diem payment system was introduced in 2003 and is the main reimbursement system for acute inpatient care in Japan. The DPC-based per-diem payment system is utilized by more than 1700 hospitals [10, 13]. The DPC data include not only claims data, but also a more detailed patient background, called the DPC format 1 [12]. The DPC Study Group database is an inpatient database, which
is a remarkable feature compared to other administrative databases in Japan; however, it has a weakness: there is non-traceability between hospitals. The JMDC claims database has been collecting receipt data from health insurance associations since 2005, and nearly half of the data have been linked to specific health check-up data [10]. Researchers are required to pay a fee to the administrator of the database to receive the data. The database consists of 454 hospitals, including those without DPC systems, with a cumulative population of approximately 13 million (as of September 2021) and 520,000 annual admissions [3, 14]. Patients are given unique identifications so that they can be followed longitudinally even if they change hospitals or visit multiple facilities. There is also family identification, making it possible to link pregnant women with their children. The weaknesses include the fact that the data are only from health insurance associations and do not include data from the late-stage elderly care system and that there are no laboratory data. The MDV database has been collecting data since 2008 [10]. The database consists of 449 acute care hospitals with an actual patient population of more than 30 million and 2.1 million inpatients per year [3, 15]. The laboratory data were available for approximately 10% of the patients. Researchers are required to pay a fee to the administrator of the database to receive the data as with the JMDC. However, unlike the JMDC, it is hospital based, making it difficult to trace patients when they are transferred [10]. These databases had the most research publications, especially those in the database list of the Japanese Society for Pharmacoepidemiology. The details of each database are provided in the Electronic Supplementary Material (ESM). The exclusion criteria were as follows: non-English studies, studies that did not use individual-level data such as review articles and editorials, studies not listed in PubMed, and studies not using data from the database.

3.1 Data Collection

Data were collected as follows: for the NDB, PubMed was used to search for “NDB & Japan, OR National Database of Health Insurance Claims and Specific Health Checkups of Japan”, and we also referred to the MHLW website (https://www.mhlw.go.jp/index.html) as in the previous review [8]. Studies using JMDC and MDV were retrieved from the comprehensive list of products on the homepage (https://www.jmdc.co.jp/, https://www.mdv.co.jp/) of each database and the DPC Study Group database was searched on PubMed under “Diagnosis Procedure Combination”. No list of products was available on the administrator’s website. In addition, we identified studies by performing a manual search of previous reviews on the NDB [8] and the references for each study.

3.2 Classification

Two researchers (J.F. and T.F.) reviewed the list of studies independently and classified them according to type of research study, design of research study, and research area. We classified the studies using the administrative claims databases into the following categories for the primary outcomes: descriptive studies, treatment effectiveness studies, and others. We classified the treatment effectiveness studies into two categories that differed in their procedural approach: exploratory treatment effectiveness studies and hypothesis evaluating treatment effectiveness (HETE) studies, which resulted in four categories of studies [16]. A HETE study was defined as a study with an a priori explicit hypothesis. In the absence of an explicit statement of the primary outcome, we used the first result listed as the primary outcome. If two or more groups were compared, but no adjustment was made for any possible confounding factors, the study was classified as a descriptive study. When two investigators (J.F. and T.F.) differed in their assessment of the classification of the primary endpoint, a consensus was reached after discussion. We also categorized the studies according to their designs: descriptive studies, cohort studies, cross-sectional studies, case–control and matched-cohort studies, quasi-experimental designs; regression discontinuity, difference in differences, instrumental variable methods, propensity score analyses, self-controlled methods; case cross-over, self-controlled case series, and others. We also classified the study designs used to examine the primary outcomes. As in the previous review [8], the classification of each research area was based on the journal in which the study was published. We used the Science Citation Index Expanded (SCIE), Emerging Source Citation Index, Social Sciences Citation Index, and Directory of Open Access Journals databases to classify the journals in which the studies were published. The SCIE is a database of the most influential scientific journals in the field, the Social Sciences Citation Index is a database of the most influential social science journals, and the Emerging Source Citation Index is a newly introduced database that includes scientific journals not listed in the SCIE but that meet certain qualifications [17]. The Directory of Open Access Journals indexes open-access peer-reviewed journals [18]. If a journal was indexed in more than one of the above databases, the journal databases were prioritized according to the order of their influence (SCIE, Emerging Source Citation Index, Social Sciences Citation Index). If a journal was categorized across multiple research areas, then the first research area was used. We only assessed the type, design, and research area for each study, therefore the full PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist was not applicable.
3.3 Statistical Analysis

We performed a descriptive analysis for the type and design of the studies according to the databases. Fisher’s exact test was used to compare the characteristics between the databases. All the analyses were performed using the Stata version 16.1 software (Statacorp LLC, College Station, TX, USA). We considered a p value of <0.05 (two-tailed) as statistically significant. A graph of the number of studies published each year was created using Microsoft Excel (Microsoft Corporation, Redmond, WA, USA).

3.4 Ethical Considerations

No ethics approval was required as there was no direct interaction with the research participants.

4 Results

4.1 Study Selection

We retrieved 812 studies (NDB: 53 studies, DPC: 341 studies from PubMed, JMDC, and MDV: 227 studies and 183 studies from their websites, and eight studies on the NDB from the MHLW website). After screening the titles and abstracts, 632 studies (NDB, 40 studies; JMDC, 185 studies; MDV, 147 studies; and DPC, 260 studies) were included in this review. A total of 28 studies were added from a previous review of the NDB [7] and seven from a manual search. Of these, 14 studies were counted as duplicates because they used both the JMDC and MDV. After the review of the full-text articles, we excluded ten studies. In total, 643 studies were included. Figure 1 shows the screening process.

4.2 Number of Publications Per Year

The number of studies published annually is shown in Fig. 2. The number has been increasing annually, notably from 78 to 171 between 2017 and 2020.

4.3 Types of Research Studies in Each Database

The types of studies are listed in Table 1. More than half of the studies were classified as HETE studies. Moreover, 70.8% (n = 182) of the studies using the DPC Study Group database and about 40% or more (n = 71, n = 52) of the studies using the JMDC and MDV databases were also classified as HETE studies. Research studies using these databases, along with exploratory studies, were classified mainly as treatment effect studies. We classified 217 studies (33.8%) as descriptive studies; when limited to studies using the NDB, 62.7% (n = 42) were classified as descriptive studies.

4.4 Design of Research Studies in Each Database

Table 2 shows the designs of the study designs that were conducted. Cohort studies were the most common study design (n = 230, 35.8%) followed by descriptive studies (n = 217, 33.8%), and then by quasi-experimental studies (n = 142, 22.1%).

4.5 Research Areas by Database

The research areas by database are shown in the ESM, with 53 (8.4%) studies in medicine, general medicine, and internal medicine, followed by 44 (6.9%) studies in pharmacology and pharmacy. However, in the NDB, most studies were conducted on infectious diseases (n = 10, 14.9%), whereas in the DPC Study Group database, most of the studies were...
A Review of Research Studies Using Data from the Administrative Claims Databases in Japan

**Fig. 2** Number of publications by year. *DPC* The DPC Study Group database, *JMDC* Japan Medical Data Center, *MDV* Medical Data Vision, *NDB* Japanese National Database of Health Insurance Claims and Specific Health Checkups

**Table 1** Types and number of studies stratified according to the administrative claims databases

|                          | Total  | NDB    | DPC    | JMDC   | MDV    | JMDC and MDVb | p valuec |
|--------------------------|--------|--------|--------|--------|--------|---------------|----------|
|                          | n = 643| n = 67 | n = 257| n = 175| n = 130| n = 14        |          |
| Descriptive studies      | 217 (33.8) | 42 (62.7) | 35 (13.6) | 73 (41.7) | 56 (43.1) | 11 (78.6) |          |
| Treatment effectiveness  |        |        |        |        |        |               |          |
| studies                  |        |        |        |        |        |               |          |
| Exploratory treatment    | 76 (11.8) | 6 (9.0) | 29 (11.3) | 22 (12.6) | 19 (14.6) | 0 (0.0)    |          |
| effectiveness studies    |        |        |        |        |        |               |          |
| Hypothesis evaluating   | 327 (50.9) | 19 (28.4) | 182 (70.8) | 71 (40.6) | 52 (40.0) | 3 (21.4)   |          |
| treatment effectiveness  |        |        |        |        |        |               |          |
| Othera                   | 23 (3.6) | 0 (0.0) | 11 (4.3) | 9 (5.1) | 3 (2.3) | 0 (0.0) < 0.001 |          |

*DPC* The DPC Study Group database, *JMDC* Japan Medical Data Center, *MDV* Medical Data Vision, *NDB* Japanese National Database of Health Insurance Claims and Specific Health Checkups

*Other studies includes machine learning studies, cost-analysis studies (not specified), validation of clinical score, or coding

*Studies using JMDC and MDV together

*p value for the prevalence of the type of research studies conducted in each database

**Table 2** Research study design and number of studies stratified according to the administrative claims databases

| Research design                  | Total  | NDB    | DPC    | JMDC   | MDV    | JMDC and MDVb | p valueb |
|----------------------------------|--------|--------|--------|--------|--------|---------------|----------|
| n (%)                            | n = 643| n = 67 | n = 257| n = 175| n = 130| n = 14        |          |
| Descriptive study                | 217 (33.8) | 42 (62.7) | 35 (13.6) | 73 (41.7) | 56 (43.1) | 11 (78.6) |          |
| Cohort study                     | 230 (35.8) | 12 (17.9) | 118 (45.9) | 57 (32.6) | 41 (31.5) | 2 (14.3) |          |
| Cross-sectional study            | 13 (2.0) | 5 (7.5) | 2 (0.8) | 5 (2.9) | 1 (0.8) | 0 (0.0) |          |
| Case–control study, matched-cohort study | 23 (3.6) | 1 (1.5) | 5 (2.0) | 11 (6.3) | 6 (4.6) | 0 (0.0) |          |
| Quasi-experimental design; regression discontinuity, Difference in differences, IV, propensity score analyses | 142 (22.1) | 5 (7.5) | 96 (37.4) | 19 (10.9) | 22 (16.9) | 0 (0.0) |          |
| Self-controlled methods; case cross-over, self-controlled case series | 6 (0.9) | 1 (1.5) | 0 (0.0) | 4 (2.3) | 1 (0.8) | 0 (0.0) |          |
| Other                            | 12 (1.9) | 1 (1.5) | 1 (0.4) | 6 (3.4) | 3 (2.3) | 1 (7.1) < 0.001 |          |

*DPC* The DPC Study Group database, *IV* instrumental variable methods, *JMDC* Japan Medical Data Center, *MDV* Medical Data Vision, *NDB* Japanese National Database of Health Insurance Claims and Specific Health Checkups

*Studies using JMDC and MDV together

*p value for the prevalence of research study designs conducted in each database
conducted in surgery (n = 25, 9.8%). Pharmacology and pharmacy were the most common research areas (n = 21, 12.1%; n = 15, 11.8%) among the research studies that used the JMDC and MDV databases.

5 Discussion

We investigated the current contexts of the types and designs of research using the administrative claims databases in Japan and elucidated the characteristics of the studies in each database. We also illustrated the suitability of the databases used for each type of research and presented the implications for future research using databases. A review of 643 studies revealed that the number of studies that used the administrative databases increased over the years. This increase has become particularly pronounced since 2017. The treatment effectiveness studies were more common in the DPC Study Group database, and the descriptive studies accounted for more than 60% of studies on the NDB. Descriptive and cohort studies were the most common study designs. Although “Medicine, General & Internal Medicine” was the most common research area, the most common research areas conducted in each database varied.

The significant increase in research since 2017 has been due to the development of databases and legal changes. The amendment of the Personal Information Protection Act in 2017 and the enactment of the Next Generation Medical Infrastructure Law in 2018 have facilitated the use of RWD in an appropriate context.

We considered that the HETE studies were conducted largely in the DPC Study Group database because it is relatively easy to collect information about underlying diseases and severity of illnesses, as well as treatments or procedures using the DPC format 1 [1, 13]. The DPC format 1 includes clinical data scores such as levels of consciousness (Japan Coma Scale), dementia severity, pneumonia severity, and other clinical scores. In addition, these types of information are also required in quasi-experimental designs, such as propensity score matching and inverse probability of treatment weighting, which have been used in many observational studies in recent years [19]. The NDB does not have information on underlying diseases or the severity of illnesses. Consequently, quasi-experimental designs are frequently used in the DPC Study Group database, probably because of the relatively easy acquisition of this information. The disadvantage of the DPC Study Group database is that its population is limited to inpatients, especially those in acute-care hospitals [13]. This drawback not only limits the scope of the study and makes it difficult to trace long-term outcomes but may also lead to a lack of generalizability and selection bias.

The NDB had the highest proportion of descriptive studies when compared with other databases; with more than 60% of the studies using the NDB being descriptive studies, similar to a previous study [8], suggesting the completeness of NDB data and the difficulty in adjusting for the severity of illness and other factors with the use of NDB. The strengths of the NDB are that it covers almost the entire Japanese population with almost all having insured medical care [11], and that because the data are collected from insurers, the traceability is ensured even if patients visit different institutions. However, using the NDB data has several disadvantages. The application system is strict, researchers must follow the procedures set by the MHLW for research using individual data, and the validity of the application must be approved by the expert committee. Data acquisition requires time, and it can take a long time (more than 180 days) from application to the provision of data [20], and the number of days from permission to information provision can range from 79 to 263 days [21]. In contrast, the JMDC can be obtained in 5–30 business days [14], and the MDV in 2 weeks [15]. In addition, the information on mortality is limited, and the structure and volume of the data make it difficult to handle, lacking information to adjust for patient risk and the severity of the illness [9]. Therefore, it may be difficult to conduct studies other than descriptive studies, and the number of studies may be limited compared with other databases.

The JMDC claims database is also an insurer-based database while the MDV is hospital based, thus the types and designs of the studies conducted were expected to be different. However, in both databases, the types of studies were mainly HETE studies, and the designs mainly those of cohort studies. Both the databases use DPC data as well as receipt data, making it possible to adjust for the underlying disease and disease severity as in the DPC Study Group database, thus the two databases may have provided similar results.

Overall, 141 studies of quasi-experimental design were conducted, most of which employed propensity score analyses. Appropriately conducted regression discontinuity or difference-in-differences may have been appropriate methods for evaluating the effect of policy/guideline changes using a large medical database [22–24]. However, regression discontinuity and difference-in-differences were used in only 11 of the studies reviewed in this study. The paucity of quasi-experimental research in healthcare has been addressed previously and may be improved by supporting the development of research capacity in the use of quasi-experimental designs, developing guidelines for implementation and reporting, making health program administrative data and clinical registers available to all publicly funded health systems, and extending validated risk-of-bias measures to include quasi-experimental designs in systematic reviews [25]. In further studies, it is anticipated that regression
discontinuity and difference-in-differences applications will be used in large-scale medical databases to evaluate the effects of policy and guideline changes. Nevertheless, most database study designs were cohort studies, suggesting that research may be conducted using methods other than quasi-experiments and that other barriers, such as access to data and the size of the data, may be challenges [10]. In addition, the prevalence of cohort studies suggests that the availability of longitudinal data may facilitate the conducting of cohort studies.

We also found that various research areas have been implemented using large-scale medical databases. Studies in the field of pharmacology and pharmacy were the second most common (44 studies, 6.9%), followed by medicine, general, and internal medicine (53 studies, 8.4%). The use of large datasets has become increasingly common in pharmacoepidemiology. In Japan, since the 2018 amendment to the Good Post-Marketing Study Practice Ordinance, pharmaceutical companies can use administrative claims databases for post marketing surveillance, which is estimated to have contributed to the rapid increase in pharmacoepidemiologic studies using the administrative claims databases. The growing use of databases in pharmacoepidemiology and the study of the efficacy and safety of drugs using RWD are expected to further increase in importance and proportion [26]. The fields in which there is little research are fields not closely related to clinical medicine, such as microbiology and biology, or areas in which outpatient care is more important, such as primary care and dermatology.

This study had some limitations. We categorized the areas in which the studies were conducted in this review; however, studies often encompassed multiple areas. This may have led to incorrect classifications of the studies. However, we developed an objective classification of the research areas using the classification of journals in literature databases.

Most studies in the areas of “critical care,” “emergency medicine,” and “surgery” were conducted using the DPC Study Group database. Because some researchers have participated in many studies in these areas, the choice of database used in the studies may have reflected the researcher’s preference. However, the DPC Study Group database is suitable for use in research studies in these areas because it provides inpatient information according to the severity of disease, as well as information on surgeries and bedside procedures.

### 6 Conclusions

Our review suggests that the types and designs of research conducted in each database correspond to the characteristics of the database. In databases consisting of DPC data (e.g., the DPC Study Group database, JMDC claims database, and MDV database), many HETE studies were conducted using information related to individuals. Studies using the NDB were mainly descriptive studies, which were thought to be influenced by the difficulties related to the availability of the data as well as the impact of completeness. Many propensity score analyses have been conducted; however, other quasi-experimental designs have not been implemented fully. Several cohort studies have been conducted because of the availability of longitudinal data. In the future, it is necessary to collect and integrate a wide variety of RWD databases to increase comprehensiveness and derive valuable evidence.

### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s40801-022-00331-5.

### Acknowledgements

We thank Dr. Masahiro Yoshida for his assistance with this study.

### Declarations

#### Funding

This work was supported by Health, Labour, and Welfare Policy Research Grants (Grant number 19IA2024).

#### Conflict of interest

The authors have no conflicts of interest that are directly relevant to the content of this article.

#### Ethical approval

No ethics approval was required as there was no direct interaction with the research participants.

#### Consent to participate

No informed consent was required as there was no direct interaction with the research participants.

#### Consent for publication

Not applicable.

#### Availability of data and material

The dataset used or generated during this study is available from the corresponding author upon reasonable request.

#### Code availability

Not applicable.

#### Authors’ contributions

All the authors contributed to the planning and critical review of the manuscript. JF drafted the manuscript. All the named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for the manuscript before submission and for this version to be published. All the authors have read and approved the final version.

#### Open Access

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission.
References

1. Kumamaru H, Fukuma S, Matsui H, Kawasaki R, Tokumasu H, Takahashi A, et al. Principles for the use of large-scale medical databases to generate real-world evidence. Ann Clin Epidemiol. 2020;2:27–32.
2. US Food and Drug Administration. Real-world evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. Accessed 20 Dec 2021.
3. Pharmacoepidemiology & Database Taskforce, Japanese Society for Pharmacoepidemiology. Survey of Japanese databases in Japan available for clinical/pharmacoepidemiology. http://www.jspe.jp/committee/020/0210/. Accessed 20 Dec 2021.
4. Penberthy LT, Rivera DR, Lund JL, Bruno MA, Meyer AM. An overview of real-world data sources for oncology and considerations for research. CA Cancer J Clin. 2022;72:287–300.
5. Nehra MS, Ma Y, Clark C, Amarasingham R, Rockey DC, Singal AG. Use of administrative claims data for identifying patients with cirrhosis. J Clin Gastroenterol. 2013;47:e50–4.
6. Imai S. An introduction to real-world data and tips for analysing it [in Japanese]. Yakugaku Zasshi. 2021;141:169–74.
7. Laurent T, Simeone J, Kuwatsuru R, Hirano T, Grahamb S, Wakahayashi 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;9:175–87.
8. Hirose N, Ishimaru M, Morita K, Yasunaga H. A review of studies using the Japanese National database of health insurance claims and specific health checkups. Ann Clin Epidemiol. 2020;2:13–26.
9. Ikemami N, Yoo BK, Hashimoto H, Ogata H, Babazono A, Watanabe R, et al. Japanese universal health coverage: evolution, achievements, and challenges. Lancet. 2011;378:1106–15.
10. Hiramatsu K, Barrett A, Miyata Y, PhRMA Japan Medical Affairs Committee Working Group 1. Current status, challenges, and future perspectives of real-world data and real-world evidence in Japan. Drugs Real World Outcomes. 2021;8:459–80.
11. 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.
12. Yasunaga H. Real world data in Japan: Chapter II. The diagnosis procedure combination database. Ann Clin Epidemiol. 2019;1:76–9.
13. Hayashida K, Murakami G, Matsuda S, Fushimi K. History and profile of diagnosis procedure combination (DPC): development of a real data collection system for acute inpatient care in Japan. J Epidemiol. 2021;31:1–11.
14. JMDC Claims Database. https://www.jmdc.co.jp/en/jmdc-claims-database/. Accessed 20 Dec 2021.
15. Medical Data Vision. About MDV database. https://www.mdv.co.jp/about-mdv-database/. Accessed 20 Dec 2021.
16. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goetttsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26:1033–9.
17. Web of Science. Search journals. https://mjl.clarivate.com/home. Accessed 12 Feb 2022.
18. Directory of Open Access Journals. https://doaj.org/. Accessed 12 Feb 2022.
19. Bärnighausen T, Tugwell P, Røttingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series-paper 4: uses and value. J Clin Epidemiol. 2017;89:21–9.
20. Ministry of Health, Labour and Welfare. Homepage for anonymized receipt information and anonymized information on specific health checkups [in Japanese]. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryouken/reseputo/index.html. Accessed 12 Feb 2022.
21. Ministry of Health, Labour and Welfare, Health and Welfare Bureau for the Elderly, Health Insurance Bureau. The role of NDB, nursing care and other DBs and their analytical platforms [in Japanese]. https://www.mhlw.go.jp/content/12401000/000350567.pdf. Accessed 12 Feb 2022.
22. Nishioka D, Takaku R, Kondo N. Medical expenditure after marginal cut of cash benefit among public assistance recipients in Japan: natural experimental evidence. J Epidemiol Community Health. 2021;76(5):505–11.
23. Zogg CK, Scott JW, Metcalfe D, Seshadri AJ, Tsai TC, Davis WA, et al. The association between Medicare eligibility and gains in access to rehabilitative care: a national regression discontinuity assessment of patients ages 64 versus 65 years. Ann Surg. 2017;265:734–42.
24. Magnani CJ, Li K, Seto T, McDonald KM, Blayney DW, Brooks JD, et al. PSA testing use and prostate cancer diagnostic stage after the 2012 US Preventive Services Task Force Guidelines changes. J Natl Compr Canc Netw. 2019;17:795–803.
25. Tugwell P, Knotternus JA, McGowan J, Tricco A. Big-5 quasi-experimental designs. J Clin Epidemiol. 2017;89:1–3.
26. Takahashi Y, Nishida Y, Asai S. Utilization of health care databases for pharmacoepidemiology. Eur J Clin Pharmacol. 2012;68:123–9.