Drugs - Real World Outcomes (2022) 9:175–187
https://doi.org/10.1007/s40801-022-00296-5

REVIEW ARTICLE

Context and Considerations for Use of Two Japanese Real-World Databases in Japan: Medical Data Vision and Japanese Medical Data Center

Thomas Laurent[1] · Jason Simeone[2] · Ryohei Kuwatsuru[3,4] · Takahiro Hirano[1,4] · Sophie Graham[5] · Ryozo Wakabayashi[1,4] · Robert Phillips[1] · Tatsuya Isomura[1,4]

Accepted: 22 February 2022 / Published online: 18 March 2022

Abstract

In Japan, an increasing interest in real-world evidence for hypothesis generation and decision-making has emerged in order to overcome limitations and restrictions of clinical trials. We sought to characterize the context and concrete considerations of when to use Medical Data Vision (MDV) and JMDC databases, the main Japanese real-world data (RWD) sources accessible by pharmaceutical companies. Use cases for these databases, and related issues and considerations, were identified and summarized based on a literature search and experience-based knowledge. Studies conducted using MDV or JMDC were mostly descriptive in nature, or explored potential risk factors by evaluating associations with a target outcome. Considerations such as variable ascertainment at different time points, including issues relating to treatment identification and missing data, were highlighted for these two databases. Although several issues were commonly shared (e.g., only month of event occurrence reported), some database-specific issues were also identified and need to be accounted for. In conclusion, MDV and JMDC present limitations that are relatively typical of RWD sources, though some of them are unique to Japan, such as the identification of event occurrence and the inability to track patients visiting different healthcare settings. Addressing study design and careful result interpretation with respect to the specificities and uniqueness of the Japanese healthcare system is of particular importance. This aspect is especially relevant with respect to the growing global interest of conducting RWD studies in Japan.

Key Points

Medical Data Vision (MDV) and JMDC are the most frequently used real-world data sources in Japan.

MDV and JMDC share common limitations with real-world data sources in other countries, though some of them are unique to Japan, including the identification of event occurrence and the inability to follow-up patients visiting different healthcare settings.

Using Japanese real-world data sources requires understanding of the uniqueness of the Japanese healthcare system.

✉ Thomas Laurent
thomas_laurent@jp-css.com

1 Clinical Study Support Inc., 2F Daiei Bldg., 1-11-20 Nishiki, Naka-ku, Nagoya 460-0003, Japan

2 Fifth Floor, Real-World Evidence, Evidera, 500 Totten Pond Road, Waltham, MA 02451, USA

3 Department of Radiology, School of Medicine, Juntendo University, Hongo 2-1-1, Bunkyo-ku, Tokyo 113-8421, Japan

4 Real-World Evidence And Data Assessment (READS), Graduate School of Medicine, Juntendo University, Hongo 2-1-1, Bunkyo-ku, Tokyo 113-8421, Japan

5 Real-World Evidence, Evidera, The Ark, 2nd Floor, 201 Talgarth Road, London W6 8BJ, UK

1 Introduction

In 1961, Japan established a universal medical insurance system that encompasses five public insurance programs. Patients can access clinics and hospitals without a family doctor's gatekeeper system or insurance-based limitations. Patients frequently visit secondary healthcare facilities with minor symptoms as outpatients, and these facilities can be accessed without a referral from primary healthcare facilities, at an affordable cost. Health insurance typically covers 70–90% of the cost of care, and the rest is paid by the insured. A fee-for-service system is predominantly used for charging medical fees. However, a lump-sum payment system, known as the Diagnosis Procedure Combination (DPC)/Per Diem Payment System (PDPS), was partially introduced for inpatient care. This payment system relies on grouping patients according to DPC categories. Each medical institution submits claims for reimbursement to the Examination and Payment Agency monthly [1].

Since April 2019, Japan has started implementing health technology assessments to address the rising costs of healthcare expenditure by targeting reimbursed drugs or devices that are anticipated to have a substantial impact on healthcare expenditures. Conversely, due to scientific misconduct, regulations on clinical trials have been strengthened after the implementation of the new *Clinical Trials Act* in 2018 [2]. Other barriers to clinical trials can be mentioned, including the high research and development costs associated with bringing a drug to market, which limit the number of new drug approvals in the Japanese market. Also, the Japanese population has a high proportion of elderly individuals with a high comorbidity burden who are generally not included in clinical trials.

In order to regulate national healthcare expenditures due to ageing, the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) was created in 2008. This database covers about 98% of data on healthcare services and is being used for academic research, but is not accessible to private companies [3]. To generate evidence for use in healthcare decision-making and reduce unmet needs in certain therapeutic areas, Japan aims to leverage real-world data (RWD) for various applications, including informing clinicians on practice patterns, and hypothesis generation for future research. As a consequence, an ongoing increase in the number of studies using RWD is being observed, including in the private sector. The overall landscape of Japanese RWD has been described elsewhere, covering the nature of the databases, their general content, and data governance and access [4]. JMDC, formerly known as Japan Medical Data Center Co., Ltd., and Medical Data Vision (MDV) are the main databases that are accessible to private companies and that do not require any institutional review board (IRB) submission for data access. Previous works have partially covered these databases, but do not provide an overall overview on the practical considerations and the different current applications when using these databases [5–10]. Hiramatsu et al. provided general information on the different types of databases in Japan, including registry, insurance-based, and hospital databases [5]. Lai et al. and Nagai et al. provided detailed information on the variables included in JMDC, although practical considerations were not addressed [6, 10]. Lee et al. discussed considerations on data collection, transformation, cleansing, and modeling when using MDV [9]. Finally, examples of drug safety and drug utilization studies using either MDV or JMDC have been described in the literature [7, 8]. In this work, we sought to further characterize the concrete advantages and current and future challenges relating to MDV and JMDC databases, particularly with respect to the Japanese healthcare context, given that they are the RWD sources predominantly accessed by pharmaceutical companies.

2 Overview of Japanese RWD Sources

As is typical with other secondary data sources, RWD data users should bear in mind that data are originally collected for different purposes other than for research. Here, we provide general information on MDV and JMDC data.

2.1 MDV

MDV is a health claims database collecting data from all insurance types containing anonymized outpatient, inpatient, and DPC data and blood test results for a limited number of institutions. MDV data were first made available in 2008, and there is around a 3-month time lag from data collection to data availability. As of September 2021, MDV accumulated data from more than 38 million patients and 23% of acute care hospitals in Japan. The main advantages of the MDV database are that it includes a representative population, has a large sample size, includes medical costs reimbursed by healthcare insurance directly in the data, and includes laboratory test results. Condition-specific hospitalization can be identified using DPC codes, which are attributed for each hospitalization in DPC data, and allows determination of the major diagnosis associated with a hospitalization. Details on the information available in DPC data, including available clinical data, have been provided by Yasunaga et al. and Lee et al. [9, 11]. In contrast, the absence of information on several clinical outcomes (e.g., death in outpatient settings, disease severity) and the absence of data linkage between prescription and diagnosis are inherent limitations

[12]. Another limitation is also the impossibility to follow patients outside DPC settings. Although some patient scales (e.g., activities of daily living) or disease staging data might be available, many missing values may be observed for measures that are deemed unnecessary for reimbursement purposes. For example, blood parameter measurements (e.g., glycated hemoglobin [HbA1c] for diabetes and estimated glomerular filtration rate for renal failure) are only available for a small proportion of institutions, and hence, results may not be generalizable to the full study sample. The number of institutions providing laboratory data may change over time since the collection of this information is governed by a contract independent of other data that may be terminated from one year to the next. Also, dosage information is not available in structured data fields, and would require extensive work for data extraction. To identify drug treatment initiation, a lookback window is required considering that MDV only includes DPC hospitals. Only month of event occurrences are provided, and therefore, it is not possible to identify populations or derive algorithms using a daily time component. Moreover, visits to other institutions outside or within the data network cannot be linked given that a unique identifier is allocated to each institution that does not allow data linkage across institutions. Finally, the number of institutions is constantly growing, and reproducing a previous study with the same condition may lead to differences.

2.2 JMDC

JMDC is a claims database that contains anonymized inpatient, outpatient, and dispensing receipts, and also medical examination data, collected from various health insurance associations [3, 7, 10]. As of September 2021, the total number of patients accumulated 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 associated costs for each visit, and there is around a 5-month time lag from data collection to data availability. This database presents several advantages, especially the availability of data on medical costs reimbursed by healthcare insurance directly in the data, the possibility of follow-up at different hospitals, and a large sample size. In general, Japanese employees tend to stay with the same employer for many years under the same insurance policy, and thus make this type of RWD more robust when compared to other countries. However, this database presents inherent limitations, including the under-representation of the elderly, since JMDC is sourced from health insurance associations for company employees and their dependents. This database shares common limitations with MDV; in particular, a low level of information on clinical outcomes (e.g., about 80% of deaths occurring outside hospital and clinical outcomes recorded using International Classification of Disease 10 codes) and absence of linkage between prescriptions and diagnoses, and only month of event occurrence is available. Although the date of medical care initiation at a medical facility is available, if an individual goes to another center for the same diagnosis, a different date of medical initiation would be reported. Although not discussed here, JMDC has created a hospital-based database collecting data from more than 460 medical institutions in order to account for the limitations of the health insurance-based JMDC database, by including DPC data and blood parameter measurement data and obtaining data on the elderly population. Further detailed information on the data provided in JMDC was reported by Nagai et al. [10].

3 Use Cases for Studies Based on MDV or JMDC Databases

We conducted a literature search using (Japanese AND "administrative claims data") or "Medical data vision" or JMDC within PubMed, and included articles published by March 2021. Overall, we found 68 and 105 articles for MDV and JMDC, respectively. We concluded that a broad scope of research questions has been addressed by leveraging the MDV and JMDC databases. A non-exhaustive list of examples of studies conducted using MDV or JMDC databases is provided in Table 1, covering a large set of research questions and therapeutic areas. Many studies focused on either product utilization patterns or descriptions of patient characteristics [17, 18]. Interestingly, two studies focused on describing drug prescriptions among hyperlipidemic patients using both databases. The authors concluded that lipid-lowering drug prescription trends were similar across databases, though some differences were observed, mainly explained by a higher proportion of elderly patients and a higher degree of poly-comorbidities among the patient population in the MDV database [23, 24]. In some cases, associations between intervention or risk factors and outcomes have been explored to generate hypotheses, but studies based on a clearly defined causal framework or focusing on a causal effect were not conducted using these databases. Although descriptive studies and studies exploring potential risk factors based on associations can be informative to generate new hypotheses, a causal inference framework is needed for multiple purposes, including replication of randomized clinical trials in the real world, comparative effectiveness and comparative safety, support of regulatory and reimbursement decisions, and clinical adoption decisions. Hypothesis generation studies are likely to continue to play an important role in RWD, because this type of evidence is much less demanding in terms of resources considering that knowledge on the data generation process is needed in causal frameworks, as

Table 1. Example of real-world evidence generation studies leveraging MDV or JMDC databases

Research question	References	Database	Method	Strength/impact	Limitations
Disease epidemiology/pharmacoepidemiology	Kobayashi 2020 [13]	MDV	Population: Inflammatory bowel disease patients Exposure: Thiopurine/anti-tumor necrosis factor-α Outcome: Incidence of malignancies	Impact No increase of the incidence of non-Hodgkin lymphoma associated with thiopurine or anti-tumor necrosis factor-α treatment in Japanese patients with inflammatory bowel disease Strength Conduction of sensitivity analysis by setting a stricter definition for the identification of the exposure	Lack of information on important confounders (disease duration and severity) Very few outcome events were observed
	Chang 2018 [14]	JMDC	Population: All patients with sufficient enrollment Exposure: None Outcome: MA-AGE, MA-NGE	Impact The study confirmed that norovirus is an important cause of MA-AGE in Japan, not only in children, but also in other age groups Strength Since only a small proportion of episodes are cause specified, the proportion attributable to norovirus was estimated using an indirect modeling method	The number of elderly individuals included in the analysis is relatively small, especially for women No information on 75 years and older Restrictive assumption of the residuals modeling method since all cause unspecified MA-AGE episodes in the database were considered due to norovirus

Table 1. (continued)

Research question	References	Database	Method	Strength/impact	Limitations
Healthcare resource use/economic burden	Yamazaki 2019 [15]	MDV	Population: Chronic hepatitis C patients Exposure: Antiviral treatment Outcome: Medical cost, healthcare resource use (including inpatient and outpatient visits)	Impact This study generated evidence that delaying antiviral treatment initiation for Japanese patients with chronic hepatitis C may increase the clinical and economic burden associated Strength Use of an algorithm to identify delayed treatment and early treatment cohort based on cirrhosis diagnosis record	Validation of the algorithm was performed with another claims database, but not in MDV Sustained virologic response status was not observed and thus could not be used for subpopulation analyses
	Inoue 2019 [16]	JMDC	Population: Asthma patients Exposure: Asthma severity (Japanese Society of Allergology guidelines) Outcome: Healthcare resource utilization (including direct costs)	Impact The results highlight a potential deviation from consensus care guidelines, and present an opportunity for further examining real-world clinical practices Strength Healthcare resource use data were generated in a large Japanese population, severe and non-severe asthma cohorts were characterized, and confirmation of age and comorbidities as relevant variables for asthma outcomes	Drug treatment was used as a proxy for disease severity since the information on disease severity was not available Smoking behavior and body mass index were not included in the analysis.

Table 1. (continued)

Research question	References	Database	Method	Strength/impact	Limitations
Product utilization pattern	Tanabe 2017 [17]	MDV	Population: Type 2 diabetes patients Exposure: Alpha-glucosidase inhibitor, biguanide, dipeptidyl peptidase-4 inhibitor, thiazolidinedione vs sulfonylurea Outcome: Treatment choice and adherence	Impact Cardiovascular disease history was not associated with treatment choice Strength Large population of diabetic patients allowed selection of patients with HbA1c data and capture of the elderly population	Analyses were restricted to patients with HbA1c, and these patients may present different characteristics when compared to those without HbA1c measurement Specialty of physicians administering treatment and some clinical information that may affect treatment selection were not available
	Matsuoka 2021 [18]	JMDC	Population: Patients diagnosed with ulcerative colitis Exposure: Corticosteroids, thiopurine, biologics Outcome: Prescription rate	Impact The study showed corticosteroid use became more appropriate as use of thiopurine and biologics increased, although there were still many cases of inappropriate use Strength This study captured the changes in corticosteroid use from 2006 to 2016 and the difference in characteristics of patients with long-term and non-long-term corticosteroid use, and identified factors associated with long-term corticosteroid use in ulcerative colitis patients	The population of elderly patients was limited Due to the unavailability of the information, the study did not address the differences in disease severity, which probably influenced corticosteroid use The reason for drug prescription could not be determined when more than one disease code was recorded

Table 1. (continued)

Research question	References	Database	Method	Strength/impact	Limitations
Characteristics of the patient population	Fuji 2017 [19]	MDV	Population: Patients with orthopedic surgeries Exposure: Multiple risk factors (including type of surgery, gender, history of venous thromboembolism, thrombophilia, age) Outcome: Pulmonary thromboembolism and deep venous thrombosis	Impact The incidence of thromboembolism and deep venous thrombosis, and the risk factors for thromboembolism and deep venous thrombosis were comparable to data obtained in previous studies Strength Validation of the outcome definitions was performed based on the clinical laboratory data from a sample of medical records and a high positive predictive value could be obtained	Inherent broader definitions of orthopedic surgeries in the database Lack of information to identify the cause of death No information on bleeding event date that was identified based on date of diagnostic imaging or examinations
	Yamada-Harada 2019 [20]	JMDC	Population: Diabetic patients Exposure: Number of risk factors (blood pressure, LDL-C, HbA1c, current smoking) Outcome: Coronary artery disease	Impact The study suggested that composite control of modifiable risk factors is important in patients with and without diabetes Strength Large sample size and long follow-up (at least 3 years). Use of health examination data (not only hospital data)	Issue of population sampling considering that only patients who had undergone physical examinations with blood tests were included Presence of unmeasured confounders such as undetected comorbidities, severity, duration of diabetes, and socio-economic status Lack of information on the symptomatic nature of coronary artery disease Target blood parameter values were assessed based on a single measurement

Table 1. (continued)

Research question	References	Database	Method	Strength/impact	Limitations
Comparative effectiveness/safety	Kohsaka 2020 [21]	MDV	Population: Patients with non valvular atrial fibrillation Exposure: apixaban, dabigatran, edoxaban and rivaroxaban versus warfarin Outcome: Risks of stroke and systemic embolism	Impact This study suggested that exposures were associated with a significantly lower risk of major bleeding and stroke/systemic embolism compared with warfarin Strength Large population and various treatment groups identified. Balancing among treatment group using stabilized inverse probability of treatment weighting was performed. Sensitivity analyses on time horizon and an alternative balancing method were used to assess the robustness of the results	Patients may present poorer health compared to the average population Due to the lack of information on the follow-up loss, the incidence of stroke may have been underestimated
	Hashimoto 2020 [22]	JMDC	Population: Pregnant women with allergic conjunctivitis Exposure: Topical ophthalmic corticosteroids Outcome: Congenital anomalies, preterm birth, low birth weight, and the composite of these three outcomes	Impact The study showed the use of topical ophthalmic corticosteroids in pregnant women with allergic conjunctivitis was not associated with congenital anomalies, preterm birth, or low birth weight Strength Availability of family identifiers to link newborn data to mother data	No possibility to confirm whether ophthalmic corticosteroids were actually used Daily frequency, duration of treatment, and dosage could not be accounted for

All studies listed adopted a cohort design

HbA1c glycated hemoglobin, *LDL-C* low-density lipoprotein cholesterol, *MA-AGE* medically attended acute gastroenteritis, *MA-NGE* medically attended norovirus-attributable gastroenteritis, *MDV* Medical Data Vision

reported by Ho et al. [25]. Moreover, hypothesis generation studies would allow the identification of potential variables to be included in a causal modeling framework as a preliminary step before conducting a causal inference study using RWDs [26]. Hence, we believe that there will be an important interplay between studies for hypothesis generation and causal inference studies.

4 Challenging and Practical Considerations when Using MDV or JMDC

The development of appropriate research questions requires researchers to take into account several aspects of these data sources to ensure that the question can be addressed when using MDV or JMDC. Several issues relating to various domains of the study should be considered when using MDV or JMDC (Table 2). In particular, those particularly relating to the specificity of the Japanese healthcare system should be carefully considered.

The nature of institutions from which data were collected is an important aspect to consider. For example, large hospitals and DPC-designated hospitals over-represent patients with severe conditions. As mentioned above, the MDV database is based on DPC data, a payment system unique to Japan, and knowledge about the patient that may regularly visit these institutions is of importance when designing a study using this database. In general, patients in MDV present more severe conditions when compared to JMDC, and thus, understanding the nature of Japanese medical institutions is crucial to clarify data coverage, and carefully considering the target population in a study before selecting a data source is advisable. Moreover, in Japan, the concept of primary care is absent, and patients can access any institution and visit doctors when they have minor ailments, without any restriction. From this perspective, the individual patient journey may not be captured entirely; also, it is for the researcher to identify whether the data for a particular therapeutic area are comprehensive enough, on a case-by-case basis. We would advise checking whether a variable is available for a specific target population before the design stage.

Although not specific to a Japanese context, few Japanese studies have implemented sequence symmetry analysis (SSA) that is underpinned by a case-only design, similarly to the case crossover design, accounting for non-time-varying confounders; this approach was shown to enable the observation of drug-related adverse event signals [27]. Considering that MDV generally does not present sufficient lookback data, we would recommend using the JMDC database for this application (Table 2). In addition, for studies requiring long-term outcome assessment, such as long-term effectiveness or safety, patient tracking may be difficult. In JMDC, it is possible to track patients across multiple institutions as patients retain the same identifier if they maintain the same insurance policy. Hence, if continuous follow-up of patients is required, the use of insurance-based claims would be recommended. The study conducted by Yamada-Harada et al. [20] in Table 1, based on JMDC, is an example of when this type of data can be more useful. Finally, for a long study period, medical practices may change overtime and may affect the relationship between the medical treatment and the outcome, for example, by redefining the recommendations for the population that may be prescribed the medical treatment. We would recommend identifying the absence of any major change in patient management guidelines, for example, by checking specific disease guidelines.

Importantly, there are different aspects to take into account at baseline for correct identification of the study sample, and for mitigating confounding (e.g., disease severity) (Table 2). Some studies implemented algorithms to handle this latter issue [28, 29]. Further, the linkage between treatment and diagnosis and the identification of treatment initiation requires assumptions, as encountered in other RWDs (Table 2). Time-related bias has been reported to be an important issue, potentially more important than randomization itself, as reported by Hernán et al. [30], and the absence of guidance for defining time zero can make the assessment inconsistent across different studies. Finally, JMDC and MDV may contain missing data for specific clinical outcomes, and we would recommend checking observed patient characteristics among patients with and without data to clarify whether using complete cases only may introduce selection bias.

To identify pharmacological treatments, MDV and JMDC databases provide information on prescription date, the medical department, the dose and the number of days of supply, cost information for claims reimbursement, and date of hospitalization. Although patients in MDV cannot be tracked outside the network of institutions included in the database, Nishimura et al. conducted a study in type 2 diabetes patients and found that the proportion of days covered for fixed-dose combinations and two-pill combinations was similar between MDV and JMDC [31]. Thus, we can suppose that the lack of data capture for other institutions in MDV may not be an issue at least in some specific applications. Overall, we would recommend collaborating with local researchers who are experienced using these data.

5 Future Challenges for RWD in Japan

Japanese real-world research studies are subject to limitations that are relatively typical of RWD sources as well as those that are Japanese specific. At present, few validation studies of diagnosis in claims databases have been conducted

Table 2. Practical considerations when using MDV or JMDC database

Domain	Issue	Considerations
Study design	Study population	Patients in MDV typically present with more severe conditions when compared to JMDC Under-representation of the elderly in JMDC, since the data are sourced from health insurance associations for company employees and their dependents
	Sequence symmetry analysis: Period settings	Testing different lengths of the time windows for run-in to ensure that incident events Testing different lengths of the time period to ensure to capture enough signal Compared to MDV, JMDC may be more suitable for this purpose since patients are followed up across different institutions
	Nested case control: Case definition	Difficult to have a precise assessment of event occurrence based only on diagnosis record (year–month data only available)
	Nested case control: Control definition	Presence of other conditions that may bias the results due to the nature of the institutions (e.g., acute care settings for MDV)
Pre-intervention	Selection of baseline characteristics	Measuring impact of the variations of lookback period from the baseline date for identification of baseline characteristics
	Confounding due to disease severity	Information on disease severity is limited generally (e.g., disease staging, patient scales) Use of proxy measurement for severity using treatment, procedure, or diagnosis records, but requires validation Misclassification when using treatments with multiple indications
	Confounding due to comorbidities	Calculation of Charlson Comorbidity Index using the ICD-10 codes
	Presence of confounding due to medical institution	Opting for an analytical framework for clustered data based on medical institution identifier
At-intervention	Identification of treatment initiation	Use of a lookback period to ensure that the treatment was initiated (e.g., 6 months or 1 year) For MDV, the first available record is considered as the reference for identifying the lookback period, but based on the assumption that treatment/procedure is prescribed at a single institution Database entry date is available, and treatment/procedure records at multiple institutions are available, though information on date for data collected before 2012 is limited (JMDC database)
	Identification of diagnosis	Combination of diagnosis records for treatment of symptoms and/or diagnosis test in a real-world practice based on clinical guidelines and expert opinions Assessing the validity of the definition to enhance evidence value
	Reason for drug prescription	Possibility to infer potential diagnosis for treatment with restricted indications, but requires knowledge on clinical guidelines and package insert for the corresponding products

Table 2. (continued)

Domain	Issue	Considerations
During follow-up	Availability of clinical outcome data including blood parameter measurements	Data available for a restricted set of institutions (MDV database) Comparing characteristics between patients with and without data to detect any potential issue of generalizability DPC-designated hospitals may lack information that is not relevant for reimbursement purposes, including clinical information, such as clinical scales, even if the variable exists in the database (MDV database) Clarifying whether the target population would present the information for the corresponding variable
	Identification of death	Only inpatient death in DPC-designated hospitals is available (MDV database) Death occurring outside of the institution missing for around 20% of patients because the payer needs to inform JMDC to obtain the information (JMDC database)
	Identification of adverse event occurrence date	Only year–month of diagnosis record available Combination with prescription and/or procedure records to identify the date of the event, but may be restricted to more severe conditions
	Identification of diagnosis-related hospitalization costs	Absence of direct data linkage between medical costs and diagnoses Considering receipts associated with hospitalization with the target diagnosis classified as the most resource consuming (for hospitalizations under diagnosis procedure combination system only)
	Hospital visits	Due to monthly data aggregation, multiple visits within the same month are not identifiable for MDV and JMDC databases Visits to other institutions cannot be linked in MDV database Visits in different institutions are identifiable in JMDC database

DPC Diagnosis Procedure Combination, *ICD* International Statistical Classification of Diseases and Related Health Problems, *MDV* Medical Data Vision

in Japan [32], and key elements to consider when planning a validation study are under discussion [33]. MDV may provide data collection on demand, such as clinical data, to compensate for limitations and address a specific research question in the near future. Although rarely conducted, in the context of RWE use for decision-making, conducting sensitivity analyses, such as quantitative bias analysis, would be informative to the decision maker to characterize the uncertainty associated with systematic errors [34].

MDV and JMDC may present potential applications to complement clinical trials. Firstly, both allow for the characterization of subpopulations of patients, based for instance on comorbidities or age, in real-world settings that are generally excluded from clinical trials based on the eligibility criteria. Secondly, we expect that clinical trial replication using RWD could provide information on the generalizability of clinical trials in terms of eligibility criteria and treatment adherence, even though the studies conducted so far lack consistency, as reported by Bartlett et al. [35]. However, identifying patients for recruitment in a clinical trial is not possible using MDV or JMDC since these databases are anonymized. In contrast, the Clinical Innovation Network, a national project for registry-oriented clinical research that includes several patient registry databases, can be leveraged for patient recruitment for clinical trials [36].

In Japan, a particular effort is being made to develop novel data sources. For instance, there is a growing need for central data aggregation of genomic and clinical information in oncology; the Center for Cancer Genomics and Advanced Therapeutics database, a repository database storing the genomic data and clinical information of patients receiving cancer gene panel tests under the national health insurance system, is under development. This is believed to be an important step to establish the next-generation infrastructure for medical care and innovation in cancer research [37]. However, at present, it is not clear whether data access will be guaranteed to private companies.

The Pharmaceuticals and Medical Devices Agency (PMDA) guidance expansion in 2018 has helped to enhance the relevance of scientific rigor of the research. In particular, this expansion has fostered discussions between PMDA and pharmaceutical companies from the design stage of studies to provide an external source of expertise, revealing the need of specific educational programs [38].

Finally, following the recent outbreak of the coronavirus disease 2019 (COVID-19) pandemic, the Japanese

government has communicated the importance of promoting RWD collection to stimulate drug development and approval processes. For example, a recent study using MDV showed a sharp drop in asthma-related hospitalizations during the COVID-19 pandemic and brought to light the importance of patient behavior and environmental factors for asthma patients [39].

In conclusion, although several data sources are available to conduct real-world research in Japan, MDV and JMDC are the main data sources that have been used by the pharmaceutical industry. Japan is the third-largest spender in health expenditure after the USA and China, and hence, growing interest in generating Japanese real-world evidence is present, especially for assessment of treatment patterns and healthcare resource utilization, though MDV and JMDC can be leveraged for other types of RWD studies. However, the importance of a careful assessment of each database's strengths and the need for sufficient knowledge on Japan's unique healthcare system were noted, especially at the study design stage. For this reason, the way that care is provided to patients differs from that in Western countries, and thus, it is particularly important to understand the differences that could be impactful on study design and interpretation, in particular, through collaboration with local researchers.

Declarations

Funding No financial support was provided for this work.

Conflict of interest TL, TH, RW, and RP are employees of Clinical Study Support Inc. SG is an employee of Evidera. JS was an employee of Evidera during the time this manuscript was developed. TI is the founder/CEO of Clinical Study Support Inc. RK has no conflicts of interest to report.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Data sharing is not applicable to this article as no datasets were generated or analyzed.

Code availability Not applicable.

Author contributions All authors contributed to the planning and critical review of the manuscript. TL, JS, TH, SG, and RW drafted the manuscript. All 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.

References

1. Ishikawa KB. Medical big data for research use: current status and related issues. Jpn Med Assoc J. 2016;59(2–3):110–24.
2. Nakamura K, Shibata T. Regulatory changes after the enforcement of the new Clinical Trials Act in Japan. Jpn J Clin Oncol. 2020;50(4):399–404. https://doi.org/10.1093/jjco/hyaa028.
3. Yasunaga H. Real world data in Japan: chapter I. Ann Clin Epidemiol. 2019;1(2):28–30. https://doi.org/10.37737/ace.1.2_28.
4. Yasunaga H, Yamana H, Rodes Sanchez M, Towse A. Data governance arrangements for real-world evidence in Japan. 2019. OHE Consulting Report, London: Office of Health Economics. https://www.ohe.org/publications/data-governance-arrangements-real-world-evidence-japan. Accessed 7 Jan 2022.
5. 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(4):459–80. https://doi.org/10.1007/s40801-021-00266-3.
6. Lai EC, Man KK, Chaiyakunapruk N, Cheng CL, Chien HC, Chui CS, Dilokthornsakul P, Hardy NC, Hsieh CY, Hsu CY, Kubota K, Lin TC, Liu Y, Park BJ, Pratt N, Roughead EE, Shin JY, Watcharathanakij S, Wen J, Wong IC, Yang YH, Zhang Y, Setoguchi S. Brief report: databases in the Asia-pacific region: the potential for a distributed network approach. Epidemiology. 2015;26(6):815–20. https://doi.org/10.1097/EDE.0000000000000325.
7. Tanaka S, Seto K, Kawakami K. Pharmacoepidemiology in Japan: medical databases and research achievements. J Pharm Health Care Sci. 2015;1:16. https://doi.org/10.1186/s40780-015-0016-5.
8. Wakabayashi Y, Eitoku M, Suganuma N. Characterization and selection of Japanese electronic health record databases used as data sources for non-interventional observational studies. BMC Med Inform Decis Mak. 2021;21(1):167. https://doi.org/10.1186/s12911-021-01526-6.
9. Lee H, Kang Y, Lee J, Myung J, Strachan L, Ishiguro Y, Lee S. Administrative claims datasets in South Korea and Japan: a comparative assessment based on extensive practical experience. J Health Tech Assess. 2020;8(1):33–42.
10. 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(6):211–8. https://doi.org/10.1002/jgf2.367.
11. Yasunaga H. Real world data in Japan: Chapter II. Ann Clin Epidemiol. 2019;1(3):76–9. https://doi.org/10.37737/ace.1.3_76.
12. Ministry of Health, Labour and Welfare. Various information of medical fees. http://shinryohoshu.mhlw.go.jp/. Accessed 26 Feb 2021.
13. Kobayashi T, Uda A, Udagawa E, Hibi T. Lack of increased risk of lymphoma by thiopurines or biologics in Japanese patients with inflammatory bowel disease: a large-scale administrative database analysis. J Crohns Colitis. 2020;14(5):617–23. https://doi.org/10.1093/ecco-jcc/jjz204.

14. Chang CH, Sakaguchi M, Weil J, Verstraeten T. The incidence of medically-attended norovirus gastro-enteritis in Japan: modelling using a medical care insurance claims database. PLoS ONE. 2018;13(3):e0195164. https://doi.org/10.1371/journal.pone.0195164.
15. Yamazaki K, Macaulay D, Song Y, Sanchez GY. Clinical and economic burden of patients with chronic hepatitis c with versus without antiviral treatment in Japan: an observational cohort study using hospital claims data. Infect Dis Ther. 2019;8:285–99. https://doi.org/10.1007/s40121-019-0234-5.
16. Inoue H, Kozawa M, Milligan KL, Funakubo M, Igarashi A, Loefroth E. A retrospective cohort study evaluating healthcare resource utilization in patients with asthma in Japan. NPJ Prim Care Respir Med. 2019;29(1):13. https://doi.org/10.1038/s41533-019-0128-8.
17. Tanabe M, Motonaga R, Terawaki Y, Nomiyama T, Yanase T. Prescription of oral hypoglycemic agents for patients with type 2 diabetes mellitus: a retrospective cohort study using a Japanese hospital database. J Diabetes Investig. 2017;8(2):227–34. https://doi.org/10.1111/jdi.12567.
18. Matsuoka K, Igarashi A, Sato N, Isono Y, Gouda M, Iwasaki K, Shoji A, Hisamatsu T. Trends in corticosteroid prescriptions for ulcerative colitis and factors associated with long-term corticosteroid use: analysis using Japanese claims data from 2006 to 2016. J Crohns Colitis. 2021;15(3):358–66. https://doi.org/10.1093/ecco-jcc/jjaa172.
19. Fuji T, Akagi M, Abe Y, Oda E, Matsubayashi D, Ota K, Kobayashi M, Matsushita Y, Kaburagi J, Ibusuki K, Takita A, Iwashita M, Yamaguchi T. Incidence of venous thromboembolism and bleeding events in patients with lower extremity orthopedic surgery: a retrospective analysis of a Japanese healthcare database. J Orthop Surg Res. 2017;12(1):55. https://doi.org/10.1186/s13018-017-0549-4.
20. Yamada-Harada M, Fujihara K, Osawa T, Yamamoto M, Kaneko M, Kitazawa M, Matsubayashi Y, Yamada T, Yamanaka N, Seida H, Ogawa W, Sone H. Relationship between number of multiple risk factors and coronary artery disease risk with and without diabetes mellitus. J Clin Endocrinol Metab. 2019;104(11):5084–90. https://doi.org/10.1210/jc.2019-00168.
21. Kohsaka S, Katada J, Saito K, Jenkins A, Li B, Mardekian J, Terayama Y. Safety and effectiveness of non-vitamin K oral anticoagulants versus warfarin in real-world patients with non-valvular atrial fibrillation: a retrospective analysis of contemporary Japanese administrative claims data. Open Heart. 2020;7(1): e001232. https://doi.org/10.1136/openhrt-2019-001232.
22. Hashimoto Y, Michihata N, Yamana H, Shigemi D, Morita K, Matsui H, Yasunaga H, Aihara M. Ophthalmic corticosteroids in pregnant women with allergic conjunctivitis and adverse neonatal outcomes: propensity score analyses. Am J Ophthalmol. 2020;220:91–101. https://doi.org/10.1016/j.ajo.2020.07.011.
23. Wake M, Onishi Y, Guelfucci F, Oh A, Hiroi S, Shimasaki Y, Teramoto T. Treatment patterns in hyperlipidaemia patients based on administrative claim databases in Japan. Atherosclerosis. 2018;272:145–52. https://doi.org/10.1016/j.atherosclerosis.2018.03.023.
24. Wake M, Oh A, Onishi Y, Guelfucci F, Shimasaki Y, Teramoto T. Adherence and persistence to hyperlipidemia medications in patients with atherosclerotic cardiovascular disease and those with diabetes mellitus based on administrative claims data in Japan. Atherosclerosis. 2019;282:19–28. https://doi.org/10.1016/j.atherosclerosis.2018.12.026.
25. Ho M, Van der Laan M, Lee H, Chen J, Lee K, Fang Y, He W, Irony T, Jiang Q, Lin X, Meng Z, Mishra-Kalyani P, Rockhold F, Song Y, Wang H, White R. The current landscape in biostatistics of real-world data and evidence: causal inference frameworks for study design and analysis. Stat Biopharm Res. 2021. https://doi.org/10.1080/19466315.2021.1883475.
26. Crown WH. Real-world evidence, causal inference, and machine learning. Value Health. 2019;22(5):587–92. https://doi.org/10.1016/j.jval.2019.03.001.
27. Yokoyama S, Wakamoto S, Tanaka Y, Nakagawa C, Hosomi K, Takada M. Association between antipsychotics and osteoporosis based on real-world data. Ann Pharmacother. 2020;54(10):988–95. https://doi.org/10.1177/1060028020913974.
28. Yamashita T, Laurent T, Kato M, Yoshihara N, Ono F. Comparison of hospital length of stay of acute ischemic stroke patients with non-valvular atrial fibrillation started on rivaroxaban or warfarin treatment during hospitalization. J Med Econ. 2020;23(12):1379–88. https://doi.org/10.1080/13696998.2020.1824384.
29. Igarashi A, Fujita H, Arima K, Inoue T, Dorey J, Fukushima A, Taguchi Y. Health-care resource use and current treatment of adult atopic dermatitis patients in Japan: a retrospective claims database analysis. J Dermatol. 2019;46(8):652–61. https://doi.org/10.1111/1346-8138.14947.
30. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016;79:70–5. https://doi.org/10.1016/j.jclinepi.2016.04.014.
31. Nishimura R, Kato H, Kisanuki K, Oh A, Onishi Y, Guelfucci F, Shimasaki Y. Comparison of persistence and adherence between fixed-dose combinations and two-pill combinations in Japanese patients with type 2 diabetes. Curr Med Res Opin. 2019;35(5):869–78. https://doi.org/10.1080/03007995.2018.1551192.
32. Koram N, Delgado M, Stark JH, Setoguchi S, de Luise C. Validation studies of claims data in the Asia-Pacific region: a comprehensive review. Pharmacoepidemiol Drug Saf. 2019;28:156–70. https://doi.org/10.1002/pds.4616.
33. Iwagami M, Aoki K, Akazawa M, Ishiguro C, Imai S, Ooba N, Kusama M, Koide D, Goto A, Kobayashi N, Sato I, Nakane S, Miyazaki M, Kubota K. Task force report on the validation of diagnosis codes and other outcome definitions in the Japanese receipt data. Jpn J Pharmacoepidemiol. 2018;23:95–123. https://doi.org/10.3820/jjpe.23.95.
34. Lash TL, Fox MP, Cooney D, Lu Y, Forshee RA. Quantitative bias analysis in regulatory settings. Am J Public Health. 2016;106(7):1227–30. https://doi.org/10.2105/AJPH.2016.303199.
35. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of using real-world data to replicate clinical trial evidence. JAMA Netw Open. 2019;2(10): e1912869. https://doi.org/10.1001/jamanetworkopen.2019.12869.
36. Matsushita S, Tachibana K, Kondoh M. The Clinical Innovation Network: a policy for promoting development of drugs and medical devices in Japan. Drug Discov Today. 2019;24(1):4–8. https://doi.org/10.1016/j.drudis.2018.05.026.
37. Mano H. Cancer genomic medicine in Japan. Proc Jpn Acad Ser B Phys Biol Sci. 2020;96(7):316–21. https://doi.org/10.2183/pjab.96.023.
38. Takahashi N, Otake R, Hokugo J, Gunji R, Taniguchi T, Nakao C, Haraguchi M, Sakamoto Y, Kani T. The survey of trends on japanese post-marketing study after enforcement of revised good post-marketing study practice (GPSP). Jpn J Pharmacoepidemiol. 2020;25(1):17–27. https://doi.org/10.3820/jjpe.25.17.
39. Abe K, Miyawaki A, Nakamura M, Ninomiya H, Kobayashi Y. Trends in hospitalizations for asthma during the COVID-19 outbreak in Japan. J Allergy Clin Immunol Pract. 2021;9(1):494-496.e1. https://doi.org/10.1016/j.jaip.2020.09.060.