



Strategies to Address Current Challenges in Real-World Evidence Generation in Japan

Thomas Laurent^{1,2} · Dimitra Lambrelli^{1,3} · Ryozo Wakabayashi^{1,2} · Takahiro Hirano^{1,2}  · Ryohei Kuwatsuru^{1,4}

Accepted: 19 April 2023 / Published online: 13 May 2023
© The Author(s) 2023

Abstract

The generation of real-world evidence (RWE), which describes patient characteristics or treatment patterns using real-world data (RWD), is rapidly growing more popular as a tool for decision-making in Japan. The aim of this review was to summarize challenges to RWE generation in Japan related to pharmacoepidemiology, and to propose strategies to address some of these challenges. We first focused on data-related issues, including the lack of transparency of RWD sources, linkage across different care settings, definitions of clinical outcomes, and the overall assessment framework of RWD when used for research purposes. Next the study reviewed methodology-related challenges. As lack of design transparency impairs study reproducibility, transparent reporting of study design is critical for stakeholders. For this review, we considered different sources of biases and time-varying confounding, along with potential study design and methodological solutions. Additionally, the implementation of robust assessment of definition uncertainty, misclassification, and unmeasured confounders would enhance RWE credibility in light of RWD source-related limitations, and is being strongly considered by task forces in Japan. Overall, the development of guidance for best practices on data source selection, design transparency, and analytical methods to address different sources of biases and robustness in the process of RWE generation will enhance credibility for stakeholders and local decision-makers.

1 Introduction

In Japan, the use of real-world evidence (RWE) for hypothesis generation and decision-making has become an area of focus to overcome the limitations of clinical trials. The Pharmaceuticals and Medical Devices Agency (PMDA) has recognized the importance of using real-world data (RWD) for regulatory purposes, and discussions regarding the acceptability of RWE for regulatory submissions have already been initiated. However, there are multiple challenges that still need to be addressed, including RWD

availability, quality, and completeness; accurate definition of clinical outcomes; and analytical approaches used in RWE studies [1]. To address some of these issues and promote the utilization of RWD throughout the drug life cycle, including post-marketing drug safety assessment, relevant guidelines have been published, and multiple initiatives have been launched in Japan [2].

✉ Takahiro Hirano
takahiro_hirano@jp-css.com

¹ Real-World Evidence and Data Assessment (READS), Graduate School of Medicine, Juntendo University, Hongo 2-1-1, Bunkyo-ku, Tokyo 113-8421, Japan
² Clinical Study Support Inc., 2F Daiei Bldg., 1-11-20 Nishiki Naka-ku, Nagoya 460-0003, Japan
³ Real-World Evidence, Evidera, The Ark, 2nd Floor, 201 Talgarth Road, London W6 8BJ, UK
⁴ Department of Radiology, School of Medicine, Juntendo University, Hongo 2-1-1, Bunkyo-ku, Tokyo 113-8421, Japan

Key Points

The real-world data landscape is rapidly evolving in Japan, with more than 20 patient-level data sources available.

The Pharmaceuticals and Medical Devices Agency is organizing two main data repositories, namely the Medical Information Database Network and the Clinical Innovation Network.

A methodological framework of how to conduct real-world evidence studies needs to be developed to ensure transparency, robustness, and reproducibility.

The number of available RWD sources is steadily increasing, with 22 patient-level databases listed in the Japanese Society for Pharmacoepidemiology database in 2022 [3]. In a previous work focusing on RWD in Japan, we described the advantages and challenges of the commercially available databases Medical Data Vision (MDV) and JMDC (formerly known as Japan Medical Data Center) [4]. Another resource, the Japanese National Database of Health Insurance Claims and Specific Health Checkups (NDB), was established in 2009 to generate data and assist in health policy planning and regulation of public healthcare expenditures. NDB includes almost all the administrative claims data collected from the insured population in Japan [5], as well as data from yearly health checkups offered to people who are covered by the National Health Insurance schemes [6]. However, NDB is currently used only by academic researchers and is not accessible to private companies [5].

Most studies using RWD from MDV and JMDC focus on describing patient characteristics and treatment patterns since such RWD sources contain little information on clinical characteristics and outcomes [4]. A recent review that assessed the types of studies conducted between 2015 and 2020 using claims databases in Japan found that descriptive studies were the most common, accounting for 63%, 43%, and 41% of the studies conducted using NDB, MDV, and JMDC data, respectively [7]. To address this issue of missing data and to create a high-quality medical information database for conducting rigorous assessments of drug safety, the Medical Information Database Network (MID-NET) was launched in April 2018, in collaboration with 23 hospitals from 10 healthcare organizations across Japan.

The strength of the MID-NET database lies in the rich clinical data and standardized coding procedures across different contributing hospitals, allowing integrated analysis. For example, an important effort is in progress to improve signal detection using this database [8]. MID-NET is also well recognized for post-authorization safety studies (PASS), especially under Good Post-Marketing Study Practice regulation. The Clinical Innovation Network (CIN) is another initiative that was launched to develop a registry-based infrastructure for improving clinical drug development. Some of the registry databases have been leveraged for multiple applications, including patient enrollment into clinical trials and PASS, though the initiative has not yet been completed. Therefore, both MID-NET and CIN may provide data of high quality and reliability [9].

Although local guidelines refer to data-related issues, including quality, completeness, suitability, and transparency, further guidance is needed to expand the framework of RWD use [10, 11]. Furthermore, inappropriate analytical methodologies may lead to biased evidence generation, undermining the credibility of RWE. Although PMDA offers expertise to pharmaceutical companies, there is no

formal process in place to systematically utilize this support, and the choice of inappropriate study design and analytical methods may negatively influence study results [12, 13]. Other challenges that could limit the interpretation of results include sensitivity analysis, analytical models, measurement of validity, and confounding. These elements are generally considered to be study limitations but are often not addressed analytically. Additionally, time-related bias is generally under-considered with respect to confounding bias. Time-related bias is potentially more important than randomization itself in the context of target trial emulation [14]; it is harder to address due to the needs of data at multiple time points, and is highly dependent on the study design [15].

This review aims to provide strategies to overcome some of these challenges, and to fill gaps in the current practices in RWE generation in Japan related to pharmacoepidemiology for regulatory purposes, by focusing mainly on data- and methodology-related issues.

2 Challenges and Solutions Relating to Data

This section discusses challenges relating to RWD and how these may be addressed (Table 1). PMDA was the primary source of evidence to assess the most recent initiatives on RWD-related considerations in Japan, and a targeted literature review was conducted to collect information on the most prominent methodological issues associated with RWE generation in Japan.

2.1 Lack of Data Transparency of RWD Sources

Sufficient transparency of RWE provides information on how findings were derived, allowing for assessment of validity and study reproducibility, but the reliability and transparency of the RWD used to generate RWE in Japan have not been adequately assessed. Several research groups have attempted to develop guidelines and best practices for the assessment of different types of data. For example, an evidence-based guideline (3X3 Data Quality Assessment) has been proposed for the assessment of electronic health record (EHR) data, focusing on three dimensions of the EHR data: patients, variables, and time [16]. A different study explored good practices and challenges, and proposed solutions for the effective use of administrative data, including aspects of data acquisition, approval processes, access and disclosure, data and metadata, research support, and data reuse [17].

Assessment of RWD should cover both non-regulatory and regulatory purposes, but generally the requirement for the latter is more stringent. Considering the future needs of RWE, RWD that meets a minimum standard accepted by

Table 1 Summary of potential solutions to overcome current challenges in real-world research

Category	Challenge	Solutions
Challenges relating to data	Lack of transparency on RWD-generation process	Use “checklist” with specific criteria to evaluate the appropriateness for regulatory-grade data quality (e.g., completeness, transparency, generalizability)
	Lack of linkage across different care settings	Data linkage using deterministic or probabilistic approach
	Lack of clinical outcome data in traditional RWD	Chart review studies; claims data sources linked to electronic health records
Challenges relating to study methodology	Lack of design transparency	Enhance design transparency, for example, clear outcome definition, causal diagrams, identification of confounders, and fit-for-purpose design by reporting key parameters into a study design diagram and tables reporting key parameters to define the study population and analysis
	Lack of clarification on “time zero” definition	Use a new user active comparator design Graphical representation of key study design variables Emulate the target trial and map key study parameters and time zero definition. Alternatively, either randomly assign the individual to one of the treatment arms or create exact copies of the individual (clones)
	Issue of handling time-varying confounding	Marginal Structural Model with inverse probability weighting or G-computation
	Robust definitions	Sensitivity analysis on the definitions of time windows, outcomes, and exposures
	Assumption of no unmeasured confounding in traditional methods	Report E-value When the E-value is too low, implement further sensitivity analysis
	Misclassification and lack of validation of code-based algorithms	Conduct quantitative bias analysis of misclassification Conduct validation studies using chart review, electronic medical data, or registry databases

regulators, a “regulatory grade,” would be needed to further expand its applications, though no consensus has been reached on the definition of the minimum accepted standard. To assess the appropriateness for “regulatory grade,” a proposed checklist includes the following elements: high quality (provenance of each datapoint must be clear, traceable, and auditable), completeness (predefined rules for data abstraction), and timing of data update [18]. A different checklist for investigators in database research focuses on following items: population covered, capture of study variable, continuous and consistent data capture, record duration and data latency, and database expertise [19]. In this regard, a flow diagram was developed allowing users to match a study objective and the data quality and quantity in the Japanese context. The process sequentially clarifies the following elements: construction process (reliability of the data collection), data related to endpoints, anonymization, data volume (sample size), and sufficiency of the follow-up period to address the research question [20].

2.2 Lack of Linkage Across Different Care Settings

In Japan, there are hospital-based (e.g., MDV) and insurance-based (e.g., JMDC) administrative claims databases,

which can cover different healthcare settings. Japan has a universal medical insurance system consisting of five public insurance programs (based on occupational status), and every resident of Japan must be enrolled in one of these programs. Patients can access clinics and hospitals without a primary care physician gatekeeper system or insurance restrictions [4]. However, the presence of different health insurance plans based on patient occupational status represents a barrier to patient follow-up when individuals change their health insurance plan. In addition, different healthcare settings, such as hospitals and clinics, are not usually linked.

Data linkage of de-identified databases, using either deterministic or probabilistic methods, has been a recent focus. These methods are considered promising, though they generally require a certain degree of personal health information [21, 22]. The deterministic approach considers linkage as binary (i.e., linked or not linked), while probabilistic data linkage is useful when linkage variables are inaccurate or not unique [23]. For instance, an algorithm was developed to perform probabilistic data linkage based solely on diagnosis codes [International Classification of Diseases, 10th Revision (ICD-10)] available in two different de-identified datasets. This method presents three main advantages: (i) it only uses diagnosis code information, (ii)

it accounts for discrepancies between the datasets, and (iii) it does not require subsample for tuning [24].

A different study performed in the USA demonstrated that a comparative effectiveness study could be reproduced when conducting imputation on missing outcome values based on probabilistic data linkage of two claim databases, using pre-index characteristics, including demographic data, comorbidities, and utilization of healthcare services [25]. Although this method has not yet been applied, it could prove to be successful in Japanese databases, for instance, linking claims data to electronic medical records (EMRs). However, transparent reporting of the data linkage method, including the quality of the data sources, linkage variables, methods, and evaluation, is crucial to inform potential linkage bias [26].

Of note, there are important concerns in terms of regulations of patient privacy. Under the Next-Generation Medical Infrastructure Act, enacted in 2018, Japanese regulation authorities have certified some business operators to conduct data linkage across different data sources. Some stakeholders have suggested using MyNumber, the Japanese national identification number, to link different databases [12].

2.3 Lack of Clinical Outcome Data in RWD

Different types of RWD, such as EHRs, administrative claims data, or disease-specific patient registries, are structured and collect data elements for specific purposes, and may vary in population coverage, variables included, and longitudinality. Previously, we highlighted the limited amount of clinical information available in widely used Japanese RWD sources, namely MDV and JDMC [4]. To address this limitation, chart review studies at single or multiple sites could be conducted. The emergence of the RWD Database (Real World Data Co. Ltd. database, containing both EMR and claims data [12], may help to fill this gap. For instance, a validation study was conducted leveraging this database to use EMR data as a gold standard for the validation of algorithms to identify cardiovascular outcomes [27]. This study developed claims-based algorithms to identify clinical outcomes and validated them through the use of EMRs. Furthermore, another study validated a claims-based (JMDC) algorithm identifying patients with Crohn's disease by using patients' medical record data from a single site. Patient identifiers from the claims database were restored to collect data from medical records [28].

The creation of databases rich in clinical outcomes may help address the gaps encountered in traditional claims data sources. As mentioned above, the MID-NET database, which is rich in clinical characteristics, has been developed to respond to the needs of conducting PASS. Multiple case studies, focusing on blood coagulability, thrombocytopenia, and renal dysfunction, identified clinical outcomes based on laboratory data that are collected in this database [2].

CIN encompasses multiple registry databases aiming to collect clinically relevant data that reflect real-world clinical practice, and studies using the Registry of Muscular Dystrophy (REMUDY) and Japanese Registry for Mechanically Assisted Circulatory Support (J-MACS) have been conducted for patient enrollment and PASS, respectively. The J-MACS database also includes quality of life data that are not available in traditional Japanese claims RWD sources. However, the validation of the data quality for each registry remains to be addressed and approved by PMDA [9].

3 Challenges and Solutions Relating to the Methodology

3.1 Lack of Design Transparency in RWE Generation

The growing importance of RWE for regulatory purposes has led to the need to build stakeholders' trust in the RWE generation process. Transparency is a necessary but not sufficient condition to ensure high-quality RWE. In real-world research the study design precedes the study implementation, but transparency is often lacking for the methodological decisions applied during study conduct and potential derivations from the planned analysis, which may result in cherry picking of the results [29]. There are several publications that provide guidance on a structured process to support the validity, transparency, and reproducibility of real-world research [30–32]. A published study proposes a sequential process from early design addressing the adequacy and feasibility of the data, analysis considerations, and reporting [31]. During design phase and prior to analysis, it is advisable to identify the critical temporal anchors (e.g., study entry date, follow-up window) and provide a design diagram [29, 32, 33]. The structured template and reporting tool for real world evidence (STaRT-RWE) has been developed as a result of a collaboration between the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the International Society for Pharmacoepidemiology to provide guidance on planning and reporting of RWE studies, including aspects of key study parameters, study design, validity assessment, sensitivity analysis [34, 35], study registration, replicability, and stakeholder involvement to increase decision-makers' confidence in RWE [30].

Study design is constrained by and strongly related to the structure of the data sources used; therefore, as a first step, we believe that transparent, detailed, structured descriptions of the data sources used, as well as reports of the limitations and advantages of the study design, would help build consistent knowledge about the appropriateness of the existing data sources in studies in Japan (Table 1). In the design phase, this information can be acquired from data dictionaries and pre-analysis investigation of data samples to assess

data completeness. In the context of evaluating treatment effectiveness, the following process has been proposed: (i) choose the estimand by specifying the hypothesis, (ii) opt for a trial emulation framework and define the variables for meeting the “exchangeability” assumption, and (iii) specify the analytical method at the design stage [36]. In this regard, examples of trial emulation using observational data were reported in the literature [37, 38]. In Japan, only one trial emulation study has been published, aiming to evaluate the effect of several antidiabetics among Japanese patients using the NDB [39]. Further pilot studies are needed to evaluate challenges and opportunities of conducting emulation trials using Japanese RWD.

3.2 Lack of Clarification on “Time Zero” Definition and Related Bias

We have previously discussed the lack of reliability and reproducibility due to the absence of guidance on the definition of “time zero” of follow-up [4]. Inappropriate definition of “time zero” due to misalignment with study eligibility and treatment assignment may result in introducing bias in real-world studies, especially selection bias and immortal time bias [14]. On one hand, setting “time zero” either after both eligibility and treatment assignment or at eligibility but after treatment assignment, introduces a selection bias in favor of prevalent users. On the other hand, setting “time zero” either before eligibility and treatment assignment or at eligibility and before treatment assignment introduces immortal time bias [14]. A recently published study using a Japanese database showed that, in the new user versus non-user design, different settings of “time zero” for non-users (i.e., at eligibility assignment, by propensity score matching, by random selection, and the cloning method) would generate substantially different results in the parameters evaluated [15]. In RWD studies comparing a treatment with an active comparator, the new user active comparator design can accommodate for this misalignment by anchoring “time zero” of follow-up to treatment initiation and eligibility [40]. Moreover, this design allows preservation of the temporality of covariates and evaluation of time-varying hazards [41]. Even though “time zero” is properly defined, it remains necessary to adjust for potential post-“time zero” selection due to censoring [14].

Determining the consistency of a treatment strategy may not always be straightforward, and in that case, two approaches have been proposed: (i) random assignment of each individual to one of the treatments, and (ii) creating exact copies of the individual clones [14]. This latter approach has recently gained attention due to the possibility of considering the time lag, for example, between diagnosis and treatment assignment as part of an intervention [42]. The

cloning method by design introduces an artificial censoring, associated with the incompatibility of a clone to be allocated to one arm. The method uses weights that account for the selection bias, and are estimated using inverse probability of censoring weighting. However, this method inflates sample size, and observations are not independently and identically distributed; hence variance estimation requires a robust method, for example, non-parametric bootstrap, which can be computationally intensive when using large databases [42]. One possibility would be to work on a subsample to address this issue when the sample is large enough.

3.3 Hurdle in Handling Time-Varying Confounding

In longitudinal studies, properly accounting for the relationship between patient characteristics, patients’ medical history, and the treatment is required to avoid bias. However, when time-varying exposure can be influenced by prior time-varying patients’ clinical characteristics and mediate the effect on the outcome, common methods for controlling confounding may not be valid [43]. A marginal structural model using inverse propensity score weighting can address the issue of mediator, and is based on estimating inverse probability of treatment weighted estimators [44]. Marginal structural model method was applied to account for the time-dependent nature of anemia to assess its effect on renal function using the JMDC database, implementing the inverse propensity score weighting method to balance for time-dependent confounding [45]. Importantly, the above-mentioned methods do not address the issue of long-term dependency, and in this regard, advanced methods such as counterfactual recurrent network were developed to handle time-dependent confounders by using balancing representations [46]. Counterfactual recurrent network implements a model that uses adversarial training and recurrent network to build a balancing representation without assuming the form of treatment assignment, resulting in removal of the association between patient history and assignment treatment. Other methods and algorithms may help address these challenges. For instance, Medical Deconfounder, which relies on a probabilistic factor model, was developed to account for common unobserved characteristics among multiple medications [47]. G-method is another method that can account for time-varying confounders affected by previous treatment, though it requires modeling both covariates and outcome, and may be sensitive to the violation of the assumption of unmeasured confounding and model misspecification [48]. Further, targeted maximum likelihood estimation would allow for implementation of a variety of machine learning algorithms [49]. However, a significant challenge is the lack of a data-driven procedure to find the most appropriate causal inference method when using RWD.

3.4 Sensitivity Analysis and Validation Study to Tackle Uncertainty of Key Parameter Definitions

Another issue is the presence of uncertainty in the operation definitions, which may have a negative impact on decision-makers. Sensitivity analyses on time windows, outcomes, and exposures may help assess the impact of those definitions on the results. Using MDV database, for instance, sensitivity analysis was conducted by setting a more stringent definition of the treatment/exposure in a study that investigated treatment effectiveness in patients with inflammatory bowel disease [50]. Further, a comparative effectiveness study among patients with non-valvular atrial fibrillation employed varying time horizons in a sensitivity analysis on the same database. We believe that conducting sensitivity analysis on study parameter definitions would be useful for stakeholders to evaluate result robustness in the presence of uncertainty on parameter definitions.

Generally, RWD are collected for purposes other than research (e.g., audit, reimbursement), and accurate diagnosis of medical conditions may be incorrectly captured in the data, resulting in potential misclassification of outcomes and exposures. Misclassification occurs in most studies, and in general, methods for sensitivity analysis consist of testing different assumptions with regards to sensitivity and specificity [51]. There are several methods to evaluate the impact of misclassification. For example, probabilistic sensitivity analysis allows a record-level correction that can handle the correction for multiple sources of bias [51]. Another study focused on outcome misclassification and proposed an adjustment method when misclassification probabilities are known beforehand [52]. One of the important issues when focusing on misclassification is setting the proper range of sensitivity and specificity parameters to conclude on the measurement validity [53].

Code-based algorithm validation studies are also scarce in Japan [12]. There are multiple considerations when conducting validation studies, such as the selection of the gold standard against which the algorithm validity is assessed. In Japan, positive predictive value (PPV) is considered as the main performance indicator when considering outcome validation [54]. There are multiple Japanese data sources that can be leveraged to conduct a validation study, and guidelines recommend using clinical measurement, registry databases, or chart review as a gold standard [54]. As mentioned above, data sources such as the Real-World Data Co. Ltd. have the potential to be used for validation, since EMR data could be used as a gold standard.

There are only a few examples of code-based validation studies using RWD in Japan. One study using JMDC claims data evaluated the sensitivity, specificity, negative predictive value, and PPV of an algorithm identifying patients with

Crohn's disease using data extracted from a medical chart review [28]. The authors concluded that using ICD-10 codes alone were not sufficient to achieve a suitable PPV, and adding prescription codes was necessary [28]. Another study aimed to validate cardiovascular outcomes among diabetic patients in claims using EMR data from the Real World Data Co. Ltd. database for validation [27]. For the three main outcomes—that is, congestive heart failure, hemorrhagic stroke, and mild or moderate chronic kidney disease—the PPV was estimated to be over 90%, despite the fact that definitions in the EMR data were based solely on ICD-10 codes, and only one of the outcomes was defined based on laboratory data [27]. As an alternative example, a validation study was carried out utilizing JMDC to evaluate distinct algorithms for identifying treated diabetic patients [55]. Hemoglobin A1c measurements from health checkup data were used as reference. PPVs above 80% were found for several algorithms [55]. The study focused on both PPV and sensitivity.

3.5 Assumption of No Unmeasured Confounding in Traditional Methods

Methods widely implemented in RWD studies, such as propensity score matching, are based on the assumption of absence of unmeasured confounding, also known as the “no unobserved confounder assumption” (NUCA). Failure to evaluate the degree of the impact of the deviation from this assumption may reduce the quality of the evidence and lower its credibility among stakeholders. To enhance the quality of evidence, the importance of investigating residual confounding was noted to better clarify the measured effect [56]. Best practices for addressing potential deviation from NUCA when estimating causal treatment effect using RWD have been suggested [57]. As the first step, an initial sensitivity analysis is conducted using E-value, which expresses the minimum strength of unmeasured confounding to nullify the observed treatment effect. E-value does not require any assumptions, nor prior information on unmeasured confounding, and the robustness of the evidence increases with larger E-values [58]. When small E-values are obtained, that is, the presence of unmeasured confounding is not implausible, further analyses can be performed to examine the deviation from NUCA. However, many of these methods are not implementable in a straightforward manner and lack guidance [57].

In Japan, RWD studies evaluating the impact of unmeasured confounding are limited. For instance, a quantitative bias analysis was conducted using E-value to assess the minimum strength of unmeasured confounding to nullify the hazard ratio of sodium-glucose cotransporter-2 inhibitors compared with other glucose-lowering drugs in Japanese patients with diabetes. Because the E-value was above two, the authors concluded that the impact of the unmeasured confounding would be limited [59].

E-value interpretation depends on the adjustment for confounders performed, and there is no clear threshold to decide whether to carry out further sensitivity analyses [60]. Depending on the level of information on confounders available in the data and the study design, methods such as instrumental variable (IV) methods [61], difference-in-difference [62], missing cause [63], trend-in-trend [64], and perturbation variable [65] may be appropriate. Studies using Japanese databases with such quasi-experimental designs accounted for 22.1% (142/643) of all the studies using RWD reported from 2015 to 2020 [7]. Few Japanese studies used quasi-experimental designs other than propensity score methods. For instance, one study examined the effectiveness of tranexamic acid in post-tonsillectomy hemorrhage by defining the use of tranexamic acid in the preceding patient as IV, using the Diagnosis Procedure Combination (DPC) database. Results between this method and the propensity-score-based method were consistent [66]. Another study had a similar approach, using IVs to evaluate the effect of dexmedetomidine on patients admitted to the intensive care unit using the DPC database. The selected IV was a proxy for hospital preference to prescribe dexmedetomidine [67]. Another quasi-experimental design, the difference-in-difference, has been implemented in Japan RWD. A COVID-19 study assessed the difference-in-difference for the change in admissions for ambulatory care sensitive conditions before and during the pandemic by comparing prefectures enforcing a state of emergency with those that did not [68].

More recently, a novel method was developed for addressing unobserved confounding when the assumption of the presence of two correlated confounders with a nonlinear condition on the exposure holds [69]. When external data is available, methods leveraging this information can be implemented, such as Bayesian twin regression [70], multiple imputation [71], and propensity score calibration [72]. However, identifying the most suitable model is challenging, especially in a causal inference framework; use of synthetic datasets generated using the “plasmode” simulation based on healthcare claims, or Wasserstein generative adversarial networks, may help assess biases among different methods [73, 74]. Overall, we would recommend testing the robustness to NUCA using E-value as the initial step (Table 1).

4 Conclusion

Coordinated efforts are being made in Japan toward optimization of RWD as the number of patient-level data sources is constantly increasing, and new databases, such as MID-NET,

strive to mitigate the lack of clinical data in the commonly used commercial RWDs. Probabilistic or deterministic linkage and the use of “checklists” for RWD assessment are acknowledged as potential solutions for data-related issues. Transparent reporting of study design is recognized as an important element to increase reproducibility and credibility with respect to stakeholders. Emphasis should be placed on carefully defining different sources of bias, especially time-related bias; addressing it analytically; and comparing different methods when no best methods have been previously defined. Robustness assessment covering the uncertainty of the definitions, misclassification, and hypothesis on unmeasured confounders would be valuable for decision-makers. Future pilot studies would provide more insight into the strengths and limitations of the databases and methodological issues. The knowledge gained can support the development of methodological guidelines for pharmacoepidemiological real-world studies in Japan.

Declarations

Funding No financial support was provided for this work.

Competing Interests TL, TH, and RW are employees for Clinical Study Support, Inc. DL is an employee for Evidera, Inc. RK has no conflicts of interest to report.

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

Code Availability Not applicable.

Authors' Contributions All authors contributed to the planning and critical review of the manuscript. TL, DL, TH 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.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

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 directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Eichler HG, Pignatti F, Schwarzer-Daum B, Hidalgo-Simon A, Eichler I, Arlett P, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. *Clin Pharmacol Ther.* 2021;109:1212–8.
- Nishioka K, Makimura T, Ishiguro A, Nonaka T, Yamaguchi M, Uyama Y. Evolving acceptance and use of RWE for regulatory decision making on the benefit/risk assessment of a drug in Japan. *Clin Pharmacol Ther.* 2022;111:35–43.
- The Japanese Society for Pharmacoepidemiology. The Japanese Society for Pharmacoepidemiology Database 2022 [Internet]. 2022. <https://sites.google.com/view/jspe-database-en2022>. Accessed 19 Apr 2023.
- 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;9:175–87.
- Yasunaga H. Real world data in Japan: Chapter I NDB. *Ann Clin Epidemiol.* 2019;1:28–30.
- Organisation for Economic Cooperation and Development (OECD). Chapter 3: Health check-ups in Japan. *OECD Reviews of Public Health: Japan: A Healthier Tomorrow.* Paris, France: OECD Publishing; 2019. p. 125–70.
- Fujinaga J, Fukuoka T. A review of research studies using data from the administrative claims databases in Japan. *Drugs - Real World Outcomes.* 2022;9:543–50.
- Yamaguchi M, Inomata S, Harada S, Matsuzaki Y, Kawaguchi M, Ujibe M, et al. Establishment of the MID-NET® medical information database network as a reliable and valuable database for drug safety assessments in Japan. *Pharmacoepidemiol Drug Saf.* 2019;28:1395–404.
- 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;24:4–8.
- Japan Pharmaceutical Manufacturers Association. Toward the promotion of RWD utilization in the pharmaceutical industry: current status, issues, discussion points, and future prospects [in Japanese]. 2020. https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000005k34-att/bd_rwd_sg3.pdf. Accessed 28 Sept 2022.
- Pharmaceuticals and Medical Devices Agency. Guidelines for the conduct of pharmacoepidemiological studies in drug safety assessment with medical information databases. 2014. <https://www.pmda.go.jp/files/000240951.pdf>. Accessed 28 Sept 2022.
- Hiramatsu K, Barrett A, Miyata Y, PhRMA Japan Medical Affairs Committee Working Group. Current status, challenges, and future perspectives of real-world data and real-world evidence in Japan. *Drugs Real World Outcomes.* 2021;8:459–80.
- Togo K, Yonemoto N. Real world data and data science in medical research: present and future. *Jpn J Stat Data Sci.* 2022;5:769–81.
- 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.
- Wakabayashi R, Hirano T, Laurent T, Kuwatsuru Y, Kuwatsuru R. Impact of “time zero” of follow-up settings in a comparative effectiveness study using real-world data with a non-user comparator: comparison of six different settings. *Drugs Real World Outcomes.* 2023;10:107–17.
- Weiskopf NG, Bakken S, Hripesak G, Weng C. A data quality assessment guideline for electronic health record data reuse. *EGEMS (Wash DC).* 2017;5:14.
- Jones KH, Heys S, Tingay K, Jackson P, Dibben C. The good, the bad, the clunky: improving the use of administrative data for research. *Int J Popul Data Sci.* 2019;4:587.
- Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (rwe): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther.* 2018;103:202–5.
- Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, Lo CR. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf.* 2012;21:1–10.
- 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;21:167.
- Curtis JR, Chen L, Bharat A, Delzell E, Greenberg JD, Harrold L, et al. Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res (Hoboken).* 2014;66:1790–8.
- Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. Privacy-preserving record linkage on large real world datasets. *J Biomed Inform.* 2014;50:205–12.
- Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J Biomed Inform.* 2015;56:80–6.
- Hejblum BP, Weber GM, Liao KP, Palmer NP, Churchill S, Shadick NA, et al. Probabilistic record linkage of de-identified research datasets with discrepancies using diagnosis codes. *Sci Data.* 2019;6: 180298.
- Blanchette CM, De Koven M, De AP, Roberts M. Probabilistic data linkage: a case study of comparative effectiveness in COPD. *Drugs Context.* 2013;2013: 212258.
- Pratt NL, Mack CD, Meyer AM, Davis KJ, Hammill BG, Hampp, Christian, et al. Data linkage in pharmacoepidemiology: a call for rigorous evaluation and reporting. *Pharmacoepidemiol Drug Saf.* 2019;29:9–17.
- Ono Y, Taneda Y, Takeshima T, Iwasaki K, Yasui A. Validity of claims diagnosis codes for cardiovascular diseases in diabetes patients in Japanese administrative database. *Clin Epidemiol.* 2020;12:367–75.
- Morikubo H, Kobayashi T, Fukuda T, Nagahama T, Hisamatsu T, Hibi T. Development of algorithms for identifying patients with Crohn’s disease in the Japanese health insurance claims database. *PLoS ONE.* 2021;16: e0258537.
- Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf.* 2017;26:1018–32.
- Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch 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.
- Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther.* 2019;105:867–77.
- Wang SV, Sreedhara SK, Schneeweiss S, REPEAT Initiative. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nat Commun.* 2022;13:5126.
- Wang SV, Schneeweiss S. Assessing and interpreting real-world evidence studies: introductory points for new reviewers. *Clin Pharmacol Ther.* 2022;111:145–9.
- Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, et al. STaRT-RWE: structured template for planning and reporting

- on the implementation of real world evidence studies. *BMJ*. 2021;372: m4856.
35. Orsini LS, Monz B, Daniel Mullins C, Van Brunt D, Daniel G, Eichler H-G, et al. Improving transparency to build trust in real-world secondary data studies for hypothesis testing—why, what, and how: recommendations and a road map from the real-world evidence transparency initiative. *Pharmacoepidemiol Drug Saf*. 2020;29:1504–13.
 36. Uemura Y, Shinozaki T, Nomura S, Shibata T. Comment on “Biostatistical considerations when using RWD and RWE in clinical studies for regulatory purposes: a landscape assessment.” *Stat Biopharm Res*. 2023;15:20–2.
 37. Braitmaier M, Schwarz S, Kollhorst B, Senore C, Didelez V, Haug U. Screening colonoscopy similarly prevented distal and proximal colorectal cancer: a prospective study among 55–69-year-olds. *J Clin Epidemiol*. 2022;149:118–26.
 38. Jin X, Ding C, Hunter DJ, Gallego B. Effectiveness of vitamin D supplementation on knee osteoarthritis - a target trial emulation study using data from the Osteoarthritis Initiative cohort. *Osteoarthr Cartil*. 2022;30:1495–505.
 39. Takeuchi Y, Kumamaru H, Hagiwara Y, Matsui H, Yasunaga H, Miyata H, et al. Sodium-glucose cotransporter-2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: target trial emulation using a nationwide administrative claims database. *Diabetes Obes Metab*. 2021;23:1379–88.
 40. Luijken K, Spekrijse JJ, van Smeden M, Gardarsdottir H, Groenwold RHH. New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: a review of reporting practices. *Pharmacoepidemiol Drug Saf*. 2021;30:960–74.
 41. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol reports*. 2015;2:221–8.
 42. Maringe C, Majano SB, Exarchakou A, Smith M, Rachet B, Belot A, et al. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. *Int J Epidemiol*. 2020;49:1719–29.
 43. Williamson T, Ravani P. Marginal structural models in clinical research: when and how to use them? *Nephrol Dial Transplant*. 2017;32:ii84-90.
 44. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc*. 2011;96:440–8.
 45. Onozawa S, Kimura T, Ito Y, Akizawa T. Estimating the causal effect of transient anemia status on renal and cardiovascular outcomes in community-dwelling patients in Japan at the beginning of impaired renal function using marginal structural modeling. *Clin Exp Nephrol*. 2022;26:178–89.
 46. Bica I, Alaa AM, Lambert C, van der Schaar M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin Pharmacol Ther*. 2021;109:87–100.
 47. Zhang L, Wang Y, Ostropolets A, Mulgrave JJ, Blei DM, Hripcsak G, et al. The medical deconfounder: assessing treatment effects with electronic health records. *Proc Mach Learn Res*. 2019;106:490–512.
 48. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol*. 2009;38:1599–611.
 49. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol*. 2017;185:65–73.
 50. 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 Crohn's Colitis*. 2019;2020:617–23.
 51. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370–6.
 52. Bakoyannis G, Yiannoutsos CT. Impact of and correction for outcome misclassification in cumulative incidence estimation. *PLoS ONE*. 2015;10: e0137454.
 53. Marshall R. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2006;35:1588–9 (author reply 9-90).
 54. Pharmaceuticals and Medical Devices Agency. Basic approach to implementing validation of outcome definitions used in post-marketing database surveys [in Japanese]. 2019.
 55. Nishioka Y, Takeshita S, Kubo S, Myojin T, Noda T, Okada S, et al. Appropriate definition of diabetes using an administrative database: a cross-sectional cohort validation study. *J Diabetes Investig*. 2022;13:249–55.
 56. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15:291–303.
 57. Zhang X, Stamey JD, Mathur MB. Assessing the impact of unmeasured confounders for credible and reliable real-world evidence. *Pharmacoepidemiol Drug Saf*. 2020;29:1219–27.
 58. Van Der Weele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-Value. *Ann Intern Med*. 2017;167:268–74.
 59. Takeuchi M, Ogura M, Minoura T, Inagaki N, Kawakami K. Comparative effectiveness of sodium-glucose cotransporter-2 inhibitors versus other classes of glucose-lowering medications on renal outcome in type 2 diabetes. *Mayo Clin Proc*. 2020;95:265–73.
 60. Huang L, Ma J, Qiu X, Suo T. Assess the application of the e-value in the unmeasured confounder evaluation of observational pharmaceutical studies. *Sci Program*. 2021;2021:1–10.
 61. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med*. 2014;33:2297–340.
 62. Abadie A. Semiparametric difference-in-differences estimators. *Rev Econ Stud*. 2005;72:1–19.
 63. Abrahamowicz M, Bjerre LM, Beauchamp M-E, Leloir J, Burne R. The missing cause approach to unmeasured confounding in pharmacoepidemiology. *Stat Med*. 2016;35:1001–16.
 64. Ji X, Small DS, Leonard CE, Hennessy S. The trend-in-trend research design for causal inference. *Epidemiology*. 2017;28:529–36.
 65. Lee WC. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach. *BMC Med Res Methodol*. 2014;14:18.
 66. Koizumi M, Ishimaru M, Matsui H, Fushimi K, Yamasoba T, Yasunaga H. Tranexamic acid and post-tonsillectomy hemorrhage: propensity score and instrumental variable analyses. *Eur Arch Otorhino-Laryngol*. 2019;276:249–54.
 67. Aso S, Matsui H, Fushimi K, Yasunaga H. Dexmedetomidine and mortality from sepsis requiring mechanical ventilation: a Japanese nationwide retrospective cohort study. *J Intensive Care Med*. 2021;36:1036–43.
 68. Kaneko M, Shimizu S, Oishi A, Fushimi K. Impact of COVID-19 infection rates on admissions for ambulatory care sensitive conditions: nationwide difference-in-difference design in Japan. *Fam Med Community Health*. 2022;10:e001736.
 69. Liu L, Hou L, Yu Y, Liu X, Sun X, Yang F, et al. A novel method for controlling unobserved confounding using double confounders. *BMC Med Res Methodol*. 2020;20:195.
 70. Zhang X, Faries DE, Boytsov N, Stamey JD, Seaman JW Jr. A Bayesian sensitivity analysis to evaluate the impact of unmeasured confounding with external data: a real world comparative effectiveness study in osteoporosis. *Pharmacoepidemiol Drug Saf*. 2016;25:982–92.

71. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behav Res.* 1998;33:545–71.
72. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol.* 2005;162:279–89.
73. Athey S, Imbens GW, Metzger J, Munro E. Using Wasserstein Generative Adversarial Networks for the design of Monte Carlo simulations. *J Econom.* 2021. <https://doi.org/10.1016/j.jeconom.2020.09.013>
74. Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. *Comput Stat Data Anal.* 2014;72:219.