



# Current Status, Challenges, and Future Perspectives of Real-World Data and Real-World Evidence in Japan

Katsutoshi Hiramatsu<sup>1</sup> · Annabel Barrett<sup>2</sup> · Yasuhiko Miyata<sup>3</sup> on behalf of PhRMA Japan Medical Affairs Committee Working Group 1

Accepted: 27 May 2021 / Published online: 19 June 2021  
© The Author(s) 2021, corrected publication 2022

## Abstract

The aim of this article is to help develop a common understanding of the current status, challenges, and future perspectives of real-world data (RWD) and real-world evidence (RWE) in Japan. RWD and RWE are very widely used terms, but standardized definitions are lacking. Given broad and growing applications of RWD/RWE from the perspective of clinical development and medical affairs, the PhRMA Japan Medical Affairs Committee Working Group 1 have proposed the following definitions: “RWD are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources” and “RWE is the evidence derived from analysis of RWD.” The key challenges for RWD and RWE in Japan include restricted access and linkage of RWD, as well as a lack of universally accepted methodological approaches, which reduces the potential for patient and healthcare benefits. These challenges for RWD/RWE are by no means unique to Japan and similar challenges exist for countries in Europe and the USA. The quality of data and analysis, study design, and the transparency of reporting should be discussed more to ensure credibility and acceptance by decision-makers as the demand for RWD and RWE increases. The future developments around Japan’s RWD and RWE are expected to include improved RWD access, data linkage, and increased acceptance by decision-makers, all supported by innovative technology. Improvements in RWD access and database linkage will enable both public and private sectors to assemble more comprehensive health information in Japan.

**Digital Features** for this article can be found at <https://doi.org/10.6084/m9.figshare.14531580>.

✉ Katsutoshi Hiramatsu  
khiramat@amgen.com

<sup>1</sup> Amgen K.K., Midtown Tower 9-7-1 Akasaka, Minato-ku, Tokyo 107-6239, Japan

<sup>2</sup> Eli Lilly and Company Limited, Berkshire, UK

<sup>3</sup> Bristol Myers Squibb K.K., Tokyo, Japan

## Key Points

As definitions of real-world data (RWD) and real-world evidence (RWE) are not standardized, the PhRMA Japan Medical Affairs Committee Working Group 1 reviewed these definitions from the perspectives of clinical development and medical affairs in Japan and propose that RWD is defined as “data relating to patient health status and/or delivery of health care routinely collected from a variety of sources” and RWE as “evidence derived from analysis of RWD.”

In Japan, challenges exist around access and linkage of RWD, as well as a lack of universally accepted methodological approaches, which reduces the potential for patient and healthcare benefits.

Improvements in RWD access and database linkage will enable both public and private sectors to assemble more comprehensive health information in Japan.

## 1 Introduction

Today, the terms real-world data (RWD) and real-world evidence (RWE) are very widely used in the medical industry, but they have a relatively short history of usage. Neither of the definitions of RWD and RWE are standardized, and the definitions and interpretations vary among agencies, organizations and individuals. Therefore, it is necessary to clarify the definition for the sake of communication.

The main reason for the existence of various definitions is thought to be related to the differences in the intended purposes of use and the viewpoints from which they are defined. Thus, Pharmaceutical Research and Manufacturers of America (PhRMA) Japan Medical Affairs Committee Working Group 1 (MAC WG1) organized information on what RWD and RWE are, considered definitions of RWD and RWE from the perspectives of clinical development and medical affairs, and proposed common definitions from the perspective of PhRMA Japan members. Having considered the definition, we decided to adopt the stance of a broad definition as this is applicable to a greater variety of purposes.

Clinical studies using RWD have been increasing in quantity worldwide, and the number of high-quality observational studies using large-scale RWD has also been increasing rapidly. The results of these studies are being used for a wider variety of purposes. Japan has lagged far behind the USA and Europe in not only randomized controlled trials (RCTs) but also clinical studies using RWD [1–5]. In recent years,

however, the number of observational studies from Japanese academia has increased significantly [3–5]. In addition, there have been a number of governance changes, the Good Post-marketing Study Practice (GPSP) was revised in 2018 to include a database study as a type of post-marketing study [6], at the same time the Clinical Trials Act was implemented [7], which tightened the standards for the conduct of interventional studies. In May of the same year, the Next-generation Healthcare Infrastructure Act (NHIA) was enforced [8], and it became possible for medical institutions to be able to provide personal medical information to certified business operators by opt-out. Since then, certified business operators have become able to anonymize medical data and provide it to users, thereby promoting the use of RWD. As a result of these changes, the importance of observational studies is rapidly increasing among Japanese pharmaceutical companies, and companies are increasing the number of personnel involved in observational research and establishing a system for conducting such research. Observational research is a rapidly developing and evolving field in Japan.

When a pharmaceutical company conducts observational research in Japan, there can be challenges, such as (1) limited access to data; (2) difficulty in linking databases; (3) poor data quality, including missing data; (4) unclear guidance on the acceptability of RWD/RWE by regulators; (5) lack of decision criteria, standards, and guidelines for RWE development; and (6) items to be reported necessary for scientific decision-making and validation in a database study have not been sufficiently established. With regard to (4)–(6), the regulatory authorities have been making progress to a certain extent with the revision of the GPSP, notification of points to keep in mind to ensure the reliability of post-marketing database surveys [9], notification of the fundamental concept of outcome definition used in post-marketing database studies [10], setting up a consultation system for the use of registries for regulatory applications [11], and two notifications, issued by the Ministry of Health, Labour and Welfare (MHLW) in March 2021, on the “Basic Principles on the Use of Registries in Approval Applications” [12] and “Points to Consider for Ensuring the Reliability When Using Registry Data for Approval Applications” [13]. Progress has been made toward the utilization of RWD in some cases, but as yet it is insufficient. We have arranged the information under the present circumstances to discuss these issues.

Efforts to make more effective and active use of RWD are also being pursued. RWD can be integrated by a variety of efforts to create even more data, and with the progress of technology, the time is approaching when analysis results can be obtained quickly and multiple studies can be conducted simultaneously. One such initiative is introduced in this article, and future prospects are described at the end. We hope that this article will help

to develop a common understanding of the current status, challenges, and future perspectives of the RWD and RWE, and contribute to future activities.

## 2 What are Real-World Data (RWD) and Real-World Evidence (RWE)?

The purpose of this section is to organize and review reports and information on RWD and RWE and provide definitions for RWD and RWE from the perspective of PhRMA Japan members. For the literature review, PubMed and Ichushi-Web (a Japanese medical literature database) were used to search scientific literature from 1 January 2010 to 31 December 2019 (date of search). The search strategy used is presented in Online Supplementary Material (OSM), Fig. 1. The definitions of RWD were assessed from the perspective of what is included and what is not included. The definitions of RWE were examined to determine whether they included a scientific approach, and if so, what words were used to represent it.

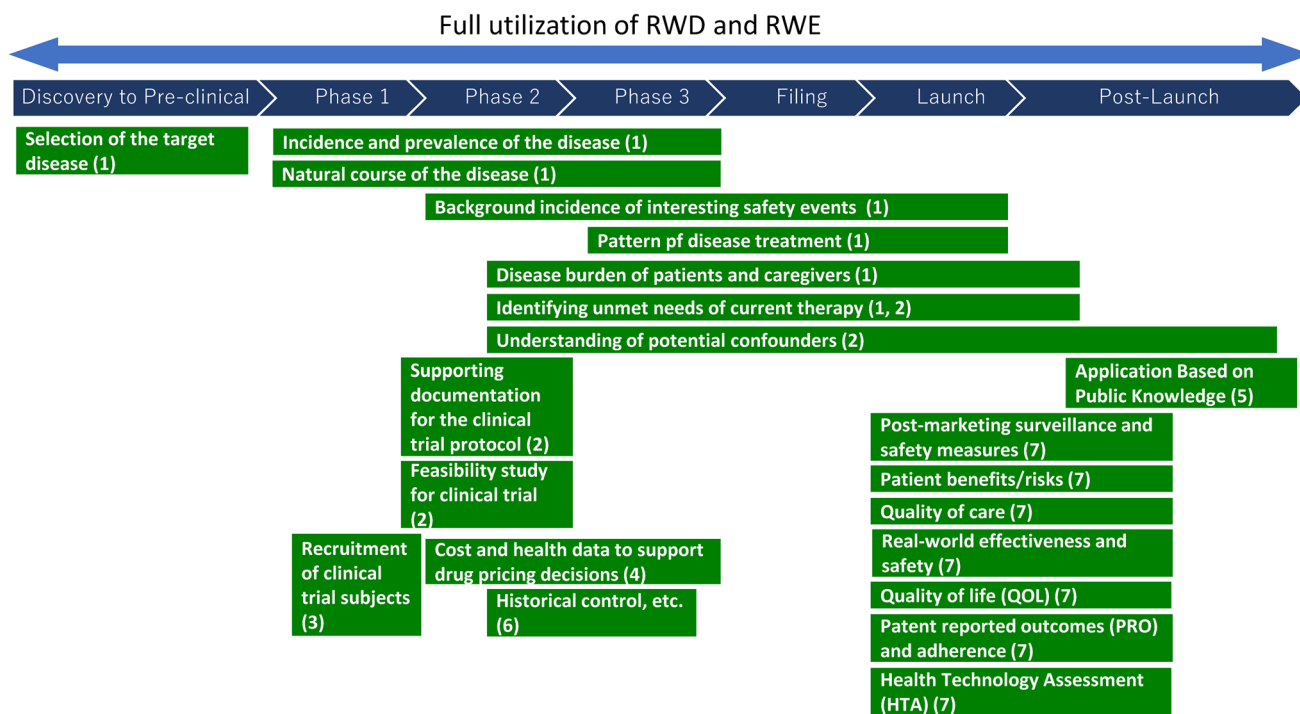
### 2.1 Perspective from Clinical Development (Pre-Launch Activities)

Purpose of use is one of the most important points when considering the definitions of RWD and RWE from the perspective of clinical development. The work of clinical development and medical affairs departments at various pharmaceutical companies can differ or overlap. Therefore, to make the following discussions easier to understand, clinical development is considered to be limited to pre-launch activities and medical affairs are limited to post-launch activities.

From the perspective of clinical development (pre-launch activities), RWD is utilized mainly for supporting evidence generation and as part of application data under an appropriate research plan/protocol (Fig. 1). Conventional application examples of RWD/RWE are shown in Table 1 [14]. However, some of these activities could be conducted as post-launch activities.

In addition to the examples in Table 1, the following examples have been initiated for the acquisition of new/additional indications:

1. In the development of pharmaceuticals and medical devices, clinical trials for rare diseases have been



**Fig. 1** Real-world data and real-world evidence utilization by pharmaceutical companies. (1) for development strategy; (2) for clinical trial design; (3) for promotion of enrolment of study participants; (4) for drug price calculation; (5) for expansion of indications; (6) for

new or additional indications; (7) for identification of unmet medical needs, closing data gaps, and informing clinical practice. *RWD* real-world data, *RWE* real-world evidence

**Table 1** Conventional application examples of real-world data/real-world evidence from the perspective of clinical development in Japan

Categories	Purpose of use
Development strategy	<ol style="list-style-type: none"> <li>1. Selection of the target disease</li> <li>2. Incidence and prevalence of disease</li> <li>3. Natural course of the disease</li> <li>4. Background incidence of interesting safety events</li> <li>5. Pattern of disease treatment</li> <li>6. Disease burden of patients and caregivers</li> <li>7. Identifying unmet needs of current therapy</li> </ol>
Clinical trial design	<ol style="list-style-type: none"> <li>1. Identifying unmet needs of current therapy</li> <li>2. Understanding of potential confounders</li> <li>3. Supporting documentation for the clinical trial protocol</li> <li>4. Feasibility study</li> </ol>
Promotion of enrolment of study subjects	<ol style="list-style-type: none"> <li>1. Recruitment of the study subjects to clinical trials</li> </ol>
Application dossier	<ol style="list-style-type: none"> <li>1. Historical control data [14, 96]</li> <li>2. Supplementary materials</li> </ol>
Drug price calculation	<ol style="list-style-type: none"> <li>1. Cost and health data to support drug pricing decisions</li> </ol>
Expansion of indications	<ol style="list-style-type: none"> <li>1. Application based on public knowledge</li> </ol>

conducted as single-arm studies only for the treatment group, and RWD has been used as a historical control.

2. Active use of disease registries in the development of medical devices is already recommended and implemented, and disease registries are actively used in drug development.
3. RWD/RWE is used for efficient development and expansion of indications in the development of drugs and medical devices with high medical needs.

## 2.2 Perspective from Medical Affairs (Post-Launch Activities)

From the medical affairs perspective (post-launch activities), RWD/RWE is primarily used to identify unmet medical needs, close data gaps and inform clinical practice (Fig. 1). Evidence is generated under an appropriate research plan/protocol. Unmet medical need assessment can concern questions around epidemiology, shortcomings in diagnosis, current standard of care, and remaining treatment gaps—such as adherence, administration difficulties, real-world effectiveness, and safety. Some of these activities may be conducted by departments other than Medical Affairs (e.g., Pharmacovigilance, Health Outcomes) and as pre-launch activities. Examples of activities are:

- Post-marketing surveillance and safety measures;
- Patient benefits/risks;
- Quality of care;
- Real-world effectiveness and safety;
- Quality of life (QOL);
- Patient reported outcomes (PRO) and adherence; and
- Health technology assessment (HTA).

As noted in the previous section, some of the activities listed may be conducted in the post-launch setting.

## 2.3 Proposed Definition for RWD

The US Food and Drug Administration (FDA) defines RWD as follows: "Real-world data are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources" (Table 2) [15]. Examples of RWD definitions by other organizations are listed for reference in Tables 2 and 3 [16–20]. As discussed below in section 2.6, the FDA states that the distinction should not be made between RWE and non-RWE based on the presence or absence of a planned intervention or the use of randomization [21]. This is the most widely held view and we support it. The context of this definition of RWD assumes that:

- RWD is used for a variety of purposes in addition to regulatory decision-making;
- There are many different types of eligible data; and
- Study designs other than conventional RCTs can collect RWD.

Based on the use of RWD/RWE described in Sects. 2.1 and 2.2. above, it is considered that the definition of RWD matches that of the FDA, and we propose using the same definition of RWD as the FDA. Strictly speaking, the data handled are not necessarily limited to patients, and may include subjects at a previous stage of disease, such as mild cognitive impairment, or potential patients who are candidates for prophylactic treatment. However, while many pharmaceutical companies uphold a patient-centric approach, patients in this case are defined in a broad

**Table 2** Definitions of real-world data by FDA, EMA, Japanese Health Science Council, JPMA Ethical Drug Product Information Summary Review Committee, and joint ISPOR-ISPE special task force

Organization	RWD definition
FDA [15]	Real-world <b>data</b> are the data relating to patient health status and/or the delivery of the healthcare routinely collected from a variety of sources RWD can come from a number of sources, for example: Electronic health records (EHRs) Claims and billing activities Product and disease registries Patient-generated data including in home-use settings Data gathered from other sources that can inform on health status, such as mobile devices
EMA [17]	RWD are defined as “routinely collected data relating to a patient’s health status or the delivery of healthcare from a variety of sources other than traditional clinical trials” Data from traditional clinical trials are especially excluded, even single-arm trials, but would incorporate data from pragmatic clinical trials if data were collected remotely through an electronic health record or other observational data source and solely under conditions of normal care
Japanese Health Science Council [18]	Variety of electronic data collected in a clinical setting that can be evaluated for safety and efficacy
JPMA Ethical Drug Product Information Summary Review Committee [19]	Includes data based on actual practice (e.g., Receipt data, DPC data, medical records, registry), data from pragmatic trials and non-interventional studies that resemble the conditions of actual practice Data may reflect some or a significant portion of the “as it is non-intervention” decisions and behaviors of practitioners and patients in clinical practice
Joint ISPOR-ISPE special task force [20]	Data obtained outside the context of RCTs generated during routine clinical practice

*DPC* Diagnosis Procedure Combination, *EMA* European Medicines Agency, *FDA* US Food and Drug Administration, *ISPE* International Society for Pharmacoepidemiology, *ISPOR* International Society for Pharmacoeconomics and Outcomes Research, *JPMA* Japan Pharmaceutical Manufacturers Association, *RCT* randomized controlled trial, *RWD* real-world data

**Table 3** Definition of real-world data by ISPOR, ABPI, RAND, IMI-GetReal [16]. Adapted with permission from Value Health 2017; 20(7):858-65. Copyright© 2017 International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. All rights reserved

Organization	RWD definition
ISPOR	Data used for decision making that are not collected in conventional RCTs
ABPI	Data obtained by any non-interventional methodology that describe what is happening in normal clinical practice
RAND Corporation	Umbrella term for different types of healthcare data that are not collected in conventional RCTs Comes from various sources in healthcare sector, and includes patient data, as well as data from clinicians, hospitals and payers, and social data
IMI-GetReal	Umbrella term for data regarding the effects of health interventions (e.g., benefit, risk, and resource use) that are not collected in conventional RCTs Collected both prospectively and retrospective from observations of routine clinical practice, and include (but are not limited to) clinical and economic outcomes, PROs and health-related QOL Can be obtained from many sources including patient registries, EMRs and observational studies

*ABPI* Association of the British Pharmaceutical Industry, *EMR* electronic medical record, *ISPOR* International Society for Pharmacoeconomics and Outcomes Research, *PRO* patient-reported outcome, *QOL* quality of life, *RCT* randomized controlled trial, *RWD* real-world data

sense. A patient in our definition is also considered to be a patient with a broad definition. "Routinely collected" is a key descriptive term for the concept of the real world and, as per FDA guidance, includes data from intervention studies and randomized studies, but does not include data from conventional RCTs (double-blind comparative studies commonly conducted as a Phase 3 study).

The proposed definition of RWD by the PhRMA Japan MAC WG1 is: “RWD are the data relating to patient health status and/or the delivery of health care routinely collected

from a variety of sources” (i.e., the same as the FDA’s definition).

Sources of RWD are described in detail in later sections, with the definition limited to the above description.

## 2.4 Points to Consider When Handling RWD

RWD in Japan must be handled under the legal framework for using personal information, which includes the Act on the Protection of Personal Information (APPI) [22] and



related government guidelines. The legislation and related governance arrangements cover two important aspects. The first aspect is that, under the revised APPI [22], medical data are defined as "Special Care-Required Personal Information" and opt-in is required for use in research. That is, the informed consent of the patient must be obtained to collect and use the data. However, such consent acquisition is often not performed in routine medical care in Japan, and the access of private companies (e.g., pharmaceutical companies) is limited mainly to access to commercial data consisting of anonymized patient data. In order to improve this situation, it is expected that it will become mandatory to explain to patients beforehand about the collection and use of data collected on a routine basis and obtain their informed consent, and that improvements will be made so that it becomes common sense for most Japanese to cooperate with research. Currently, when academic researchers or academic societies handle medical data for the purpose of academic research, it is not applicable under Article 76 (1) of the APPI [22], and obtaining prior patient consent is not essential.

The second aspect is the anonymization of RWD. That is, under the NHIA, medical institutions can provide business operators certified by the competent ministries and agencies (e.g., Cabinet Office; Ministry of Education, Culture, Sports, Science and Technology; MHLW; and Ministry of Economic, Trade and Industry) with medical data without patients opting-in, if they used the "opt-out" procedure. This allows certified business operators to collect individual patient data directly from healthcare providers and link all the data at an individual level. This law has the potential to dramatically boost research using RWD. In December 2019, the Life Data Initiative (LDI; as a "Certified Anonymizing Medical Data Producer") and NTT DATA Corporation (as a "Certified Business Operator Handling Medical Data") were

certified for the first time, operating since January 2020 [23]. More recently, in June 2020, the Japan Medical Association Medical Information Management Organization (J-MIMO) was designated a "Certified Anonymizing Medical Data Producer," and Integrated Clinical Care Informatics, Inc. (ICI) and NS Solutions Corporation, a "Certified Business Operator Handling Medical Data," respectively [24]. However, it should be noted that no major commercial database vendors, such as the Japan Medical Data Center Inc. (JMDC) or Medical Data Vision Corporation, Ltd (MDV), have been certified at this time. MDV collect anonymized medical data from medical institutions, but the government does not evaluate whether the anonymized medical data meets their standards. There is a potential risk that these commercial databases may be judged inappropriate for use. Hopefully, more certified business operators will operate soon and the environment handling RWD will be improved.

## 2.5 Available RWD Sources

In order to accurately define or correctly understand RWD, it is essential to consider the data sources available and the study designs that collect RWD. The FDA illustrated some of the data sources in their definition of RWD (Table 2) [15]. However, study designs are not referred to in the FDA's definition of RWD, but are mentioned in their definition of RWE (Table 4) [15]. Here, we first consider the data sources for RWD, and the study designs that collect RWD are discussed in section 2.6.

There are various data source classification methods, and new data sources will be created in the future. It is difficult to list all RWD sources, but "records of health, medical data and personal information (e.g., occupation, annual income) obtained from procedures that may occur in daily life" is

**Table 4** Definitions of real-world evidence by FDA, EMA, Health Science Council, JPMA Ethical Drug Product Information Summary Review Committee, and joint ISPOR-ISPE special task force

Organization	RWE Definition
FDA [15]	Real-world <i>evidence</i> is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective)
EMA [17]	RWE is defined as the information derived from analysis of RWD
JPMA Ethical Drug Product Information Summary Review Committee [19]	Derived from RWD; since RWE may be subject to various biases, careful handling is required when using it in advertising materials
Joint ISPOR-ISPE special task force [20]	Obtained from analyzing RWD

EMA European Medicines Agency, FDA US Food and Drug Administration, ISPE International Society for Pharmacoepidemiology, ISPOR International Society for Pharmacoeconomics and Outcomes Research, JPMA Japan Pharmaceutical Manufacturers Association, RWD real-world data, RWE real-world evidence

considered to be applicable to most. It should be noted that possible procedures in daily life may include intervention or randomization, as stated by the FDA (see Sect. 2.6).

Examples of RWD data sources are provided in Table 5. Primary data are data collected for study purposes and secondary data are data collected for purposes other than research. Hybrid data includes both primary and secondary data.

### 2.6 Study Designs that Can Collect RWD

The results of a literature review and a series of stakeholder interviews reported on the research designs that could collect RWD. Fifty-three percent of respondents stated that research designs other than RCTs would collect RWD [16]. Note that responses saying research designs other than RCTs included the responses saying *other than conventional RCTs* (this term is considered to refer to the double-blind comparative study commonly conducted in Phase 3).

The FDA refers to study designs as described in an article by Sherman and others [21] and in their published definition of RWE (Table 4) [15]. The FDA’s view is summarized below:

- The distinction should not be based on the presence or absence of a planned intervention or the use of randomization (the most important matter).
- Real-world research and the concepts of a planned intervention and randomization are entirely compatible.

- Randomization other than conventional RCTs (double-blind comparative study) is a study design that collects RWD.
- Pragmatic clinical trials and large simple trials collect effectiveness data, not efficacy data.
- Pragmatic clinical trials are encouraged to be designed to be flexible.
- Large simple trials are encouraged to broaden the selection criteria and, if possible, to use claims data.

As mentioned above, we believe that "study designs other than conventional RCTs can collect RWD" and the following are examples of research designs that can collect RWD:

- Large simple trials;
- Pragmatic clinical trials;
- Observational studies; and
- Supplements to RCTs.

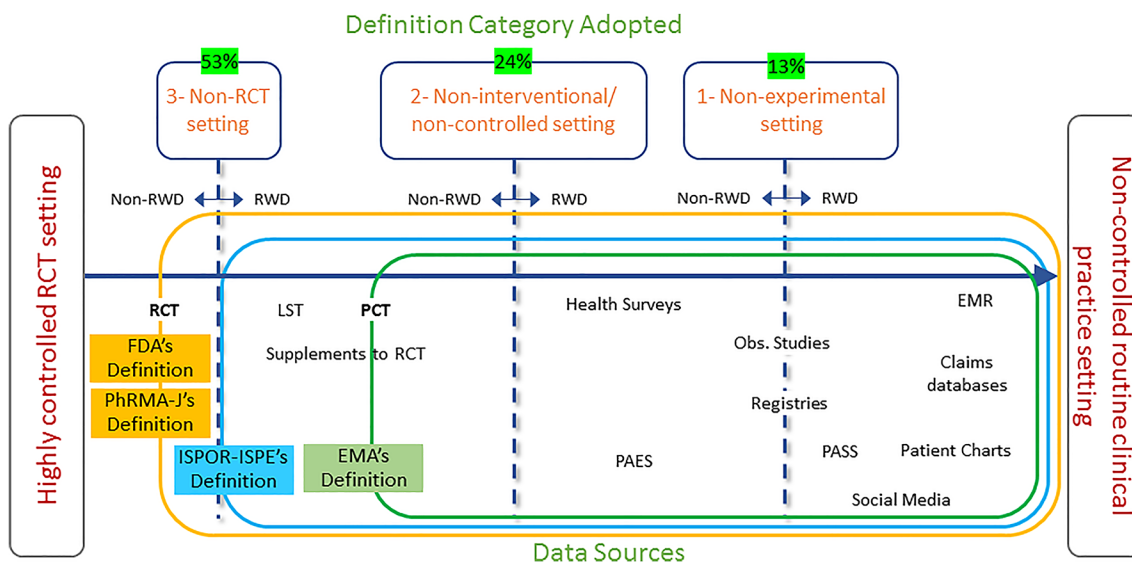
However, there is still controversy as to whether or not RWD/RWE is generated when considering individual examples other than RCTs. For example, data from pragmatic RCTs are considered to be RWD/RWE, while data from conventional RCTs are not. Therefore, we did not refer to the exemplified study designs as designs that collect RWD, but rather as designs that *can* collect RWD.

Figure 2 is very useful in understanding the differences between the definitions of the current RWD, and shows the difference in the data sources handled.

**Table 5** Data sources for real-world data

Categories	Examples
Main data sources	EMRs Insurance claims and payments Drug registries, disease registries Data generated by patients, including household use Data collected from other sources that report health conditions, such as mobile devices Health survey Medical institutions and clinicians Social media Pharmacy data Clinical database
Primary data	Healthcare services and clinical trials Data obtained from within medical institutions (e.g., clinical observations; procedures, tests, and prescribed drugs; and interviews, surveys, and QOL assessments) Data obtained outside medical institutions (e.g., community interviews or surveys; wearable devices; home measurement equipment for blood pressure, blood sugar, and blood oxygen; and patient-reported outcomes)
Secondary data	Medical records (paper, electronic) during individual patient care Medical care fee claims involving pecuniary exchange Data collected by the government on a daily basis (e.g., birth, death, marriage, divorce, and address details [in some countries]; vaccination records; results of disease screening; cancer registry; occupational health; and other)
Hybrid data	Data containing both primary and secondary data

EMR electronic medical record, QOL quality of life



**Fig. 2** Real-world data definitions and data sources. Adapted from Makady et al. [16] with permission from Value Health 2017; 20(7):858-65. Copyright© 2017 International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. All rights reserved. *EMR* electronic medical record, *LST* large simple trial, *Obs.* Observational, *PAES* post-authorization efficacy

study, *PASS* post-authorization safety studies, *PCT* pragmatic clinical trial, *RCT* randomized clinical trial, *RWD* real-world data, *EMA* Europe Medical Agency, *FDA* US Food and Drug Administration, *ISPOR* the International Society of Pharmacoeconomics and Outcomes Research, *ISPE* the International Society of Pharmacoepidemiology

## 2.7 Proposed Definition for RWE

The FDA defines RWE as “clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD” (Table 4) [15]. Definitions of RWE by other organizations are also listed in Table 4 [15, 17–20]. RWE is not merely a result of the accumulation of RWD, but rather the result of scientific assessment under an appropriate research plan. The FDA and the joint special task force between the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the International Society for Pharmacoepidemiology (ISPE) both use the term analysis to describe a scientific assessment. On the other hand, as described in Sects. 2.1 and 2.2, our activities are broader than those of the FDA, and we often analyze RWD for the purposes of HTA, disease burden of patients and caregivers, selection of the target disease, natural course of the disease, incidence and prevalence of the disease, background incidence of interesting safety events, and recruitment of clinical trial subjects. We consider these to be RWE, and propose a definition of RWE as follows:

The definition of RWE by PhRMA Japan MAC WG1: “RWE is the evidence derived from analysis of RWD.”

Clarifying the definition of RWE allows us to clearly set out the scope of RWE in our activities and avoid miscommunication.

## 2.8 Points to Consider When Developing RWE

Analysis for the purpose of RWE is challenging. Whilst randomization in RCTs can reduce group bias, including unobserved factors, many studies using RWD cannot control for bias due to the study design. Therefore, it is important to use statistical methods to adjust for confounding factors after data collection is complete. Although there are some statistical methodologies for dealing with bias (e.g., propensity-score matching, instrumental variables), they cannot control for unobserved confounding factors or make very strong assumptions. Thus, when interpreting RWE, bias, confounding factors and study limitations should be carefully considered. When reporting results, either in a study report or a published manuscript, it is necessary to clearly state how potential confounding factors and biases were considered and addressed and describe the study limitations. When RWE is described as a material for healthcare professionals, it is essential that the study limitations are clearly specified.

## 2.9 Databases in Japan

Details on Japanese databases are described in the report “Fiscal year 2018, Research and Survey of Drug Discovery Needs toward the Creation of Innovative Therapeutic Drugs” by the Japan Health Science Foundation [18], the



Office of Health Economics (OHE) report (Data Governance Arrangements for Real-World Evidence in Japan) [25], and the “Database Survey Applicable to Clinical Epidemiology and Pharmacoepidemiology in Japan” by the Pharmacoepidemiology and Database Task Force of the Japan Society of Pharmacoepidemiology [26], which are very helpful.

The following sections describe the major databases in Japan, broken down by major data source types.

### 2.9.1 Registries

A registry, sometimes called a “disease registry,” “patient registry,” or “disease registry system,” is a system that collects and registers detailed data about patients with a particular disease from many medical institutes. There are two types: the first in which a research question is determined in advance and necessary data is collected according to a research plan; and the second in which a research question is not determined in advance and data are widely collected, set later, and necessary data are extracted and used. Participation is obligatory for some registries and voluntary for others. The advantage of a registry is that it often collects disease-specific clinical and outcome information.

There is a Clinical Innovation Network (CIN) project [27] led by the MHLW and promoted by the Japan Agency for Medical Research and Development (AMED). The CIN project aims to revitalize the clinical development of new drugs and medical devices in Japan by utilizing patient registries and cohort studies for clinical development.

The main objectives of the CIN are as follows:

- To establish CIN promotion centers to promote the use of patient registries and support efficient clinical development of pharmaceuticals and medical devices in Japan;

- To collect information on patient registries in Japan, build and publish a search system for such information, and operate it continuously; and
- To provide information and consultation to researchers, companies, and patients concerning the construction and utilization of patient registries.

In the past, each registry was operated by a different organization and information was not centralized, making it difficult for researchers and companies to use it for clinical development. However, a major achievement of the CIN is the centralization of information and the development of a registry search system. The registry search system has a rich search function, and approximately 500 patient registries in Japan are available [28]. In addition to free-text searching, searches can be carried out by ICD-10 classification of the target disease or disease region, the presence or absence of image data, genome data, and omics data.

The CIN aims to improve the efficiency of medical research and development. The intended use of the registries in the CIN are summarized in Table 6 [29]. The key disease registries in Japan are described below:

#### 1. National Cancer Registry

This is a mandated registry under the “The Act on Promotion of Cancer Registry” [30]. The National Cancer Registry is a system in which all patients diagnosed with cancer are registered, and the National Cancer Center tabulates, analyzes, and manages the data [31]. Information on more than 850,000 cases is collected annually. Data on the prevalence, mortality, survival, and clinical practice of cancer necessary for planning and evaluation of cancer countermeasures in Japan have been aggregated and analyzed. The results of aggregation and analysis are published on the National Cancer Center website and the Cancer Statistics Digest (cancer statistics digest) of the *Japanese Journal of Clinical Oncology* (Journal of the Japanese Society of Clinical Oncology).

**Table 6** The intended use of registries in Clinical Innovation Network [29]

Purpose	Contents
1 Market research	Considering the possibility of developing products based on the number of patients and regional distribution
2 Development of clinical trial plan	Develop a highly accurate clinical trial plan based on the number of patients expected to participate in the clinical trial and their natural history
3 Investigating the feasibility of clinical trials	Study highly accurate feasibility based on the number of clinical trial candidates and the expected number of clinical trial candidates
4 Clinical trial recruitment	Effectively incorporate patients from patient panels into clinical trials and studies as candidates
5 Control group data for clinical trials	Understand the natural history of the patient and use it as control group data for clinical trials (at the time of additional indications)
6 Intervention group data	Data from off-label use as intervention group data in regulatory applications
7 Post-marketing surveys	Grasp of the occurrence of adverse drug reactions
8 Study of clinical trial optimization	Optimization of the use of drugs and medical devices in clinical practice, measurement of medical economic effects, and evaluation of clinical trial costs

## 2. National Clinical Database (NCD)

The Japan Surgical Society's NCD was created in cooperation with ten surgical societies, mainly the Japanese Surgical Society, in order to understand the current state of surgical care in Japan [32–34]. Over 1.5 million cases are entered annually, covering more than 95% of the operations performed by general surgeons in Japan. It has also become possible to evaluate surgical risks based on big data, and a system that displays surgical risks instantly by inputting patient information and surgical procedures on the internet (risk calculator) has started operation. It is also possible to compare the results of each institution with the national average, which has helped improve medical care at individual institutions. In addition to surgery, it also functions as cancer data registry, and national data on breast cancer, pancreatic cancer and liver cancer are being collected.

## 3. Japan CardioVascular Surgery Database (JCVSD)

The JCVSD is divided into adult and congenital sections. The purpose of the JCVSD is to conduct nationwide research on the types and risks of cardiovascular surgery performed in Japan, and the degree of difference in the risks of surgery between patients with good preoperative conditions and those with severe conditions [35, 36].

## 4. J-DREAMS

The Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System (J-DREAMS) has constructed a database of diabetes patients in Japan [37]. J-DREAMS is studying ways to improve the quality of diabetes treatment in Japan by grasping the actual conditions of diabetes treatment and investigating what kind of patients are prevalent and what kind of treatment is best for which patients [38]. The system is operated jointly by the National Center for Global Health and Medicine and the Japan Diabetes Society.

## 5. JROAD, JROAD-DPC

The Japanese Registry Of All Cardiac and Vascular Diseases (JROAD) is a disease registry whose objective is to conduct a survey on the actual conditions of cardiovascular treatment in Japan and to prepare basic data for improving the quality of cardiovascular treatment based on the obtained data [39, 40]. As one of its projects, JROAD-DPC, a database of medical care from admission to discharge, is being constructed for cardiovascular diseases (e.g., acute coronary syndromes, cardiac failure) in diagnosis procedure combination (DPC) participating facilities [41]. Over a 4-year period, a total of 3.6 million data items were collected, including 160,000 myocardial infarction cases and approximately 500,000 heart failure cases. The secretariat is established in the National Cerebral and Cardiovascular Center and it is operated by the Japanese Circulation Society (JCS). The limitations of JROAD-DPC include: that readmission due to hospitalization in the same hospital is traceable, but readmission due

to hospitalization in other facilities is difficult; that there is a dissociation between the general clinical classification of ST-segment elevation myocardial infarction and the classification possible by DPC; and that information on the index of severity is scarce except for myocardial infarction (Killip classification) and heart failure (New York Heart Association classification). Currently, only members of the JCS have access to the data, but the JCS have begun to consider allowing companies to access it.

## 6. Japan Stroke Data Bank

The Japan Stroke Data Bank is a database designed to understand the actual state of stroke care in Japan [42, 43]. More than 150,000 cases have been accumulated and it is operated by the National Cardiovascular Center.

## 7. J-CKD-Database

The objective of the J-CKD-Database is to build a comprehensive nationwide chronic kidney disease clinical efficacy database that enables longitudinal studies, such as prognostic surveys [44, 45]. It is operated by the Japanese Society of Nephrology.

## 8. National Database of Rheumatic Diseases in Japan (NinJa), NinJa-BioBank

NinJa is a database established to investigate the degree of improvement in the condition and physical function of patients with rheumatoid arthritis in Japan through drug treatments and orthopedic interventions, and the occurrence of various adverse events; data on more than 15,000 patients have been accumulated. NinJa is operated by Sagami National Hospital [46]. The NinJa-BioBank conducts translational research using synovial membrane and bone marrow blood, while expanding the research system using tissues that can be collected at the time of joint surgery for rheumatoid arthritis patients, with the aim of clarifying the pathophysiology of rheumatism, which remains even after drug treatment. It is implemented as a joint project mainly by the National Hospital Organization (NHO).

## 9. SCRUM-Japan

The SCRUM-Japan project studies genetic changes in cancer in order to deliver optimal therapeutic drugs to patients [47, 48]. Using the latest high-quality genetic panel tests, researchers simultaneously study multiple cancer-causing genetic changes, including rare changes, to determine which treatment is best for each patient and which new drug trials can be registered. Two projects are under way: LC-SCRUM-Asia (formerly LC-SCRUM-Japan) for patients with lung cancer and MONSTAR-SCREEN (formerly GI-SCREEN-Japan) for patients with a wide range of solid tumors. It is operated by the National Cancer Center East Hospital. As of 2018, collaborative research has been conducted with 17 pharmaceutical companies, with a further three companies joining in the last 2 years (OSM, Fig. 2) [49]. Research representatives have started discussions with the Pharmaceuticals and Medical Devices Agency (PMDA)

and MHLW on the criteria required for the approval of new drugs.

#### 10. Japan Trauma Data Bank

The Japan Trauma Data Bank collects and analyzes trauma data to improve the quality of trauma care [50–52]. It is operated by the Japan Society for Acute Medicine and Japan Injury Association.

### 2.9.2 Insurance-Based

Administrative claims databases utilize statements sent from medical institutions to health insurance societies for billing purposes. These statements, describing procedures performed and drugs used, are referred to as “Receipts” in Japan. The advantage of health insurance claims data is high patient traceability. For medical practices that are eligible for reimbursement to an individual, even if the medical institution is different, all data are aggregated into the individual’s Receipt data. Disadvantages include few clinical test results, low reliability of “death” data, and an overwhelmingly small number of elderly people. In Japan, attempts to apply health insurance claims databases to clinical research have been slow. At present, health insurance claims databases have been gradually developed, and their application in clinical research is becoming popular. The strength of Receipt data and DPC data is that they are relatively easy to integrate with data of other facilities because it is a common format nationwide, unlike EMR data.

#### 1. National Database of Health Insurance Claims and Specific Health Check-up (NDB)

The NDB is a huge database built by MHLW based on the “Act on Assurance of Medical Care for Elderly People” [53–55]. It has collected anonymized Receipt information and specific medical examination/health guidance information (i.e., clinical test data, interview data, and health guidance for metabolic syndrome check-ups) from all over Japan since 2009. MHLW began providing NDB data to third parties in 2013, and makes NDB data available to researchers with restrictions. To date, data disclosure to private companies has not been conducted, but it has been decided to disclose by the Japanese Government and preparations are underway to implement this.

The strengths of NDB are that the target population includes almost all people in Japan, it covers almost all insured medical care, and it collects data from insurers to ensure the traceability of patients even when patients visit different hospitals. The NDB is one of the largest and most comprehensive databases in the world, and it is characterized by a large amount of data on the elderly. On the other hand, the use of NDB data has disadvantages, such as no information on death, a strict application system, a long time required from approval of use to data acquisition, difficulty

in data handling, and limitations of the data itself (lack of information to adjust patient risk and severity).

#### 2. JMDC Claims Database (Insurance-Based)

The insurance-based JMDC Claims Database has collected Receipt data from health insurance societies since 2005, and nearly half are linked to specific health examination data [56, 57]. It is provided by JMDC for a fee and can also be used by private companies. The cumulative population is about 5.6 million (as of June 2018). A unique ID is assigned to each patient so that patients can be followed up longitudinally even if they have been transferred to another hospital or visited multiple institutions. Unlike NDB data, there is also family identification, so it is possible to link pregnant women and their children. Weaknesses are that the number of patients included is smaller and less generalizable than that of the NDB, that the data are from health insurance societies only and do not include data from the medical care system for the elderly in the latter stage of life, and that laboratory data are not available.

#### 3. Medi-Scope

Medi-Scope is a specific health examination and reception database provided for a fee by JMIRI, and is also available to private companies [53, 58]. Data have been collected since 2010 and include almost all items (except for organ transplants and comments) for all Receipt types (including dental Receipts). The cumulative patient number is about 6.66 million (as of 2018). Similar to the JMDC claims database, a unique identifier is assigned to each patient, and patients can be followed longitudinally even if they have been transferred to another hospital or visited multiple institutions. A distinct advantage from the JMDC database is the abundance of data items. The downside, like the JMDC, is that the database is relatively small, and since there are only health insurance union data, it does not include data from the medical care system for the elderly.

### 2.9.3 Hospital-Based

Hospital-based databases integrate EMRs, Receipts, and DPC data held by medical institutions. Advantages include more detailed data, such as daily data and laboratory test results. Disadvantages are that it is difficult to integrate data from different medical institutions when patients are treated at more than one medical institution, and that follow-up is often impossible when patients are transferred.

#### 1. EBM Provider®

EBM Provider® is a medical care database provided to industries for research purposes by MDV for a fee [53, 59]. It is a database of inpatient and outpatient care for acute care institutions, and data have been collected since 2008. Currently, a database of medical information collected from more than 750 acute-phase medical institutions and a database limited to about 250 medical institutions with

permission for secondary use are being constructed. The advantages are that the actual number of patients exceeds 29 million (overlapping counts for visits to multiple medical institutions; 29.84 million as of December 2019; 10.45 million from 1 January 2018 to 31 December 2018), which is larger than that of the JMDC database, that the data of the late-stage elderly are included, and that laboratory data are available for about 10% of the subjects. The weaknesses are that the target patients are limited to data on patients with serious diseases who visit acute-phase medical institutions, that the follow-up period for patients is shorter than that of the JMDC, and that it does not identify patients and cannot track patients across hospitals, such as transfers, because the MDV collects anonymized data at each DPC hospital.

### 2. Medical Information Database Network (MID-NET)

The PMDA operates the MID-NET as part of its work on safety measures for pharmaceuticals [53, 60, 61]. The MID-NET cooperates with 23 hospitals and has collected data since 2009; it currently has information on approximately 4 million patients. Data extraction is requested from each medical institution on demand and submitted in the Standardized Structured Medical Information eXchange (SS-MIX) 2 format. The data include test value results in addition to the Receipt data. Advantages include near-real-time information availability, data quality control, and data reliability. Weaknesses are that it is small and not generalizable, making it unsuitable for research on rare diseases.

### 3. RWD database

The Health, Clinic, and Education Information Evaluation Institute (HCEI), based on a contract with local governments nationwide, has been collecting anonymized medical data since 2003 [53, 62]. The data currently consist of medical examination records for about 150,000 students from 120 local governments in nine grades (first grade elementary school to third grade junior high school), reception data for about 19 million students from 174 medical institutions nationwide, DPC data, and EMR data. Real World Data, Co., Ltd. has been commissioned by HCEI to support the construction and operation of the database [53, 63, 64]. Efforts are being made to improve the quality of medical care and public health through surveys on the actual conditions of medical care, the effects of drugs, and safety.

### 4. Medical Information Analysis Databank (MIA) and NHO Clinical Data Archives (NCDA)

The NHO has been operating the MIA since 2010, which has accumulated Receipt and DPC data from all 143 NHO hospitals and has also been operating NCDA since 2016, which are secondly available in SS-MIX 2 format [53, 65, 66]. Access to MIA and NCDA is restricted to staff of the NHO or those who conduct a collaborative study with the staff of the NHO. Approval needs to be obtained by the

NHO, before a collaborative study can be conducted with the staff of the NHO [65].

### 5. Tokushukai Medical Database

The Tokushukai group consists of about 340 medical and nursing care facilities including 70 hospitals [67] and is capable of unifying information management within the group [68]. Clinical research is being conducted using an integrated database of collected Receipt data, DPC data, and EMR information since 2009.

### 6. DPC Database

The DPC system was introduced in 2003 based on a Cabinet decision and is a comprehensive evaluation system for medical fees for acute inpatient care. DPC is a diagnosis group classification unique to Japan, and is also a calculation method of medical expenses according to classification that is defined by the patients, diagnosis, and procedure. It determines the hospitalization cost per day for each classification. The DPC Research Institute collects anonymized DPC data from medical institutions after obtaining individual informed consent from DPC hospitals nationwide, and provides the data to researchers [69]. The number of hospitals using the DPC system is 1,730 (as of April 2018), and other hospitals calculate medical expenses by the conventional piece-rate method. The data collected by the DPC Data Research Team, supported by the Health and Labour Sciences Research Grants, exceeds 7 million cases per year. Hundreds of clinical research papers utilizing the DPC database have been published. The DCP database is one of the most successful sources of RWD in Japan and has had the greatest impact on healthcare [70].

### 7. Electronic Medical Record (EMR) Database

EMR data can acquire information such as vital signs, blood test results, and image diagnosis data. Databases derived from EMRs are expected to have more useful information to adjust for confounding factors due to the structuring of text data in medical records. For example, a validation study can be conducted by comparing Receipt data (administrative claims) with EMRs as the gold standard for databases that store both EMRs and Receipt data, such as the database owned by Real World Data, Co., Ltd. In addition, although it is necessary to use the disease name with Receipt or DPC data in order to extract patients with chronic renal failure, the estimated glomerular filtration rate can be accurately defined using EMR data. For this reason, a database is being constructed in which data derived from an EMR is introduced in addition to Receipt and DPC data.

## 2.9.4 Pharmacy-Based

Pharmacy-based databases use prescribing statements (or “Prescribing Receipts”) sent to the health insurance societies. There are four types of Receipt: inpatient, outpatient, dental, and prescribing. The advantage of the Prescribing



Receipt data is that the prescription and the prescribing record can be regarded as almost identical, since a prescription in Japan is effective for only 4 days. The disadvantage is that Prescribing Receipt data do not include the results of the clinical test, the diagnosis, or the reason for the prescription.

#### 1. IQVIA National Prescription Audit (NPA) data

IQVIA NPA data is a Prescribing Receipt database managed by IQVIA Solutions Japan Co., Ltd., with data sources of approximately 9,200 dispensing pharmacies throughout Japan [53]. It covers about 18.7% of the outside prescription of Japan. Data items include drug names, quantities, dosage and administration, as well as patient age and gender.

#### 2. Japan Medical Information Research Institute (JMIRI) Pharmacy Claims Database (DB)

JMIRI Pharmacy Claims DB is a dispensing Receipt database with a source of about 2550 dispensing pharmacies from all over Japan managed by the JMIRI [53, 71]. Data items include information on drugs, quantities, dosage, and administration, as well as the name of the department to which the prescribing physician belongs and the classification by the number of beds.

### 2.9.5 Others

Personal Health Records (PHRs) include information about an individual's health, medical care, and nursing care. Currently, the government is working on the utilization of medical data, mainly PHR, with the aim of providing excellent services suited to the health condition of individuals by managing and utilizing their own information in a chronological manner over a lifetime.

In the PHR model that the Ministry of Internal Affairs and Communications has been researching since 2016, each person first acquires an application that matches his or her life stage. Applications are distributed by the affiliates and organizations involved. Through these applications, personal medical and health information is collected in chronological order with the informed consent of the user. AMED has solicited the development of PHR applications, and several PHR models have been adopted and are being implemented. In the future, such data will be used as a database.

In addition to the government-led PHR initiative, some private companies directly obtain health information from patients and potential patients and build databases for commercial use. The advantage of using these data is that it is possible to obtain information on the patient's condition other than at the time of visit and admission, and to obtain subjective outcomes from the patient.

#### 1. National Health and Wellness Survey (NHWS)

The NHWS is a patient database based on the questionnaire for general consumers over total 2 million persons of the main ten nations managed by Social Survey Research Information Co., Ltd. Data collection began in 1998 (2008

for Japanese data), and a questionnaire survey of more than 250,000 adults is conducted every year [53, 72]. Data elements include prevalence, diagnosis, and treatment of more than 165 diseases and conditions, as well as information on QOL, work productivity and activity impairment, severity and comorbidities, drugs used, and attitudes and behaviors toward medical care.

#### 2. Subject Volunteer Database

This database is managed by 3H Medi Solution Inc. and is based on information from about 750,000 members (as of February 2021) registered in their clinical trial information site, established in 2009 [73]. Data items include information on past history, current illness, drugs used, and blood sampling data.

## 2.10 Available RWD Databases in Japan

The Japan Society for Pharmacoepidemiology's "Pharmacoepidemiology and Database Task Force" has investigated the characteristics of databases applicable to clinical epidemiology and pharmaceutical epidemiology in Japan, and they are available in Japanese and English at the following links:

[http://www.jspe.jp/mt-static/FileUpload/files/JSPE\\_DB\\_TF\\_J.pdf](http://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_J.pdf) (Japanese)

[http://www.jspe.jp/mt-static/FileUpload/files/JSPE\\_DB\\_TF\\_E.pdf](http://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_E.pdf) (English)

Table 1 in the OSM summarizes the databases available in Japan, including the database name, summary of content, administrator, total number of registered patients, and data period, as reported by the Committee on Drug Epidemiology as of 25 September 2019 [53].

## 3 Challenges for RWD and RWE in Japan

The objective of this section is to outline the key challenges that RWD and RWE present for Japan and to review approaches that can address them.

### 3.1 Access to RWD in Japan

Access to RWD such as administrative claims data, national health check-up data, DPC data, and EMR data are restricted in Japan. A recent Office of Health Economics (OHE) report outlines the current state of data governance arrangements in Japan for these types of insurer/hospital-based RWD [25]. Attention should have been paid to compliance with both the APPI [22] and the "Ethical Guidelines for Medical and Health Research Involving Human Subjects (Ethical Guidelines)" [74] when trying to access RWD in Japan to conduct a study. When using routinely collected medical



data that has not been anonymized, the APPI stipulates that informed consent should be obtained from study subjects due to handling “Special Care-Required Personal Information” [22]. The Ethical Guidelines [74] provide explanations when applying the APPI [22] to clinical studies, especially observational studies (e.g., descriptions of handling of personal information in protocols and the handling of personal information acquired in association with the implementation of studies [including safety control measures]). In the “Ethical Guidelines for Human Medical Research Guidance” document [75], the term “personal information” appears more than 300 times, showing that the protection of personal information is a significant matter. The APPI [22] and Ethical Guidelines [74] have complemented each other so far.

The “Ethical Guidelines for Medical and Health Research Involving Human Subjects” [74] and the “Ethical Guidelines for Human Genome and Gene Analysis” [76] were integrated on 23 March, 2021. The reasons for this were that the procedures set forth in both guidelines have many common points, and that many studies have been conducted in compliance with both guidelines. The integrated ethical guidelines, the “Ethical Guidelines for Life Science and Medical Research Involving Human Subjects (Integrated Ethical Guidelines)” [77], inherit the privacy principles and procedures of the Ethical Guidelines [74]. The APPI [22] and the Integrated Ethical Guidelines [77] also complement each other with regard to the protection of personal information in clinical research.

Legislation and related governance arrangements in Japan cover two important aspects: first, patient consent for collecting and using routinely collected data, and second, anonymization of routinely collected data. The APPI does not apply to the use of personal information, when academic researchers or academic societies handle medical data for the purpose of academic research (article 76 of the APPI [22]). As a result of this, the majority of this type of RWD in Japan is only open to academic researchers and societies. The private sector only has access to commercially available databases which use anonymized medical data. Specifically, the most frequently used commercial databases are the JMDC Claims Database, Medi-Scope, and EBM provider® (administered by MDV). These databases are generally smaller compared with databases provided by the MHLW available to academic groups (e.g., NDB, DPC database, long-term care insurance database). Like many other countries, this means that a lower volume of research than maybe optimal is carried out by the private sector, which could have otherwise been beneficial to Japanese patients and public health.

In addition, regarding the protection of personal information, there is an issue that applicable laws are different for each research entity. Private businesses (e.g., private universities, academic societies, private hospitals, and private companies) need to comply with the APPI [22], government

administrative organs and national research institutes need to comply with the “Act on the Protection of Personal Information for Administrative Organs,” independent administrative agencies and national universities need to comply with the “Act on the Protection of Personal Information for Independent Administrative Agency,” and local governments, public universities, public research institutions, and public medical institutions need to comply with the “Personal Information Protection Ordinance.” When medical data are handled by academic researchers or academic societies for the purpose of academic research, the APPI [22] is not applicable, but it may be applicable to other acts and ordinance, and therefore judgment needs to be made individually, and attention needs to be paid to applicable laws.

The General Data Protection Regulation (GDPR), a new European initiative for the protection of personal information, is a rule for the processing and transfer of personal data, which sets stricter rules than those set forth in the “Personal Information Protection Act” in Japan. In the European GDPR, public interest in data utilization takes precedence, whereas in Japan, public interest does not prevail.

### 3.2 Linkage of Databases in Japan

Another challenge for RWD is the ability to link different databases. A single database may have limited outcomes. Therefore, linking data from several databases pertaining to an individual can add considerable value. For example, the NDB does not contain mortality data, so linking this database to one with mortality information would create a more comprehensive dataset. An existing capability could be extended by developing and enabling central linkage of different datasets. In 2006, the MHLW started the SS-MIX, which enables medical institutions to share EMR. The PMDA’s MID-NET has used this to establish a linked EMR and health insurance claims database, which also includes DPC and specimen test results. Like many other countries, there remain key challenges around linking data due to patient anonymity and consent and the fact that there are multiple database custodians.

In 2018, the NHIA was implemented to improve data linkage and governance. In accordance with the NHIA, so far, the Cabinet Office and other organizations certified LDI and J-MINO as business operators that collect, organize, and anonymize medical data and provide anonymized medical data, and also certified NTT Data Inc., ICI, and NS Solutions as business operators that collect and handle anonymized medical data entrusted by LDI or J-MINO. Whether this will lead to a set of nationally agreed and implemented standard rules to optimize interoperability of health record systems is still unknown [25].

### 3.3 Evaluation of RWD

Administrative and claims data sources have the advantages of being free from recall bias, providing easily accessible data and usually being relatively large in size. The weaknesses of these data are the lack of information on potential confounders, disease detail, and the uncertainty of diagnosis [78–80]; misclassification can also be a problem.

RWD should be evaluated as “fit for purpose” based not just on the quality of the data, but also on its relevancy of the data [81]. A “fit for purpose” evaluation depends on the context of the research question that is being answered and how the characteristics of the data impact the resulting RWE. A particular RWD source may be fit for purpose in one setting but not suitable in another context. The quality of the data can depend on the accuracy, completeness, provenance, and transparency of any data processing (i.e., how the data moves from the point of collection into the databases). Relevancy considers whether the RWD are representative of the population of interest, and whether critical data fields representing covariates and outcomes of interest are present (or can be derived from present data fields).

In Japan, the MHLW is under an obligation to maintain up-to-date and accurate records of the NDB and DPC databases. Ultimately, it is important for researchers to understand the characteristics and limitations of any database they work with and to validate outcomes and conduct sensitivity analyzes where appropriate.

### 3.4 Acceptance of RWD/RWE

As discussed above, there is growing interest in the use of RWD from various parties; however, there are barriers to the perceived credibility of both RWD and RWE [78]. For example, to demonstrate treatment effect, traditional RCTs are seen as the gold standard for clinical evidence, having high internal validity. Randomization minimizes the chance of bias from patient selection, treatment assignment, and outcome evaluation. Most RWE studies, which are often based on data from claims databases and EMRs, lack randomization. Patients receive treatment based on routine care and therefore comparing outcomes of patients receiving different treatment is subject to confounding bias. Although RWE studies have higher external validity than RCTs, internal validity is low, which can be a concern for decision-makers.

For decision-makers to fully accept the potential of RWE, results and the process leading to these results need to be transparent. RWD can vary in quality and content, and the study design and analysis need to be appropriately executed.

There is a lack of universally accepted methodological standards, although we do now see growing support to adopt common standards and guidelines for various aspects of RWD/RWE.

### 3.5 Standards and Guidelines for RWE Development

The FDA are leading the drive for the use of RWE, publishing a framework outlining the implementation of the RWE program at the end of 2018 [82]. The FDA’s framework looks at how RWD is defined, collected and analyzed, and also provides guidance on RWE study designs.

Outside of the regulatory space, ISPOR and ISPE are actively working to improve the standards and practice of the collection and analysis of RWD. Their work includes recommendations on good procedural practices for confirmatory treatment effect RWE studies [83]. The recommendations include the following:

1. A priori declaration that a study is confirmatory or “hypothesis evaluating” (i.e., study is testing explicit a priori hypotheses in a specific population) or exploratory (i.e., study primarily serves as a step to learn about possible treatment effectiveness).
2. Post a confirmatory study protocol and analysis plan on a public study registration site prior to conducting the study.
3. Publish confirmatory study results with attestation of conformance with or deviation from the original study protocol and analysis plan.
4. Provide opportunities to replicate confirmatory studies.
5. Perform confirmatory studies on different data sets than the one used to generate the hypotheses to be tested.
6. Authors should work with individuals to address methodological criticisms of their study.
7. Include key stakeholders in designing, conducting and disseminating the research.

### 3.6 Reporting Standards

The ISPOR/ISPE group have also worked to develop reporting guidelines with the aim to reproduce and facilitate validity assessment for database studies [20]. They include recommendations on the reporting of scientific decision-making during database study execution to enable potential replication, which would facilitate a robust assessment of the validity of these types of studies. Specific parameters for reporting to increase reproducibility include:

1. Data source: provider, extraction date, sampling, source data range, data type, linkage, cleaning, and any data model conversion;

2. Study design, including diagram;
3. Inclusion/exclusion criteria, study entry date, sequencing of exclusions, enrolment window, codes, care setting, and washouts;
4. Exposure definition, type of exposure, and risk window;
5. Reporting on follow-up time, follow-up window, and censoring criteria;
6. Reporting on outcome definition, event date, codes, and frequency;
7. Reporting on covariate definition, including time window;
8. Reporting on control sampling; and
9. Statistical software used.

The REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement gives guidance on the reporting of studies that have been conducted with routinely collected health data (RWD) such as administrative claims, EMRs, and disease registries [84]. The RECORD statement builds on the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines, which focus on reporting recommendations for observational studies in epidemiology [85].

### 3.7 Summary

There are challenges for RWD around data access and linkage, and for RWE there are challenges around its acceptance by decision-makers. These challenges for RWD/RWE are by no means unique to Japan and similar challenges exist for countries in Europe and the USA. As the demand for RWD and RWE increases, we need to focus on the quality of data, data relevance, the quality of analysis, study design, and the transparency of the entire process and reporting to ensure credibility and acceptance by decision-makers.

## 4 Future Perspectives

### 4.1 Advent of the Era of Rapid and Simultaneous Generation of Multiple RWEs

Research using RWD has been conducted in the USA since the 1980s. In Japan, until the middle of 2000, RCTs were frequently performed in the field of hypertension and other areas. The DPC started in 2003 as a tool for comprehensive payments, and in 2006 the DPC Data Research Team supported by the Health Science and Labour Research Grants began to actively conduct clinical research using DPC data (including more patient information than US Medicare, such as height, weight, smoking history, and stage of cancer). There have been hundreds of clinical studies using DPC data to-date, making it the most widely reported RWD sources

in Japan. Since 2010, the government and medical societies have made significant moves to improve the environment for clinical research using RWD, and the number of available RWD databases (NDB, information on nursing care Receipts, and disease registries managed by medical societies) has increased, as well as the use of existing RWD databases. In the latter half of the 2010s, the CIN project, led by MHLW and promoted by AMED, was started, the environment for efficient clinical development in Japan was improved, and the utilization of patient registries and cohort studies for clinical development was promoted. Along with the revision of the GPSP Ordinance, the MID-NET operation started to be used for post-marketing surveys conducted by pharmaceutical companies in addition to research conducted by medical institutions.

Research using RWD has become active in Japan, primarily because of the high cost of conducting RCTs and because RCTs are often impractical in terms of ethics and feasibility. Research using RWD has been expanding, mainly in academia, to compensate for the lack of RCTs. At present, in order to efficiently carry out clinical development and post-marketing surveillance, we are at the stage of working on the utilization of patient registries and cohort studies for clinical development and the utilization of MID-NET. Research using RWD in patient registries and cohort studies, as well as other RWD databases, is expected to increase in the future, including in the private sector.

The issues of RWD and RWE in Japan are pointed out in section 3 as follows: (1) access to RWD; (2) linkage of databases; (3) evaluation of RWD; (4) acceptance of RWD/RWE; (5) standards and guidelines for RWE development; and (6) reporting standards. Furthermore, the following challenges are also pointed out: (1) missing data; (2) the need for computable phenotypes; (3) lack of standardization; (4) interoperability issues with proprietary health information systems; and (5) data quality [86–88].

The following are our proposals and predictions of future prospects for Japan after the environment surrounding RWD and RWE has been improved and these issues have been resolved.

#### *Access to RWD and Data Quality*

- Ultimately, we would like to see that all Japanese people provide informed consent, including electronic informed consent, for the use of data collected in routine clinical practice, taking pride in cooperating in research and understanding the benefits of doing this. There have been issues with implementing this approach so far. First, patients may not have sufficient knowledge regarding the handling of personal information and research ethics, while in other cases the significance of providing data is not clear and patients may not be aware of the benefits of research using RWD.

In an attempt to resolve these issues, the Japanese government have adopted an approach whereby businesses can anonymize personal information; however, this approach uses a large amount of resources (e.g., time, cost, and labor). In addition, this approach may lead to higher RWD usage fees, a loss of immediacy of data utilization, and a lack of important information being collected. Instead, we propose that the Japanese government implement measures such as education to the public on personal information and research ethics, as well as the use of research, which could increase understanding and cooperation. The resources required for data anonymization would be reduced and the anonymization work would be simplified, and the resources could instead be used for education and getting informed consent. The costs associated with education and informed consent may be lower than those required for anonymization, thereby leading to a more effective use of resources. By realizing these measures, more RWD will be generated, and databases may be increasingly accessible to the private sector, allowing for faster creation of higher quality RWE.

#### *Database Linkage*

- We would also like to see Japan become one of the most advanced digital nations in the world and adopt a system to store RWD, such as medical information, linked to an individual identifier, similar to that used in Estonia [89, 90]. We propose that medical data from an individual patient, which may exist separately in different databases, be linked by a national identifier (My Number) to create a large, high-quality, lifetime health record for each person. The MHLW has started a new project in 2021, in which everyone can use a My Number Card instead of a health insurance card at a medical institution, and they are working on making this available at almost all medical institutions by 2023 [91]. This movement will provide impetus for our proposal to link RWD data with My Number. With informed consent and the use of My Number, data can be linked much more easily. Innovative information technology will allow for an inclusive system that can link large amounts of data with immediate access, while maintaining data security. Medical information in databases is tied to My Numbers, and data are shared by multiple medical institutions. Thanks to this system, they can receive appropriate treatment based on the medical data. If a person suddenly falls ill while traveling within Japan and goes to a local hospital, as a result, the labor, time, and cost for linking the databases of the RWD can be greatly reduced, and the overall utilization of the RWD can be facilitated. As a result, all Japanese people will be able to fully benefit from the use of digital technology and RWD, and their quality of life will be improved.
- We propose that the My Number identifier be also linked to family information. This allows the relationship between the onset of hereditary diseases and family history can be studied with reliable information.
- We propose that the right to use the internet be guaranteed to all people in Japan, and a system be established in which all people can input necessary personal health data, which is used for treatment and diagnosis. PHR data are appropriately incorporated into the EMR of the medical institution.
- We propose that international standardization of medical care databases be advanced. At the same time, information and communication technology (ICT) will make it possible not only to integrate with RWD databases in Japan, but also with overseas RWD databases, enabling the creation of RWE from RWD from both Japan and overseas. This leads to generating RWE that cannot be created from only Japanese RWD. For example, although the relationship between drug exposure and adverse effects in pregnant women, fetuses, and newborns are often inconclusive in conventional database studies in Japan due to the small number of cases, the number of cases that can be evaluated through integrated analysis with the similar databases in other Asian countries can be increased, and the accuracy of the analysis results can be improved.

#### *RWD Evaluation, RWD/RWE Acceptance*

- We propose that the use of RWD and RWE in the development of pharmaceuticals and medical devices be discussed by industry, government, academia and patients, and there be clear guidance from MHLW that is consistent with guidance from regulatory authorities in other countries. This guidance has enabled pharmaceutical companies to reduce development costs and obtain faster approval of new drugs, thereby increasing their contribution to patients and society.

#### *Innovation of Supporting Technology*

- We would also like to see that ICT and rapidly evolving digital health technologies allow the necessary RWD to be extracted and linked from a variety of RWD databases and quickly prepared for analysis. This greatly increases the amount of data that can be handled and the speed of research. Large ICT companies will contribute to the innovation of these supporting technologies. The ability to do several complex things at once will change the way we think about research. From the era in which one study using RWD creates RWE for one purpose, to the



era in which one study attempts to analyze and comprehensively interpret RWE from various angles and simultaneously creates a large amount of RWE.

- We would also like to see that RWD, which are updated daily, are used to create RWE in real time. This will lead to an era of monitoring changes in RWE. Real-world impacts can be examined in real time for the true effects expected (e.g., decreased mortality by anticancer drug), and the impact on medical care and society can be evaluated more directly.

#### *Research Framework using RWD*

- Research using RWD is not limited to conventional corporate sponsored studies and investigator-initiated studies, but will be diversified to include collaborative study among academia, contract research organizations, application companies, patient groups, local public bodies, such as prefectures and cities, and research institutions outside of Japan, and research using RWD has become more popular.

#### *Indicators Used in Medical Affairs*

- RWD, RWE, or their effects on them will be among the indicators used in the activities of Medical Affairs.

Various efforts are needed to reach the future prospects described above. An example of the challenge and the response to the problem are considered.

The Observational Health Data Sciences and Informatics (OHDSI) project is considered an effective solution tool for big data standardization and the utilization of RWD. OHDSI is an international voluntary open science community that promotes large-scale observational medical data analysis using a common data format [92]. OHDSI supports the joint development of evidence to promote better healthcare and aims to create a world where observational research provides a comprehensive understanding of health and disease. OHDSI was launched in the USA in 2014 and now has participants worldwide. Although it is open, medical data are protected by each participating organization, and personal information is not released outside the participating organizations. In just 5 years since its inception, the OHDSI global network has grown to include an estimated more than 600 million individuals with the exception of duplicate counts [93]. Following in the footsteps of Europe, China, and South Korea, a new community was established in Japan in the autumn of 2019 [94]. The OHDSI project uses the Observational Medical Outcomes Partnership (OMOP) Common Data Model [95], which organizes disparate medical data in a common format for easier analysis.

The common data format, the opt-out, and the anonymous processing of RWD promoted by the government are effective means to solve the current problems. On the other hand, with regard to the former, the merits of data standardization are obvious, but if the standardization is carried out excessively, it becomes impossible to collect data that is not applicable to the standardized items, and there is a risk that the collection of data becomes worse because it takes time and effort to create data in a new format. Rather, in order to make it easier to collect data, there is an opinion that it is better to secure a certain degree of freedom in data formats and richer data can be aggregated by ICT and digital health technology. In the latter case, there is also a concern that the anonymization of data will result in higher costs for the collection and use of RWD, resulting in less data being available.

In the case of Japan, the major issues that need to be resolved are the promotion of linkage among existing RWD databases (e.g., linking NDB and mortality data) and the creation of a system in which patients opt-in on a daily basis and cooperation is obtained from most Japanese people for the collection and use of data collected on a daily basis.

The larger and more complex the RWD database becomes, the more it tends to draw attention and awareness to the technical theory of database handling and to lose consciousness of the contribution to patients. Care must be taken to ensure that this never happens.

## 5 Conclusions

Research using RWD is increasing in Japan and provides important additional evidence for the purpose of drug development, understanding patient outcomes and disease, and medical decision-making. There are challenges in Japan regarding access to RWD sources and linkage of different databases, and various efforts are being made to address these issues. The OHDSI project is one of those efforts, using a common data format to facilitate data analysis of large observational studies internationally.

Developments in Japan's RWD and RWE are expected in the following areas:

- *RWD access* Opt-in will become entrenched in society, creating richer, less biased RWD datasets that will be accessible to pharmaceutical companies and enable faster creation of higher quality RWE.
- *Database linkage* A system for storing RWD such as medical information linked to a "My Number" identifier is being adopted, and linkage between databases will be facilitated resulting in more informative databases being constructed.
- *Innovation of supporting technology* ICT and digital health technologies enabling rapid data preparation



for analysis. This will greatly increase the amount of data that can be handled, increasing the overall speed of research. The way of thinking about research has changed, where one study can attempt to analyze data from various angles and comprehensively interpret it. We are in the era of rapid and simultaneous creation of a large amount of RWE.

Finally, it should be noted that although technical theories dealing with databases tend to attract attention and awareness, it is necessary to keep in mind that awareness of contributions to patients should not be lost.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40801-021-00266-3>.

**Acknowledgements** The authors are members of the Pharmaceutical Research and Manufacturers of America (PhRMA) Japan Medical Affairs Committee Working Group 1 (MAC WG1) on Real-World Evidence. The authors express their deep gratitude to the other members of the same WG1, the PhRMA Japan MAC as well as Yukiko Kimura, PhD, ex-member of the WG1, and the PhRMA USA for their valuable and scientific comments and suggestions in the preparation of the manuscript. The authors also thank Sarah Greig, PhD, of inScience Communications, Springer Healthcare, for editing and styling the manuscript prior to submission. This medical editing assistance was funded by PhRMA.

## Declarations

**Funding** The investigative work and writing activities to develop this review are conducted as PhRMA Japan MAC WG1 members' activities, and there is no special funding per se. The editorial fees and the article processing charge were funded by PhRMA, a non-profit organization.

**Conflicts of interest/competing interests** KH is an employee of Amgen K.K. and was an employee of Amgen Astellas BioPharma K.K. at the time of planning and writing this manuscript and a stockholder of Amgen Inc. AB is an employee and stockholder of Eli Lilly and Company Limited. YM is an employee of Bristol Myers Squibb K.K. and was an employee of Janssen Pharmaceutical K.K. and a stockholder of Johnson & Johnson at the time of planning and writing this paper.

**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. KH and AB 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.

**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

1. Yoneoka D, Hisashige A, Ota E, Miyamoto K, Nomura S, Segawa M, et al. Are Japanese randomized controlled trials up to the task? A systematic review. *PLoS One*. 2014;9:e90127. <https://doi.org/10.1371/journal.pone.0090127>.
2. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16:495. <https://doi.org/10.1186/s13063-015-1023-4>.
3. Ministry of education, culture, sports, science and technology - Japan. White Paper on Science and Technology 2018. Part 1: For further reinforcing the fundamental capability for STI [in Japanese]. 2018. [https://www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2018/06/06/1405921\\_002.pdf](https://www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2018/06/06/1405921_002.pdf) [original Japanese version]; [https://www.mext.go.jp/component/english/\\_icsFiles/afieldfile/2019/09/04/1420912\\_005.pdf](https://www.mext.go.jp/component/english/_icsFiles/afieldfile/2019/09/04/1420912_005.pdf) [provisional English translation version]. Accessed 9 Apr 2021.
4. Yasunaga H. Clinicoepidemiological studies using the DPC data: challenges for the future [in Japanese]. *Iryo To Shakai*. 2016;26:7–14. <https://doi.org/10.4091/iken.26.3>.
5. Li J, Gao XH, Bian Q, Guo ZY, Mei XB, Yu G, et al. Comparative study of scientific publications in urology and nephrology journals originating from USA, China and Japan (2001–2010). *PLoS ONE*. 2012;7:e42200. <https://doi.org/10.1371/journal.pone.0042200>.
6. Ministry of Health, Labour and Welfare. Ordinance no. 171: Ministerial ordinance on standards for post-manufacturing and marketing investigation and testing of pharmaceuticals [in Japanese]. 2006. [https://elaws.e-gov.go.jp/search/elawsSearch/elaws\\_search/lsg0500/detail?lawId=416M60000100171](https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=416M60000100171). Accessed 9 Apr 2021.
7. Ministry of Health, Labour and Welfare. Clinical research law (law no. 16 of Heisei 29) [in Japanese]. <https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000163413.pdf>. Accessed 9 Apr 2021.
8. Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labour and Welfare, Ministry of Economy, Trade and Industry. Ordinance no. 1 of 1991: Law enforcement regulations on anonymously processed medical information to contribute to research and development in the medical field [in Japanese]. [https://elaws.e-gov.go.jp/search/elawsSearch/elaws\\_search/lsg0500/detail?lawId=430M60000582001](https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=430M60000582001). Accessed 9 Apr 2021.
9. Pharmaceuticals and Medical Devices Agency. Regarding reliability assurance in post-marketing database surveys of pharmaceutical products [in Japanese]. <https://www.pmda.go.jp/files/000223003.pdf>. Accessed 9 Apr 2021.
10. Pharmaceuticals and Medical Devices Agency. Used in post-marketing database surveys. Formulation of the basic concept regarding the implementation of validation of outcome definitions [in Japanese]. <https://www.pmda.go.jp/files/000235927.pdf>. Accessed 9 Apr 2021.

11. Pharmaceuticals and Medical Devices Agency. Consultation on product registry utilization for pharmaceuticals / regenerative medicine, etc [in Japanese]. <https://www.pmda.go.jp/review-services/f2f-pre/consultations/0101.html>. Accessed 9 Apr 2021.
12. Ministry of Health, Labour and Welfare. Basic principles on the use of registries in approval applications [in Japanese]. <https://www.mhlw.go.jp/hourei/doc/tsuchi/T210324I0010.pdf>. Accessed 9 Apr 2021.
13. Ministry of Health, Labour and Welfare. Points to consider for ensuring the reliability when using registry data for approval applications [in Japanese]. <https://www.mhlw.go.jp/hourei/doc/tsuchi/T210324I0020.pdf>. Accessed 9 Apr 2021.
14. Pharmaceuticals and Medical Devices Agency. PMDA's Clinical Innovation Network (CIN) initiatives [in Japanese]. 2020. <https://www.mhlw.go.jp/content/10801000/000621593.pdf>. Accessed 9 Apr 2021.
15. US Food and Drug Administration. Real-world evidence. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>. Accessed 9 Apr 2021.
16. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health*. 2017;20:858–65. <https://doi.org/10.1016/j.jval.2017.03.008>.
17. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin Pharmacol Ther*. 2019;106:36–9. <https://doi.org/10.1002/cpt.1426>.
18. Subcommittee on the Pharmaceutical and Medical Device System of the Health Science Council. Utilization of health science foundation medical real world data FY 2018 (FY2018) research report on drug discovery technology part 2. 2019. Accessed 9 Apr 2021.
19. Guideline for the preparation of product information brochure of ethical drugs and the preparation of advertisements of ethical drugs inserted in scientific journals (papers) [in Japanese]. In: Japan Pharmaceutical Manufacturers Association Review Committee for Product Information Brochure: Tokyo, Japan, 2013.
20. 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. <https://doi.org/10.1002/pds.4297>.
21. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med*. 2016;375:2293–7. <https://doi.org/10.1056/NEJMsb1609216>.
22. Personal Information Protection Commission - Japan. [in Japanese]. <https://www.kantei.go.jp/jp/it/privacy/houseika/hourituan/030307houan.html>. Accessed 9 Apr 2021.
23. Office of Medical Information Infrastructure, Japan Agency for Medical Research and Development, Cabinet Office Life Science Division, Research Promotion Bureau, Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labour and Welfare, Ministry of Economy, Trade and Industry. About certification of business operator based on next-generation medical infrastructure law [in Japanese]. 2019. <https://www8.cao.go.jp/iryoun/nintei/nintei/jigyousha/pdf/shiryuu.pdf>. Accessed 9 Apr 2021.
24. Cabinet Office. Japan Medical Association Medical Information Management Organization, ICI Co., Ltd. and NS Solutions Co., Ltd. 2020. <https://www8.cao.go.jp/iryoun/nintei/nintei/jigyousha2.html>. Accessed 9 Apr 2021.
25. 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 9 Apr 2021.
26. Japanese Society for Pharmacoepidemiology. Database survey applicable to clinical epidemiology and drug epidemiology in Japan [in Japanese]. <http://www.jspe.jp/committee/020/0210/>. Accessed 9 Apr 2021.
27. Japan Agency for Medical Research and Development. Project for promoting Clinical Innovation Network. 2018. <https://www.amed.go.jp/en/program/list/05/01/015.html>. Accessed 9 Apr 2021.
28. Clinical Innovation Network Promotion Support Project. Request for cooperation in the CIN National Land Group 2019 Patient Registry and Cohort Study Survey (3rd). <https://cinc.ncgm.go.jp/cin/G001.php> [in Japanese]; <https://cinc.ncgm.go.jp/cin/en/G001.php> [in English]. Accessed 9 Apr 2021.
29. Clinical Innovation Network. Registry creation and operation guide [in Japanese]. [https://cinc.ncgm.go.jp/?page\\_id=540](https://cinc.ncgm.go.jp/?page_id=540). Accessed 9 Apr 2021.
30. Ministry of Health, Labour and Welfare. Cancer registration: about the Act on Promotion of Cancer Registration, etc (act no. 111 or 2013) [in Japanese]. [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou\\_iryuu/kenkou/gan/gan\\_toroku.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/kenkou/gan/gan_toroku.html). Accessed 9 Apr 2021.
31. Japanese Association of Cancer Registries. Cancer registry in Japan. [http://www.jacr.info/eng/CRIJ\\_eng.pdf](http://www.jacr.info/eng/CRIJ_eng.pdf). Accessed 9 Apr 2021.
32. National Clinical Database. <http://www.ncd.or.jp/>. Accessed 9 Apr 2021.
33. Seto Y, Kakeji Y, Miyata H, Iwanaka T. National Clinical Database (NCD) in Japan for gastroenterological surgery: brief introduction. *Ann Gastroenterol Surg*. 2017;1:80–1. <https://doi.org/10.1002/ags3.12026>.
34. Baba H. National Clinical Database (NCD) in Japan: clinical and social significance. *Ann Gastroenterol Surg*. 2019;3:462–3. <https://doi.org/10.1002/ags3.12282>.
35. Japan Cardiovascular Surgery Database Organization. JCVSD [in Japanese]. 2020. <http://jcvsd.umin.jp/index.html>. Accessed 9 Apr 2021.
36. Motomura N, Takamoto S. Development and expanding use of Cardiovascular Surgery Database in Japan. *JMAJ*. 2010;53:319–20.
37. J-DREAMS. Japanese compREhensive database project based on an Advanced electronic Medical record System [in Japanese]. <http://jdreams.jp/>. Accessed 9 Apr 2020.
38. Sugiyama T, Miyo K, Tsujimoto T, Kominami R, Ohtsu H, Ohsugi M, et al. Design of and rationale for the Japan Diabetes comprehensive database project based on an Advanced electronic Medical record System (J-DREAMS). *Diabetol Int*. 2017;8:375–82. <https://doi.org/10.1007/s13340-017-0326-y>.
39. Kanaoka K, Okayama S, Yoneyama K, Nakai M, Nishimura K, Kawata H, et al. Number of board-certified cardiologists and acute myocardial infarction-related mortality in Japan - JROAD and JROAD-DPC registry analysis. *Circ J*. 2018;82:2845–51. <https://doi.org/10.1253/circj.CJ-18-0487>.
40. The Japanese Circulation Society. The Japanese Registry Of All cardiac and vascular Diseases (JROAD). <https://jroadinfo.nccv.go.jp/> Accessed 9 Apr 2021.
41. Yasuda S, Nakao K, Nishimura K, Miyamoto Y, Sumita Y, Shishido T, et al. The current status of cardiovascular medicine in Japan – analysis of a large number of health records from a nationwide claim-based database. *JROAD-DPC Circ J*. 2016;80:2327–35. <https://doi.org/10.1253/circj.CJ-16-0196>.
42. Kobayashi S. Significance and development of Japanese Stroke Data Bank [in Japanese]. *Nihon Rinsho*. 2006;64:25–31.

43. Japan Stroke Data Bank (JSDB). [http://strokedatabank.ncvc.go.jp/en/?doing\\_wp\\_cron=1614085427.7640581130981445312500](http://strokedatabank.ncvc.go.jp/en/?doing_wp_cron=1614085427.7640581130981445312500). Accessed 9 Apr 2021.
44. J-CKD-Database. <http://j-ckd-db.jp/english/index.html>. Accessed 9 Apr 2021.
45. Nakagawa N, Sofue T, Kanda E, Nagasu H, Matsushita K, Nan-gaku M, et al. J-CKD-DB: a nationwide multicentre electronic health record-based chronic kidney disease database in Japan. *Sci Rep.* 2020;10:7351. <https://doi.org/10.1038/s41598-020-64123-z>.
46. Sagamihara National Hospital. Rheumatoid Arthritis Research Department [in Japanese]. <https://sagamihara.hosp.go.jp/rinken/section.html>. Accessed 9 Apr 2021.
47. National Cancer Center Hospital. SCRUM-Japan [in Japanese]. <http://www.scrum-japan.ncc.go.jp/index.html>. Accessed 9 Apr 2021.
48. Tsuchihara K. SCRUM-Japan, a nation-wide cancer genome screening and patient registry program. *Juntendo Med J.* 2018;64:191–7. <https://doi.org/10.14789/jmj.2018.64.JMJ17-WN11>.
49. Otsu A, Okamoto W, Fuse M, et al. From an academic perspective: new initiatives in the SCRUM-Japan Registry. In: Presented at: The 8th Joint Cancer New Drug Development Symposium. 2018.
50. Japan Trauma Databank. [in Japanese]. <https://www.jtcr-jatec.org/traumabank/index.htm>. Accessed 9 Apr 2021.
51. Nakahara S, Sakamoto T, Fujita T, Koyama T, Katayama Y, Tanabe S, et al. Comparison of registry and government evaluation data to ascertain severe trauma cases in Japan. *Acute Med Surg.* 2017;4:432–8. <https://doi.org/10.1002/ams2.302>.
52. Yokota J. Japan Trauma Data Bank (JTDB) managed by Japan Trauma Care and Research (JTCR) [in Japanese]. *Nihon Rinsho.* 2016;74:329–36.
53. Japanese Society for Pharmacoepidemiology. Databases available for pharmacoepidemiology researchers in Japan (information obtained from survey answers as of August 2020). 2020. [http://www.jspe.jp/mt-static/FileUpload/files/JSPE\\_DB\\_TF\\_J.pdf](http://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_J.pdf) [original Japanese version]; [http://www.jspe.jp/mt-static/FileUpload/files/JSPE\\_DB\\_TF\\_E.pdf](http://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_E.pdf) [English version]. Accessed 9 Apr 2021.
54. Shinichiro K, Tatsuya N, Tomoya M, Yuichi N, Tsuneyuki H, Hiroki M, et al. National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): outline and patient-matching technique. *bioRxiv.* 2018:280008. <https://doi.org/10.1101/280008>.
55. Hirose N, Ishimaru M, Morita K, Yasunaga H. A review of studies using the Japanese National Database of Health Insurance Claims and Specific Health Checkups. *Ann Clin Epidemiol.* 2020;2:13–26. [https://doi.org/10.37737/ace.2.1\\_13](https://doi.org/10.37737/ace.2.1_13).
56. JMDC Inc. JMDC Claims Database. <https://www.jmdc.co.jp/en/jmdc-claims-database>. Accessed 9 Apr 2021.
57. Nagai K, Tanaka T, Kodaira N, Kimura S, Takahashi Y, Nakayama T. Data resource profile: JMDC claims databases sourced from Medical Institutions. *J Gen Fam Med.* 2020;21:211–8. <https://doi.org/10.1002/jgf2.367>.
58. JMIRI. Medi-Scope [in Japanese]. [https://www.jmiri.jp/medi\\_scope](https://www.jmiri.jp/medi_scope). Accessed 9 Apr 2021.
59. Medical Data Vision. About MDV database. <https://en.mdv.co.jp/about-mdv-database/>. Accessed 9 Apr 2021.
60. Pharmaceuticals and Medical Devices Agency. MID-NET (Medical Information Database Network) [in Japanese]. <https://www.pmda.go.jp/safety/mid-net/0001.html>. Accessed 9 Apr 2021.
61. 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. <https://doi.org/10.1002/pds.4879>.
62. HCEI. HCEI database overview [in Japanese]. <https://www.hcei.or.jp/page/database>. Accessed 9 Apr 2021.
63. Real World Data. [in Japanese]. <https://rwdata.co.jp>. Accessed 9 Apr 2021.
64. Scientist. Real World Data Co. Ltd. <https://app.scientist.com/providers/real-world-data-co-ltd>. Accessed 9 Apr 2021.
65. Koba Y, Horiguchi Y, Okada C, Fushimi K. Action of the hospital usability test that utilized practice information [in Japanese]. *Iryo.* 2012;66:684–7.
66. Horiguchi H. Clinical research using the database of the National Hospital Organization [in Japanese]. *Progress Med.* 2018;38:127–30.
67. Tokuhshukai Medical Group. <https://www.tokushukai.or.jp/en/>. Accessed 9 Apr 2021.
68. Iwakami M, Kobayashi S. Clinical research using Tokushukai Medical Database [in Japanese]. *Progress Med.* 2018;38:131–5.
69. Ministry of Health, Labour and Welfare. Homepage about providing related information such as anonymous medical treatment [in Japanese]. [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou\\_iryuu/iryuhoken/dpc/index.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuhoken/dpc/index.html). Accessed 9 Apr 2021.
70. Yasunaga H. Trends in clinical research using DPC database 1) General remarks/Oncology area [in Japanese]. *Progress Med.* 2018;38:137–40.
71. JMIRI. Patient / pharmacist survey service [in Japanese]. <https://www.jmiri.jp/service/primarysearch>. Accessed 9 Apr 2021.
72. Kantar. National Health and Wellness Survey (NHWS). <https://www.kantar.com/expertise/health/da---real-world-data-pros-claims-and-health-records/national-health-and-wellness-survey-nhws>. Accessed 9 Apr 2021.
73. 3H Medi Solution. <https://global-3h.com/>. Accessed 9 Apr 2021.
74. Ministry of Health, Labour and Welfare. Ethical guidelines for medical and health research involving human subjects. <https://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000153339.pdf> [Japanese version]; <https://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000080278.pdf> [English version]. Accessed 9 Apr 2021.
75. Ministry of Health, Labour and Welfare. Ethical guidelines for human medical research guidance [in Japanese]. <https://www.mhlw.go.jp/content/10600000/000757222.pdf>. Accessed 9 Apr 2021.
76. Ministry of Health, Labour and Welfare. Ethical guidelines for human genome and gene analysis [in Japanese]. <https://www.mhlw.go.jp/content/10600000/000757268.pdf>. Accessed 9 Apr 2021.
77. Ministry of Health, Labour and Welfare. Ethical guidelines for life science and medical research involving human subjects [in Japanese]. <https://www.mhlw.go.jp/content/000757566.pdf>. Accessed 9 Apr 2021.
78. Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. *Biol Pharm Bull.* 2015;38:53–7. <https://doi.org/10.1248/bpb.b14-00543>.
79. Sakai M, Ohtera S, Iwao T, Neff Y, Kato G, Takahashi Y, et al. Validation of claims data to identify death among aged persons utilizing enrollment data from health insurance unions. *Environ Health Prev Med.* 2019;24:63. <https://doi.org/10.1186/s12199-019-0819-3>.
80. Hara K, Tomio J, Svensson T, Ohkuma R, Svensson AK, Yamazaki T. Association measures of claims-based algorithms for common chronic conditions were assessed using regularly collected data in Japan. *J Clin Epidemiol.* 2018;99:84–95. <https://doi.org/10.1016/j.jclinepi.2018.03.004>.
81. Duke University Margolis Center for Health Policy. Determining real-world data's fitness for use and the role of reliability. 2019. [https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd\\_reliability.pdf](https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf). Accessed 9 Apr 2021.

82. US Food and Drug Administration. Framework for FDA's real-world evidence program. 2018. <https://www.fda.gov/media/120060/download>. Accessed 9 Apr 2021.
83. 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. <https://doi.org/10.1002/pds.4295>.
84. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 2015;12:e1001885. <https://doi.org/10.1371/journal.pmed.1001885>.
85. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147:573–7. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>.
86. Kaneyama T, Sakaguchi M, Nakajima A, Aoki K, Shiragasawa C, Niwa S, et al. Real world data—current scientific activities in pharmaceutical industry. *Regul Sci Med Prod.* 2017;7:225–36. <https://doi.org/10.14982/rsmp.7.225>.
87. Kimura E. Considerations of the observational research database utilizing Real World Data. *J Natl Inst Pub Health.* 2018;67:179–90. [https://doi.org/10.20683/jniph.67.2\\_179](https://doi.org/10.20683/jniph.67.2_179).
88. National Academies of Sciences, Engineering, Medicine. Real-world evidence generation and evaluation of therapeutics: proceedings of a workshop. Washington, DC: The National Academies Press; 2017.
89. e-Governance Academy. e-Estonia: e-Governance in practice. <https://ega.ee/wp-content/uploads/2016/06/e-Estonia-e-Governance-in-Practice.pdf>. Accessed 9 Apr 2021.
90. e-Estonia.com. Welcome to the electronic nation "e-Estonia": the most advanced digital society in the world [in Japanese]. <https://e-estonia.com/wp-content/uploads/2828-e-estonia-introduction-presentation-jap-estonian-design-team-19121622.pdf>. Accessed 9 Apr 2021.
91. Ministry of Health, Labour and Welfare. About use of insurance card of my number card [in Japanese]. [https://www.mhlw.go.jp/stf/newpage\\_08277.html](https://www.mhlw.go.jp/stf/newpage_08277.html). Accessed 9 Apr 2021.
92. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inf.* 2015;216:574–8. <https://doi.org/10.3233/978-1-61499-564-7-574>.
93. Banda JM. Fully connecting the Observational Health Data Science and Informatics (OHDSI) initiative with the world of linked open data. *Genom Inf.* 2019;17:e13. <https://doi.org/10.5808/GI.2019.17.2.e13>.
94. Observational Health Data Sciences and Informatics. OHDSI Japan. [in Japanese] <https://www.ohdsi-japan.org/>. Accessed 9 Apr 2021.
95. Observational Health Data Sciences and Informatics. OMOP common data model. <https://www.ohdsi.org/data-standardization/the-common-data-model/>. Accessed 9 Apr 2021.
96. Pharmaceuticals and Medical Devices Agency. Japan's response to COVID-19 – From the perspective of international cooperation [in Japanese]. 2020. <https://www.pmda.go.jp/files/000238063.pdf>. Accessed 9 Apr 2021.