



Real world data and data science in medical research: present and future

Kanae Togo¹ · Naohiro Yonemoto¹

Received: 13 December 2021 / Revised: 18 February 2022 / Accepted: 20 March 2022 /
Published online: 13 April 2022
© The Author(s) 2022

Abstract

Real world data (RWD) are generating greater interest in recent times despite being not new. There are various purposes of the RWD analytics in medical research as follows: effectiveness and safety of medical treatment, epidemiology such as incidence and prevalence of disease, burden of disease, quality of life and activity of daily living, medical costs, etc. The RWD research in medicine is a mixture of digital transformation, statistics or data science, public health, and regulatory science. Most of the articles describing the RWD or real-world evidence (RWE) in medical research cover only a portion of these specializations, which might lead to an incomplete understanding of the RWD. This article summarizes the overview and challenges of the RWD analysis in medical fields from methodological perspectives. As the first step for the RWD analysis, data source of the RWD should be comprehended. The progress of the RWD is closely related to the digitization, especially of medical administrative data and medical records. Second, the selection of appropriate statistical and epidemiological methods is highly critical for an RWD analysis than those for randomized clinical trials. This is because it contains greater varieties of bias, which should be controlled by balancing the underlying risk between treatment groups. Last, the future of the RWD is discussed in terms of overcoming limited data by proxy confounders, using unstructured text data, linking of multiple databases, using the RWD or RWE for a regulatory purpose, and evaluating values and new aspects in medical research brought by the RWD.

Keywords Real world data · Medical database · Observational study · Causal inference · Bias

✉ Kanae Togo
kanae.togo@pfizer.com

¹ Health and Value, Pfizer Japan Inc., Tokyo, Japan

1 Introduction

In the field of medical research, randomized clinical trials (RCTs) are the golden standard to estimate the causal inference between treatment interventions and outcomes. Real world data (RWD) are the data relating to patient health status and/or the delivery of the healthcare, routinely collected from a variety of sources regardless of the size of data (Hiramatsu et al., 2021; US Food & Drug Administration, 2018). Data from observational studies are also considered as RWD. The RWD are mutually complementary to limitations of RCTs, such as a small sample size of trials, participants with limited age groups including a smaller number of people, or no older people and minors, participants with limited complications, short follow-up duration, etc.

The RWD are generating greater interest in recent times although the RWD has been used since more than decades ago. Progress of the RWD is closely related to digitization, especially of medical administrative data and medical records. The UK established the first European electronic health records (EHR) database of Clinical Practice Research Datalink (CPRD) in 1987. In the US, the HITECH Act was enacted in 2009, providing funds toward encouraging medical practices to better adopt and make meaningful use of the EHR (Menachemi & Collum, 2011). In Japan, electronic claim of public health insurance was legalized in 1976, finally becoming an obligation in 2011. At present, the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) cover 99% of claims in Japan.

Various purposes of analysis using the RWD (hereafter, RWD analysis) in medical research include effectiveness and safety of medical treatment, epidemiology such as incidence and prevalence of disease, burden of disease, quality of life and activity of daily living, and medical costs. As an example of comparative effectiveness, long-term survival advantage among patients who underwent coronary-artery bypass grafting (CABG) was shown compared with percutaneous coronary intervention (PCI), using a large study population of 189,793 patients in total, from claims and patient registry databases (Weintraub et al., 2012). These days, the RWD and real-world evidence (RWE) are used for regulatory submission (Feinberg et al., 2020), as well as for other activities during clinical development and post-launch of drugs in the pharmaceutical industry (Fig. 1) (Togo et al., 2019).

The RWD in medicine is a mixture of digital transformation, statistics, data science, public health, reimbursement, pricing of products and regulatory science. Unfortunately, most of the articles describing the RWD in medical research cover only a portion of these specializations and that might lead to RWD's incomplete understanding. Further, describing various sources of the RWD and relating analytic issues as biases (systematic errors) remain insufficient as well. Therefore, this article summarizes the overview of data source of the RWD and challenges of the RWD analysis in medical fields from methodological perspectives.

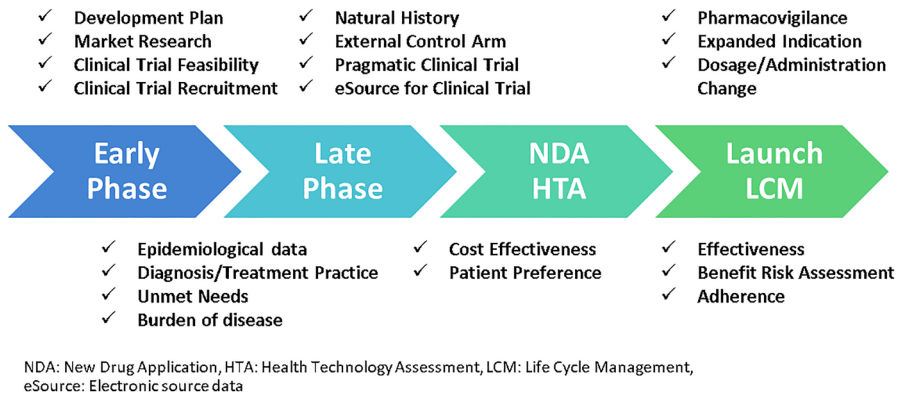


Fig. 1 Use of real-world data during clinical development and post-launch in pharmaceutical industry

2 Source of RWD

There are two approaches to collecting RWD: primary and secondary data collection (Schneeweiss & Patorno, 2021). Primary data are collected directly from study participants for the purposes of the study, and may be collected either retrospectively (via patient charts or other data sources), or prospectively (Mueller et al., 2018). Secondary data are obtained from existing health care data collection infrastructures, such as administrative claims databases, EHR databases, existing patient registries or study cohorts, or individual patient medical records.

Major data sources of secondary data are presented in Table 1 (Nabhan et al., 2019). Each database of secondary data has strengths and limitations. The overview of medical databases in Japan is annually surveyed by the Japanese Society for Pharmacoepidemiology and provided on the website (Pharmacoepidemiology & Database Taskforce, Japanese Society for Pharmacoepidemiology, 2021). The largest medical database in Japan is the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB), which covers approximately 99% of health insurance claims in Japan. Although its use was limited to certain organizations for public welfare and academic institutions, NDB has been open to private companies for research of public health since October 2020 (Kaneyama et al., 2017; Kohsaka et al., 2021).

When selecting an appropriate data source for a research, protection of data privacy is one of the key elements, as well as strengths and limitations of each data source. The RWD includes medical data with sensitive personal information. Therefore, data privacy has to be protected for any types of data source in compliance with national data protection laws such as the US Health Insurance Portability and Accountability Act (HIPAA), the EU General Data Protection Regulation (GDPR), and the Japan Act on the Protection of Personal Information (Personal Information Protection Commission, Government of Japan, 2020; Office of Civil Rights, Department of Health and Human Services, 2002; The European Parliament & the Council of the European Union EUR-Lex, 2019; Wirth et al., 2021). In Japan, medical data

Table 1 Major data sources of secondary data (Nabhan et al., 2019)

Data source	Description
Administrative claims database	A health insurance claim is a request made for direct payment or reimbursement for medical services from hospitals, clinics, pharmacy. Claims data are systematic and well-structured. Large claims databases are available in many countries. However, claims are recorded to maximize the reimbursement and the data sometimes might be unrepresented as the disease name of clinical practice
EHR database (Evans, 2016)	An EHR is an individual patient health record. A typical EHR may include a patient's medical history, diagnoses, treatment plans, immunization dates, allergies, radiology images, pharmacy records, and laboratory and test results. Although EHR databases are more likely to capture important health information about patients than administrative data, most of that information is unstructured
Patient registry	A patient registry is defined as an organized system that collects data and information on a group of people defined by a particular disease or condition, and that serves a pre-determined scientific, clinical and/or public health (policy) purpose (European Medical Agency, 2020). In addition to its use as secondary data, a registry-based study is another way of leveraging registry system. Registry often includes clinical outcomes, but missing data is common
Wearable, sensor	Sensors and/or software apps on smartphones and tablets that can collect health-related data remotely i.e., outside of the healthcare provider's office (Izmailova et al., 2018). These provide monitoring treatment response, including the monitoring of efficacy and safety of treatment, and monitoring of patient-reported outcomes and/or quality of life measures

is regarded as sensitive information and consent from patients (opt-in consent) is required to use such data for secondary purposes unless it is anonymized according to the act. However, for academic research, medical information regarding public health can be used with opt-out consent to provide patients an opportunity to refuse inclusion in the research (Ministry of Education, Culture, Sports, Science and Technology et al., 2021). Although database linkage enables large data creation, a wide range of personal information makes data linkage difficult, or at times, even impossible, because of data privacy.

3 Statistical and epidemiological methods for the RWD analysis

Although the RWD can be leveraged for various research questions, the selection of appropriate statistical and epidemiological methods is highly critical than those for RCTs. In traditional clinical trials, randomization has long been an essential tool for minimizing bias by balancing underlying risk between treatment groups (Sherman et al., 2016). Among dozens of biases that have been defined, the major biases are the selection bias and the information bias (Table 2) (Rothman et al., 2008). A bias needs to be prevented by adequate designing of the study, since bias once identified,

Table 2 Biases in the RWD analysis

	Description	Example of limitation due to secondary data
Selection bias	Bias that results from procedures used to select participants and from factors that influence study participation	Insurance claims data, based on worker's association, consists of relatively young and healthier people. The population of EHR depends on the types of hospitals (Rothman et al., 2008)
Information bias (Measurement error, misclassification)	Systematic error such as self-reporting or recall bias, measurement error bias, confirmation bias (investigator belief) (Althubaiti, 2016; Rothman et al., 2008). Immortal time bias is also a kind of information bias where immortal time is a span of time in the observation or follow-up period during which the study outcome could not have occurred (Suissa, 2008)	Diagnosis for reimbursement, validity of defined outcome, misclassification of drug exposure especially time-varying exposure due to limited data
Confounding	Confounding is related independently to both, the exposure and the outcome. It may create an apparent association or mask a real one (Strom et al., 2019). Confounding by indication can occur when participants, who undergo treatment, are more likely to have an underlying condition	Limited data of potential confounders

cannot be reverted. Confounding is another challenge in the RWD analysis. It can be controlled either in the study design or in its analysis.

Basic principles to prevent a bias and confounding in design are random allocation (generally not applicable in RWD), subject selection or localization, stratification, and matching. Common epidemiological study designs using these principles are cohort studies of new-user design, nested case–control studies, and self-control studies (Table 3) (Strom et al., 2019). Furthermore, study designs for regulatory decision making in combination with clinical trials emerge (Table 3) (Baumfeld et al., 2020), whereas the purpose of those designs is to support the evidence from clinical trials rather than to prevent bias. The epidemiological study designs are adaptive to the studies for regulatory decision making as well.

Statistical methods should be carefully selected in the RWD analysis for causal inference and estimation of treatment effect, as well as the study design. We summarize some statistical methods.

3.1 Stratification, matching, and weighting using propensity score

Propensity score is the probability of treatment assignment ($Z=1$ for treated, $Z=0$ for control) conditional on measured baseline covariates (X) and the propensity score for subject i ($i=1, \dots, n$) is $e_i = Pr(Z_i = 1|X_i)$ (Rosenbaum & Rubin, 1983). The propensity score is often applied to stratification and matching to balance large numbers of covariates. However, the balancing between treatment groups using propensity score requires the assumption of no unobserved confounders which is rarely true in research using the RWD. Other difficulty is the distribution of propensity scores. It often does not overlap adequately between treatment groups when the treatment choice is strongly related with patient backgrounds.

Inverse probability weighting (IPW) using propensity score, or inverse probability of treatment weighting (IPTW) is an alternative method to estimate treatment effect with covariate adjustment. The inverse probability of treatment weight is defined as $w_i = Z_i/e_i + (1 - Z_i)/(1 - e_i)$. The weight can be highly large when a subject has very low propensity score. A number of alternatives are proposed to stabilize weights such as $w_i = Pr(Z = 1)Z_i/e_i + Pr(Z = 0)(1 - Z_i)/(1 - e_i)$, where $Pr(Z = 1)$ and $Pr(Z = 0)$ are the marginal probability of treatment and control in the overall sample (Austin & Stuart, 2015).

A large number of RWD studies with application of propensity have been published, but only a few reviews suggested the pitfall and the guidance (Austin, 2007; Yao et al., 2017; Zakrisson et al., 2018).

3.2 Structural models and doubly robustness

The difficulty of a RWD analysis, in addition to that caused by non-randomization, is bias due to time-varying exposures and confounders. These effects post-treatment initiation are not considered in Intent-to-Treat approach, which is generally employed in RCTs. Marginal structural models produce consistent estimates of the average causal treatment effects even in the presence of treatment changes, time-dependent

Table 3 Study designs using the RWD (Baumfeld et al., 2020; Strom et al., 2019)

	Description	Limitation
Epidemiological study design		
Cohort studies of new-user design	Identifying patients who start a new drug and begin follow-up after initiation. Those patients have been evaluated by physicians who concluded that the patients might benefit from a newly prescribed drug. This makes treatment groups highly similar in characteristics that might not have been observed in the database	Secondary data requires the assumption that the first use of the drug after a certain period of no drug use is regarded as a new use
Nested case-control studies	Selecting all cases in the cohort, then randomly selecting one or more controls from risk-set for each case. The antecedent exposure is compared between cases and controls	Whenever the disease to be investigated is changed, controls need to be re-selected, as well as the cases. Whereas, the case-control study can easily consider many exposures
Self-control studies	This is a design where comparisons between exposures are made within subjects, thus significantly attenuating the problem of confounding	This design can only deal with an acute adverse event, a short wash-out period of exposure and requires precise data of exposure
Study designs for regulatory decision making in combination with clinical trials		
External control arm for clinical trials	Participants are selected from a cohort of RWD to be balanced with subject of a clinical trial in backgrounds for comparing key efficacy or safety outcomes	Differences in unmeasured confounders between the arms of clinical trial and the external control remain even after adjusting all measured confounders
Pragmatic studies	A randomized clinical trial incorporating pragmatic design elements. The intervention should be delivered as in clinical practice (Ford & Norrie, 2016)	This is the most feasible in health care systems with reliable and accessible electronic health records that capture the events of interest, which is at present are challenging in many countries
Long-term follow-up studies	Post-marketing requirement studies of safety and effectiveness outcomes of interest require longer follow-up durations	Regulators or companies may prefer RCTs due to feasibility (e.g., level of measurement and/or monitoring)

confounders, and missing at random study dropout as application of inverse-probability weighting (Faries et al., 2014). Structural nested mean model is also applied to time-varying exposures and confounders, and are fitted using g-estimation. This model is better tailored for dealing with failure of the standard assumptions of no unmeasured confounders (Vansteelandt & Sjolander, 2016). Several applications of time-depending confounders have been reported in a wide range of diseases (Clare et al., 2019; Yang et al., 2014). Of these, one example is the comparative effectiveness study of angiotensin receptor blockers (ARBs), in patients with chronic heart failure (CHF), using a national cohort of beneficiaries of the US Department of Veterans Affairs medical care system (Desai et al., 2012). A marginal structural model in reducing mortality in CHF included the time-dependent confounder of hospitalization, which lies on the causal pathway from treatment to death. ARB treatment history and hospitalization were defined on monthly basis.

The idea of doubly robustness is to combine outcome and exposure modelling in a fashion that provides a valid estimate if either model is correct. For example, matching using propensity scores in a model for exposure, and then regressing outcomes on exposure and covariates in the matched sample. Robins et al. (1994) developed an improved augmented inverse probability weighting (AIPW) using the process of double robustness property involving 2 basic steps: first, fitting a propensity score model, and then fitting models that estimate the outcome Y_i under treatment and control conditions, $f(Z|X_i)$ (Kurz, 2021; Qi & Sun, 2014; Robins et al., 1994). The average treatment effect of AIPW can be estimated by

$$\frac{1}{n} \sum_i \left[\frac{Z_i Y_i}{e_i} - \frac{Z_i - e_i}{e_i} f(1|X_i) \right] - \frac{1}{n} \sum_i \left[\frac{(1 - Z_i) Y_i}{1 - e_i} - \frac{Z_i - e_i}{e_i} f(0|X_i) \right].$$

3.3 Instrumental variable

Instrumental variable is the approach without assuming potential unobserved confounders. Instrumental variables naturally create quasi-random treatment choice and is related to the interested treatment. The instrumental variable estimator is simple: $(E[Y|T = 1] - E[Y|T = 0]) / (E[Z|T = 1] - E[Z|T = 0])$, where T is the instrumental variable (Strom et al., 2019). An example of instrumental variable is site-level preference for the use of embolic protection devices (EPD) for assessing causality of EPD use during transcatheter aortic valve replacement (TAVR) on in-hospital stroke (Butala et al., 2021). However, the difficulty in finding a valid instrumental variable is the reason for its relatively limited use (Faries et al., 2014).

3.4 Machine learning, AI

At times, the RWD is a large and high-dimensional data. Machine learning methods are used for identifying groups with disease prognostic or treatment response from a large data (Bakouny & Patt, 2021; Bica et al., 2021) In addition, machine learning and deep learning are leveraged for propensity score estimation. Application of

AI to the RWD is an intensive research field to use complex RWD including texts, images, voice data, etc. Disease diagnosis from image data and outcomes, such as treatment response and adverse effects, extracted from physician notes are popular and practical themes using AI.

3.5 Sensitivity analysis

Although it may be strange to deal with sensitivity analysis along with the methods mentioned above, it is worth referring to sensitivity analysis to understand the robustness of a study's findings in a RWD analysis. Sensitivity analysis for quantifying a bias is sometimes called bias analysis. For unobserved confounders, external adjustment is to adjust relative risk using external evidence. For misclassification of disease or exposure, statistics can be adjusted using sensitivity and specificity estimated in a validation study or external literatures. For assessing the effect of a bias in the study design, definitions of baseline period, outcomes, and exposure can be varied in the sensitivity analysis (Rothman et al., 2008). Causal effects estimated in observational studies are not binary signals, with or without statistical significance, but are numerical quantities. To quantify the effect as unbiasedly and precisely as possible, multiple studies using the different RWD sources, and meta-analysis of them could provide a highly reliable estimate (Hernán, 2021). The practical implementation, and quite a few assumptions of the sensitivity and bias analysis are continuously discussed.

4 Discussion

There are several approaches to overcome the limited data of a RWD database. Although administrative database has limited information, high dimensional propensity score approach can be exploited by applying proxy confounders of variables created from a large quantity of structured data, such as disease and prescription records (Bosco-Lévy et al., 2021). For the EHR, unstructured text data including physician's progression notes and reports of test results is drawing attention. Clinical outcomes are derived from unstructured medical records using retrospective review and automated analysis, using natural language processing and AI. This is critical for the progress of comparative effectiveness research using the RWD. Linking multiple databases enables patient follow-up for a long period of time, or covers a wide range of personal information. The data privacy regulation in each country makes data linking difficult or at times, even impossible. As a countermeasure for medical research, in Japan, "Act on Anonymized Medical Data That Are Meant to Contribute to Research and Development in the Medical Field" (Next Generation Medical Infrastructure Law) was enforced in 2018 that allows certified enterprises to deal with identifiable medical information from multiple facilities.

In drug development, pharmaceutical companies and regulatory authorities consider utilizing an external control arm for a non-randomized clinical trial of a single arm when randomization may not be feasible or ethical (Nishioka et al., 2021;

US Food & Drug Administration, 2018) Bias is greater problematic in comparison between treatment arms from clinical trial and the RWD, than the comparison between treatment arms within a database of the RWD. However, there is a strong need of the external control arm in rare disease area, and regulatory guidance has been issued in several countries. In Japan, the Ministry of Health, Labour, and Welfare (MHLW) and the Pharmaceuticals and Medical Devices Agency (PMDA) have been working to promote the RWD use for regulatory decision making. They have issued several guidelines, including the RWD use for post-marketing surveillance and registry data use for drug applications (Nishioka et al., 2021; Ishii et al., 2021).

The RWD provides new aspects on medical research. Research using the RWD provides results with an increased speed. It is based on larger data than research using primary data collection, with a relatively low cost, once the platform is established. Demands of rapid RWE in post-marketing safety surveillance are increasing considerably after the COVID-19 pandemic (Naidoo et al., 2021). For example, a study to explore the frequency and severity of myocarditis after COVID-19 vaccination used the data until May 24, 2021 from the database of Clalit Health Services in Israel, and the results were published in the *New England Journal of Medicine* just after about 4.5 months from the data period (Barda et al., 2021). Another aspect is new technologies as means of collecting the RWD novel outcomes created by digital devices, such as mobility and sleep by wristwatch type wearable and ECG by skin patch, yielding new values of drug effectiveness. These data from new technologies provide opportunities to apply new methodologies of data science dealing with large data of intensive longitudinal time periods (Izmailova et al., 2018).

There still exists large potential for improvement in the ways of demonstrating the reliability or degree of bias, and the uncertainty about the evidence obtained from the new types of RWD. We would rather quantify them for considering the effect on decision making, than emulating the inference of RCTs. Moreover, the RWD reflect “real world” of local regions, and it is important to understand the local healthcare system and clinical practices to evaluate values for reimbursement and pricing. Therefore, Statistics and data science have opportunities for contributing toward presenting new methodologies for the RWD.

5 Conclusion

The RWD is not new, but rapidly evolving in terms of data source, digital devices for data collection, application fields, and regulations. Statistics and data science should be updated corresponding to these rapid changes for various sources and settings in the RWD. We emphasize on improving methodologies for explaining relevance of the obtained evidence including biases and uncertainty.

Author contribution 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 this version to be published.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability statement No data was generated and analyzed.

Declarations

Conflict of interest KT and NY are employees of Pfizer Japan and shareholders of Pfizer.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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/4.0/>.

References

- Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, 9, 211–217. <https://doi.org/10.2147/JMDH.S104807>
- Austin, P. C. (2007). Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. *The Journal of Thoracic and Cardiovascular Surgery*, 134(5), 1128–1135. <https://doi.org/10.1016/j.jtcvs.2007.07.021>
- Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28), 3661–3679. <https://doi.org/10.1002/sim.6607>
- Bakouny, Z., & Patt, D. A. (2021). Machine learning and real-world data: More than Just Buzzwords. *JCO Clinical Cancer Informatics*, 5, 811–813. <https://doi.org/10.1200/CCI.21.00092>
- Barda, N., Dagan, N., Ben-Shlomo, Y., Kepten, E., Waxman, J., Ohana, R., Hernán, M. A., Lipsitch, M., Kohane, I., Netzer, D., Reis, B. Y., & Balicer, R. D. (2021). Safety of the BNT162b2 mRNA Covid-19 vaccine in a nationwide setting. *The New England Journal of Medicine*, 385(12), 1078–1090. <https://doi.org/10.1056/NEJMoa2110475>
- Baumfeld, E. A., Reynolds, R., Caubel, P., Azoulay, L., & Dreyer, N. A. (2020). Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiology and Drug Safety*, 29(10), 1201–1212. <https://doi.org/10.1002/pds.4932>
- Bica, I., Alaa, A. M., Lambert, C., & van der Schaar, M. (2021). From real-world patient data to individualized treatment effects using machine learning: Current and future methods to address underlying challenges. *Clinical Pharmacology and Therapeutics*, 109(1), 87–100. <https://doi.org/10.1002/cpt.1907>
- Bosco-Lévy, P., Debouverie, M., Brochet, B., Guillemin, F., Louapre, C., Maillart, E., Heinzlef, O., Lignot, S., Diez, P., Abouelfath, A., Lassalle, R., Blin, P., & Droz-Perroteau, C. (2021). Comparative effectiveness of dimethyl fumarate in multiple sclerosis. *British Journal of Clinical Pharmacology*. <https://doi.org/10.1111/bcp.15071> Advance online publication.
- Butala, N. M., Makkar, R., Secemsky, E. A., Gallup, D., Marquis-Gravel, G., Kosinski, A. S., Vemulapalli, S., Valle, J. A., Bradley, S. M., Chakravarty, T., Yeh, R. W., & Cohen, D. J. (2021). Cerebral embolic protection and outcomes of transcatheter aortic valve replacement: Results from the Transcatheter Valve Therapy Registry. *Circulation*, 143(23), 2229–2240. <https://doi.org/10.1161/CIRCULATIONAHA.120.052874>
- Clare, P. J., Dobbins, T. A., & Mattick, R. P. (2019). Causal models adjusting for time-varying confounding: A systematic review of the literature. *International Journal of Epidemiology*, 48(1), 254–265. <https://doi.org/10.1093/ije/dyy218>
- Desai, R. J., Ashton, C. M., Deswal, A., Morgan, R. O., Mehta, H. B., Chen, H., Aparasu, R. R., & Johnson, M. L. (2012). Comparative effectiveness of individual angiotensin receptor blockers on risk

- of mortality in patients with chronic heart failure. *Pharmacoepidemiology and Drug Safety*, 21(3), 233–240. <https://doi.org/10.1002/pds.2175>
- European Medicines Agency. (2020). *Guideline on registry-based studies*. <https://www.ema.europa.eu/en/guideline-registry-based-studies>. Accessed 1 Dec 2021.
- Evans, R. S. (2016). Electronic health records: Then, now, and in the future. *Yearbook of Medical Informatics*. <https://doi.org/10.15265/IYS-2016-s006>
- Faries, D. E., Haro, J. M., Leon, A., & Obenchain, R. L. (2014). *Analysis of observational health care data using SAS*. SAS Institute.
- Feinberg, B. A., Gajra, A., Zettler, M. E., Phillips, T. D., Phillips, E. G., Jr., & Kish, J. K. (2020). Use of real-world evidence to support FDA approval of oncology drugs. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 23(10), 1358–1365. <https://doi.org/10.1016/j.jval.2020.06.006>
- Ford, I., & Norrie, J. (2016). Pragmatic trials. *The New England Journal of Medicine*, 375(5), 454–463. <https://doi.org/10.1056/NEJMra1510059>
- Hernán, M. A. (2021). Causal analyses of existing databases: No power calculations required. *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2021.08.028> Advance online publication.
- Hiramatsu, K., Barrett, A., Miyata, Y., PhRMA Japan Medical Affairs Committee Working Group 1. (2021). Current status, challenges, and future perspectives of real-world data and real-world evidence in Japan. *Drugs Real World Outcomes*, 8(4), 459–480. <https://doi.org/10.1007/s40801-021-00266-3>
- Ishii, M., Sakamoto, W., Togo, K., Nakazawa, T., Shimazaki, T., Tanaka, M., Ote, T., & Matsuzawa, H. (2021). Regulations and case study of real world data for regulatory decision-making in Japan and the US. *Regulatory Science of Medical Products*, 11(1), 27–41. <https://doi.org/10.14982/rsmp.11.27> Only Japanese text available.
- Izmailova, E. S., Wagner, J. A., & Perakslis, E. D. (2018). Wearable devices in clinical trials: Hype and hypothesis. *Clinical Pharmacology and Therapeutics*, 104(1), 42–52. <https://doi.org/10.1002/cpt.966>
- Kaneyama, T., Sakaguchi, M., Nakajima, A., Aoki, A., Shiragasawa, T., Niwa, S., Matsushita, Y., Miyazaki, M., Yoshinaga, T., & Kimura, T. (2017). Real world data: Current scientific activities in pharmaceutical industry. *Regulatory Science of Medical Products*, 7(3), 225–236. <https://doi.org/10.14982/rsmp.7.225> Only Japanese text available.
- Kohsaka, S., Morita, N., Okami, S., Kidani, Y., & Yajima, T. (2021). Current trends in diabetes mellitus database research in Japan. *Diabetes Obesity and Metabolism*, 23(Suppl. 2), 3–18. <https://doi.org/10.1111/dom.14325>
- Kurz, C. F. (2021). Augmented inverse probability weighting and the double robustness property. *Medical Decision Making: an International Journal of the Society for Medical Decision Making*. <https://doi.org/10.1177/0272989X211027181>
- Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*, 4, 47–55. <https://doi.org/10.2147/RMHP.S12985>
- Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labour and Welfare, Ministry of Economy, Trade and Industry (2021) Ethical Guidelines for Medical and Biological Research Involving Human Subjects. https://www.lifescience.mext.go.jp/bioethics/seimeikagaku_igaku.html. Accessed 1 Feb 2022 (**Only Japanese text available**)
- Mueller, S., Gottschalk, F., Groth, A., Meeraus, W., Driessen, M., Kohlmann, T., & Wilke, T. (2018). Primary data, claims data, and linked data in observational research: The case of COPD in Germany. *Respiratory Research*, 19(1), 161. <https://doi.org/10.1186/s12931-018-0865-1>
- Nabhan, C., Klink, A., & Prasad, V. (2019). Real-world evidence-what does it really mean? *JAMA Oncology*, 5(6), 781–783. <https://doi.org/10.1001/jamaoncol.2019.0450>
- Naidoo, P., Bouharati, C., Rambiritch, V., Jose, N., Karamchand, S., Cilton, R., & Leisegang, R. (2021). Real-world evidence and product development: Opportunities, challenges and risk mitigation. *Wiener Klinische Wochenschrift*, 133(15–16), 840–846. <https://doi.org/10.1007/s00508-021-01851-w>
- Nishioka, K., Makimura, T., Ishiguro, A., Nonaka, T., Yamaguchi, M., & Uyama, Y. (2021). Evolving acceptance and use of RWE for regulatory decision making on the benefit/risk assessment of a drug in Japan. *Clinical Pharmacology and Therapeutics*. <https://doi.org/10.1002/cpt.2410>. Advance online publication
- Office of Civil Rights, Department of Health and Human Services. (2002). *Standards for privacy of individually identifiable health information: final rules*. Fed Regist, 67:53182–53272. <https://aspe.hhs.gov/standards-privacy-individually-identifiable-health-information>. Accessed 1 Dec 2021.

- Personal Information Protection Commission, Government of Japan. (2020). Act on the Protection of Personal Information. https://www.ppc.go.jp/files/pdf/APPI_english.pdf. Accessed 1 Dec 2021.
- Pharmacoepidemiology & Database Taskforce, Japanese Society for Pharmacoepidemiology. (2021). Survey of Japanese databases in Japan available for clinical/pharmacoepidemiology. <http://www.jspe.jp/committee/020/0210/>. Accessed 1 Dec 2021.
- Qi, L., & Sun, Y. (2014). Missing data approaches for probability regression models with missing outcomes with applications. *Journal of Statistical Distributions and Applications*, 1, 23. <https://doi.org/10.1186/s40488-014-0023-3>
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of American Statistical Association*, 89(427), 846–866. <https://doi.org/10.2307/2290910>
- Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins.
- Schneeweiss, S., & Paterno, E. (2021). Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocrine Reviews*, 42(5), 658–690. <https://doi.org/10.1210/endrev/bnab007>
- Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., LaVange, L., Marinac-Dabic, D., Marks, P. W., Robb, M. A., Shuren, J., Temple, R., Woodcock, J., Yue, L. Q., & Califf, R. M. (2016). Real-world evidence: What is it and what can it tell us? *The New England Journal of Medicine*, 375(23), 2293–2297. <https://doi.org/10.1056/NEJMsb1609216>
- Strom, B. L., Kimmel, S. E., & Hennessy, S. (2019). *Pharmacoepidemiology*. John Wiley & Sons.
- Suissa, S. (2008). Immortal time bias in pharmaco-epidemiology. *American Journal of Epidemiology*, 167(4), 492–499. <https://doi.org/10.1093/aje/kwm324>
- The European Parliament and the Council of the European Union EUR-Lex. (2019). *General Data Protection Regulation (EU) 2016/679 (GDPR)*. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Accessed 1 Dec 2021.
- Togo, K., Kawamatsu, S., Kiguchi, R., & Imai, Y. (2019). Utilizing real world data in drug development: Expectations from pharmaceutical companies. *Japanese Journal of Pharmacoepidemiology*, 24(1), 19–30. <https://doi.org/10.3820/jjpe.24.19>
- US Food and Drug Administration. (2018). *Framework for FDA's Real-World Evidence Program*. <https://www.fda.gov/media/120060/download>. Accessed 1 Dec 2021.
- Vansteelandt, S., & Sjolander, A. (2016). Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. *Epidemiologic Methods*, 5(1), 37–56. <https://doi.org/10.1515/em-2015-0005>
- Weintraub, W. S., Grau-Sepulveda, M. V., Weiss, J. M., O'Brien, S. M., Peterson, E. D., Kolm, P., Zhang, Z., Klein, L. W., Shaw, R. E., McKay, C., Ritzenthaler, L. L., Popma, J. J., Messenger, J. C., Shahian, D. M., Grover, F. L., Mayer, J. E., Shewan, C. M., Garratt, K. N., Moussa, I. D., Dangas, G. D., et al. (2012). Comparative effectiveness of revascularization strategies. *The New England Journal of Medicine*, 366(16), 1467–1476. <https://doi.org/10.1056/NEJMoa1110717>
- Wirth, F. N., Meurers, T., Johns, M., & Prasser, F. (2021). Privacy-preserving data sharing infrastructures for medical research: Systematization and comparison. *BMC Medical Informatics and Decision Making*, 21(1), 242. <https://doi.org/10.1186/s12911-021-01602-x>
- Yang, S., Eaton, C. B., Lu, J., & Lapane, K. L. (2014). Application of marginal structural models in pharmacoepidemiologic studies: A systematic review. *Pharmacoepidemiology and Drug Safety*, 23(6), 560–571. <https://doi.org/10.1002/pds.3569>
- Yao, X. I., Wang, X., Speicher, P. J., Hwang, E. S., Cheng, P., Harpole, D. H., Berry, M. F., Schrag, D., & Pang, H. H. (2017). Reporting and guidelines in propensity score analysis: A systematic review of cancer and cancer surgical studies. *Journal of the National Cancer Institute*, 109(8), djw323. <https://doi.org/10.1093/jnci/djw323>
- Zakrisson, T. L., Austin, P. C., & McCredie, V. A. (2018). A systematic review of propensity score methods in the acute care surgery literature: Avoiding the pitfalls and proposing a set of reporting guidelines. *European Journal of Trauma and Emergency Surgery : Official Publication of the European Trauma Society*, 44(3), 385–395. <https://doi.org/10.1007/s00068-017-0786-6>