

**Guidance on Using Real-World Evidence to
Support Drug Development and Regulatory
Evaluation
(For Trial Implementation)**



CDRI
三方协调委员会

Center for Drug Evaluation, NMPA

January, 2020

Table of Contents

1. INTRODUCTION	1
1.1 Background and Purpose.....	1
1.2 Progress in the Development of Regulations or Guidelines by Domestic and Foreign Regulatory Agencies.....	3
2. DEFINITIONS IN REAL-WORLD RESEARCH	4
2.1 Real-World Data	6
2.2 Fitness of Data	7
2.3 Real-World Evidence	9
3. REAL-WORLD EVIDENCE SUPPORTING DRUG REGULATORY DECISIONS	10
3.1 Provide Efficacy and Safety Evidence to Support the Registration of New Drugs	10
3.2 Provide Evidence for Label Modifications to Marketed Drugs.....	11
3.3 Provide Evidence for Post-marketing Requirements or Drug Re-evaluation.....	11
3.4 Summary and Clinical Development of Distinguished Veteran Traditional Chinese Medicine Doctors' Experience Prescription and Traditional Chinese Medicine Preparation from Medical Institutions.....	12
3.5 Other Applications to Support Regulatory Decision Making	15
4. THE BASIC DESIGN OF REAL-WORLD RESEARCH	16
4.1 Pragmatic Clinical Trials.....	16
4.2 Single-arm Trials Using Real-world Data as Control.....	18
4.3 Observational Studies.....	19
5. EVALUATION OF REAL-WORLD EVIDENCE	20
5.1 Real-world Evidence and the Clinical Questions it Supports	20
5.2 Transform Real-world Data to Real-world Evidence	20
6. COMMUNICATION WITH THE REGULATORY AGENCY	21

References	23
Appendix 1: Glossary for Real-World Research	25
Appendix 2: Examples for Real-World Evidence Application	28
Example 1: Use real-world evidence to support new indications	28
Example 2: Use real-world evidence to support extended drug combination	28
Appendix 3: Common Statistical Analysis Methods for Real-World Research	30
A3.1 Descriptive and Unadjusted Analyses	30
A3.2 Adjusted Analyses	30
A3.3 Missing Data Consideration	34
A3.4 Sensitivity Analyses and Quantitative Analysis of Bias	36



ARL
三方协调委员会

Guidance on Using Real-World Evidence to Support Drug Development and Regulatory Evaluation (For Trial Implementation)

1. INTRODUCTION

1.1 Background and Purpose

Randomized Controlled Trials (RCTs) are generally considered the gold standard for evaluating drug safety and efficacy and are widely used in clinical research. With strictly controlled trial eligibility criteria and the utilization of randomization, RCTs minimize the impact of other factors that potentially affect the evaluation of treatment effect, and hence result in more definitive conclusions and derive more reliable evidence. However, RCTs have limitations: 1) challenges exist when extrapolating the results of an RCT to real clinical practice, such as: stringent entry criteria may reduce the representativeness of the trial population to the target population, the standard trial interventions may not be completely consistent with real-world clinical practice, the limited sample size and short follow-up time leads to insufficient evaluation of rare adverse events, etc.; 2) for certain disease areas, such as some rare and major life-threatening diseases that lack effective treatments, conventional RCTs may be difficult to implement; 3) conventional RCTs may be inefficient in time and cost. Therefore, how to use real-world evidence (RWE) during drug R&D and registration to evaluate drug safety and efficacy has become a common and challenging question for global regulatory agencies, the pharmaceutical industry, and academia.

First, there is a need to clarify the definition, scope, and content of RWE on a conceptual level.

Secondly, whether real-world data (RWD) are fit for answering scientific questions of clinical interest, whether and how RWD can provide sufficient support in generating RWE, etc all face questions that need to be discussed and addressed (e.g., data sources, data standards, data quality, data sharing mechanism, data infrastructure, etc.) and, in the meanwhile, put forward an urgent need for the formulation of guidelines.

Third, the approaches for utilizing RWD need to be streamlined. RWE stems from the correct and thorough analysis of RWD. The analysis methods used are mainly for causal inference, which often requires more complex models, assumptions, and even applications of artificial intelligence and machine learning. All these bring higher requirements for relevant personnel.

Forth, the scope of RWE application remains to be clarified. RWE and evidence from conventional RCTs both can serve as an integral part of the evidence needed for regulatory decisions, supporting the formation of a comprehensive, complete and rigorous evidence chain to further improve the efficiency and scientific validity of drug development and regulation. Therefore, it is necessary to clearly define the scope of RWE application according to the stage of drug development, and in the meanwhile adopt appropriate adjustment as the actual conditions evolve over time.

In the review of the above, this guideline aims to, in support of drug development and regulatory decision making, provide clarity on the definition of RWE, guide the collection and fitness evaluation of RWD, outline the use and scope of RWE, explore the basic principles for the evaluation of RWE, and consequently provide a reference for the industry and regulatory authority when utilizing RWE to support drug development. This guideline only represents the current views and understandings and will be revised and improved on an ongoing basis as the research continues.

1.2 Progress in the Development of Regulations or Guidelines by Domestic and Foreign Regulatory Agencies

In 2009, the American Recovery and Reinvestment Act played a significant role in promoting Comparative Effectiveness Research (CER). Accordingly, the use of real-world research (RWR, or real-world study RWS) was further expanded given CER's context of the real-world environment.

In December 2016, the United States passed the 21st Century Cures Act (the Act), encouraging the Food and Drug Administration (FDA) to accelerate the development of pharmaceutical products by conducting research in the use of real-world evidence. Under the support of the Act, during 2017-2019 the FDA issued a series of guidelines, namely *Use of Real-World Evidence to Support Medical Device Regulatory Decisions*, *Guidelines for the Use of Electronic Health Record Data in Clinical Research*, *Framework for Real World Evidence Solutions*, and *Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics Guidance for Industry*.

In 2013, the European Medicines Agency (EMA) participated in the GetReal Initiative and committed to develop new methodologies in collecting and integrating RWE for earlier consideration in the decision-making process of drug development and healthcare. In 2014, EMA also launched the Adaptive Licensing Pilot to assess the feasibility of using RWD including data from observational research to assist regulatory decision-making. In 2017, the Heads of Medicines Agencies (HMA) and EMA jointly established a Big Data working group, aiming at enhancing regulatory decision-making with improved evidence standards. There,

RWE is a subset of big data, including data from electronic health records, registry systems, hospital records, and health insurance.

At the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), Japan's Pharmaceuticals and Medical Devices Agency (PMDA) proposed a new discussion topic on the technical requirement for more efficient use of RWD to conduct post-marketing pharmacoepidemiology research.

In fact, the global use of RWD to evaluate the safety of medical products has accumulated a wealth of practical experience. For example, in 2008, the US FDA launched a sentinel program to use existing electronic health data to actively monitor the safety of medical products in the market.

In China, the systematic use of RWE to support regulatory decision-making is still at the initial stage. However, the national regulatory agency has already begun to utilize RWE in the review practices (refer to Appendix 2 for cases).

2. DEFINITIONS IN REAL-WORLD RESEARCH

Real-world research refers to a process of, based on preset clinical questions, collecting data related to the health of research subjects in a real-world setting (i.e., the RWD) or utilizing the summary of the RWD, through analyses, to obtain clinical evidence of the drug usage and their potential benefit-risk (i.e., the RWE) (as was shown in Figure 1).

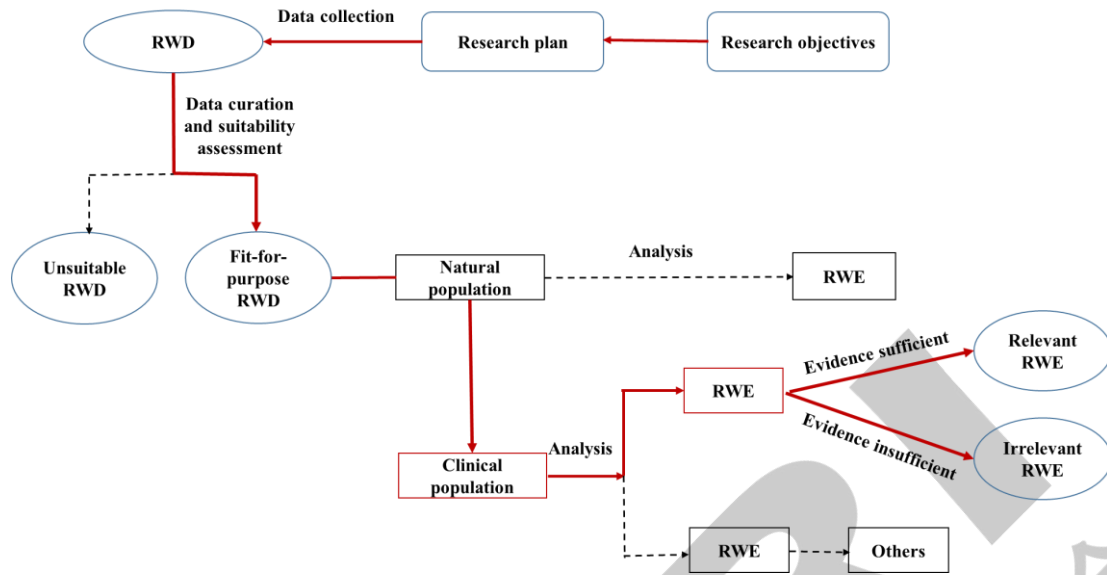


Figure 1 Pathways for RWR to support drug regulatory decisions (solid line)

The RWE generated by RWR can be used to support drug development and regulatory decisions, as well as other scientific purposes (such as clinical decisions that are not intended for registration purposes). This guideline mainly focuses on the RWR, based on the clinical population, that is used to support drug regulatory decisions. In individual cases, such as vaccines or other preventive care products for the healthy population, the guideline may also cover a wider range of the natural population.

Generally, the RWR can be categorized into either non-interventional (observational) studies or interventional studies. The former includes retrospective and prospective observational studies that offer no administration of any intervention. Patients' diagnosis and treatment, disease management, and information collection are completely from regular medical practice. On the other hand, the biggest difference in the latter is the active administration of certain interventions, such as in the situation of a Pragmatic Clinical Trial (PCT). Due to the diversity of RWR, the complexity of design and analysis methods, and the uncertainty in

result interpretation, higher requirements are put forward for the evaluation of drug safety and effect as well as for regulatory decisions.

2.1 Real-World Data

2.1.1 Definition

Real-world data (RWD) refer to a variety of data, collected on a daily basis, that are related to the patient's health status and/or diagnosis and healthcare processes. Not all RWD, but only those that satisfy fitness criteria, can be analyzed to generate RWE.

2.1.2 Source of real-world data

Common sources of RWD include, but are not limited to:

(1) Health Information System (HIS): similar to the electronic health record, digital patient records include structured and unstructured data fields, such as patient demographics, clinical characteristics, diagnosis, treatment, laboratory tests, safety and clinical outcomes.

(2) Medical Insurance System: structured data such as basic patient information, medical service utilization, diagnosis, prescriptions, billing, medical expenses, and planned health care.

(3) Disease Registry System: a database of patients with specific (usually chronic) diseases, often derived from a cohort registry of the disease population in the hospital.

(4) China ADR Sentinel Surveillance Alliance (CASSA): the use of electronic data from medical institutions to establish an active monitoring and evaluation system for the safety of drugs and medical devices.

(5) Natural population cohort and special disease cohort database: the (to be) established natural population cohort and special disease cohort database.

(6) Omics-related databases: databases that collect information on the physiology, biology, health, behavior, and possible environmental interactions of patients, such as pharmacogenomics, metabolomics, and proteomics.

(7) Death registration database: a database formed by death registries jointly confirmed by hospitals, centers for disease control and prevention (CDC), and department of household registration.

(8) Patient-reported outcome: self-assessed or measured data that are self-administered by the patient.

(9) Mobile devices: medical mobile devices such as wearable devices that measure relevant data.

(10) Other specific data sources: related data generated by medical institutions in certain geographic regions according to corresponding policies and regulations by importing drugs approved overseas to address urgent medical needs and for specific medical purposes; databases created for special purposes, such as statutory infectious disease databases, national immunization program databases, etc.

2.1.3 Data standards

Uniform data standards ensure the predictability and consistency of the submitted data and allow the information to be shared with other databases. The submitted data should have uniform standards for the planning of data standards, the collection, coding and storage of data, the format of data to be analyzed, the verifiability and traceability of data, and the format of electronic submission.

2.2 Fitness of Data

The fitness (fit-for-purpose) of real-world data is mainly assessed by data relevance and reliability.

2.2.1 Relevance

To assess whether RWD are closely related to clinical questions of interest, important factors to be considered include, but are not limited to:

(1) The inclusion of important variables and information related to clinical outcomes, such as drug exposure, patient demographic and clinical characteristics, covariates, follow-up duration, outcome variables, etc.;

(2) Whether the definition of clinical outcome is accurate and the corresponding clinical meaning is clear;

(3) Whether the patients in the RWD are representative of the target population of the study;

(4) Whether sufficient sample size and follow-up duration exist to demonstrate the effectiveness and obtain sufficient potential safety events.

2.2.2 Reliability

The reliability of real-world data is mainly evaluated in terms of data completeness, accuracy, transparency, and quality assurance.

(1) **Completeness:** missing data problems, including missing variables and missing variable values, are inevitable in the real-world setting. When the proportion of missing exceeds a certain limit, especially in relation to key variables of the research. For example, when important prognostic covariates are unobserved or missing, there is an increasing level of uncertainty about the research conclusion. In this case, careful consideration should be given to whether the data can support the generation of RWE.

(2) **Accuracy:** the accuracy of the data is critically important, and often needs to be verified against authoritative sources. Data elements and algorithms for transforming data should be correct. The accuracy of the

data is also reflected in how consistent and reasonable the data are. Consistency means that relevant data standards, formats, and calculation methods need to be consistent within the database; the reasonability of data speaks to the uniqueness of variable values, reasonable data distributions, the expected interdependence between variables, and whether the time-varying variables change as expected.

(3) Transparency: the source of data and the entire process of data collection, curation, and governance should be transparent and traceable. In particular, key exposure information, covariates, and outcome variables should be able to trace back to source data. The transparency of data also includes the accessibility of data, the sharing of information between databases, and the methods for protecting patient privacy.

(4) Quality Assurance: The reliability of RWD needs to take data quality into account. Quality assurance approaches include, but are not limited to: whether a clear process and qualified personnel exist for data collection; whether a common definition framework is used, i.e., a data dictionary; whether a common time frame is followed when collecting key data points; whether a research plan, protocol, and analysis plan related to the collection of RWD are established; and whether the technical methods used for data collection are adequate, including integration of data from various sources, drug use and laboratory test records, follow-up records, the link to insurance data, and data security.

2.3 Real-World Evidence

Real-World Evidence (RWE) refers to the clinical evidence on the use and potential benefit-risk of a drug, obtained through appropriate and sufficient analysis of fit-for-purpose RWD. RWE includes evidence from retrospective/prospective observational studies or interventional studies such as pragmatic clinical trials.

3. REAL-WORLD EVIDENCE SUPPORTING DRUG REGULATORY DECISIONS

RWE may be used to support drug development regulatory decisions, including a variety of stages from pre-marketing clinical development to post-marketing re-evaluation. For example, it may provide efficacy or safety evidence to support the approval of a new product for marketing; it may also provide evidence for label modification for an approved product, including extension or modification of the indication, modification of the dose, regimen, or route of administration, addition of new target populations, inclusion of comparative efficacy information, inclusion of safety information, etc.; or it may serve as part of post-marketing requirements to support regulatory decision making, etc.

The following describes certain examples within the scope of RWE application to support drug regulatory decisions. However, it does not preclude other reasonable applications.

3.1 Provide Efficacy and Safety Evidence to Support the Registration of New Drugs

For different diseases, depending on the disease characteristics, accessibility of treatments, target populations, treatment effect, and other factors related to clinical research, it is possible to obtain drug effect and safety information through RWR and provide supportive evidence for the registration and marketing of new drugs.

Common RWR that provides efficacy and safety evidence for the registration and marketing of new drugs include: randomized clinical trials using RWD-based efficacy outcomes or safety information, including PCT designs, etc.; rare and major life-threatening diseases that lack effective treatment options, and as such have to consider single-arm clinical trials with RWD-based external control.

3.2 Provide Evidence for Label Modifications to Marketed Drugs

For drugs that are already marketed, adding new indications usually need to be supported by RCTs. However, when RCTs are not feasible or do not represent optimal study designs, RWE generated from PCTs or observational studies to support indication expansion may become more feasible and justified.

In terms of pediatric medication development, the use of RWE to support the expansion of the target population is also one of the potentially suitable situations for drug regulatory decisions.

In summary, the use of RWE to support label modifications for marketed drugs mainly include the following situations:

- Addition or modification of an indication;
- Change in dose, dosage form, or route of administration;
- Addition of new suitable patient population;
- Addition of results from comparative effectiveness studies;
- Addition of safety information;
- Other modifications to the package insert.

3.3 Provide Evidence for Post-marketing Requirements or Drug Re-evaluation

Due to factors such as limited sample size, short study duration, strict enrollment criteria, and standardization of intervention, drugs approved based on RCTs usually have limited safety information, lack of generalization of efficacy conclusions, less optimal drug regimen, and insufficient health economic benefits. As a result, there is a need to use RWD for more comprehensive assessments of these aspects of the approved drugs, and to refine the decision-making based on the evidence from real-world medical practice on a continuous basis.

3.4 Summary and Clinical Development of Distinguished Veteran

Traditional Chinese Medicine Doctors' Experience Prescription and Traditional Chinese Medicine Preparation from Medical Institutions

For the clinical development of drugs with human-use experience (e.g., distinguished veteran traditional Chinese medicine doctors' experience prescription and traditional Chinese medicine preparation from medical institutions, etc.), given that the prescription and manufacturing process are established and stable, attempts can be made to combine RWR and RCTs to explore new pathways for clinical development.

There can be multiple ways to utilize RWE to support the clinical development of traditional Chinese medicine with human-use experience. Different R & D strategies shall be selected according to the characteristics of the product, existing clinical utilization, and data fitness. For example, the possibility of using (retrospective and prospective) observational studies to replace routine Phase I and/or Phase II clinical trials in initial exploration of clinical efficacy and safety can be explored; based on observational studies, drug efficacy may be further confirmed through PCTs or RCTs, and thereby provide supportive evidence for product registration. If, based on evaluation, there exists fit-for-purpose RWD with high quality and the RWE generated by well-designed observational studies is considered scientific and sufficient, then such RWE may be used as evidence to directly support the drug registration after communication with the regulatory authority.

There can also be multiple development pathways by combining observational research with RCTs or PCTs. Figure 2 and Figure 3 are two of the possible pathways. Figure 2 illustrates the path based on the combination of observational studies and RCTs. Specifically, stage 1 starts with retrospective observational studies. At this stage, effort should be

made to collect as much as possible existing real-world data related to the use of the product including all possible covariates, develop data cleaning rules, identify possible controls, assess data quality, and conduct comprehensive and detailed analyses using appropriate statistical methods. If the retrospective observational studies show that the drug has potential benefits for patients in clinical use, it may proceed to the next stage of the development, otherwise, the process should be terminated. In stage 2, prospective observational studies can be conducted. Based on the stage 1 research, this second stage can be more carefully designed in terms of several aspects, including data acquisition and its system, data quality control, data cleaning rules, and clearer definition of controls. Once this stage 2 prospective observational research has progressed to a certain phase, and if the results of data analysis are consistent with the results of stage 1 retrospective observational studies by continuing to show clinically meaningful benefits, a third stage of RCT can be conducted in parallel. If needed, a pilot RCT may be conducted first to acquire sufficient information to support the design of the primary RCT. However, if existing evidence from previous observational studies is deemed sufficient, a confirmatory RCT may be designed and conducted directly. In terms of timing, the duration of the RCT may be covered by the stage 2 prospective observational studies, which can be completed prior to the start of the RCT, or at the same time as the RCT, or even extended for some time after the end of the RCT, in order to accumulate more RWE or for other purposes such as adding new indications or expanding the targeted population.

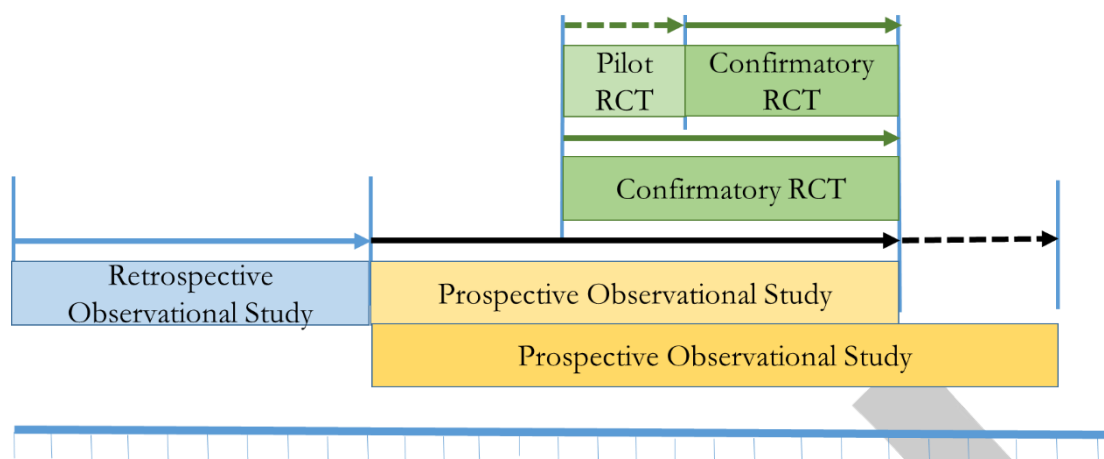


Figure 2 One of the pathways of clinical research and development of traditional Chinese medicine with existing human-use experience

Another potentially possible pathway, which combines observational studies with PCTs, is outlined in Figure 3. In the first stage, retrospective observational studies are conducted first. If it is concluded that the drug has potential benefits in clinical practice, it may proceed to the second stage, otherwise, the process should be terminated. The second stage consists of a PCT research, which provides evidence that can be used to support the evaluation of the drug's clinical efficacy and safety.

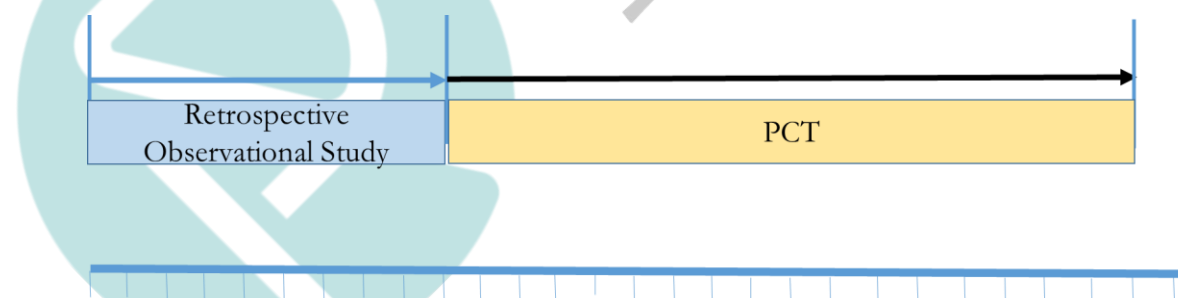


Figure 3 Another pathway of clinical research and development of traditional Chinese medicine with existing human-use experience

The clinical development of traditional Chinese medicine with human-use experience is not limited to the two pathways mentioned above. Instead, it should adopt appropriate strategies, based on the characteristics

of the product, information from basic research (e.g., toxicology studies), clinical use, and efficacy data accumulated from previous clinical practice.

3.5 Other Applications to Support Regulatory Decision Making

3.5.1 Guiding clinical trial design

Compared with other potential applications, using RWE to guide clinical trial design has more practical utilization. For example, the two potential pathways for the clinical development of traditional Chinese medicines described previously have used the RWE generated by retrospective observational studies, including the disease natural history, the disease prevalence in the target population, the effectiveness of standardized treatments, and the distribution and variation of key-related covariates in the target population, to provide a basis for the next stage study design. A more general application of RWE is to provide valid references for inclusion and exclusion criteria, parameters for sample size estimation, and determination of non-inferiority margins, etc., and hence support the assessment of trial design during regulatory evaluation.

3.5.2 Accurately identify the target population

Precision medicine aims to better predict the therapeutic benefits and risks of drugs to specific populations (subgroups), and RWE provides the possibility for the development of precision medicine. For example, due to the limited sample size, regular clinical trials often ignore or have limited power to consider subgroup effects in the research plan. This prevents important information on potential treatment responders or high-risk populations with serious side effects from being fully recognized. Since RWD often consist of big data of various types, through a thorough analysis of RWD, the treatment benefits and risks in different subgroups can be more adequately assessed, and hence RWE can be obtained to support more precise identification of the target population.

The identification of biomarker is critical for preclinical and early clinical studies of targeted therapies. Using real-world information such as omics data, public gene bank information, and related clinical data in population cohorts, real-world evidence can be generated through various contemporary data mining techniques such as machine learning, which can in consequence support the precise identification of population for targeted therapies.

4. THE BASIC DESIGN OF REAL-WORLD RESEARCH

4.1 Pragmatic Clinical Trials

Pragmatic Clinical Trials (PCT), also known as practical clinical trials, refer to clinical trials that are designed and conducted in an environment close to real-world clinical practice. They represent a type of study between RCTs and observational studies. Unlike RCTs, PCT interventions can be either standardized or non-standardized; subjects in the PCTs can be randomized or allocated per pre-defined criteria; the inclusion criteria for the subjects are often less restrict and considered more representative of the target population; the evaluation of intervention outcomes may not be limited to clinical efficacy and safety; PCT generally uses clinical endpoints, and avoids the use of surrogate endpoints that may be used in conventional RCTs; multiple control groups can be considered at the same time to reflect different standard treatment in clinical practice; placebo control is generally not considered; in most cases, no blinding is imposed, but sufficient attention should be paid in how to estimate and adjust the resulting measurement bias; data collection usually relies on the daily medical records of patients. On the other hand, unlike observational studies, PCTs are intervention studies, although the interventions are often designed with additional flexibility.

For example, a patient-centered study evaluating the benefit and long-term effectiveness of different doses of aspirin used a randomized PCT design that included patients with atherosclerotic cardiovascular disease and at high risk for ischemic events. Subjects were randomly assigned to two different doses of aspirin treatment groups (plus daily health care). The primary endpoint was a composite endpoint (integrating all-cause death, hospitalization due to non-fatal myocardial infarction, and hospitalization due to stroke) based on data from electronic health records and insurance claims databases.

The following factors should also be considered when designing a PCT: ① Are the data collected fit to support the production of RWE? ② Are therapeutic areas and interventions, etc., in line with various forms of routine clinical practice? ③ Is there a sufficient number of cases to be evaluated (especially when clinical outcomes are rare)? ④ Is the evaluation and reporting of endpoints consistent across participating trial centers and even different databases? ⑤ Is the randomization method to be used to control bias? ⑥ When blinding is not feasible, the possible impact of unblinding on outcome variables (particularly patient-reported outcomes) should be considered. Endpoints that are not impacted by treatment assignment (e.g., stroke, tumor size, etc.) may be used to minimize potential bias caused by unblinding.

Since a PCT needs to consider the impact of all potential factors, including various biases and confounding factors, its study design and statistical analysis are usually complicated, and the required sample size can be much larger than a regular RCT design. PCTs, when randomization is utilized, will reduce the impact and biases of the confounders and thus provide a generally speaking robust causal inference. Since PCTs are

conducted in a setting close to real clinical practice, the evidence obtained by PCTs is considered relatively good RWE in most cases.

4.2 Single-arm Trials Using Real-world Data as Control

Single-arm clinical trials are also a method to assess the efficacy and safety of an investigational drug. For example, recruitment can be challenging in clinical trials for certain rare diseases; randomized controlled trials often introduce ethical issues for certain major life-threatening diseases that lack effective treatment options. Therefore, in the above two cases, external control using RWD from the natural disease cohort can be considered.

External controls are primarily used in single-arm trials and can be either historical or parallel controls. Historical external controls use RWD obtained earlier as the benchmark and should consider the impact on data comparability introduced by potential differences in disease definition, diagnosis, classification, natural history, and available treatment methods over time. On the other hand, parallel external controls use disease registration data conducted in the same period for comparison. Using external controls should consider the impact of comparability of the target population on corresponding RWE. For data from patients receiving other interventions, sufficient covariates should exist to support a correct and adequate statistical analysis.

The use of external controls has limitations, mainly including different healthcare environments, changes in medical technology over time, different diagnostic criteria, different outcome measures, different baseline conditions of patients, diverse interventions, data quality, etc. These limitations result in additional challenges in the comparability of research subjects, the accuracy of research results, the reliability and extrapolation of research conclusions.

To address these limitations, it is first necessary to ensure that the collected data meet the fitness requirements of RWD. Secondly, the use of parallel external controls is generally superior to historical controls. Parallel external controls can use disease registration modes to ensure that data records are as complete and accurate as possible. Third, appropriate methods shall be adopted for statistical analysis, such as the Propensity Scores (PS) method and Virtual Matched Control method. Finally, sensitivity analyses and quantitative analyses of bias should be conducted to evaluate the impact to the analysis results by model assumptions as well as the (un)known and (un)measured confounding factors.

4.3 Observational Studies

The data collected from observational studies are close to the real world, but their most notable limitations are the existence of various biases as well as the challenges in data quality and confounding factors identification. These challenges leave the study conclusion with large uncertainty.

Whether the data collected from observational studies are fit to generate RWE to support regulatory decisions depends on a few areas of focus: ① Data characteristics: e.g., data source and quality, study population, data collection of exposure and related endpoints, consistency in data records, data curation process, description of missing data, etc.; ② Study design and analysis: e.g., the existence of appropriate positive control, the existence of unmeasured confounding factors and variation in measurement, whether the analysis method is rigorous, transparent, and compliant with regulatory requirements, etc.; ③ Robustness of results: pre-definition of sensitivity analyses, quantitative analysis of bias, and statistical diagnostic methods.

The primary method for observational studies is causal inference (see Appendix 3).

5. EVALUATION OF REAL-WORLD EVIDENCE

The evaluation of RWE should follow two main principles: whether RWE can support the clinical questions to be answered; and whether existing RWD can generate the required RWE through the scientific design of research, rigorous execution, and appropriate statistical analyses.

5.1 Real-world Evidence and the Clinical Questions it Supports

Prior to the decision to use any evidence including RWE, the clinical questions under evaluation should first be clearly defined. For example, the safety considerations for the use of drugs in combination with other drugs after marketing; the expansion of indication for approved products; and robust and reliable history or external controls for single-arm clinical trials for a rare disease. Secondly, whether the use of RWE can answer the clinical questions of interest needs to be considered, as assessed from four aspects: scientific validity (e.g., with interpretable results, reasonable assumptions, well-controlled type I error, etc.), regulatory requirements (potential conflict with other regulatory requirements, regulatory requirements for special disease areas, etc.), ethical considerations (ethical issues without using RWE?), and operational feasibility (e.g., an independent statistician to ensure blinding and avoid bias during matching; any other operational challenges, etc.). The above considerations are important criteria for evaluating the use of RWE.

5.2 Transform Real-world Data to Real-world Evidence

A few key factors need to be considered: ① The research environment and data acquisition need to be close to the real world, such as a more representative target population, diversity of interventions

compatible with clinical practice, or natural selection of interventions; ② Use of appropriate controls; ③ More comprehensive evaluation of drug effectiveness; ④ Effective bias control, such as the use of randomization, harmonization of measurement and evaluation methods, etc.; ⑤ Appropriate statistical analyses, such as the correct use of causal inference methods, reasonable handling of missing data, adequate sensitivity analyses, etc.; ⑥ The transparency and reproducibility of evidence; ⑦ Reasonable interpretation of results; ⑧ Consensus among the relevant parties.

It should be noted that all study designs, assumptions, and specific definitions and methodologies relevant to the generation of RWE should be clearly defined in advance in the study protocol. Post-hoc supplementary data citation, definition, analysis, and interpretation are generally not acceptable for regulatory decisions.

6. COMMUNICATION WITH THE REGULATORY AGENCY

The use of RWE for the purpose of drug registration requires sufficient communication with the regulatory agency for drug evaluation, to ensure that both parties reach consensus on the use of RWE and conduct of RWR.

When an applicant plans to use RWE to support drug registration, before conducting the study, the applicant should initiate a consultation application, by following the established communication procedure, to discuss with the regulatory agency for drug evaluation about the study objectives, the study design, the feasibility of using RWE, and the data collection and analysis methods.

After completing the real-world study and prior to submitting the application materials, the applicant should also apply to communicate with the regulatory agency to confirm the execution, results and conclusions of the study, as well as the requirements for the application materials.



References

- [1] Sun Yuxi, Wei Fengfang, Yang Yue. Real world evidence is used for opportunities and challenges in drug supervision and health decision-making. *China Pharmacovigilance*, 2017; 14(6): 353-358.
- [2] Yilong Wu, Xiaoyuan Chen, Zhimin Yang and others (Jieping Wu Medical Foundation, China Thoracic Oncology Research Collaboration Group). Real world guidelines for research. 2018.
- [3] The General Office of the CPC Central Committee and the General Office of the State Council on Issuing the Opinions on Deepening the Reform of the Evaluation and Approval Systems and Encouraging Innovation on Drugs and Medical Devices. 2017.
- [4] ADAPTABLE Investigators. . Aspirin Dosing: a Patient-Central Trial Assessment Benefits and Long-Term Effectiveness (ADAPTABLE) study protocol. http://pcornet.org/wp-content/uploads/2015/06/ADAPTABLE-Protocol-Final-Draft-6-4-15_for-post_06-26-.pdf[J]. Published June, 2015, 5.
- [5] Berger M, Daniel G, Frank K, et al. A frame work for regulatory use of real-world evidence[J]. White paper prepared by the Duke Margolis Center for Health Policy, 2017, 6.
- [6] Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for europe[J]. *Clinical pharmacology and therapeutics*, 2019, 106 (1) : 36.
- [7] Dreyer NA. Advancing a framework for regulatory use of real-world evidence: when real is reliable[J]. *Therapeutic innovation & regulatory science*, 2018, 52 (3) : 362-368.
- [8] Egger M, Moons K G M, Fletcher C, et al. GetReal: from efficacy in clinical trials to relative effectiveness in the real world[J]. *Research synthesis methods*, 2016, 7 (3) : 278-281.
- [9] Ford I, Norrie J. Pragmatic trials[J]. *N Engl J Med*, 2016, 375 (5) : 454-463.
- [10] Institute of Medicine 2009. Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press. <https://doi.org/10.17226/12648>.
- [11] James S. Importance of post-approval real-word evidence[J]. *European Heart Journal-Cardiovascular Pharmacotherapy*, 2018; 4 (1) :10-11.
- [12] Kohl S. Joint HMA/EMA task force on big data established[J]. *Eur J Hosp Pharm*, 2017, 24 (3) : 180-190.
- [13] Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data[M]. Springer Science & Business Media, 2011.
- [14] Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews[J]. *Value in health*, 2017, 20 (7) : 858-865.
- [15] Olariu E, Papageorgakopoulou C, Bovens S M, et al. Real world evidence in Europe: a snapshot of its current status[J]. *Value in Health*, 2016, 19 (7) : A498.
- [16] Roland M, Torgerson D J. Understanding controlled trials: What are pragmatic trials?[J]. *BMJ*, 1998, 316 (7127) : 285.
- [17] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us[J]. *N Engl J Med*, 2016, 375 (23) : 2293-2297.
- [18] Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials[J]. *JAMA*, 2014, 311 (23) : 2381-2382.
- [19] US Food and Drug Administration. Framework for FDA’s real-world evidence program. December

2018[J]. 2019.

- [20] Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12 (13) -EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
- [21] Von Elm E, Altman D G, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies[J]. Annals of internal medicine, 2007, 147 (8) : 573-577.



AARI
三方协调委员会

Appendix 1: Glossary for Real-World Research

Patient Registry: A system of collecting standard clinical and other data, using observational research approaches, to evaluate specific disease, condition, or specific outcome in the exposed population, for one or more predefined scientific, clinical, or policy objectives.

Single-arm/One-arm Clinical Trial: A non-randomized clinical trial where only the experimental group is set up. A single-arm trial usually uses external controls based on historical data or in a parallel manner.

Observational Study: A study that explores the outcomes of a treatment or exposure in natural or clinical populations without active intervention, based on specific research objectives.

Retrospective Observational Study: An observational study with target population identified at the study start, and based on historical data (generated before the start of the study).

Prior Event Rate Ratio (PERR): For two treatment groups (A: exposed, and B: non-exposed) and certain events, the PERR refers to the ratio of 1) post-treatment event rate ratio of exposed vs. non-exposed groups, and 2) pre-treatment event rate ratio of exposed vs. non-exposed groups. This ratio is used to estimate the treatment effect after eliminating the impact of unmeasured confounders.

Clinical Population: Populations undergoing medical treatment and observation and/or clinical investigation, including those participating in drug clinical trials.

Clinical Trial: An interventional clinical study in which one or more interventions (possibly including placebo or other controls) are prospectively assigned to human subjects to assess the impact of these interventions on health-related biomedical or behavioral outcomes.

Prospective Observational Study: An observational study with target population identified at the study start, where exposure/treatment and outcome data are to be collected prospectively.

Comparative Effectiveness Research: A research method suitable for most research types, by considering individuals or the population in an environment as close as possible to the real world, that evaluates through

comparison the clinical effectiveness and safety, social effects, or economic benefits of a particular intervention. Such evaluation helps key stakeholders such as patients, doctors, policymakers, and service consumers to improve healthcare services so that the most appropriate interventions or strategies can achieve the optimal outcomes in the most appropriate target population and timing.

Pragmatic Clinical Trial (PCT)/Pragmatic Trial: Also known as practical clinical trials, are a type of clinical trials that are designed and conducted in an environment as close as possible to the clinical real world. It is a type of research between RCTs and observational studies.

Data Standard: A set of rules on how to construct, define, format, or exchange specific types of data between computer systems. Data standards allow the submission of information to be predictable and consistent, and in forms that information technology systems or scientific tools can use.

Data Curation: The management and processing of original (raw) data so that they become fit to the statistical analyses corresponding to specific clinical research questions. This includes, at a minimum, several aspects such as data collection (possibly from multiple data sources), data security processing, data cleaning (logic edit checks, outlier management, and data integrity processing, etc.), data import and structure (common data models, normalization, natural language processing, medical coding, derivative sites, etc.), and data transfer.

Randomized Controlled Trial (RCT): A clinical trial that utilizes a randomization method in subject assignment to experimental and appropriate control groups.

External Control: The control in clinical trials is established based on data outside the scope of the study, such as real-world data, to evaluate the effects of the interventions under investigation. External controls can be historical data or data obtained during the same period of time in a parallel manner.

Medical Claims Data: A compilation of information on medical claims submitted to insurance companies for reimbursement of medical expenses for treatments and other interventions by healthcare providers.

Causal Inference: An inferential action, often based on real-world data, that characterizes the causal relationship between interventions or

exposures to clinical or health outcomes, taking into account the effects of various covariates and measured or unmeasured confounders and controlling possible biases. Appropriate statistical models and analytical methods should be used to establish the conclusions and causal relationship.

Real-World Data (RWD): A variety of routinely collected data that are related to the patient's health status and/or diagnosis and healthcare processes. Not all RWD, but only those that satisfy fitness criteria, can be analyzed to generate RWE.

Real-World Research/Study (RWR/RWS): A process of, based on preset clinical questions, collecting data related to the health status and/or the diagnosis and treatment of research subjects in a real-world setting (i.e., the real-world data) or utilizing the summary derived from the RWD, through analyses, to obtain clinical evidence of the drug usage and their potential benefit-risk (i.e., the real-world evidence).

Real-World Evidence (RWE): The clinical evidence on the use and potential benefit-risk of a drug, obtained through adequate and sufficient analysis of fit-for-purpose real-world data.

Natural Population: Also known as the full population, which includes both clinical populations and non-clinical populations.

Intermediate Variable: A variable in the middle of a causal relationship, impacted by drug exposure while also impacting the outcome, or is associated with the outcome. The former is also called a mediator.

Appendix 2: Examples for Real-World Evidence Application

Example 1: Use real-world evidence to support new indications

The Sponsor has initiated a study of a marketed drug, based on real-world data, to evaluate its effectiveness and safety in reducing clinical osteoporotic fractures in Chinese women. The study followed good practice in real-world research, and the study protocol was made public in advance. The real-world data are from a source that well represents the study population, with a sample size of more than 40,000. The primary endpoint of the study was verified through the review of medical records and analyzed primarily using propensity score matching. In the meanwhile, several approaches such as inverse probability weighting method and high dimensional propensity score adjustment were used for sensitivity analyses and quantitative assessment of the impact from unmeasured confounders. The results of this real-world study were similar to those of global RCTs and could be replicated by real-world data from different sources and different research institutions.

Example 2: Use real-world evidence to support extended drug combination

Bevacizumab is a humanized monoclonal antibody of the vascular endothelial growth factor (VEGF). In 2015, Bevacizumab in combination with chemotherapy (carboplatin and paclitaxel) was approved in China for the front-line treatment of patients with late-stage unresectable metastatic or recurrent non-squamous non-small cell lung cancer. However, the real-world use of chemotherapy with bevacizumab is not limited to carboplatin or paclitaxel, but also includes pemetrexed in combination with platinum, gemcitabine and cisplatin. In October 2018, bevacizumab was approved to extend the treatment regimen to a combination with platinum-based

chemotherapy, based on the strong supporting evidence from three real-world studies. These studies retrospectively analyzed patient data from three hospitals and showed that the combination of Bevacizumab with platinum-based dual drug chemotherapy significantly prolonged PFS and OS compared with chemotherapy alone, consistent with global population data without new safety issues. In addition, related real-world studies have also provided data in different patient subgroups such as those with EGFR mutations or brain metastases, confirming the effectiveness and safety of bevacizumab combination therapy from multiple perspectives



Appendix 3: Common Statistical Analysis Methods for Real-World Research

Compared with RCT studies, statistical analysis methods in real-world research are mainly methods for causal inference, where special attention is required to control or adjust for confounding effects, in order to avoid biased estimation of treatment effect. The following only provides a general description of some commonly used causal inference methods. See related references for method use and technical details (without excluding the reasonable use of other methods).

A3.1 Descriptive and Unadjusted Analyses

For real-world research, correct and effective descriptive statistical analysis can play an important role. For example, in disease registry cohort studies, stratified descriptive statistics of relevant covariates by levels of exposure factors can help to examine their distribution balance; in propensity score matched datasets, summary statistics by group of relevant covariates with different exposure factors can help to identify imbalances in residuals after the matching, etc. Real-world research often needs to account for possible confounders from a large number of covariates. Therefore, it is necessary to carry out extensive and comprehensive exploratory analyses of relevant subject characteristics through descriptive analyses.

A3.2 Adjusted Analyses

(1) Selection of covariates

For causal inference methods that adjust for covariates, covariate selection methods are broadly divided into two categories. One is, based on a causal network established from the exposure-to-outcome relationship, to identify risk factors, confounders, intermediate variables, time-varying

confounders, collider variables, and instrumental variables. Risk factors and confounders should be included as covariates in the model, while the inclusion of intermediate variables, collider variables, and instrumental variables should be avoided. However, for complex situations such as time-variant treatment or confounders, intermediate variables, and collision node variables may need to be adjusted. To control the additional bias introduced in such situations, attention should be paid to the use of reasonable statistical analysis methods. In practice, when a part of the causal structure is known, existing covariate selection methods can be used, based on relevant professional background knowledge, to adjust all observed baseline variables that may be associated with the outcome, known outcome-related risk factors, and all direct dependent variables for treatment or outcome. Another type of covariate selection method is based on high-dimensional variable selection, which empirically learns the correlation between variables from the data, and selects the variables related to the treatment factors and/or outcome variables. These two types of methods can also be used in combination, i.e., first, use the professional experience to identify a set of variables, and then use appropriate empirical learning methods to further select the covariates to be included in the final analysis model. This has the advantage of limiting reliance on empirical learning, reducing the risk of over-adjustment while also reducing the confounding effect. It should be noted that the covariate selection process must be open and transparent.

(2) Adjusted analysis using regression models

The potential impacts of confounding factors are often adjusted using various types of regression models, in order to estimate the effect of drug exposure. Generally, the variables to be adjusted are those that are simultaneously related to the treatment factors and outcome measures of

the study, and are located on the causal pathway before the treatment factors. The choice of regression models should be considered: whether the model assumptions are valid, whether the selection of independent variables is appropriate, and whether there is a need to use composite covariates (such as Propensity Score or Disease Risk Score), exposure variables, and the incidence rate of response variables (outcome events).

(3) Propensity score

A propensity score is defined as the probability that an observed subject receives a certain treatment (or exposure) under the observed covariate condition. Propensity score is a method of adjusting for confounding effects in the presence of multiple covariates, which comprehensively summarizes the between-group balance of all observed covariates, and effectively controls for confounding effects. Propensity-Score Matching, Stratification/Subclassification, Inverse Probability of Treatment Weighting (IPTW), and the method of including Propensity Score as the sole covariate in the statistical model for adjustment analysis are all commonly used approaches for estimation of causal effect.

When we use propensity scores for causal effect estimation, it is necessary to determine whether the distribution of covariates in patients with similar propensity scores is balanced between groups, and how much coincidence exists between groups in terms of the distribution of propensity scores. Remediation such as restricting study subjects to overlapping range of the propensity score distributions across groups may be considered in case of poor coincidence, but it should be noted that the resulting changes in the target population may generate causal effect estimates that are not applicable to the original target population. Note that the propensity score matching method can only adjust for the impact of covariates that are known and observed, and that for the unknown or

unmeasured covariates, it needs to be evaluated through sensitivity analyses. In addition, the traditional regression adjustment method and propensity score matching method each has advantages and disadvantages. The former does not guarantee that the study covariates are balanced, and the latter may lead to a decrease in sample size. Hence, further sensitivity analyses are very necessary.

(4) Disease risk score (DRS)

Disease risk scores (DRS) are similar to propensity scores and are a composite measure based on all covariates. A disease risk score is defined as the probability of an outcome event under the assumption of no treatment/exposure or specific covariate conditions. The method of estimating DRS generally falls into one of the two categories: one uses all observations of the study sample in fitting a regression model, and then take exposure (default as no exposure) and covariates and independent variables and study outcomes as the dependent variable to predict the DRS score; the other approach is to estimate DRS using only the samples without exposure, and then replace the covariates from all study samples with values in the DRS model, and calculate the corresponding DRS prediction for all study samples.

For studies where outcome events are common but treatment (exposure) factors are rare or there may be multiple levels of exposure, the DRS approach is a good option to balance baseline disease risk across groups. In particular, in case of multiple levels of treatment (exposure) factors, where some of them are rare, it is often recommended that the DRS method is considered instead of the PS method.

(5) Instrumental variables

The above-mentioned traditional methods of multiple regression, propensity score, and disease risk score can only control for confounding

factors that are measured, but not for those that are unknown or unmeasured. Instead, the instrumental variables can be used to control for unmeasured confounders, thereby estimating the causal effect of treatment and outcome without specifically adjusting for confounders/covariates.

A variable may be referred to as an instrument variable if it is related to a treatment factor and the effect on the outcome variable can only be achieved through its effect on the treatment factor, while being independent of confounders to exposure and outcome.

The biggest challenge in using instrumental variables to estimate causal effects lies in the identification of suitable instrumental variables. First, instrumental variables cannot be associated with any observed or unobserved confounders of treatment and outcome. Second, instrumental variables cannot have a direct effect on the outcome but only an indirect impact through the treatment-to-outcome pathway. Finally, instrumental variables need to be highly correlated with the treatment factor. Once instrumental variables are identified, the estimation of causal effects can utilize a two-stage least-squares approach.

A3.3 Missing Data Consideration

The missing data problem is often inevitable in real-world studies. Not only the outcome variables, but covariates may also be missing. Investigators and the Sponsor should optimize the trial design to minimize the rate of missing.

Prior to conducting the primary analysis, an attempt should be made to assess the reason for missing. Generally, there are three types of missing mechanisms: Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR). Missing completely at random means that the missing data are independent of the

measured or unmeasured covariates and outcome variables. Missing at random refers to the case that the missing data are independent of the potential outcome conditional on the measured covariates and outcome variable. Finally, if the data are missing not at random, the missing data may depend on the value of the missing data themselves, and may also be related to the measured covariates and outcome data.

For missing data problems, selecting the appropriate methods for imputation and analysis is an effective way to avoid bias and information loss, as otherwise the missingness will reduce the sample size and impact the study efficiency. Appropriate imputation methods should be determined using appropriate assumptions based on missing mechanisms and the corresponding clinical question. In general, for MCAR, analyzing only samples with complete data is reasonable; for MAR, statistical models can be constructed for prediction-based imputation, such as multiple imputation (MI), traditional regression model methods, Markov Chain Monte Carlo (MCMC) methods, Fully Conditional Specification (FCS), etc. In case of MNAR, Pattern Mixture Models (PMM) methods can be used to construct different statistical models for missing data and non-missing data. In addition, the single value imputation method can also be considered, with the advantages of simple principle and easy implementation and the disadvantages being that the results may not be valid even under MAR and the variability of missing values is not considered. As such, it is generally not recommended for primary analysis.

In observational studies with missing covariates, according to the specific pattern of missingness, a number of existing statistical methods may be considered, including complete data analysis, multiple imputation, and propensity score.

It needs to be clarified that the assumption on any of the three types of missing mechanism are generally not verifiable and can only be justified through a correct description and understanding of the data collection process.

In reality, it is difficult to identify the best or uniquely suitable method to address missing data problems, and there does not exist a method that can produce the same robust unbiased estimate as the original complete data. The key to identifying the best strategy for addressing missing data lies in the appropriate design and implementation of the research.

A3.4 Sensitivity Analyses and Quantitative Analysis of Bias

The various causal inference methods described above each has its own applicable conditions and assumptions, such as interchangeability, consistency, and positive correlation of unobserved covariates, and therefore it is necessary to perform sensitivity analyses on these assumptions to evaluate the robustness of the causal inference results. For example, for two patients with identical baseline covariates, their unobserved covariates may result in completely different probabilities of receiving treatment. Sensitivity analyses may assess the impact of unobserved covariates on the bias in the estimation of treatment effect, and assist in determining the upper and lower limits of the estimated efficacy based on the probability of receiving treatment.

The quantitative analysis of bias should ensure that the analysis process is transparent and credible, by following the steps listed below:① Combining causal structural models and observational data to identify possible biases;②Calculate the magnitude of bias and its impact on the interpretation of causal effects using causal graphs that include hypotheses;

③Combining the objectives of the study and the bias model, evaluate the magnitude and level of uncertainty of the bias by using the distribution of bias parameters.

Finally, like other confirmatory studies, the interpretation of analysis results for real-world studies should be as comprehensive, objective, accurate, and adequate as possible, not only emphasizing statistical significance (such as p-values and confidence intervals), but also focusing on clinical meaningfulness; not only depend on the final conclusion, but also on the logic and integrity of the entire evidence chain that supports the conclusion; not only depend on the overall effect, but also on the subgroup effect; not only to control the measured or measurable confounding factors, but also to control potential unmeasured or unmeasurable confounding factors (such as using prior event rate ratio adjustment); in addition, the control and impact of various possible biases and confounding factors need to be elaborated as much as possible.

