

**Guidance on the Design and Protocol
Development of Real-World Studies for Drugs
(DRAFT)**

*English Translation by: Jie Chen, Hongbo Yuan,
Haijun Tian, Kristijan Kahler*

*Disclaimer: The English version is for information only
and not an official translation and under any dispute the
Chinese version will prevail.*

**Center for Drug Evaluation (CDE)
National Medical Products Administration of China**

July 2022

Table of Contents

1. Introduction.....	1
2. Main Types of designs for RWS.....	1
2.1. Observational study design.....	2
2.2. Pragmatic clinical trial study design.....	4
2.3. Single-arm study design	5
3. Framework of RWS Protocols	8
3.1. Protocol synopsis.....	8
3.2. Study background.....	8
3.3. Study objectives.....	8
3.4. Hypotheses.....	9
3.5. Overall design.....	9
3.6. Study population.....	9
3.7. Treatment or intervention	10
3.8. Study endpoints/outcome variables	11
3.9. Baseline variables and important covariates.....	12
3.10. Observational/ follow-up period and time points	12
3.11. Data curation and data management plan	13
3.12. Bias considerations	13
3.13. Statistical analysis plan.....	14
3.14. Quality assurance.....	19
3.15. Ethical considerations.....	19
3.16. Registration.....	19
3.17. Protocol amendment	19
3.18. Implementation.....	19
4. Other Considerations for RWS designs	20
4.1. Feasibility of RWS	20
4.2. Representativeness of the target population	20
4.3. Hybrid study design.....	20
4.4. Method of virtual controls	21
4.5. Estimands.....	21
4.6. Target Trial Emulation	23

5. Communication with Regulatory Agencies	24
References	25
Appendix 1: Glossaries	27
Appendix 2: Chinese-English Vocabulary	31

Guidance on the Design and Protocol Development of Real-World Studies for Drugs

1. Introduction

The “Guidance on Using Real-World Evidence to Support Drug Development and Regulatory Evaluation” [1] and the “Guidance on Using Real-World Data to Generate Real-World Evidence” [2], issued by the National Medical Products Administration (NMPA) of China, provided foundations for conducting real-world studies (RWS) to support drug development and regulatory decision-making. In alignment with these two guidance documents, the NMPA also released guiding documents on using RWS to support regulatory decision-making for drugs to be used in pediatrics and rare diseases, etc. [3-9]

This guidance provides sponsors with fundamental considerations and technical requirements for the design of RWS and the development of the RWS protocols during drug development.

This guidance applies to the RWS that are intended to generate clinical evidence during drug development to support regulatory decision-making. Scenarios for using RWS have been provided in the Guidance on Using Real-World Evidence to Support Drug Development and Regulatory Evaluation [1]. This guidance can also serve as a reference to formulate an RWS for non-regulatory purposes.

2. Main Types of designs for RWS

While observational (or non-interventional) study designs are more common in RWS, interventional study designs can also be used, such as pragmatic clinical trials (PCTs) [10]. Single-arm studies are a special case of clinical study designs, in which the investigational arm can be either

interventional or non-interventional and real-world data (RWD) are often used as an external control.

2.1. Observational study design

Observational study designs largely include cohort studies, case-control studies, and cross-sectional studies [11]. Cohort studies are the recommended designs for RWS aiming at causal inference of treatment effect. Unless otherwise specified, observational studies in the rest of this guidance refer to cohort studies.

According to the time points at which the study is initiated and data are generated, cohort studies can be classified as retrospective, prospective, or retro-prospective cohort studies. Retrospective cohort studies involve the use of data generated before the study initiation (historical data); prospective cohort studies use data generated after the study initiation (prospective data); retro-prospective cohort studies use both historical and prospective data.

The following three critical areas must be addressed at the design stage: study cohort representing target population, causal inference, and data quality. Other areas for considerations are detailed in Section 3 on study protocol framework.

- **Cohort representing target population.** The cohort that represents the target population of the study should be defined based on the clinical questions of interest. The characteristics of the target population help define the longitudinal cohort starting from the initiation of treatment to the end of observational period. Specifically, given the clinical questions that determine the study objectives, the target population is usually defined by eligibility criteria (including new users, those who did not use the study treatment during a wash out period before entering the cohort, or non-new users) and should be reflected by available data sources [11,17]. Data can

be sourced from multiple centers; if data are sourced from a single center, it should be thoroughly evaluated for its representativeness of the target population and for possible extrapolation of research results. Important variables of the target population include treatment (including treatment cohort and control cohort), baseline and time dependent covariates, and outcome measures. The sample size of an RWS should meet the minimum sample size requirements for statistical inference, but usually does not have an upper limit, especially for retrospective studies. The cohort start-time, length of observational/follow-up period, and observational/visit time points should be defined based on the disease characteristics and the requirements of clinical evaluation.

- **Causal inference.** Causal inference is challenging due to the uncertainty and complexity of causal relationships among variables in observational studies [14, 25]. The choice of analytical approaches may also impact study conclusions. To avoid result-driven bias, the study should pre-define the primary statistical hypothesis, analysis datasets, analytical methods and corresponding assumptions (including both causal and statistical assumptions) during the design stage. To ensure accuracy and robustness of the study results, the study should consider methods to identify and control potential biases such as confounding bias, selection bias, and information bias, and specify the missing data handling strategy and any associated assumptions. In addition, the study should fully consider plans and strategies for sensitivity analyses and quantitative bias analyses to assess the impact of assumptions and potential biases (e.g., violation of model assumptions, various possible sources of biases).

- **Data quality.** Data quality assurance involves steps to ensure high quality of data to be collected and used for analyses. First, a data curation plan (for existing data) or a data management plan (for prospectively

collected data) needs to be formulated in advance to ensure that the generated data meet the fit-for-use requirements (see “Guidance on Using Real-World Data to Generate Real-World Evidence” [2]). Second, specific measures should be developed and implemented to ensure the accuracy of observed variable values, for example, measures to ensure the consistency of measurement tools, units of measurement, and evaluation methods.

2.2. Pragmatic clinical trial study design

Pragmatic clinical trials (PCTs) refer to clinical trials that are designed and conducted in settings close to routine clinical practice. PCTs are a type of interventional studies between traditional randomized controlled trials (RCTs) and observational studies. Comparing with traditional RCTs, PCTs have some special features such as (1) the intervention can be standardized or non-standardized, (2) the intervention can be assigned by randomization or participant self-selection, (3) inclusion and exclusion criteria can be relatively less restrictive, (4) endpoints are not limited to clinical efficacy and safety, (5) clinical endpoints, rather than surrogate endpoints, are generally preferred, (6) more treatment groups can be considered to reflect a variety of treatment or dose options in clinical practice, (7) placebo control groups are usually not used, (8) blinding may not be feasible, which may prompt measures for bias assessment and subsequent adjustment, and (9) data collection generally relies on either abstraction from medical records or scheduled follow-up visits for which the time windows are usually wider than those in RCTs. Unlike observational studies, PCTs are interventional, although there is considerable flexibility in the design and conduct of the studies.

PCT design should focus on the following aspects: (1) whether the collected data are fit-for-use to generate relevant real-world evidence (RWE) to answer the research question, (2) whether the interventions in a

particular therapeutic area are in line with routine clinical practice, (3) whether there are enough evaluable cases (especially when the clinical outcomes are rare), (4) whether the endpoint evaluation and reporting methods are consistent across study sites, (5) whether randomization is used to control selection bias, (6) when blinding is not possible, the impact of unblinding on outcome variables, especially on patient reported outcomes (PRO), should be assessed; to reduce the impact of unblinding, objective endpoints (such as stroke, death, etc.) are usually used, and (7) analytical methods for observational studies can be used for the analysis of PCTs [13, 20-21, 23-24].

For randomized PCT (P-RCT), the choice of treatment strategy (e.g., a single-treatment strategy or a continuous-treatment strategy) and the dataset for the primary analysis of efficacy need to be pre-specified. Comparing to RCTs which often use ITT/mITT (modified or adjusted ITT) for the primary analysis, P-RCTs may consider whether it is more reasonable to use Per-protocol Set (PPS) [13], or some other appropriate analysis set for the primary analysis to reflect the best interest of patients and to account for, e.g., treatment strategy change, dose change, drug withdrawal, and treatment switches. The sample size estimation should also consider the above-mentioned factors.

2.3. Single-arm study design

The prerequisite for considering a single-arm study is whether an RCT is infeasible (e.g., an extremely rare disease) or is unethical (e.g., a life-threatening disease with no efficacious treatment or a disease that is relapsed, refractory, or incurable with existing therapies). The investigational arm in a single-arm study can be either interventional (single arm trial) or non-interventional (single arm observational study) and is often compared with an external control, such as a historical control,

concurrent external control, or a fixed threshold value [7-8]. To reduce bias, subjects in the external control should be comparable with the study subjects in the investigational arm with respect to population characteristics (e.g., baseline demographic and other variables, clinical characteristics, etc.), diagnostic criteria, use of prior and concomitant medications, measurement and evaluation methods of endpoints, and any other factors that may potentially impact the probability of treatment assignment and outcomes or prognoses. Moreover, single-arm study designs should also consider at least the following important areas.

2.3.1. Treatment group

The treatment group can be interventional, which is more common with pre-defined treatment regimens that should be strictly followed during the study conduct. The treatment group can also be non-interventional with no pre-defined standardized treatment schedule or regimens, which can add a level of complexity, as patients may receive concomitant medications in routine care. Therefore, it is important to clearly define the targeted treatment of interest.

2.3.2. External controls

- **Historical control.** When using existing data as a control, the study should consider the impact of population heterogeneity, consistency in variable definitions and diagnoses, classifications and stages of the disease, and available treatment options in different historical periods, on the estimated treatment effect.

- **Concurrent external control.** Concurrent controls can be selected from external cohorts of patients with similar natural history of the disease or from other external RWD that are collected simultaneously and prospectively with the treatment arm.

- **Fixed threshold.** The determination of a threshold value for a

control effect should be based on sufficient evidence. The choice of such a value should first consider national standards, industry standards, and expert consensus; otherwise, a threshold value can be determined by an integrated analysis of relevant information including, but not limited to, published literature, study reports, and other research data.

- **Mixed control.** Historical controls and concurrent controls can be pooled together to form a control arm. These external controls can be derived from existing RWD or obtained from past relevant clinical studies (observational or interventional). The study should first assess the fit-for-use and representation of the external data and pre-specify if weights will be used when integrating different data types. It is recommended to pre-define sensitivity analyses to assess the impact of such weights on the research results.

2.3.3 Other considerations

There is greater uncertainty in causal inference results for single-arm studies with external controls due to potential confounding, population heterogeneity, and various sources of possible biases. To overcome these limitations or reduce their impact, in addition to the above considerations, attention should also be given to the following: (1) objective primary endpoints such as hospitalization are preferred; (2) inclusion and exclusion criteria and screening process should be clearly defined and strictly followed; (3) collected data should meet the fit-for-use requirements for RWD; (4) concurrent external controls are preferred over historical controls; (5) statistical analysis methods for the primary analysis, such as appropriate use of multi-variable modeling, propensity score (PS), virtual control methods, instrumental variable methods, etc., should be defined in advance; (6) matching criteria should be pre-defined in the protocol if matching is used; and (7) sensitivity analysis and quantitative bias analysis

should be performed to examine the impact of unmeasured confounding factors, heterogeneity effect, violations of model assumptions, and other possible biases on the analysis results.

3. Framework of RWS Protocols

The main bodies of RWS protocol framework are similar across different types of study designs and some differences pertaining to specific study designs are explained in corresponding sections as follows. However, this recommended framework for RWS protocols does not preclude special considerations for some specific RWS.

3.1. Protocol synopsis

The study synopsis is a concise summary of key components of the study protocol and is often presented in a tabular form. These key components include title, study objectives, hypotheses, overall design, study population (including diagnostic criteria, inclusion and exclusion criteria, etc.), treatments (defining both treatment and control groups), endpoints, baseline characteristics and important covariates, safety outcomes, observational period and time points for treatment and endpoint measurement, data sources, data curation and management, sample size determination and justification, statistical and sensitivity analyses, bias control, etc [17, 25-26].

3.2. Study background

The study background should include a brief summary of relevant studies in the literature that support the need of the current study. The rationale of choosing an RWS (such as infeasibility and ethical risks when conducting an RCT) and the positioning of the study (such as providing evidence to support regulatory decision making, or exploratory analyses, etc.) should be fully explained [16].

3.3. Study objectives

The study objectives (including primary objectives, secondary objectives, and exploratory objectives, if any), as determined by the research question, should be clearly defined and include information about the target population, treatments (including controls), and outcomes.

3.4. Hypotheses

The study hypotheses are formulated according to the study objectives.

3.5. Overall design

The overall study design including elements such as multi- or single-center, observational or interventional, single- or two-arms/multiple-arms, etc., should be described. For observational studies, details on either retrospective or prospective design should be described.

For interventional studies, the following should be further explained: (1) whether randomization is used, and if so, the randomization scheme and its implementation process, (2) whether blinding is used, and if yes (single- or double-blind), the implementation method, and (3) for open label design, whether blinded endpoint assessment is used, and if yes, the implementation method.

For single-arm studies, it is necessary to indicate whether the study group is interventional or observational, and what type of external control is used and why.

3.6. Study population

3.6.1. Diagnostic criteria

If different diagnostic criteria are available for the disease under study, the specific criteria used in the study should be described in detail, including the rationale and relevant references. The content of the diagnostic criteria can be presented as an appendix of the study protocol if it is lengthy. Where appropriate, specific diagnosis codes (e.g., ICD 9/10) should also be provided. If the diagnostic criteria are standard and well-

known to medical professionals, detailed explanation in the protocol is not necessary.

3.6.2. Inclusion and exclusion criteria

Inclusion and exclusion criteria (IEC) should be defined to precisely describe the target population of the study. In general, the IEC are relatively less restrictive in observational studies than in interventional studies. The IEC should be chosen to avoid biases, such as selection bias and immortal-time bias. When necessary, the rationale for important exclusion criteria should be explained with an assessment of their impact on the study results.

3.7. Treatment or intervention

In this guidance, population cohorts receiving investigational products or treatment strategies are referred to as “treatment groups” in observational studies or as “experimental groups” in interventional studies (such as PCT). Population cohorts receiving non-study products or non-treatment strategies (e.g., comparators or usual care) are referred to as “control groups”.

3.7.1. Treatment/experimental group

For the treatment or experimental group, the protocol should specify the treatment regimens in detail, including the trade name and manufacturer of the product, dosage, administration route, schedule, and duration, etc. If the treatment is a physical therapy (such as radiotherapy or laser therapy), specific treatment modalities and parameters should be provided. Treatment strategies and treatment patterns in observational studies are generally determined by clinical practice and should be considered in data collection and analysis and result interpretation.

Unlike in observational studies, the treatment regimens in interventional studies should be pre-specified based on some standard to

form relatively standardized treatment strategies.

3.7.2. Control group

Real world studies usually use active comparators or standard of care as a control arm. Subjects in the control arm should receive a treatment regimen or strategy with demonstrated therapeutic effect commensurate with that in clinical practice at the time of study conduct or data collection

Concurrent controls are always preferred over historical controls if fit-for-use RWD are available. To minimize selection bias in retrospective studies, all the patients meeting the IEC in the pre-defined data collection period should be considered in the study for both treatment and control groups. If the number of patients is too large and the burden for data curation is too high, a random sample can be taken to select a representative set of patients. In either retrospective or prospective studies, the selection criteria for the control group patients (e.g., PS matching) should be clearly defined.

The choice of controls for interventional studies is similar to that in RCTs.

The choice of controls for single-arm studies is discussed in Section 2.3.2; see also [15].

3.7.3. Concomitant treatments

Concomitant treatments are common in RWS. The possible (expected) concomitant treatments should be described in the study protocol. Unforeseen concomitant treatments and their impacts on study results should also be fully discussed in the analysis.

3.8. Study endpoints/outcome variables

3.8.1. Effectiveness endpoints

The primary and secondary (especially key secondary) effectiveness endpoints should be pre-defined. Each effectiveness endpoint definition

should include the name or description of the endpoint, time point or period at which the endpoint is measured, measurement methods and tools, calculation methods, and evaluation methods, etc. When necessary, an independent endpoint adjudication committee can be set up, for which a detailed implementation process such as standard operating procedures (SOPs) should be described. Note that surrogate endpoints are usually not used as primary endpoints in RWS; otherwise, a justification should be provided [12].

3.8.2. Safety endpoints

Refer to relevant (sections of) guidelines for clinical trials for the definition and handling of safety endpoints.

3.8.3. Exploratory endpoints (if any)

If needed, exploratory endpoints, such as pharmacoeconomic endpoints, etc., can be defined in the study.

3.9. Baseline variables and important covariates

Baseline variables and important covariates, as well as their units of measurement and time of observation, should be defined in the study protocol. The selection of these variables can be based on existing research findings (e.g., the variables/factors that affect prognosis as described in medical guidelines, expert consensus, published literature, conference reports, etc.) and/or expert knowledge from study team. Important covariates should be documented with rationale based on domain knowledge and causal diagrams. For identified important covariates, it is recommended to define in the protocol or statistical analysis plan the role of each covariate, such as effect modifier, risk factor, time-independent or -dependent covariate, intermediate variable, collider variable, instrumental variable, etc.

3.10. Observational/ follow-up period and time points

The protocol should define the observational or follow-up period, the start/end time and time interval during which endpoints are measured for each study group.

3.11. Data curation and data management plan

The methods and process of data curation and data management should be well documented. Historical data, regardless of data source (e.g., medical records or data obtained from different clinical studies), should go through a consistent and well-defined data curation process to meet analysis requirements. For prospectively collected data, a data management process that is scientifically rigorous and regulatorily compliant should be pre-specified and implemented. The data curation or data management plan can be presented as an appendix of the protocol if the volume of the plan is large.

The source of data should be described, e.g., research centers from which the data were collected, the start and end time points of data collection, date of data extraction, system used and record format for data storage, etc. If the data are derived from previous studies, the form of recording and storage of the original data should be described to ensure traceability of the data.

Detailed requirements for data curation and data management can be found in the “Guidance on Using Real-World Data to Generate Real-World Evidence” [2].

3.12. Bias considerations

Bias is a special challenge that needs to be addressed in RWS. Various sources of potential biases and their effects should be fully discussed in the protocol, and effective measures to control these biases should also be developed and documented. Commonly encountered biases include: (1) information bias due to inaccurate or inconsistent recording of data

measurement or collection or their evaluation methods, (2) selection bias introduced by inappropriately selecting subjects, loss to follow-up, subject withdrawal, missing records, etc., (3) treatment-effect heterogeneity bias when treatment effect is correlated by treatment status, (4) confounding bias due to insufficient balance or control of confounding variables in the analysis, and (5) result-driven bias caused by selecting the most favorable result among those generated by, e.g., different analysis methods or different analysis sets that are not determined in advance. In addition, some specific information biases can occur, e.g., immortal-time bias or lead time or zero-time shift bias that may arise when determining survival time, publication bias in meta-analysis based on literature, recall bias due to recalling error of past information, and survivor bias arising from the inclusion of only prevalent user [21].

3.13. Statistical analysis plan

To avoid result-driven bias and ensure transparency of the study, it should be emphasized that the statistical analysis plan (SAP) for the primary analysis should be developed in parallel with the study protocol. This differs from what is usually done for traditional RCTs where the SAP can be completed after the finalization of the protocol and before the database lock. The SAP for the primary analysis can be presented as an appendix of the protocol if it is lengthy. In addition to the key elements from the protocol such as study objectives, target population, endpoints and their definitions, an independent full SAP should also include the following topics.

3.13.1. Sample size estimation

The sample size of an RWS is usually determined for the primary analysis with many factors such as the type of the study (e.g., number of arms), type of comparison (superiority or non-inferiority), statistical

analysis method, expected effect size or parameter of outcome variable and its statistical distribution, significance level (one-sided or two-sided), statistical power, randomization ratio, multiplicity adjustment, dropout rate, treatment compliance, etc. In addition to the above-mentioned factors, an RWS should also consider the impact of analytical methods used for confounding adjustment. There are two commonly used methods handling confounding in sample size estimation: (1) Estimation method based on multivariable models. For example, if the primary analysis uses a logistic regression model, the generalized coefficient of determination R^2 for covariates and treatment grouping is needed to determine the effective sample size. (2) Match-based estimation method. If the primary analysis is based on matched cohorts (e.g., the PS-matching method), the sample size is generally determined for the after-matching cohorts, and then the final sample size is estimated according to the expected matching discount rate.

In addition, there are some empirical methods, e.g., the sample size needed based on an empirical estimate of the number of positive events (events per variable, EPV) or multiples of the number of covariates that may be included in the multivariable model.

For single-arm studies with external controls, the sample size of the control group should not be less than, or can be several times larger than, that of the treatment group. In addition, a high proportion of missing data is quite common and should be taken into account in effective sample size determination.

3.13.2. Analysis sets

Different analysis sets may address different research questions. The RWS protocol should define appropriate analysis sets based on corresponding research questions and study objectives. If randomization is used, the analysis set for effectiveness should ideally be defined based on

the randomized cohorts. If the target population is a subset of the analysis set, this subset should be labeled as the corresponding target population.

3.13.3. Missing data

Missing data are quite common in RWS. During the data curation process, missing records should be traced and documented as much as possible to improve data quality. For the primary analysis and related sensitivity analysis, methods used to address missing data and their rationale should be explained in the SAP.

3.13.4. Descriptive analysis

Descriptive analyses should characterize the main features of variables including baseline and endpoint variables. Descriptive statistics should be appropriately chosen according to the distributional characteristics of the variables.

3.13.5. Analysis of heterogeneity

Potential heterogeneity factors, such as study center, age, gender, disease severity, etc., should be defined in advance to provide scientific rationale for subgroup or stratified analyses. The SAP should include methods (e.g., analytical models) used to evaluate the presence or absence of heterogeneity and the significance level (e.g., $\alpha = 0.10$) used to test the interactions of the treatment with heterogeneity factors, in which the study objectives and clinical significance of heterogeneity should be taken into consideration.

3.13.6. Primary analysis

The primary analysis is usually conducted on the primary endpoints (sometimes also on key secondary endpoints) to answer the most important questions of interest in the study. The primary analysis should be stated in detail in the SAP and should include, but not limited to, the following: (1) statistical hypothesis, (2) statistical models and associated assumptions for

adjusted and unadjusted analyses, (3) covariates to be included in the adjusted analysis and pre-defined rules of variable selection based on observed data during the analysis, including identification of time-independent and/or -dependent confounders, risk factors, intermediate variables, and factors causing potential heterogeneity, (4) if PS matching is used, the matching ratios, matching methods including specific parameter settings (such as caliper value), and methods used to assess the matching performance, and (5) potential competing risks in the analysis of survival data. In addition, model assumptions, such as nonlinear relationship or non-proportional hazards, etc., should be assessed. For PCTs, it is recommended that the handling of covariates in the primary analysis should be the same as in observational studies, regardless of whether randomization is used, because the control of baseline comparability in PCTs (particularly for cluster randomization) is far less stringent than that in RCTs. For specific causal inference methods, please refer to the appendix of the “Guidance on Using Real-World Evidence to Support Drug Development and Regulatory Evaluation” [1] and other relevant literature. In case of limited sample sizes that permit only descriptive analyses, the SAP should provide corresponding explanation.

3.13.7. Subgroup analysis

Subgroup analyses, if required, should be pre-defined based on previous research findings and domain knowledge on heterogeneity factors. Subgroup analyses may also be performed when some key covariates significantly interact with treatment. For details on subgroup analysis, please refer to the “Guidance for Industry on Subgroup Analyses in Confirmatory Clinical Trials” [9].

3.13.8. Sensitivity analysis

The robustness of study conclusions is important, and can be assessed

using sensitivity analyses which should be prespecified and conducted for different assumptions, including, but not limited to, assumptions on unmeasured confounding variables, different mechanisms of data missingness, different definitions of analysis set, different analytical methods, different combinations of covariates in the analytical models, etc.

3.13.9. Quantitative bias analysis

The potential impact of bias on study conclusions requires special attention in RWS. It is highly recommended to thoroughly explore various sources of possible biases and determine their potential impact. This can be done by using quantitative bias analyses to determine the direction, magnitude, and uncertainty of various sources of biases and their impact on study conclusions [18-19]. For example, data can be analyzed for different analysis sets defined with different criteria and the results can be compared to determine whether there is a selection bias; in the hybrid study design (Section 4.3), differences in treatment effects between internal and external data can be compared to determine if there is treatment-effect heterogeneity bias, and the heterogeneity bias parameters can be used for bias correction. The distribution of bias parameters shows the direction, magnitude, and uncertainty of bias, and the critical point analysis to examine the impact of various sources of possible biases can also be considered as a method for quantitative bias analysis. In addition, sensitivity analysis and quantitative bias analysis can also be discussed together.

3.13.10. Safety analysis

There are obvious limitations in actively monitoring safety events using RWS, especially retrospective study data, which may need to be complemented by external evidence, such as safety information of investigational drugs from other studies and from adverse event reporting

and monitoring systems. If the study objective is to address whether the investigational product has a better safety profile than the control product, sufficient information on the safety of the control product should also be provided. If the primary study objective is to answer safety questions, please refer to relevant regulatory guidelines and the literature.

Statistical analyses of safety endpoints and associated assumptions should be described in the SAP. The corresponding output formats (statistical tables and graphics) can be specified after finalization of the SAP but before the initiation of formal safety analysis.

3.14. Quality assurance

In general, quality assurance of RWS is similar to that of RCTs. However, special attention should be paid to the quality control during the data curation process. For details, refer to the “Guidance on Using Real-World Data to Generate Real-World Evidence” [2].

3.15. Ethical considerations

Ethical requirements for RWS can follow the requirements for RCTs. Retrospective studies may use a general consent to exempt from obtaining patients’ informed consent.

3.16. Registration

Registrations of the study in public website(s) should be provided.

3.17. Protocol amendment

Significant changes to or deviations from the original protocol, e.g., changes in data curation or primary statistical analysis in the SAP, may occur and such changes or deviations require a protocol amendment. The amended protocol should be submitted to regulatory authorities for agreement.

3.18. Implementation

A general implementation plan for clinical trials can be adopted,

which may also include special features of implementation for the proposed RWS.

4. Other Considerations for RWS designs

4.1. Feasibility of RWS

Before designing a study, the sponsor should thoroughly assess the feasibility of conducting the proposed RWS. The assessment should include, but not limited to, (1) whether a traditional RCT is feasible, (2) whether an RWS is a better option than an RCT, (3) whether available RWD are sufficient to support the proposed study in terms of both quality and quantity (sample size) to generate reliable and robust RWE, (4) the position of the RWS in the overall drug development plan and the role of generated RWE within the totality of evidence, and 5) whether the study is endorsed by the regulatory agency. Note that the sponsor should communicate with the regulatory agency regarding the feasibility of an RWS and implement it only after reaching agreement with the agency.

4.2. Representativeness of the target population

It is important to ensure that the study cohorts represent the target population to which the study results will be applied. A good representativeness of selected cohorts to the target population can be achieved through random sampling. However, the study samples for RWS are often obtained through convenient sampling due to practical considerations and may not appropriately represent the target population. Therefore, additional analyses should be conducted to assess the representativeness of study samples to evaluate the potential impact on the study conclusions in the target population (external validity).

4.3. Hybrid study design

In this guidance, a hybrid study refers to a study that uses internal data and external RWD to form a study arm (including a control arm). A single-

arm study design with external controls is a special case of a hybrid study design. Hybrid study designs can also be used in PCTs. The hybrid study designs require statistical assumptions on the merge of internal and external data, either at group level or individual subject level, to ensure comparability of the internal and external populations. Comprehensive sensitivity analyses and quantitative bias analyses should be carried out. If a Bayesian method is used, simulations should be performed to investigate the impact of prior distributions and other relevant parameter settings on study conclusions. Since the valid sample size of external data may be impacted by the degree of overlapping population between internal and external data and treatment-effect heterogeneity, a hybrid study should include sufficient subjects in the current study to ensure robustness and reliability of the study results.

4.4. Method of virtual controls

Virtual controls can be used in effectiveness assessment of a product in disease areas for which no effective treatment is available [22]. The basic idea of virtual controls is based on counterfactual concept, i.e., (1) use existing RWD and key covariates considered in the single-arm study to establish a prediction model without the study drug, (2) plug the values of key covariates obtained in the single-arm study into the prediction model and calculate the predicted outcomes (i.e., virtual controls) and corresponding summary statistics without the study drug, (3) calculate the actual outcomes and corresponding summary statistics with the study drug, and (4) compare the summary statistics with and without the study drug to estimate the treatment effect. This method requires a large sample size to build and validate the prediction model with high accuracy robustness of prediction.

4.5. Estimands

ICH E9(R1) summarizes the estimands for clinical trials with five important attributes: target population, treatment, endpoint, intercurrent events, and population-level summary. Unlike in traditional RCTs, defining an estimand in an RWS requires additional considerations due to population heterogeneity, flexibility of treatment regimens/strategies/policies, a wide variety of intercurrent events, challenges in selecting and defining study endpoints, and complexity of conducting sensitivity analysis. This guidance does not have specific requirements for the construction of estimands but encourages sponsors to actively explore the feasibility of applying such concepts in RWS. The following are some special considerations for defining estimands in RWS.

4.5.1. Heterogeneity of study population

Unlike in RCTs, the study population in RWS is generally more heterogeneous due to relatively loose IEC. The study population often includes not only subjects with more diversified demographics and clinical characteristics, geographic locations, and study centers/sites, but also subjects who are unwilling to participate or are often under-represented in RCTs (e.g., ethnic minorities, elderly, and those residing in remote areas). Therefore, the estimand in an RWS should be defined by taking into consideration of the heterogeneity of the target population.

4.5.2. Flexibility of treatment

Treatment exposure in RWS is often complex because of different available doses, use of concomitant medications, and variation in treatment cycles, etc. Patients' adherence to and preference of treatment options should also be considered when defining an estimand.

4.5.3. Variety of intercurrent events

In addition to treatment-induced intercurrent events (ICEs) (e.g., intolerability, lack of efficacy) that are commonly seen in RCTs, some

ICEs often encountered in RWS can be induced by patient behavioral (e.g., preference for certain treatment, convenience use of a treatment, doctor-patient relationship, etc.) and non-behavioral factors (e.g., change of medical insurance policy affecting the use of current treatments, improvement of health condition, etc.). These ICEs should be considered when defining an estimand in RWS.

4.5.4. Endpoint selection

Real world studies usually use clinical endpoints (or outcomes), preferably single-measured and easily observable clinical endpoints (such as death or hospitalization), rather than surrogate endpoints [12]. If a study uses a composite clinical endpoint, it is important to ensure that each component is accurately recorded.

4.5.5. Sensitivity analysis

Causal inference in RWS is complex due to confounding and biases. To ensure the precision and reliability of the effect estimates, sensitivity analyses are required.

In addition to the issues discussed above, there may be other challenges that need to be addressed in defining an RWS estimand, such as data fusion in hybrid studies, censoring of individual survival time in observational studies, etc.

4.6. Target Trial Emulation

Target trial emulation is an approach for conducting an RWS which uses existing RWD to emulate a well-designed RCT. Key components of this RCT consist of specifications of the eligibility criteria, treatment strategies, treatment assignment procedure, visiting time points and follow-up period, outcomes, causal contrasts of interest and analytical strategy. Sequentially, the analysis set is created for an RWS and causal inference methods are used to derive the research results. The target trial approach

facilitates the identification and prevention of unnecessary biases, such as immortal time bias or prevalent user bias etc. It also provides a reasonable framework to clarify the decision making in the observational study. Target trial emulation should consider appropriate scenarios and ensure the availability of fit-for-use RWD with sufficient sample sizes and high possibility of emulating RCT. Currently its application still needs further consensus, but it is an approach that may be explored.

5. Communication with Regulatory Agencies

Clear specifications of an RWS protocol and transparency of its implementation are critical for an RWS. Sponsors should thoroughly discuss with the regulatory agency before initiation of the study about the RWS protocol including feasibility and rationale of the study, data curation and/or management plan, sample size determination, and the SAP, etc. During the study, if there are major changes in the original protocol, such as changes in data curation plan, changes of basic models used for the primary analysis in the SAP, etc., the sponsor should first communicate with the agency to obtain agreement about the changes and then initiate protocol amendment to reflect these changes.

References

1. NMPA. Guidance on Using Real-World Evidence to Support Drug Development and Regulatory Evaluation. Center for Drug Evaluation, National Medical Product Administration, China, 2020.
2. NMPA. Guidance on Using Real-World Data to Generate Real-World Evidence. Center for Drug Evaluation, National Medical Product Administration, China, 2021.
3. NMPA. Guidance on Real-World Studies to Support Pediatric Drug Development and Regulatory Evaluation. Center for Drug Evaluation, China National Medical Product Administration, 2020. Accessed: April 11, 2022.
4. NMPA. Guidance on Clinical Development of Traditional Chinese Medicine Compound Preparation Based on Human Experience (trial). Center for Drug Evaluation, China National Medical Product Administration, 2021. Accessed: April 11, 2022.
5. NMPA. Guidance on Statistical Principles for Clinical Trials in Rare Diseases. Center for Drug Evaluation, China National Medical Product Administration, 2021. Accessed: April 11, 2022.
6. NMPA. Guidance on Clinical Development of Drugs for Rare Diseases. Center for Drug Evaluation, China National Medical Product Administration, 2022. Accessed: April 11, 2022.
7. NMPA. Guidance on Technical Communications of Single-Arm Clinical Trials for Anti-Cancer Drug Registration. Center for Drug Evaluation, China National Medical Product Administration, 2020. Accessed: April 11, 2022.
8. NMPA. Guidance on Technical Communications of Pre-Pivotal Single-Arm Clinical Trials for Anti-Cancer Drugs to Support Marketing Authorization. Center for Drug Evaluation, China National Medical Product Administration, 2020. Accessed: April 11, 2022.
9. NMPA. Guidance for Industry on Subgroup Analyses in Confirmatory Clinical Trials. Center for Drug Evaluation, China National Medical Product Administration, 2020.
10. Berger M, Daniel G, Frank K, et al. A framework for regulatory use of real-world evidence[J]. White paper prepared by the Duke Margolis Center for Health Policy, 2017.06.
11. Boslaugh S. Encyclopedia of Epidemiology. [M] SAGE Publications. 2008.
12. Chen EY, Raghunathan V, Prasad V. An overview of cancer drugs approved by the us food and drug administration based on the surrogate end point of response rate[J]. JAMA Intern Med .2019;179(7):915-921
13. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials[J]. N Engl J Med, 2017, 377(14): 1391-1398.
14. Hernán MA, Robins JM. Causal Inference[M]. Boca Raton: Chapman & Hall/CRC 2019.
15. ICH E10: Choice of Control Group and Related Issues in Clinical Trials. 2000
16. James S. Importance of post-approval real-world evidence[J]. European Heart Journal-Cardiovascular Pharmacotherapy, 2018; 4(1):10-11.
17. Last JM. A Dictionary of Epidemiology[M]. 4th Edit. Oxford University Press. 2001.
18. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data[M].

Springer Science & Business Media, 2011.

19. Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis[J], *International Journal of Epidemiology*, 2014, 43(6): 1969–1985.
20. Roland M, Torgerson DJ. Understanding controlled trials: What are pragmatic trials? [J]. *BMJ*, 1998, 316(7127): 285.
21. 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.
22. Strayhorn JM. Virtual controls as an alternative to randomized controlled trials for assessing efficacy of interventions [J]. *BMC Medical Research Methodology* 2021; 21(3): 1-14.
23. Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials[J]. *JAMA*, 2014, 311(23)ua ua: 2381-2382.
24. US FDA. Framework for FDA’s real-world evidence program. December 2018. 2019.
25. Velentgas P, Dreyer NA, Nourjah P, et al. 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; 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
26. Von Elm E, Altman DG, 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.

Appendix 1: Glossaries

Single-Arm/One-Arm Study: A non-randomized clinical study that only sets up an experimental group or a treatment group, and may use external controls, such as historical controls, concurrent external controls, or a fixed target value as a control.

Quantitative Bias Analysis (QBA): A class of methods that can be used to assess the sensitivity of study results to various possible sources of systematic error (such as misclassification, uncontrolled confounding, selection bias, etc.). A QBA method can also be used to evaluate the direction and degree of biases on the effect estimation and therefore provide guided, bias-corrected analysis results.

Estimand: A precise description of the treatment effect that reflects the clinical question posed by a given clinical study objective. It summarizes at a population level what the outcomes would be in the same patients under different treatment conditions being compared.

Observational Study: a.k.a. non-interventional study, in which no active intervention is applied, and the study aims at exploring the causal relationship between treatment and outcome in the target population based on a specific clinical question.

Retrospective Observational Cohort Study: An observational study that identifies the target population at the start of the study and is based on historical data (data collected before the study initiation).

Target Trial Emulation: A real-world study method that is based on

analysis dataset generated from existing real-world data sources and is designed according to a good RCT design which can support causal inference study conclusions.

Bias: Any tendency leading to results or conclusions that systematically (as opposed to randomly) deviate from the truth and may be present in the study design, data collection, analysis, interpretation, reporting, publication, and/or review of data.

Prospective Observational Study: An observational study in which the target population is identified at the start of the study, and the treatment, outcome, and other relevant data that are pre-defined and collected prospectively during the study period

Pragmatic Clinical Trial/Pragmatic Trial (PCT): sometimes called a practical clinical trial, for which the design and conduct of trial are as close as possible to the real-world clinical practice. The PCT is a type of clinical studies that are between traditional RCTs and observational studies.

Data Curation: A processing of raw data for the purpose of statistical analysis based on specific clinical questions. Data curation includes at least data capture and collection (may include multiple data sources), data security processing, data cleaning (logical judgment and abnormal data processing, data completeness processing, etc.), data import and structuring (common data model, normalization, natural language processing, medical coding, derivative points, etc.), data transmission and other related processing steps.

External Control: Subjects from external data other than those in a clinical study are used to constitute a control group to evaluate the effect of a treatment or intervention being studied. External control data can be historical data, data obtained from concurrent controls, or a fixed target value.

Method of Virtual Control: A method based on the concept of counterfactual to assess the treatment effect, conducted by establishing a prognostic prediction model without the trial drug based on the existing RWD and the key variables considered in the single arm trial, plugging the covariates abstained in the single arm trial into the prediction model and calculate the predicted outcomes (i.e., virtual controls) without the trial drug, and lastly calculating the actual outcomes with trial drug and comparing them predicted outcomes from virtual controls to assess the treatment effect.

Causal Inference: A class of theories and methods that are used to characterize, based on real-world data, the causal relationship between an intervention or treatment and a clinical or health outcome using appropriate statistical models and analytical methods that eliminate or minimize the effects of various covariates, measured/unmeasured confounding factors and possible bias, and thus draw inference on causal relationship between an intervention or treatment and a clinical or health outcome.

Real-World Data: Data derived from various sources reflecting patient's health status and/or diagnosis and health care that are collected in routine practice. Not all real-world data can be used to generate real-world evidence and only real-world data that satisfies fit-for-purpose

requirements can potentially be used to generate real-world evidence.

Real-World Study: A study intended to obtain clinical evidence, per pre-defined clinical question, on the use and potential benefit-risk of a medical product in real-world settings. This is done through the collection and analysis of RWD related to the health status and/or diagnosis, treatment, and health care of study subjects, or through the aggregate data derived from these RWD.

Real-World Evidence: Clinical evidence on the use and potential benefit-risk of a medical product obtained through appropriate and adequate analysis of fit-for-purpose real-world data.

Intermediate Variable: A variable in the causal pathway between a treatment and an outcome, i.e., a variable that is affected by the treatment and itself affects the outcome at the same time, or a variable associated with the outcome; the former is also called a mediator.

Appendix 2: Chinese-English Vocabulary

English	Chinese
Case-control Study	病例对照研究
Causal Inference	因果推断
Cohort Study	队列研究
Collider Variable	碰撞变量
Confounder	混杂因素
Cross-sectional Study	横断面研究
Data Curation	数据治理
Data Management	数据管理
Derived Variable	衍生变量
Estimand	估计目标
Events per Variable, EPV	每个协变量所需阳性事件数
Immortal-time bias	恒定时间偏倚
Instrumental Variable	工具变量
Intermediate variable	中间变量
Lead-time Bias	领先时间偏倚
New User	初治病例
Observational Study	观察性研究
Patient Reported Outcome, PRO	患者报告结局
Pragmatic Clinical Trial, PCT	实用临床试验
Propensity Scores, PS	倾向评分

English	Chinese
Prospective Study	前瞻性研究
Publication Bias	发表偏倚
Quantitative Bias Analysis, QBA	定量偏倚分析
Real World Data, RWD	真实世界数据
Real World Evidence, RWE	真实世界证据
Real World Research/Study, RWR/RWS	真实世界研究
Recall Bias	回忆偏倚
Retrospective Study	回顾性研究
Standard Operation Procedure, SOP	标准操作程序
Statistical Analysis Plan, SAP	统计分析计划
Survivor Bias	幸存者偏倚
Target Trial	目标临床试验
Target Trial Emulation	模仿目标临床试验
Time-dependent Variable	时依变量
Traceability	可追溯性
Zero-time Shift Bias	起点时间偏倚