attachment1

Medical device real-world study design and statistical analysis

Registration Review Guidelines

(Draft for comments)

This guiding principle aims to standardize and reasonably guide the application of real-world data in the clinical evaluation of medical devices, and provide technical guidance for applicants to conduct real-world research and technical review by regulatory authorities.

This guiding principle is a guiding document for registration applicants and technical reviewers, but it does not include administrative matters involved in registration approval, nor is it enforced as a regulation. This guideline must be used under the premise of complying with relevant regulations and mandatory standards. Guiding Principles. If there are other methods that can meet relevant regulatory requirements, they can also be used, but detailed research information and verification information need to be provided.

This guiding principle is formulated based on the current regulations and standards system and the current level of understanding. With the continuous improvement of regulations and standards, as well as the continuous development of science and technology, the relevant content of this guiding principle will also be adjusted in a timely manner.

## 1. Scope of application

This guiding principle applies to real-world research on medical devices and does not apply to real-world research on in vitro diagnostic reagents managed as medical devices. This guideline is based on the "Technical Guidance Principles for the Clinical Evaluation of Medical Devices with Real-World Data (Trial)" (hereinafter referred to as the "General Principles") and combined with the current accumulated experience to further refine the real-world research design and statistical analysis of medical devices. general requirements. At the current stage of development, real-world evidence in the clinical evaluation of medical devices is mainly used as a supplement to existing clinical evidence and cannot replace the existing clinical evaluation path.

## 2. Real-world research types and their applications

### pragmatic randomized controlled trial

randomized controlled clinical trial ( pRCT ) refers to a study that uses a randomized, controlled design to compare the treatment results of different intervention measures in clinical practice in a real or close to real medical environment. It focuses on the effectiveness of the intervention measures in routine Effects in clinical practice . pRCT combines the advantages of randomization and real-world data, and can better control confounding and bias. Its research results can provide high-quality real-world evidence for the evaluation of the effectiveness of intervention measures. The purpose of pRCT is to measure the effect of a certain intervention in routine clinical practice. Its study population inclusion criteria are usually broader and the population representativeness is better. However, the heterogeneity of the population is usually high and the test power is lower. It is suitable for those who need to generate data. Application of clinical evidence to wider populations and clinical situations. For example, pragmatic randomized controlled trials can provide data on the safety and effectiveness of devices in different subgroups of patients, providing more effective information for device benefit and risk assessment.

### observational real-world research

#### descriptive research design

Common descriptive research designs include cross-sectional designs, case report and case series designs, etc., which are not used for statistical causal inference.

In cross-sectional studies, all measurements are completed at a specific time point and are mainly used to describe the basic characteristics, health status, disease recovery and other distribution of patients who have received a certain intervention, providing clues for subsequent research. Cross-sectional design can be used for descriptive statistical research on device adverse events, such as investigating the occurrence of pressure ulcers related to a certain type of medical device at a certain point in time; describing the clinical use effects of devices for groups of people with different characteristics. Case reports are used to describe the detailed clinical characteristics of one or a few clinical cases, and usually do not describe the central tendency or dispersion of things . Case series research is the summary and summary of multiple case data.

#### queue design

In the evaluation of device safety and effectiveness, a cohort study is to divide a specific population into different groups according to whether they use the device to be studied, track and observe the occurrence of outcomes in each group, and compare the differences in outcome incidence rates between groups to determine An observational research method to determine whether there is a causal association and the degree of association between a device and an outcome. Prospective cohort studies allow for greater control over data quality because research planning is developed prior to data collection and implementation. In a retrospective cohort study, the data already exists and the study completion time is short, but the data quality is poorly controllable. It is recommended to verify the data integrity and data accuracy before use. Two-way cohort study refers to the retrospective cohort study observing "now" and then continuing the prospective follow-up. Cohort designs are the most commonly used design type in observational real-world studies and are used in a wide range of situations. There are currently more and more cohort studies conducted based on registration databases. For example, the CathPCI registration database is used to conduct a retrospective cohort study comparing the Mynx vascular closure device with other similar marketed products; the National Joint Registry database is used to compare the performance of joint prostheses designed by different manufacturers. Renovation rate, etc.;

#### Case-control and its derivative designs

A case-control study uses patients who have experienced an outcome event as the case group, and patients who have not experienced an outcome event as a control group. The proportion of the case group and the control group using the device to be studied is compared to study the association between the device to be studied and the outcome event. sex. When the incidence rate of the clinical outcome to be observed for the device to be studied is low, a pragmatic randomized controlled trial is used, or when the sample size required for the cohort design is too large to be feasible, a case-control design may be considered .

Case-control derivative designs include nested case-control and case-cohort designs . Both derivative designs combine the advantages of cohort studies and case-control studies and collect information on exposure factors , confounding factors, etc. before the outcome occurs ; the experimental group and the control group are derived from For the same population, population comparability is better ; there is no need to measure the entire population of the cohort. At this stage, more and more nested case-control designs are carried out based on registration databases.

### control for single-arm trials

External control refers to finding a group of research subjects with similar characteristics from other trials or historical cases, and synthesizing a control group that receives different interventions. Real-world data as an external control for single-arm trials is one of the external control designs . This guideline does not include the specific usage scenarios of this type of design, as well as research design , statistical analysis and other requirements , and separate guiding principles will be formulated for related content .

## 3. Real-world research plan design considerations

### Research background and purpose

the expected scope of application of the product and the technical characteristics of the product, combined with existing non-clinical and clinical data, clarify the safety and effectiveness issues to be solved in the real-world research and clarify the purpose of the research .

### feasibility assessment

After the research purpose is determined, the applicant needs to evaluate whether it has the objective conditions to conduct real-world research, mainly considering whether the existing experience and knowledge accumulation are sufficient to determine in advance the confounding variables that affect clinical outcomes, and whether the required variable data are accessible and the amount of data and the adequacy of data quality.

Whether the variables that affect clinical outcomes can be determined in advance depends on the current accumulation of clinical knowledge and experience about the disease, diagnosis and treatment methods, and devices. For areas where research experience and knowledge accumulation are insufficient, applicants cannot ensure the identification of variables that affect the outcomes. For all variables with important effects, the existence of unmeasured or unadjusted confounding variables cannot be ruled out when conducting observational real-world research. The bias of the research results cannot be determined, and the robustness of the conclusions cannot be guaranteed.

Secondly, the availability and quality of variable data need to be assessed. In the real world, missing data is common. Data derived from the real world may lack out-of-hospital follow-up data, outcome indicators (such as functional scores, pain scores, etc.), imaging examinations, etc. In addition, it is also necessary to consider whether the real-world data observation time meets the research purpose, whether the sample is representative and whether the sample size can ensure sufficient test performance, whether the quality of the existing data can meet the requirements of statistical analysis, etc.

### Determine the appropriate type of real-world research design

Applicants should select the appropriate research design type based on the determined research purpose and refer to the content in Chapter 2. As mentioned above, there are different types of real-world research designs, including experimental pRCT , observational cohort study, case-control, nested case-control design, etc. Each design has different characteristics and is suitable for different application scenarios.

### Research flow chart

Considering the differences in the implementation process of different types of real-world research, it is recommended that the implementation process be presented in the form of a flow chart. The flow chart presents the specific steps in the research process (such as population screening, acceptance of intervention measures, inspection and inspection, etc.) in chronological order.

### Define the study population

The research protocol needs to stipulate in advance the definition of the research population, make the definition clear and unambiguous, avoid ambiguity and vague expressions, and specify clear inclusion and exclusion criteria in advance.

For retrospective study designs, attention should be paid to the representativeness of the data source used for the intended applicable population. For people with multiple records of device use, clear criteria for inclusion in the study need to be defined in advance, such as no record of using the same device in the 6 months before using the device, or exposure defined as the first use of the device. The use of different diagnostic criteria will lead to the inclusion of people who do not actually meet the research requirements. The diagnostic results under different criteria will also affect the accuracy of statistical results. It is recommended to verify whether different clinical institutions have adopted the same diagnostic criteria during the design stage.

### instrument exposure

pRCT and traditional RCT use random methods to determine device exposure. However, for observational real-world studies, the specific device used by patients is not determined by random grouping, but in the real world based on actual conditions (such as doctor preference, patient condition) ), there is a risk of selection bias, and attention should be paid to assessing this.

For retrospective real-world data, it is usually judged through the records of the devices used in the data (such as expense lists, surgical records) to determine what specific devices were used by the patients. It is necessary to verify whether the device information records are complete and accurate, including at least the device manufacturer, Model specification information. It is recommended to verify the accuracy of the device exposure information, such as including patients who are known to have used and not used the device, determine whether the device is used through retrospective data, and finally compare the determination results with whether the patients actually use the device to verify the retrospective data to determine the device. Exposure accuracy.

### control group

In a pragmatic randomized controlled design, a control group is formed by random grouping. For observational real-world designs, such as cohort studies, case-control studies, etc., it is necessary to form a control group in an appropriate manner according to the purpose of the research and the type of design. The basic principle of setting up a control group is to ensure that the distribution of confounding variables is balanced among groups as much as possible . Depending on the purpose of the study, the control group can be a similar product already on the market from a single or multiple manufacturers, other non-device interventions, or a placebo group.

In the cohort study design, according to the purpose of the research, patients who have used other similar marketed products can be selected to form the control group, or patients who have not used similar devices but have used drugs or other diagnostic and treatment methods can be selected to form the control group; in case-control design , usually a matching method is used to construct a corresponding control group for the case group. The specific matching methods of case-control, nested case-control, and case-cohort are different. For details, see the content of Chapter 2 "Case-Control and Its Derivative Design " .

### Evaluation index

Clearly stipulate in the plan the observation purpose, definition, observation time point, indicator type, measurement method, calculation formula (if applicable), judgment criteria (applicable to qualitative indicators and grade indicators), etc. of each evaluation indicator, and clearly define the main evaluation indicators. and secondary evaluation indicators.

For retrospective real-world studies, care should be taken to ensure that different clinical institutions define outcomes the same and do not miss out on patient outcome events. It is recommended to choose objective indicators as much as possible, such as death, and avoid subjective evaluation indicators such as functional scores that have a greater risk of measurement bias.

### data collection

It is recommended to develop a complete case report form and variable dictionary, and collect and record data based on the case report form and variable dictionary.

For retrospective real-world studies, the basic information of the data source to be used must be listed, including the field information contained, the number of patients, missing data, data recording accuracy and other data quality information. The steps and methods of data cleaning must be clearly stated in the plan. If multiple databases are involved, the specific linking methods must be clearly stated in the plan, as well as the verification method for the accuracy of the data linking. Provide clear and specific definitions when extracting diagnosis and clinical outcome data . When data is not extracted manually , the specific algorithm for extracting each variable must be provided and the accuracy of its judgment must be verified. Pay attention to the rationality of the follow-up time design when constructing the algorithm for research outcome screening. In principle, there should be a sufficiently long time interval between the occurrence time of the target clinical outcome and the implementation of the intervention ( compared with the natural course of the disease). If the intervention occurs during the intervention Outcomes that occur within a short period of time after implementation may be unrelated to the intervention and may introduce new biases if not distinguished.

### Identify confounding variables to adjust for

Real-world research designs that do not use randomization need to predetermine the confounding variables that need to be adjusted. In principle, all confounding variables need to be identified so that confounding bias can be controlled during the design and statistical analysis stages. Usually, it can be judged as a confounding variable according to the following three criteria: (1) There is a causal relationship between the variable and the outcome variable; (2) There is an association between the variable and the grouping variable (exposure variable); (3) The variable is not causally related to the grouping variable and the outcome variable. Intermediate variables in the path. It is recommended to first develop a reasonable variable screening process and determine confounding variables based on accumulated professional knowledge and clinical experience. It is recommended to discuss and confirm with a team of clinical experts. For variables that ultimately cannot be confirmed as to whether they should be included, sensitivity analysis can be performed on inclusion and non-inclusion situations. List the reasons for including or not including all variables in the adjustment and provide supporting information. In the actual operation process, it is recommended to adopt a conservative attitude towards variable screening, and only remove variables that are unrelated to treatment allocation and outcome variables from the model. However, care should also be taken to avoid including collision node variables, instrumental variables, and intermediate variables. In order to clearly show the causal relationship between variables, a directed acyclic graph can be used.

### Follow-up time

Typically, the time of exposure to the device of interest is designated as the start of the study, and the end of the study is determined by whether the follow-up or observation period is sufficient to answer the study question. For implantable devices, the initial follow-up time is usually the first day after the completion of the implantation surgery. For devices with multiple treatments as a complete course of treatment, the initial follow-up time needs to be the first day after the end of the treatment course. Pay attention to the treatment process. Security incidents in the network also need to be observed.

### Calculate sample size and test power

For retrospective real-world studies, test power can be estimated based on the available sample size. For prospective real-world studies, the sample size can be calculated based on estimated parameters . Different research designs have different ways of estimating sample size. For example, cross-sectional studies can estimate sample size based on the expected width of the confidence interval, while research designs with control groups can estimate sample size based on comparative differences between groups, relative risks, odds ratios, etc.

For real-world studies involving variable adjustment, since there are many parameters that need to be estimated in advance, some parameter estimates may lack support from literature data. Compared with traditional randomized controlled clinical trials, sample size estimation is more complicated and requires more factors to be considered. For example, for stratified adjustment statistics based on propensity scores, the size of the effect size within the stratum (such as effectiveness, odds ratio, incidence rate, etc.), the probability of allocation to each stratum of test groups, the degree of overlap of propensity scores, etc. need to be considered.

### QC

#### Data quality

be evaluated in terms of representativeness, completeness, accuracy, authenticity, consistency and repeatability . For specific evaluation content, see Chapter 3 of the General Principles. Applicants must evaluate the quality of the data sources used in accordance with the above 6 Evaluate each aspect and present the evaluation results of each dimension in the form of a table.

#### Risk of bias

Bias may exist at all stages of real-world research design, implementation, analysis, and reporting. Applicants can describe in detail the measures used to control different risks of bias in the real-world research plan from three aspects: selection bias, information bias, and confounding bias. . For observational real-world studies, you can refer to the ROBINS-I evaluation tool for non-randomized interventional clinical studies to assess the risk of bias of the overall study. Here are just some of the types of bias found in real-world research:

##### Study population lacks representation

In the design stage, it is also very important to set reasonable entry criteria. The study entry criteria setting considers whether the included population can represent the expected scope of application of the product. pRCT usually adopts looser entry criteria, so it is less subject to the inclusion criteria. the impact of selection bias. For prospective studies, consecutive enrollment is recommended to avoid patient selection. For certain devices that are easily affected by clinical institutions and physician levels, a multicenter design is recommended. For studies with controlled settings, especially case-control designs, measures need to be taken to avoid admission bias in the design, such as the experimental group and the control group being determined by sampling from the same population.

##### confounding bias

Confounding bias means that the degree of correlation (association) between exposure factors and intervention measures is distorted or interfered by other factors, so that the relationship between the research variables presented and the evaluation indicators or outcome variables is not true, but biased with the superposition of confounding effects. relation.

Randomization is a powerful means of controlling confounding, balancing both measurable and unmeasured confounding factors. Since the vast majority of real-world study designs ( except pRCTs ) do not use randomization, other methods such as restrictions, paired and stratified designs can be considered during the analysis stage to control for confounding. During the analysis stage, various adjustment statistical methods (such as stratified analysis, multivariable regression analysis, propensity score-based adjustment methods, etc. ) can also be applied to control confounding.

##### deviation from intervention

In real-world studies, interventions may deviate midway through treatment due to various reasons, such as patients actively requesting to change treatment methods, doctors changing treatment strategies, etc., interventions with multiple treatments (such as hemodialysis) or interventions with long treatment times. (e.g., ventilators, extracorporeal membrane oxygenators), there is a greater likelihood of intervention deviation. When conducting real-world research, it is necessary to consider in advance the degree of risk of such bias in the device to be studied. If there is a non-negligible risk of bias in intervention measures, when selecting a real-world data source, it is necessary to consider whether the data source is detailed and accurate. Document the treatments used and any changes that occur during treatment.

In clinical practice, there may also be errors in the recording of intervention measures, such as errors in the manufacturer, model and specification of the device used, leading to information bias related to the intervention measures. When it is suspected that there is a possibility of recording errors, consider using the patient's imaging system The implant shape, marker point characteristics, price on the bill and other other information are verified.

##### measurement bias

In real-world research, accurate and precise measurements are important measures to reduce information bias. Imposing blinding can help overcome measurement bias caused by subjective factors of applicants or subjects. When blinding is difficult, objective hard endpoints (such as death, etc.) should be chosen as much as possible. During the implementation process, develop detailed operation manuals, train staff, standardize data collection procedures and monitor data collection activities, and use unified methods to collect, measure and interpret information; under applicable conditions, a third-party independent data monitoring committee can be set up Or unify standards and standardize the measurement results of indicators; when it is suspected that the data measurement is inaccurate, carry out data verification. In addition to the above commonly considered measures, corresponding measures need to be specified based on the specific types of measurement bias that may occur.

Measurement bias from subjects: Sufficient training is required so that subjects can correctly understand the questions and answer them accurately.

Measurement bias for the source of evaluators: This measurement bias can be reduced by using multiple evaluators for parallel measurements. Although in real-world studies, more often one person (i.e., the attending physician) completes the relevant measurement or evaluation activities, a certain In some cases (e.g. image-based measurements), measurements can be taken again by another evaluator afterwards.

Measurement bias in the source of evaluation tools: use measurement methods with proven reliability and validity, use precise instruments, etc.

##### recall bias

Try to avoid collecting information through the recall of the research population during the design stage, and try to record the data in documents as soon as it is generated. The nested case-control design can avoid the recall bias caused by the traditional case-control method of obtaining intervention measures, baseline data, etc. through recall.

In some cases, reviewing the patient's other health information may help confirm whether the patient's recollection is accurate. For example, if a patient recalls having pain or inflammation after receiving an intervention, the patient's health records, medication records, and electronic medical records on the corresponding dates can be reviewed to see if there is any relevant information for further support.

##### Selection bias caused by loss to follow-up

It is necessary to set up adequate measures to prevent loss to follow-up as much as possible in real-world research plans, including remedial measures that can be adopted after loss to follow-up, such as supplementing relevant data through additional follow-up methods (such as phone calls, home visits), and integrating with other data sources. (such as medical insurance data, death registration data, etc.) links, etc .;

In view of possible missing data when using retrospective data, the methods and principles for handling missing data need to be clarified in advance in the research plan. For missing data, the reasons for loss to follow-up need to be investigated as clearly as possible. If the loss to follow-up is not related to the intervention or outcome, it can be filled in according to the imputation methods and principles specified in the plan. A conservative approach can also be used for imputation, for example, the experimental group is imputed as invalid and the control group is imputed as valid.

##### reporting bias

Selectively presenting favorable results will cause selection reporting bias. The best way to avoid reporting bias is to pre-specify it in the protocol or statistical analysis plan. It is recommended that the protocol be pre-specified on public websites (such as China Clinical Trial Registration Center, ClinicalTrials.gov, etc.) register.

For real-world research using retrospective data, applicants must set up measures to ensure that researchers do not have access to outcome data before formal statistical analysis, to prevent researchers from conducting data mining in order to obtain expected statistical results before the start of the study. For example, when applying statistical analysis methods based on propensity scores, a two-stage design can be adopted. In the first stage, it is necessary to build an outcome data firewall, identify independent statisticians, determine confounding variables, and establish a propensity score estimation model. After a satisfactory balance of confounding variables is achieved in the first stage, a statistical analysis plan will be formulated in the second stage.

##### Unmeasured confounding bias

If all confounders have been collected and modeled correctly, and the sample size is sufficient, estimation bias can be reduced or eliminated through appropriate analytical methods. However, in practice, it is difficult to obtain data on all confounding factors, and some confounding factors are not or cannot be measured. The resulting bias is called unmeasured confounding bias. The effect size of unmeasured confounding is difficult to estimate, and sensitivity analysis can be attempted to assess its potential impact on conclusions.

#### Assess direction and magnitude of bias

Bias is directional, that is, the effect size of an intervention is underestimated or overestimated. Bias also varies in degree. Some relatively small biases may not affect the conclusion of the study. After completing the study, it is recommended to review and summarize any remaining biases during the study and assess the impact on the strength of the evidence.

How to assess bias varies depending on the specific study. For example, for selection bias caused by loss to follow-up, comparing the characteristics of the study population who were lost to follow-up with the characteristics of the study population who were not lost to follow-up, it may be due to the discovery that the intervention was lost to follow-up due to poor effects, thus determining the bias. existence and direction of bias. For measurement bias, some statistical indicators (such as intraclass correlation coefficient, coincidence rate, etc.) can be used to compare the measurement values of different people and different clinical institutions to help evaluate measurement bias.

### Ethical review and informed consent

Ethical review and informed consent for real-world research must comply with relevant regulations and guidelines such as the Declaration of Helsinki of the World Medical Assembly and the Measures for Ethical Review of Life Sciences and Medical Research Involving Humans .

## 4. Statistical analysis of real-world research

### statistical analysis plan

Real-world research needs to include a detailed and specific statistical analysis plan. This part needs to clarify the specific statistical methods and parameter settings used, as well as the reasons and basis for the statistical methods and parameter settings. Compared with traditional randomized controlled clinical trials, real-world studies more often involve statistical analysis methods that adjust for confounding effects due to control bias. Using different analysis methods for the same data, the result values are usually different. If the statistical analysis results are close to the set research success value, there may be different research conclusions using different methods. Even if the same statistical method is used, the results may be different due to subtle differences such as parameter selection. Therefore, the statistical analysis plan needs to be specified in as much detail as possible in advance. The level of detail in the statistical analysis plan must be such that when the analysis is performed according to the plan, there will be no situation where analysts can freely choose analysis methods and parameters.

### Analyze data sets

Different data sets are defined in advance according to different analysis purposes, such as effectiveness data sets, safety data sets, subgroup analysis data sets, etc.

### Confounding adjusted statistical analysis

#### Stratified analysis

Stratified analysis is a commonly used method to control confounding factors. The specific dividing principles for the number of strata need to be specified in advance in the plan and the specific statistical methods used, such as the Mantel-Haenszel method . If other weighted statistical methods are used, the method must be clearly defined. Provenance.

#### multivariable regression analysis

Multivariable linear regression analysis needs to specify the independent variables to be included in advance, and the identified confounding variables need to be included in the model. When the specific independent variables cannot be determined in the design stage, specific rules need to be clarified in the plan and determined according to these rules. Variables included in the model are unique. The number of research subjects (and cases) included in multivariable regression analysis needs to meet the number of cases required for the parameters to be estimated in the model. Generally, the number of research subjects is at least 20 to 30 times the number of covariates. When the outcome event occurs The number of patients is recommended to be at least 10 times the number of covariates. In addition, it is necessary to test whether the basic assumptions of the model used are established, such as independent residuals, zero expectation, homogeneous variances, distribution assumptions, linear assumptions, proportional hazard assumptions such as Cox regression, etc.; properly handle multicollinearity, interaction Function; the linear regression model needs to have acceptable goodness of fit, and the indicators for evaluating the goodness of fit of the model need to be clearly defined in advance (multiple determination coefficient, residual mean square, Malos Cp statistic, Akaike information criterion and Bayesian information criteria, etc.), acceptable thresholds and the basis for their determination.

Multicollinearity testing parameters need to be clarified in advance, such as correlation coefficients, variance expansion factors, condition numbers based on eigenvalues, etc. The thresholds for determining whether there is multicollinearity and the basis for setting the thresholds should be clarified in advance, as well as the subsequent processing of multicollinearity. Principles need to be reasonably and adequately stated. Since multicollinearity relationships more complex than pairwise regressors cannot be well detected, it is not recommended to use only correlation coefficients to test for multicollinearity.

Whether to include interaction terms requires consideration of professional knowledge and experience, as well as statistical analysis. In terms of professional knowledge and experience, if there is already prior information showing that there is an interaction between the regressors, or professional knowledge can determine that there is an interaction between a variable and another variable, the interaction term needs to be included in the model. If the interaction term is statistically significant, but it is impossible to judge professionally whether the interaction is real , it is recommended to perform statistical analysis with or without inclusion as a sensitivity analysis .

For multivariable regression models that use relative risk statistical indicators, such as odds ratio and hazard ratio, the relative values are not as intuitive as the clinical significance of rate difference, mean difference, etc. It is necessary to pay attention to the clinical significance of the magnitude of these indicators. explain. The threshold for determining whether the research hypothesis is valid should be clearly stated in the protocol in advance.

#### Adjustment method based on propensity score

The research proposal based on the propensity score is designed as two independent stages. The main content of the first stage includes identifying all confounding variables, estimating the sample size, building an outcome data firewall, identifying independent statisticians, and establishing a propensity score estimation model. There were iterations until a satisfactory balance of covariates was achieved, and the entire process was blinded to the outcome data; the main purpose of the first phase was to achieve balance between the experimental and control groups. The second stage is to apply propensity scores to estimate the effect size of the intervention, including stratification, matching, inverse weighting, regression and other methods.

In the first stage of the propensity score (i.e., the stage of finding the balance of the propensity score between groups), some characteristic groups were eliminated because they were not evenly distributed among the groups. At this time, attention should be paid to the extrapolation of the research conclusions after excluding some groups. , it is necessary to have a clear record of the process and a sufficient discussion of the extrapolation of the research conclusions. It is recommended that all subjects in the study device group be included in the analysis population, and a control group data source can be added if the propensity score is not balanced.

The method and related parameters used for effect size estimation need to be clarified in advance. For the stratified method, it is necessary to clarify in advance the specific division of stratification, as well as the threshold and basis for determining the balance between covariate groups, the calculation method of layer weight coefficient, etc.; for the matching method, it is necessary to clarify in advance the matching ratio of the experimental group and the control group (such as 1 :1 or 1:n), the matching method used (such as exact matching method, nearest neighbor matching method, caliper matching method, etc.), the matching success determination threshold and its basis. Usually, a single patient data is only used for one matching .

Inverse weighted sum regression is not recommended. Inverse weighting determines the individual weight coefficient based on the propensity score to weight the effect value. When the propensity score is close to 0 or 1, the weight will be too large or too small, and inverse weighting has requirements for the accuracy of the propensity score model. Very high. The regression method directly incorporates the propensity score into the model, assuming that the regression model of the outcome variable, group, and propensity score is correct, which is usually difficult to verify. Like inverse weighting, it is more sensitive to the accuracy of the propensity score model, and it is difficult for the regression method to remain blind to the outcome data during the design stage.

#### Other adjustment methods

Adjustment methods used to control confounding also include more complex statistical methods such as marginal structural models, instrumental variables, and structural equation models, which currently have few practical applications in clinical evaluation of medical devices.

### Handling missing data

A variety of reasons can lead to missing data, such as poor compliance, lack of improvement, side effects, poor treatment experience, and external factors unrelated to the study. To properly handle missing data, you need to first explore and clarify the missing data mechanism, and use corresponding statistical methods according to the missing mechanism. for processing. Missing data that are not related to all measured or unmeasured variables are missing completely at random, which will not introduce bias; missing data that are related to measured variables but are not affected by unmeasured data are missing at random; missing data are not related to themselves Relevant data become non-random missing, such as missing data due to poor efficacy or side effects.

The best strategy for dealing with missing data is to prevent them through sound study design and high-quality conduct. For prospective real-world studies, the study can be shortened by selecting experienced and responsible researchers, setting up a variety of different follow-up methods, using positive controls, using easily measurable outcome indicators, reasonable data collection forms, and fast data entry methods . time, training, etc. to reduce missing data; for retrospective real-world studies, the degree of missing data in the database needs to be assessed. If there are many missing data, or the missing data cannot be determined to be completely random, it is not recommended to use this data source to conduct real-world research.

Currently, missing data imputation methods can be divided into two categories: single imputation and multiple imputation methods, and specific processing methods for missing data need to be specified in advance. It is recommended to use a conservative carry-forward method for the single imputation method. For example, all missing values in the experimental group should be filled in as invalid, and all missing values in the control group should be filled in as valid. It should be noted that the single imputation method will reduce the variance and the parameter accuracy will be overestimated, which is reflected as: Confidence intervals narrow. For multiple imputation, the imputation model, analysis model, number of interpolations, merging rules and corresponding determination basis need to be specified in advance. Model diagnostic methods and diagnostic indicators for interpolation and analysis models are specified in advance, such as the proportion of missing information, relative increased variance, relative efficiency, and parameter stability.

### Subgroup analysis

If the population included in the study is heterogeneous, subgroup analysis may be considered to explore whether the effect sizes are consistent in different subgroups. If they are inconsistent, attention should be paid to the extrapolation of the research conclusions to the expected population at this time to avoid the product having significant efficacy in some populations but being ineffective in other populations. The wider the inclusion and exclusion criteria, the easier it is for heterogeneity to occur. Subgroup analysis can be determined in advance in the protocol based on previous research experience and knowledge. Failure to specify it in advance will reduce the credibility of the subgroup analysis results.

### sensitivity analysis

Sensitivity analysis is used to assess the robustness of study results and may be required in a number of different situations, especially for observational real-world studies. Common application situations of sensitivity analysis in real-world studies include violation of model assumptions, statistical methods (such as whether to include interaction terms, treatment of collinear variables, etc.), variable selection, assessment of the impact of unmeasured confounding, assessment of the impact of intervention deviation, and data Missing filling, conflicting data processing, outliers, inconsistent variable definitions, different population subgroups, baseline imbalance, etc. All sensitivity analysis results need to be reported and cannot be reported selectively. If the results of the sensitivity analysis are inconsistent with the results of the main analysis, it indicates that the research conclusion is not robust enough and additional research may need to be conducted for further verification.

## 5. Research report

Research reports must follow the overall principles of completeness , accuracy , and standardization . There are differences in the content of different types of real-world research reports . The content of pRCT research reports can refer to the CONSORT guideline for pragmatic trials. Observational studies such as cohort design and case-control design can refer to the STROBE guideline. You can also refer to other applicable documents , such as the STaRT -RWE checklist . , to ensure that the clinical report elements are complete . Based on the above considerations , special attention should be paid to the following :

1. Screening flow chart: Provide a screening flow chart for the corresponding research objects, explaining how to gradually screen out qualified analysis objects from the original database during the research process, giving the sample size of the original database, the number of research objects excluded in each step, and the corresponding exclusions Reasons and sample size of research subjects finally included in the analysis;
2. Description of the basic characteristics of the data source: including population representativeness, data quality, etc.; the data source’s own quality control measures, high-level literature published based on the data source, and other information.
3. Description of the basic characteristics of the population: describe in detail the baseline characteristics of the research subjects, the baseline characteristics of each group of cases when they were selected, and whether the baseline data is balanced.
4. Variable data extraction: Provide specific definitions of variables such as exposure, outcome, confounding factors and effect modifiers. If automatic methods are used to extract variable data from observational databases, list the specific extraction algorithms for each variable and provide the accuracy of the corresponding algorithm. Sexual verification information.
5. Statistics on concomitant medication and concomitant treatment: record and count the use of other intervention measures (drug therapy and other diagnostic and treatment methods).
6. Carry out statistical analysis according to the plan and present all statistical analysis results, including: main analysis results, effectiveness analysis results, safety analysis results, subgroup analysis results, interim analysis results, and sensitivity analysis results;
7. Handling of missing data: Describe the quantity and specific circumstances of missing data, list the handling of missing data, and comply with the provisions of the plan. If it is inconsistent with the data missing processing method specified in the plan, provide reasonable reasons.
8. Conflicting data processing: statistics describing important information of data from different sources, sensitivity analysis results for contradictory data
9. Risk of bias and control : List all possible biases in a table, list the measures to reduce/eliminate bias one by one, and list the residual risk of bias after taking the measures.
10. Discussion and conclusions : Discuss study limitations and consider potential sources of bias or imprecision. Discuss the direction and magnitude of potential bias ; discuss the extrapolability (external validity) of the study results ; provide a careful and comprehensive interpretation of the results in light of the study objectives, limitations, multiple analytical methods, results from similar studies, and other relevant evidence .

## 6. References

[1] State Food and Drug Administration. Technical Guiding Principles for the Clinical Evaluation of Medical Devices with Real-World Data (Trial): Notice of the State Food and Drug Administration on the Release of Technical Guiding Principles for the Clinical Evaluation of Medical Devices with Real-World Data (Trial) . 2020 Year No. 77[Z].

[2] Gao Pei, Wang Yang, Luo Jianfeng, et al. Technical specifications for statistical analysis of research on evaluating treatment outcomes based on real-world data [J]. Chinese Journal of Evidence-Based Medicine. 2019;19(7):787-793.

[3] Wen Zehuai, Li Ling, Liu Yanmei, et al. Technical specifications for pragmatic randomized controlled trials. Chinese Journal of Evidence-Based Medicine [J]. 2019;19(7):794-802.

[4] Nie Xiaolu, Peng Xiaoxia. Reporting specifications for observational studies using routinely collected health data - RECORD specifications. Chinese Journal of Evidence-Based Medicine. 2017;17(4):475-487.

[5]Yue LQ, Campbell G, Lu N, Xu Y, Zuckerman B. Utilizing national and international registries to enhance pre-market medical device regulatory evaluation. Journal of Biopharmaceutical Statistics [J] . 2016;26 *(* 6 ):1136- 1145.

[6] Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions [J] . *BMJ* . 2016;355:i4919.

[7]Wang SV, Pinheiro S, Hua W, et al. STaRT -RWE: structured template for planning and reporting on the implementation of real world evidence studies [J] . *BMJ* . 2021;372:m4856.