Guideline on Enrichment Strategies and Designs in Clinical Trials

(For Public Review)

English Translation by: Jie Chen

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

Center for Drug Evaluation, NMPA

March 14, 2023

Table of Contents

1

Guideline on Enrichment Strategy and Design in Clinical Trials

1. Overview

The purpose of clinical trials is to demonstrate the efficacy and safety of an investigational drug in subjects. However, due to the complexity of pathophysiological characteristics of subjects and mechanism of action, the therapeutic effects may not be the same among different subjects, thus affecting the efficiency of clinical trials. In order to maximize the benefit of subjects from investigational drugs and improve the efficiency of clinical trials, the concept of enrichment strategies has emerged.

Enrichment refers to the prospectively and precisely defining the target population in a clinical trial that maximizes the benefit from the investigational drug based on certain characteristics of subjects (such as demographics, pathophysiology, histology, genomics and proteomics, etc.). There are many subject-selection enrichment strategies in clinical trials, for example, subjects can be enriched based on their responses to the study drug, or their insensitivity to existing drugs, or their likelihood to have endpoint events.

This guideline describes the principle and method of commonly used enrichment strategies and designs, their advantages and disadvantages, and explains the key considerations from practical applications and regulatory

perspectives. In this guideline, "enrichment strategy" primarily refers to the method used to select subjects who may benefit from randomized controlled clinical trial, but can also be extended to the single-arm trial using external (historical or parallel) control.

The guideline is applicable to confirmatory clinical trials for the purpose of supporting drug registration and marketing authorization and can also be used as a reference for clinical trials with non-registration purposes.

2. Applicability of Enrichment Strategy and Design

Broadly speaking, the concept of enrichment is used in all clinical trial designs, which can be reflected in the inclusion and exclusion criteria of subjects with the purpose being to enroll subjects who most likely respond to the investigational drug, so as to improve the efficiency of clinical trials. For example, when studying a cholesterol-lowering drug to reduce the incidence of cardiovascular events, clinical trials may only include patients whose total cholesterol concentration in the blood is higher than a threshold. In fact, different enrichment strategies and designs may be chosen based on the disease area, mechanism of action of the drug, and the response of the subjects. The applicability of enrichment strategies can be assessed from the aspects of scientific validity, interpretability of trial results, and generalizability in medical practice.

(1) Scientific validity: This includes scientific rationale for screening

subjects, sensitivity and specificity of screening methods that meet certain requirements, measures to avoid bias (such as randomization, blind method, etc.) in the design of the trial, and the control of type I errors.

(2) Interpretability of trial results: This refers to that the efficacy of the investigational drug in the enriched population can be explained in terms of the pathophysiology, genomics, genetics, or drug mechanism of action of the disease; if it cannot be explained due to limited knowledge of biology, medicine, or pharmacology, the efficacy of the investigational drug in the enriched population needs to be reproducible to certain extent.

(3) Generalizability in medical practice: Enrichment strategies should be able to be widely used in clinical practice in order to timely and accurately identify patients who respond or are sensitive to investigational drugs. Sometimes, the generalization of a patient screening method is not possible due to its complexity, low sensitivity, high cost, etc., or the screening method is time-consuming and cannot enrich patients at the beginning of treatment, which will affect the generalizability of enrichment strategies and methods.

3. Commonly used enrichment strategies and designs

According to the main research question and implementation process of clinical trials, different enrichment strategies can be used. There are five commonly used types of enrichment strategies: homogeneous enrichment, prognostic enrichment, predictive enrichment, hybrid (prognostic and

predictive combined) enrichment, and adaptive enrichment.

In practice, the enrichment strategies and designs are usually chose n according to some markers related to the mechanism of drug action. Here "markers" are defined as various characteristic variables such as ep idemiological factors (e.g., demographics), past medical history, family h istory, clinically observed variables (e.g., disease severity), laboratory tes ts (e.g., pathophysiology, drug metabolism), genomics and proteomics rel ated to subject prognosis or response to drug treatment. According to th e different roles of markers, they can be divided into prognostic, predict ive, and hybrid markers. In addition, in some disease areas, there may be no clear marker, and enriched subjects are generally selected based o n their response to treatment during screening, or data from other clinic al trials and literature reports.

3.1 Enrichment for Homogeneity

Enrichment for homogeneity refers to a strategy of reducing the het erogeneity of subjects to improve the power of clinical trials. The simpl est and most practical way to reduce heterogeneity is to select subjects with stable disease as much as possible, accurately define the selected s ubjects, and accurately measure the status of the disease and related var iables. For example, in trials of hypertension drugs, in order to screen out subjects with relatively stable blood pressure, subjects' blood pressur e may be measured for a period of time before enrollment to exclude s

ubjects with large variations in blood pressure.

In general, to more accurately define an enriched population, the following aspects should be considered, in addition to the conventional inclusion and exclusion criteria:

(1) Inclusion criteria: Inclusion criteria should be more carefully defined to ensure consistent baseline characteristics among enrolled subjects.

(2) Exclusion criteria: Exclude those subjects who (1) are too sensitive to placebo; (2) have unstable baseline test results, such as subjects with unstable conditions or symptoms during the primary screening period; (3) may die prematurely due to a concomitant disease; (4) take drugs with similar therapeutic effects to the test drug; (5) may not tolerate the test drug treatment; (6) may withdraw from the study early due to complications.

(3) Compliance: Subjects with good compliance should be included, i.e., subjects who do not withdraw for non-medical reasons (e.g., inconvenience to go to the study site), and subjects who are able to adhere to the treatment according to the trial protocol, so as to reduce differences due to excessive withdrawal of subjects or use of different treatment methods. Patient compliance identification and selection must occur prior to randomization.

(4) Training: The investigators and clinical trial coordinators shall receive relevant training to ensure that the subject enrollment strictly follows the protocol and the study is conducted in accordance with the protocol.

3.2 Prognostic enrichment

Prognostic enrichment refers to a strategy of enrolling high-risk patients based on their prognostic markers increase the power of the study. In general, high-risk patients are more likely to observe an endpoint event of interest or disease progression (especially those who are more likely to have a prognostic outcome or disease progression). This strategy mainly increases the absolute effect of the trial, not the relative effect. For example, in a clinical trial aiming to reduce the incidence of an endpoint events, after a period of treatment, the incidence of the endpoint events is reduced from 10% to 5% in the high-risk population and from 1% to 0.5% in the low-risk population. Although the relative effects of both are reduced by 50%, the former obviously requires less sample size or shorter follow-up time to observe the efficacy of the investigational drug. There are two commonly used prognostic enrichment designs.

(1) Event-based enrichment design

In studies with the reduction of the incidence of endpoint events as the primary objective, the investigational drug is generally considered more effective in reducing more events in higher risk population. Therefore, enrolling high-risk subjects should be considered. In general, when the sample size is unchanged, the high-risk population is more likely to have more endpoint events than the low-risk population, and the incidence of endpoint events is greatly reduced after treatment, so the test is more powerful. This strategy is often used in studies of drugs for anti-tumor and cardiovascular diseases, such as breast or

ovarian cancer prevention in female population with BRCA1/2 mutation; and in studies of hypolipidemic drugs, patients with high concentration of low-density lipoprotein (LDL), low concentration of high-density lipoprotein (HDL) and high concentration of C-reactive protein (CRP) in blood are selected for trials. In some disease areas, such as Alzheimer's disease and various cancer drug studies, high-risk patients can also be screened by genomic or proteomic screening.

(2) Slowing disease-progression-based enrichment design

Prognostic enrichment designs can also be used to study an experimental drug that can slow disease progression, such as clinical trials for Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, chronic obstructive pulmonary disease, and malignancies, where subjects with potentially faster disease progression can be enrolled. For example, rheumatoid arthritis patients with the following characteristics tend to have faster disease progression: rheumatoid factor positive, certain clinical characteristics (such as multiple joints affected, diseases other than joints, subcutaneous nodules, limited activities, etc.) and abnormal laboratory results (such as decreased hemoglobin); patients with chronic obstructive pulmonary disease (COPD) may have faster disease progression: a history of recent onset (at least one attack in the past year) or higher plasma fibrinogen. In the study of anti-tumor drugs, common prognostic markers include histological grade, vascular invasion, molecular subtype, and metastatic tumor nodules.

It should be noted that if there is an interaction between a prognostic marker and the trial drug, i.e., the study drug has an effect on both marker-positive and marker-negative patients, but the efficacy is different, the prognostic marker can also play a predictive role, and such markers are usually called hybrid markers.

3.3 Predictive enrichment

Predictive enrichment refers to an enrichment strategy in which subjects who are most likely to respond to the test drug are selected based on their physiological or disease characteristics. For example, in targeted anti-cancer therapy, subjects may respond to a treatment based on drug-related targeted genes or proteins, or physiological functions (e.g., renin hypertension/hypotension, chronic heart failure score). Adopting this strategy can increase both the absolute effect and relative effect of the test drug, so that a higher power can be obtained with a smaller sample size. This enrichment strategy is useful when only a small percentage of subjects with a disease respond to the test drug (e.g., only a subset of subjects have drug-acting receptors). In practice, subjects can be selected based either on the investigator's knowledge of the disease (e.g., various markers) or on previous trial data and results.

(1) Enrichment design based on pathophysiological characteristics

Subjects whose pathophysiological characteristics of the disease could suggest a better response to the test drug. Pathophysiologically based

enrichment indicators can be biomarkers (such as gene mutation affecting tumor growth, gene/protein expression level), imaging characteristics, and / or clinical characteristics related to disease phenotype (such as disease staging, typing, etc.). Depending on the nature of the enrichment markers, the strategies can be classified into the following categories:

① **Gene or protein marker-based:** Anti-tumor drugs usually target relevant receptors, enzymes, hormones, or other endogenous active substances on or inside the tumor cell surface, for which an enriched population can be selected based on one or more corresponding gene or protein markers. For example, trastuzumab is mainly used to treat breast cancer patients who are human epidermal growth factor receptor 2 (HER-2) protein-positive. Some cell receptors that initially act as protein markers, but are later confirmed as tumor gene markers (such as EGFR and BRAF gene mutations) have been used to define the pathophysiological status and to select the subjects who may benefit from target therapy. The accuracy and precision of marker detection test is essential when using a gene or protein marker in an enrichment design. If the diagnostic test is not accurate, it will not only affect the effect of enrichment design that may reduce the study power, but may also increase the type I error in non-inferior trials.

② **Drug metabolite-based:** The metabolic capacity of different subjects to the same investigational drug could be different. Enrolling subjects who can produce a sufficient amount of active metabolites can improve the efficiency of

clinical trials. In some cases, higher doses are given to patients who can produce less active substance, helping them to produce enough amount of active substance so that the efficacy of the drug is more likely to be observed. Patients who are completely unable to produce the active ingredient should be excluded from the trial.

③ **Tumor metabolites-based:** A clinical trial of antineoplastic drug may select subjects by measuring the amount of tumor metabolites in the tissue or blood. For example, only those subjects with metabolic reactions were enrolled, or grouped by the degree of metabolic reaction in cancer patients, and the primary analysis can be performed on subjects with metabolic reactions.

(2) Enrichment design based on evidence of response to study drug

Such an enrichment design may allow selection of potentially suitable subjects based on their response to the study drug (or similar drugs in the past) during the screening period.

① **Screening subjects who respond:** For clinical trials in which the subjects responding to the investigational drug cannot be identified based on the markers prior to the study, a reasonable screening period should be set during which all subjects are given the investigational drug. Subjects who respond to treatment are selected based on a predetermined primary or surrogate endpoint and then are randomized. Selecting responders can be performed using a two-stage randomized withdrawal design. In the first stage, subjects are tested to determine whether they can respond to the study drug (which can be done in

a single-arm or randomized trial); in the second stage, responders are randomized to receive the test drug (continue to use the experimental drug) or placebo (withdraw the experimental drug) and non-responders are excluded from the study. For example, a study investigating a cholesterol lowering drug can use the randomized withdrawal design, in which subjects with high cholesterol are enrolled in the first stage and responders (based on cholesterol reduction) to the test drug are selected to be randomized to receive the test drug or control drug in the second stage study. may also be used for selecting subjects who respond. The design is generally divided into two stages, that is, the first stage tests whether the subject responds to the test drug (single-arm open trial or randomized controlled trial can be used), the second stage responds to the test drug, the subject is randomly assigned to the test group (continue to use the test drug) or the placebo group (withdraw the test drug), and the subject who does not respond withdraws from the trial (Figure 1). To judge whether the patient has responded to the study drug, evaluation can be performed based on some surrogate indicators such as symptoms, signs, laboratory tests and disease recurrence.

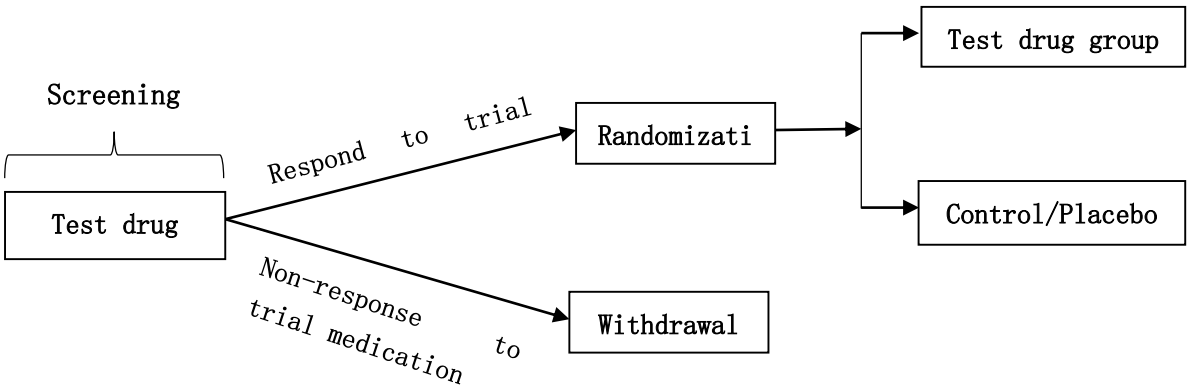

Figure 1 Flow Chart of Randomized Withdrawal Study

The randomized withdrawal design improves the efficiency of clinical trials by selecting subjects who respond to the test drug; at the same time, the long-term efficacy or safety of the study drug can be studied among subjects who continue using the drug in the second stage and the withdrawal effect can be studied among subjects who withdraw from the trial in the second stage. On the other hand, this design is more ethical, that is, the trial can be terminated in a timely manner once treatment has failed and can be used in pediatric drug research. This design by screening subjects in the first stage and randomized them in the second stage based on their prior responses may screen more subjects and stratified randomization can be done according to their degrees of responses. The trial can first study drug effect among subjects with strong responses and if positive outcome is observed, then among subjects with weak responses. However, the design may not be suitable to study test drugs with long-term residual effect or lethal or harmful withdrawal effect, or drugs with a long duration between treatment initiation and subject responses.

② **Selecting subjects based on historical data or literature reports:** Subjects can be enrolled based on the characteristics identified in previous studies, i.e., if little or no significant treatment effect was observed in the overall population, but a significant effect may likely be achieved in a subpopulation, enrollment can be restricted only to the sub-population. For example, the combination isosorbide dinitrate/hydralazine hydrochloride is a drug for treatment of severe heart failure, and previous studies have found that its therapeutic effect on African-Americans is significantly better than that on Caucasians, then the subsequent randomized placebo-controlled trial enrolled 1050 African-American patients with heart failure which demonstrates the effectiveness of the combination drug in this sub-population of patients with heart failure.

(3) Enrichment by selecting non-responders to existing drugs

In addition to selecting subjects who respond to the test drug, subjects who do not respond to the existing drug may be considered for a trial in order to better show the treatment effect of the test drug that has a different mechanism of action from the existing drug.

Enrichment for selecting non-responders is appropriate for clinical trials with the following conditions: the investigational drug has a different mechanism of action from an existing drug, or the investigational drug is at least slightly more effective than an existing drug. If no selection of subjects is performed, a larger sample size may be required to show efficacy of the

investigational drug; on the contrary, if only subjects who do not respond to the existing drugs are selected, because the response rate of the control group is very low, a smaller sample size may be necessary to demonstrate that the experimental group is superior to the control group. It must be noted that for certain potentially life-threatening, progressive diseases, randomizing subjects who do not respond to the control drug may be unethical.

3.4 Composite enrichment

Composite enrichment refers to an enrichment strategy using hybrid markers (e.g., prognostic and predictive markers) simultaneously to reduce subject heterogeneity. For some disease areas, the occurrence, development and prognosis of diseases may have complex mechanisms and heterogeneity could be high among subjects. Therefore, it is unlikely to enrich the subjects using a single marker, while the use of multiple markers or a composite marker (such as a comprehensive score) for patient enrichment can effectively reduce the heterogeneity of subjects and thus improve the study efficiency.

It should be noted that individual markers that compose of the composite marker should be listed when using the composite marker score and their roles and relationships should be elucidated. If different individual markers are given different weights, the biological principle should be described in detail.

3.5 Adaptive enrichment

Adaptive enrichment strategy refers to a strategy of mid-course modification on the target study population (e.g., changing the inclusion and

exclusion criteria, adjusting sample size) based on the results of pre-defined interim analysis in the study protocol, on the premise of ensuring the rationality and integrity of the trial.

When the efficacy of the investigational drug is uncertain in subjects with positive and negative markers, the trial can enroll both marker-positive and -negative subjects before the interim analysis and adaptative enrollment can be made according to the results of interim analysis. If the interim analysis shows that the efficacy in marker-negative subjects is much lower than that in marker-positive subjects, then enrollment can be restricted only to marker-positive subjects and enrollment of marker-negative subjects should be reduced or completely stopped. The trial can also first enroll marker-positive subjects and then marker-negative subjects if the interim analysis shows therapeutic effects in marker-positive subjects; otherwise, the trial should stop.

In general, if the relationship of a marker to the therapeutic effect is uncertain, then it is necessary to enroll marker-negative subjects, which can help assess the benefits and risks of the drug when used in the full population. When the predictability of a marker is uncertain, the primary analysis can be performed in the full population; if the marker-positive population and the full population are primary analysis populations, the test level α shall be split according to certain rules. In either case, the testing hypothesis should be specified in the protocol beforehand and the type I error needs to be controlled.

4. Related Considerations of Enrichment Strategy and Design

4.1 Sensitivity and specificity of marker detection

When screening tests are used to select subjects, the reliability of the screening method must be taken into account in order to more accurately select subjects who are at high risk or most likely respond to the test drug. Ideally, a screening test should have a high sensitivity for selecting subjects at high risk or who respond to test drug and a high specificity for identifying subjects at low risk or who do not respond to test drug.

When biomarkers are used to screen subjects, if the threshold value of predictive markers cannot be accurately given, the sensitivity and specificity of different threshold points of markers can be analyzed by Receiver Operating Curve (ROC) analysis, and the screening effect can be measured by the area under ROC curve. With regard to the determination of predictive marker thresholds, it is generally possible to first give a preliminary threshold in the early research stage and then adjust it through trials with larger samples to obtain a more reliable threshold.

4.2 Whether the subjects with positive and negative markers are included

The enrichment design can enroll either only marker-positive subjects or both marker-positive and marker-negative subjects. However, the key issue in the enrichment design is the proportion of marker-positive and marker-negative subjects to be enrolled. In general, the following enrichment strategies can be considered:

(1) Enrolling only marker-positive subjects

If the mechanism of action or available data show that the investigational drug has significant efficacy in marker-positive subjects, but has less or no efficacy in marker-negative subjects, then the trial should not enroll marker-negative subjects.

(2) Enrolling both marker-positive and -negative subjects

If the mechanism of action or available data suggest that marker-positive markers subjects may have better efficacy than those with marker-negative subjects, the trial can enroll both marker-positive and -negative subjects if the test drug is less toxic. This strategy has the advantage of providing a reasonable benefit-risk estimate in a non-enriched population.

If a marker can be identified before the start of the trial, stratified randomization can be implemented within stratum, and the primary analysis can be restricted to marker-positive subjects. In practice, the primary analysis can also be performed in the full population, or simultaneously in the full population and in marker-positive subjects, with appropriate control of type I error.

In general, if the threshold for a marker or the magnitude of response for marker-negative subjects is uncertain, it is necessary to include marker-negative subjects.

4.3 Inclusion population and analysis set

The main concerns of using enrichment strategies are the applicability and extrapolability of the findings, that is, when using an enrichment design, it is

important to consider whether this enrichment strategy can be used in medical practice to identify subjects who respond to the study drug and whether the drug has similar efficacy in a wider patient population. Note that it is equally important to study patient populations who do not meet the enrichment criteria. It should also be noted that the enrolled subjects and the primary analysis set identified in the trial may be different (the latter may be a subset of the former), but these must be clearly defined in the study protocol. When genetic or other test results are not immediately available and patients need timely treatment, the overall population can be selected to provide more safety information, but the primary efficacy analysis can be a subset of the study population.

4.4 Different effects of enrichment strategies on superiority and non-inferiority trials

The use of markers to select subjects has a different impact on superiority and non-inferiority trials. For superiority trials using predictive enrichment, if the screening method is insensitive, more subjects need to be recruited for screening in order to enroll a pre-specified sample size of enriched subjects; if the screening method has a lower specificity, the sample size of enriched subjects may be large or the trial mat last longer in order to obtain the sufficient number of endpoint events. Nonetheless, it does not inflate the type I error in superiority trials.

However, for non-inferiority trials, the accuracy of screening method will not only affect the sample size or duration of the trial, but may also inflate the

type I error rate. For example, a non-inferiority trial using a prognostic enrichment strategy may result in a lower efficacy estimate of the positive control group than in previous studies if the screening method of the positive control is different from that in previous studies, thereby increasing the type I error. The impact on type I error using predictive enrichment strategies is more complex in non-inferiority trials, depending on whether the marker is related to the efficacy of the test drug and the active comparator, or to the efficacy of only one of the treatments. Therefore, the screening method for selecting subjects in non-inferiority trials should preferably be consistent with the screening method of positive control in previous studies, or both screening methods have similar sensitivity and specificity.

4.5 Control of type I errors

For the enrichment design including both enriched population and non-enriched population, different hypothesis testing strategies may be considered according to the accuracy of screening method and the subject responses to treatment. If there are multiple hypotheses, such as hypotheses in marker-positive population and on overall population, multiplicity adjustment needs to be considered. If there is only one hypothesis, such as hypothesis in marker-positive population, there is no need to consider the problem. The distribution of type I error α under different assumptions can be set according to the degree of response of the marker-positive population to the drug, the proportion of the positive population in the overall population, and the sample size required

according to the prespecified power of the test. When hypothesis testing is performed for overall population and enriched population, independent or sequential testing strategy may be adopted for hypothesis testing.

5. Regulatory Considerations

5.1 Clarifying the enriched population

Whether, when, and what enrichment strategy is used in clinical trials depends mainly on whether the enriched population can be accurately identified, which has a clear impact on the specification of product label and subsequent medical practice. If the enriched population cannot be accurately identified using the enrichment strategy and design, it may not accurately define the patient populations who respond to the treatment, thus failing to accurately guide drug use in clinical practice.

5.2 The efficacy of non-enriched populations should not be neglected

After the efficacy and safety of the investigational drug in the enriched population have been confirmed, the corresponding information of the investigational drug in the non-enriched population should also be considered. Further studies in non-enriched populations may provide a more comprehensive picture of the benefit-risk profile of the test drug and thus provide a basis for its use in a broader patient population.

For drugs approved based on prognostic enrichment in high-risk populations, different outcome measures may be used in subsequent trials

among low-risk populations, such as mortality endpoint in high-risk populations and a composite outcome measure in low-risk populations to help improve trial efficiency.

5.3 Predetermine the study protocol and communicate with the regulatory authorities

In general, subject selection should be preplanned and determined prior to trial start. If characteristic variables or markers are known enrichment can be implemented when screening subjects. When the effect or distribution of characteristic variables or markers in the study population is uncertain, adaptive enrichment may be considered, that is, enrichment can be adopted during the course of the trial according to the interim analysis of the accumulated data. Regardless of the strategy and design used, the adjustment methods and processes should be described in advance in the study protocol to ensure their integrity and validity, and adequately communicated to the regulatory authorities.

References

1. FDA. Enrichment Strategies for Clinical Trials to Support Determination of effectiveness of Human Drugs and Biological Products: Guidance for Industry. The US Food and Drug Administration: Silver Spring, 2019.
2. EMA. Points to consider on multiplicity issues in clinical trials. European Medicines Agency: EMA, 2002.
3. EMA. Guideline on the investigation of subgroups in confirmatory clinical trials. European Medicines Agency: EMA, 2019.
4. ICH. ICH E5. Ethnic Factors in the Acceptability of foreign Clinical Data E5 (R1). ICH Harmonised Tripartite: ICH, 1998.
5. ICH. ICH E5. Implementation Working Group Questions & Answers (R1). ICH Harmonised Tripartite: ICH, 2006.
6. Wang S, Hung HMJ, O Neill RT. Genomic Classifier for Patient Enrichment: Misclassification and Type I Error Issues in Pharmacogenomics Noninferiority Trial. Stat Biopharm Res, 2011, 3:310-319.
7. Temple RJ. Special study designs: early escape, enrichment, studies in non-responders. Communications in Statistics - Theory and Methods, 1994, 23:499-531.
8. Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med, 2004, 351:2049-2057.
9. Temple R, Stockbridge NL. BiDil for heart failure in black patients: The U.S. Food and Drug Administration. Ann Intern Med, 2007, 146:57-62.
10. Institute OM. Small Clinical Trials: Issues and Challenges. Washington, DC. The National Academies Press; 2001.
11. D 'Agostino RB. The Delayed-Start Study Design. New Engl J Med, 2009, 361:1304-1306.
12. Priscilla Velengtas, Penny Mohr, Messner DA. Making informed decisions: Assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research. National Pharmaceutical Council ed. Washington, DC; 2012.
13. Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. Nat Rev Clin Oncol, 2014, 11:81-90.
14. Wang SJ, Hung HM, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. Biom J, 2009, 51:358-374.
15. Temple R. Enrichment of clinical study populations. Clin Pharmacol Ther, 2010, 88:774-778.
16. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. Pers Med, 2010, 7:33-47.
17. Wang SJ, O 'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharm Stat, 2007, 6:227-244.
18. Liu A, Liu C, Li Q, et al. A threshold sample-enrichment approach in a clinical trial with heterogeneous subpopulations. Clinical trials (London, England), 2010, 7:537-545.
19. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined effect subset. J Natl Cancer Inst, 2007, 99:1036-1043.

20. D 'Amico AV, Chen MH, Roehl KA, et al. Preoperative PSA velocity and the risk of deat h from prostate cancer after radical prostatectomy. N Engl J Med, 2004, 351:125-135.
21. Fan C, Oh DS, Wessels L, et al. Concordance among Gene-Expression – Based Predictors f or Breast Cancer. New Engl J Med, 2006, 355:560-569.
22. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study (4S). Lancet, 1994, 344:1383-1389.
23. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med, 2008, 359:2195-2207.
24. EMA. Quali cation Opinion of Alzheimer's Disease Novel Methodologies/biomarkers for BM S-708163. European Medicines Agency: EMA, 2011.
25. EMA. Quali cation opinion of low hippocampal volume (atrophy) by MRI for use in regula tory clinical trials - in pre-dementia stage of Alzheimer's disease. European Medicines Agenc y: EMA, 2011.
26. Amur S, LaVange L, Zineh I, et al. Biomarker Qualification: Toward a Multiple Stakeholde r Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clin Phar macol Ther, 2015, 98:34-46.
27. FDA. Quali cation of Biomarker – Plasma Fibrinogen in Studies Examining Exacerbations a nd/or All-Cause Mortality in Patients with Chronic Obstructive Pulmonary Disease. US Food and Drug Administration: Silver Spring, 2016.
28. FDA. Quali cation of Biomarker – Total Kidney Volume Studies in Treatment of Autosoma l Dominant Polycystic Kidney Disease. US Food and Drug Administration: Silver Spring, 20 16.
29. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet, 2005, 365:1687-1717.
30. Loo E, Khalili P, Beuhler K, et al. BRAF V600E Mutation Across Multiple Tumor Types: Correlation Between DNA-based Sequencing and Mutation-specific Immunohistochemistry. Ap pl Immunohistochem Mol Morphol, 2018, 26:709-713.
31. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melano ma with BRAF V600E mutation. N Engl J Med, 2011, 364:2507-2516.
32. Kowanetz M, Zou W, Gettinger SN, et al. Differential regulation of PD-L1 expression by i mmune and tumor cells in NSCLC and the response to treatment with atezolizumab (anti-PD -L1). Proc Natl Acad Sci USA, 2018, 115: E10119-E10126.
33. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibi tor immunotherapy. Nat Rev Cancer, 2019, 19:133-150.
34. Schrock AB, Ouyang C, Sandhu J, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. Ann Oncol 2019, 30:1096-1103.
35. Hughes RA, Donofrio P, Bril V, et al. Intravenous immune globulin (10% caprylate-purified chromatography) for the treatment of chronic inflammatory demyelinating polyradiculoneurop athy (ICE study): a randomised placebo-controlled trial. Lancet Neurol, 2008, 7:136-144.
36. Parkinson Study Group. A controlled, randomized, delayed-start study of rasagiline in early Parkinson disease. Arch Neurol, 2004, 61:561-566.

37. Singh BN. Comparative efficacy and safety of bepridil and diltiazem in chronic stable angin a pectoris refractory to diltiazem. The Bepridil Collaborative Study Group. Am J Cardiol, 19 91, 68:306-312.
38. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design i n the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther, 2009, 86:97-100.

Appendix 1: Glossary

Sensitivity: one of the basic indicators to evaluate the accuracy of diagnostic test and screening test. In an enrichment study for a drug clinical trial, sensitivity indicates the probability that a subject who is at high risk for an endpoint event or who responds to the drug can be correctly identified.

Adaptive Enrichment Design: According to the predetermined plan and based on the interim analysis results of clinical trial data, on the premise of ensuring the rationality and integrity of the trial, it is allowed to adaptively update the inclusion and exclusion criteria during the trial and select the adaptive design for the subjects who may benefit from the treatment.

Randomized Withdrawal Design: In such a design, all subjects receive the investigational drug during the initial open-label period, then subjects who do not respond to the drug withdraw from the trial, and subjects who respond (enriched) are randomized to receive the investigational drug or placebo in the second phase of the trial.

Specificity: It refers to one of the basic indexes for evaluating the accuracy of diagnostic test and screening test. In an enrichment study for a drug clinical trial, specificity indicates the probability of being able to correctly identify

subjects who are at low risk for an endpoint event or who do not respond to the drug.

Heterogeneity: In clinical trials, heterogeneity is reflected in two levels: individual and group. The former usually refers to that there are different characteristics among subjects. Different nature or status of individual subjects may lead to different responses to treatment; the latter usually refers to that subjects from different centers, ethnicities and regions have different characteristics, which may lead to different responses to treatment for different subjects.

Predictive Enrichment: refers to a study strategy or design that selectively includes the subjects who may respond to the treatment. These subjects have common biological and histopathological characteristics with predictive significance and can more sensitively display the investigational drug.

Prognostic Enrichment: A research strategy or design that selectively includes subjects who are more likely to experience an endpoint event, such as death or disease worsening, thereby reducing the sample size required to achieve a statistically significant effect.

Appendix 2: Comparison Table between Chinese and English

Chinese	English
Targeted therapy	Target Therapy
Pathophysiology	Pathophysiology
Proteomics	Proteomics
Single arm trial	Single-arm Trial
Low Risk Population	Low-risk Population
Multiplicity	Multiplicity
Non-Enriched Population	Non-Population enriched
Non-inferiority trial	Non-inferiority Trial
Analysis Set	Analysis Set
Composite Outcome Measures	Composite Endpoint
Enrichment strategy	Enrichment Strategies
Enriched Population	Enriched Population
High Risk Population	High-risk Population
Complex enrichment strategy	Mixed Enrichment Strategies
Benefit-risk ratio	Benefit-risk Ratio
Genomics	Genomics

Chinese	English
Gene mutation	Gene Mutation
Hypothesis test	Hypothesis Test
Generalizability	Generalizability
Sensitivity	Sensitivity
Target population	Target Population
Bias	Bias
Confirmatory clinical trial	Confirmatory Clinical Trial
Screening test	Test Screening
Biomarker	Biomarkers
Biomarker positive	Biomarkers Positive
Biomarker negative	Biomarkers Negative
Adaptive enrichment strategy	Adaptive Enrichment Strategies
Subject Diagnostic Characteristics	Receiver Operating Characteristic (ROC)
Random withdrawal	Randomized Withdrawal
Specificity	Specificity
Surrogate Indicator	Surrogate Marker
Homogenization Enrichment Strategy	Reducing Heterogeneity Strategies

Chinese	English
External control	External Control
Analysis	Subgroup Analysis
Population	Subpopulation
Type I error	Type I Error
Compliance	Compliance
Superiority test	Superiority Trial
Predictive enrichment strategy	Predictive Enrichment Strategies
Prognostic enrichment strategy	Prognostic Enrichment Strategies
Endpoint Event	Endpoint
Tumor metabolites	Tumor Metabolite
Primary Efficacy Analysis	Primary Efficacy Analysis

Appendix 3: Study cases for enrichment design

Example 1: Prognostic Enrichment – Cardiovascular Study

In cardiovascular studies, outcome events may be more easily observed in high-risk subjects (such as those with AMI, stroke, cholesterol level, very severe CHF and undergoing angioplasty, etc.). The Scandinavian Simvastatin Survival Study (4S) is a trial of lipid-lowering drugs with the primary aim of assessing whether simvastatin can improve survival in patients with coronary heart disease by lowering serum cholesterol. The study was a randomized double-blind placebo-controlled multicenter clinical trial that enrolled 4444 patients with angina or previous myocardial infarction (MI), all of whom had high total cholesterol (TC) levels. During a mean follow-up of 5.4 years, cardiovascular mortality was significantly reduced with simvastatin as compared with placebo (relative risk RR 0.70, 95% CI: 0.58 – 0.85).

Example 2: Predictive enrichment – melanoma study

BRAF kinase inhibitors are a type of targeted drugs for the treatment of melanoma, and exon 15 (V600E) of the BRAF gene can be used as a predictive biomarker. The BRAF gene is known to encode a cytoplasmic serine/threonine kinase, an enzyme that regulates the mitogen-activated protein kinase signal transduction pathway that controls several important cellular functions including cell growth and division (proliferation). It has been found that BRAF

V600E is mutated in a variety of tumors, such as melanoma, colorectal cancer, papillary thyroid carcinoma, hairy cell leukemia and Langerhans cell hyperplasia. In a phase III clinical trial of melanoma that enrolled 675 subjects with metastatic or unresectable BRAFV600E mutation who were treated with the BRAF kinase inhibitor vemurafenib or the chemotherapeutic drug dacarbazine, the response rate was 48% in subjects treated with vemurafenib targeted agents and only 5% in subjects treated with dacarbazine chemotherapy; the relative risk of death was reduced by 63% in subjects treated with vemurafenib.

Example 3: Predictive enrichment – MSI study

Microsatellite instability (MSI) is a biomarker that responds to immune checkpoint inhibitors. PD-1/PD-L1 pathway is a signaling pathway that regulates T cell activation and plays an important role in tumorigenesis and progression. In practice, the expression level of PD-L1 protein is usually detected by immunohistochemical method, which is used as a predictive marker and selects subjects with high expression, but its response rate to PD-1/PD-L1 inhibitors is only 10-20%. However, a 50% response rate was achieved in subjects with tumors with high-grade microsatellite instability (MSI-HIGH). Based on this, the FDA approved pembrolizumab for the treatment of subjects with MSI-HIGH-type or mismatch-repair deficient colorectal and endometrial cancers.

Example 4: Randomized withdrawal design - study of pregabalin for fibromyalgia

A clinical trial investigating the efficacy of pregabalin in the treatment of subjects with fibromyalgia used a two-stage randomized withdrawal design to compare the difference in time to loss of therapeutic response (TLTR) between pregabalin and placebo. The first phase was an open-label trial in which subjects with fibromyalgia were all treated with pregabalin and observed for 6 weeks. At 1 – 3 weeks, subjects received escalating doses of pregabalin to decide their optimal dose; at 4 – 6 weeks, subjects were maintained at this optimal dose. After completion of the open-label treatment in Stage I, subjects had to have at least a 50% pain reduction and at least a "marked improvement" in their self-evaluation on the PGIC scale in order to enter the double-blind, placebo-controlled trial in Stage II. Of the 1051 subjects, 566 entered the second phase after the first phase of treatment, of whom 287 were randomly assigned to placebo and 279 to pregabalin. After 26 weeks of treatment in the second stage, a significant difference in the time to loss of therapeutic response (LTR) was observed between the two groups ($p < 0.0001$). At the end of the trial, 61% (178) of placebo and 32% (90) of pregabalin achieved LTR.