Blue Paper for ICH E17 Implementation

Peking University

R&D-Based Pharmaceutical Association Committee under the China Association of Enterprises with Foreign Investment

December 2023

Table of Contents

Chapter I
Background and Objectives

Section I ICH E17

With the increasing globalisation of drug development, it has become important that data from multi-regional clinical trials (MRCTs) can be accepted by regulatory authorities across regions and countries as the primary source of evidence, to support marketing approval of drugs (medicinal products). But in the traditional drug development model, clinical trials are usually conducted independently in different regulatory regions to comply with their respective regulatory requirements. This practice not only increases redundant trials and additional costs, but also causes unnecessary delays in drug launch in some countries.

To cope with this challenge, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) issued the E5 guideline titled *Ethnic Factors in the Acceptability of Foreign Clinical Data* in 1998, the purpose of which is to facilitate the registration of medicines among ICH regions by recommending a framework for evaluating the impact of ethnic factors upon a medicine's effect while minimizing duplication of clinical studies and supplying medicines expeditiously to patients for their benefit. In this document, ethnic factors are defined as those factors relating to the genetic and physiologic (intrinsic) and the cultural and environmental (extrinsic) characteristics of a population. Despite the clear intent of the E5 guideline, many clinical trials are still conducted independently in different countries, continuously causing high costs and delayed drug launch.

To address this challenge, in 2017 ICH subsequently released the E17 guideline titled *General Principles for Planning and Design of Multi-Regional Clinical Trials*. The purpose of this guideline is to describe general principles for the planning and design of MRCTs with the aim of increasing the acceptability of MRCTs in global regulatory submissions. The guideline should be used together with other ICH guidelines to achieve the shift from regional thinking to global thinking, from retrospective strategy to prospective strategy, and from bridging trials to simultaneous global development.

ICH E17 includes 7 basic principles for designing MRCTs: ① Encouraging the use of MRCTs in drug development, with potential differences across regions be carefully considered; ② Early identification of the intrinsic and extrinsic factors that may affect drug development, and collection of relevant factors in the exploratory and confirmatory clinical trials; ③ Estimation of the overall treatment effect and approaches to sample size allocation to regions; ④ Pre-specified pooling of regions or subpopulations, based on established knowledge about similarities; ⑤ A structured exploration to examine the consistency across regions and subpopulations; ⑥ High-quality design and conduct of MRCTs in accordance with the ICH E6 guideline; ⑦ Encouraging sufficient and efficient communication among sponsors and regulatory authorities. Further understanding of the basic principles of ICH E17 can be based on the training materials on ICH E17 provided and some investigators' findings, and in combination with ICH E5 [1].

With the significant improvement of China's new drug development and clinical study levels, more and more clinical trial institutions in China have started to participate in MRCTs. This development trend not only effectively shortens the delay time for drug launch in China and overseas, improves Chinese patients' access to new drugs and helps China's medical institutions and investigators gain access to the latest knowledge and technology in global drug development, but is also of great significance to more in-depth exploration and analysis of the efficacy and safety of new drugs in the Chinese population.

Despite China's increasing participation in MRCTs and many benefits brought thereby, there are still a series of challenges when implementing the ICH E17 guideline worldwide. First and foremost, different requirements of regulatory authorities for clinical trials in different countries add to the complexity in the design and conduct of global clinical trials. For example, there may be different criteria and requirements for efficacy assessment and safety monitoring of drugs across countries, leading to differences in data collection and analysis. Secondly, quality control on a global scale is also a major challenge, as differences in healthcare infrastructure and resources in different countries and regions may affect the quality of clinical trials and the reliability of results. Furthermore, ICH member countries have gradually begun to implement the *ICH E9 (R1): Estimands and Sensitivity Analysis in Clinical Trials*, making the conduct of MRCTs even more challenging. In addition, although the ICH E17 guideline provides the basic principles and framework, there are still many pending issues in the specific operation and implementation process, such as trial design, simultaneous trial initiation, consistency evaluation of efficacy and safety, pooling strategy, and sample size allocation to regions of the investigational drug worldwide. To cope with these issues and challenges, there is an urgent need for appropriate industry consensus to guide the details of corresponding design, execution, and interpretation of results.

Section II Overall Challenges in Implementing ICH E17

Currently, MRCT-based simultaneous global development has become one of the fastest and most effective methods in new drug development. Participation in MRCTs can significantly accelerate patient access to innovative drugs in certain regions, and enhance the influence of investigators from those regions in the global pharmaceutical arena. However, the conduct of MRCTs is accompanied by a series of unique challenges. These challenges not only impact the design and statistical analysis of MRCTs, but also involve differences in clinical diagnosis and practice across regions, clinical trial operation management and investigator capacity, regulatory harmonization across regions, and ethical concerns, etc. Specifically, these challenges include, but are not limited to the following aspects.

1.2.1 Design and Statistical Aspects

The purpose of an MRCT is to assess treatment effects simultaneously in different regions, which involves considerations in many aspects, posing a great challenge to its design. Firstly, it is necessary to pre-define homogeneous regions and the pooling strategy based on characteristics of the disease, intrinsic or extrinsic factors affecting disease or treatment outcomes, as they may affect the evaluation of overall treatment effects and the selection of statistical methods during consistency evaluation, as well as the interpretation of results. Moreover, the overall sample size of a study and sample size allocation to regions are also key points to consider at the design stage of the trial, as the determination of the sample size is closely associated with the primary endpoint, the method of data analysis, and the regulatory requirements of individual regions. For example, for the time-to-event outcome variable, the sample size will depend on the opening time of regional study sites and the number of events in the target region; if a random-effects model incorporating variability among regions is used in the analysis model, a larger sample size is required than that for a fixed-effects model. In addition, prompt communication with regulatory authorities is required at the design stage of an MRCT. In the event that regulatory authorities in different regions have different approval requirements, a handling method that meets different approval requirements should also be pre-specified. Last but not least, many other factors need to be considered at the design stage of an MRCT. It is recommended to consider the objectives of the trial, primary/key endpoints, primary analysis population, method of the primary analysis, pre-defined (pooled) regions, intrinsic and extrinsic factors of regional differences, randomization issues, overall sample size, sample size allocation to regions, benefit-risk evaluation, consistency evaluation, non-inferiority/equivalence trial objectives and other factors.

1.2.2 Clinical Diagnosis and Treatment Aspects

As mentioned earlier, the purpose of MRCTs is to evaluate treatment effects simultaneously in different regions, so intrinsic and extrinsic factors that have a significant impact on drug development programme should be identified at the planning stage of MRCTs. Intrinsic or extrinsic factors that may affect treatment effects are mentioned in the ICH E5 guideline, and extrinsic factors include a number of considerations for clinical diagnosis and treatment: large regional differences in disease definitions, methods of diagnosis, disease severity, common concomitant medications, criteria for admission, understanding of certain endpoints, or regulatory policies and biomarker distributions may affect the implementation of MRCTs. In this regard, clear inclusion and exclusion criteria and study steps can be defined to minimize relevant differences. Furthermore, challenges in the execution of MRCTs also include patients' demographic characteristics (age, gender, ethnicity, etc.), baseline disease status, weight and BMI, treatment compliance, diet, smoking and alcohol, environment, cultural background, socioeconomic factors (e.g., contraceptive use, route of administration preference), educational level, available health services, placebo effect, and sensitivity of subjects' responses to different drugs by intrinsic factors that contribute to the high heterogeneity of the study population.

1.2.3 Clinical Trial Operation Aspects

During the execution of clinical trials, it is of utmost importance to ensure the quality of trial data obtained from MRCTs; only with high-quality data is it possible to determine the treatment effect of the investigational drug in the overall population with relatively larger heterogeneity and to analyze the applicability of that overall treatment effect to individual regions. In order to ensure the smooth conduct of an MRCT and to obtain higher-quality data, the following aspects should be considered at the planning stage of the MRCT: selection of sites and regions, localized translation of English SOPs and training, key points and precautions in patient recruitment, patient screening, dispensing of drugs to clinical trial sites and the pathway for sites to supply them to patients during the clinical trial, randomization scheme of the overall population of the MRCT across regions, training of investigators from sites conducting the MRCT, verification of the data obtained from the clinical trial, compliance with the protocol during trial conduct, retention of documentations during the trial, reasons for subject dropout and discontinuation of the study, reporting normality, relevance of benefits, etc.

1.2.4 Regulatory Aspects

The importance of communication with regulatory authorities is emphasized in the ICH E17 guideline. During the design, planning, conduct and analysis of a trial, there are points to be agreed upon with regulatory authorities. This is because the requirements for the trial may vary from one regulatory authority to another, such as the selection and definition of the primary endpoint, how to measure the primary endpoint, the selection of treatment for the trial's comparator, and whether all regional regulatory authorities recognize the use of this treatment. What's more, the main idea of MRCTs is to first investigate whether drugs have treatment effects in the overall population in multiple regions, and then to evaluate the treatment effects in individual regions when trial results show that there are treatment effects in overall subjects in all regions. However, the primary concern of each regulatory authority is still the treatment effect in the local population. Therefore, regulatory recognition of data from other regions in MRCTs and the requirements for data from the region may vary from region to region, posing a major challenge in conducting MRCTs.

1.2.5 Ethical Issues

Both traditional clinical trials and MRCTs are interventional studies conducted in human populations, and thus both require attention to issues related to ethics. As specified in the *Declaration of Helsinki*, "In every biomedical research project involving human subjects, concern for the interests of the subject must always prevail over the interests of science and society". Therefore, the primary concern in MRCTs is the benefit to the patients in each region; if

regional benefit-risk differences exist, stringent vigilance should be exercised during the design, execution and analysis of MRCTs to avoid causing relatively larger risks to patients in certain regions.

Besides, when a medical study involves subjects who are capable of giving consent, each potential subject must be adequately informed of any information related to the study such as study objectives, methods, possible conflicts of interest, anticipated benefits and potential risks, as well as the right to refuse to participate in the study or to withdraw the informed consent at any time without being unduly influenced thereby. That is, prior to participation in the trial, subjects will sign the informed consent form to be informed of relevant information on the clinical trial and their rights. However, for MRCTs, there are some challenges in the development of the informed consent form, such as whether different language versions of the informed consent form used in multiple regions can convey the unified content in light of different cultural settings and language environments across different regions and whether understanding differences can be avoided when developing the informed consent form.

Meanwhile, this often brings about new difficulties in ethics review conducted by the ethics committee: choosing whether to establish a unified central ethics committee or regional ethics committees. The central ethics committee can better ensure the unified criteria for ethics review in entire MRCTs across regions, while it is common in practice for multiple regional ethics committees to conduct ethics review of MRCTs. However, the major difficulty arising from establishing regional ethics committees lies in the possibility to guarantee identical review criteria of regional ethics committees and how to develop unified criteria so that regional ethics committees can conduct homogeneous review of ethics-related issues to reduce heterogeneity across regions.

Additionally, the challenges in the conduct of ethics review of MRCTs from an ethical point of view include how to ensure the transparency of the study in each region, and how to ensure the quality of the study in each region at a comparable and higher level, the quality of reviewing the study in each region, the process of data collection, the security of the data, and the compliance with Good Clinical Practice (GCP), etc.

Section III Main Challenges and Strategies

While possible challenges in implementing ICH E17 are discussed in the previous section at a macro level, practical challenges in implementing E17 will be listed and appropriate strategies to address these challenges will be discussed in this section at the execution level. During the execution of MRCTs, centered on consistency evaluation, considerations should be given to the identification of ethnic factors, sample size allocation, selection of comparators, implementation of other ICH guidelines and the pooling strategy in MRCTs, communication with regulatory authorities, and other major issues. An in-depth analysis of these key issues is made in this section to probe into their impact and possible strategies in MRCTs.

Firstly, based on the discussion on consistency evaluation, we will focus on the importance of ethnic factors in clinical trials and new ideas that may be utilized; next, attention will be paid to rational allocation of the sample size and its impact on consistency evaluation across regions; in this section, the complexity of consistency evaluation and the difficulties encountered in the selection of comparators will also be analyzed, especially in the context of possible differences in standard of care across different regions and the challenges in implementing the *ICH E9 (R1): Estimands and Sensitivity Analysis in Clinical Trials* in MRCTs; subsequently, we will discuss the implementation of the pooling strategy. Last but not least, the importance of communicating with regulatory authorities will be discussed with a view to providing comprehensive insights and solutions to support the effective and reliable design and conduct of MRCTs. The main challenges and strategies in practice are presented below.

1.3.1 Consistency Evaluation

The core of evaluating the consistency in clinical results is the overall benefit-risk evaluation based on the totality of evidence. When trial results show that the investigational drug has treatment effects in all study subjects in all regions, it is necessary to further evaluate the consistency in treatment effects across regions. In other words, there are no clinically relevant differences in treatment effects among all regions included.

This consistency evaluation is not based solely on the numerical similarity of overall and regional results, but should also further confirm consistency conclusions and post-marketing medical expectations and evaluate overall benefits and risks as well as regional benefits and risks through an analysis of the totality of evidence chain consisting of biological interpretability, statistical assessment, efficacy and safety results, anticipated risks, results of early trials, and other internal or external data.

As regulatory authorities in various countries vigorously advance the new drug development process, more and more complex and innovative trial designs are being applied in clinical settings. For example, adaptive design, which allows trial adjustment to be made based on the results of the interim analysis as pre-specified in the protocol, accelerates drug launch while increasing the complexity of trials. In this case, the planned simultaneous conduct of MRCTs will further add to the complexity of trials. Therefore, there is an urgent need for relevant consensus to clarify how to plan interim analysis and allocate the sample size when combining adaptive design with MRCTs, in an effort to evaluate the consistency across regions. In addition, due to a large number of unmet clinical needs in the area of rare disease drug development, the regulatory authorities in various countries have introduced incentives to encourage rare disease drug development. MRCTs provide an effective method of rapid recruitment of subjects with rare diseases. However, compared to MRCTs in general, the small number of patients with rare diseases, insufficient historical and comparable information, the absence of effective treatment in current single-arm trials and other characteristics make it more challenging to use pooling strategies and allocate the sample size. How to conduct consistency evaluation under such circumstances is also a major issue that urgently needs to be solved.

1.3.2 Identification of Ethnic Differences

An MRCT is a clinical trial conducted in multiple regions and involves the recruitment of subjects from multiple regions, with ethnicity being a key topic. Since it is assumed in MRCTs that treatment effects apply to the entire target population, that the sample size in each region is insufficient for descriptive evaluation to judge treatment effects, and that descriptive consistency evaluation should be used to assess treatment effects, factors that may influence treatment effects between regions (e.g., ethnic factors) should be identified early. Ethnic factors are a key issue in the planning and design of MRCTs, and the ICH E17 guideline also recommends careful consideration of the impact of regional differences, intrinsic and extrinsic factors (described as ethnic factors in the ICH E5 guideline) at the planning stage [2] to determine the role that MRCTs can play in drug development strategies. In the future, with the gradual increase in the number of trial sites in the Asia Pacific region to be included in MRCTs, more and more patients from China, Japan, and South Korea will participate in the trials. As such, it will be even more important to consider intrinsic and extrinsic ethnic factors during the design of MRCTs. Currently, more and more investigators have been proposing the use of real-world data (RWD) [3], [4] to assess ethnic differences between regions. Although some clinical trials exploring ethnic differences may be exempted, the statistical methods and conduct details of the specific application are still a major challenge and need to be further investigated.

1.3.3 Selection of Comparators

In principle, the comparators set in MRCTs should be identical in all participating regions. However, possible differences in standard of care in different regions, the approved dosage form, dosage and administration, and manufacturing quality of the drug may also vary, giving rise to the complexity in selecting comparators and

the comparability between the two groups at the design stage of MRCTs. In this case, strategies of stratified randomization may be considered, such as stratified randomization by control drug in case of different chemical drugs. Plus, a number of investigators have also proposed that RWD [3], [4] can be used as an external control for clinical trials, thereby reducing the sample size of the trials. However, RWD is characterized by diverse sources, highly heterogeneous quality, etc. The use of RWD as external data in MRCTs will further increase the complexity of the entire trial design, analysis, and interpretation of the results. Hence, further exploration is needed to determine whether RWD can be used.

1.3.4 Sample Size Allocation

The overall sample size of MRCTs is determined based on the effect size and variability of the primary treatment effect in the overall population across multiple regions. However, the overall treatment effect in populations across multiple regions varies greatly, which may entail a greater sample size. At the same time, the estimation of the treatment effect in the overall population may be more challenging due to the differences in the disease characteristics, drug mechanism of action, as well as the impact of intrinsic or extrinsic factors across regions. In that case, relevant contributing factors and parameter data can be obtained from early exploratory trials of the investigational drug conducted in relevant populations to support sample size determination. As mentioned earlier, some regulatory authorities still require the use of local clinical trial data for registration purposes. Therefore, an approach based on local significance or preservation of effect may be used for sample size allocation to regions, or local sample sizes may also be allocated at a certain proportion. If such allocation approaches are required for multiple regions in MRCTs, it will increase the overall sample size of MRCTs and defeats the purpose of planning and designing MRCTs. The aforementioned pooling strategies, if implemented well, may reduce certain requirements of regulatory authorities for local sample sizes, facilitate the flexibility in sample size allocation in MRCTs and contribute to consistency evaluation.

1.3.5 Estimands and Sensitivity Analysis

Currently, several ICH member countries have begun to implement the *ICH E9 (R1): Estimands and Sensitivity Analysis in Clinical Trials*. However, to date, practical experience with the guideline has been limited, and there is no standard operating document applicable to ICH E9 (R1). As a result, it will be more challenging to consider the implementation of the ICH E9 (R1) guideline during the conduct of MRCTs.

The ICH E9 (R1) guideline mainly covers two parts: the construction of estimands and the imputation of missing data. At the design stage of MRCTs, estimands can be considered in terms of the five attributes one by one to maintain consistency across regions in these five attributes as much as possible. However, due to factors such as ethnic differences between regions, differences may exist in clinical issues between the overall population and the regional populations. When setting MRCTs' estimands, it is necessary to consider them one by one. For more details, see 6.2.3 Multi-Regional Clinical Trials in E9 (R1) Blue Paper V2 [9]. The imputation method of missing data should, in principle, be as consistent as possible across regions, but the details of actual use still need to be further explored, such as whether the same imputation method can still be used when the pattern of missing data is not the same in one region as in other regions, and whether the level of evidence obtained can be questioned if different imputation methods are used; and whether it is appropriate to impute the data of individual regions with the global data.

1.3.6 Implementation of Pooling Strategies

Pooling strategies facilitate sample size allocation and consistency evaluation [1]. In 2020, the European Federation of Pharmaceutical Industry Associations (EFPIA) conducted a survey among its member companies to assess the challenges in implementing the E17 guideline in some ICH member and observer countries. As per the

EFPIA survey, there has been a positive trend towards reduced data requests for local clinical data by a number of regulatory authorities since the finalization of the ICH E17 guideline, which tallies with the original purpose of the ICH E17 guideline [5], [6].

The premise of pooling strategies is a rational analysis of intrinsic and extrinsic factors that may affect the conduct of trials. However, pursuant to local laws and regulations in many regions, regulatory authorities still require local clinical trial data for registration purposes, resulting in insufficient practice of pooling strategies. Ethnic factors may cover multiple intrinsic and extrinsic factors. As such, it is important to participate in early studies or early MRCTs as early as possible to accumulate relevant scientific knowledge and data to determine the impact of ethnic factors on treatment effects. Meanwhile, it is also necessary to try to draw on the experience from available classic case studies to assist in developing pooling strategies. Upon the adequate assessment of the impact of regional differences on the efficacy and safety of drugs, without the influence of other clear intrinsic and extrinsic factors, the pooling of the Asian population in MRCTs can be considered. In particular, patients from China, Japan, and South Korea can be pooled as the East Asian population.

1.3.7 Communication with Regulatory Authorities

As per the aforementioned EFPIA survey, the proportion of EFPIA member companies communicating with regulatory authorities was less than 1/3 in countries such as Brazil, South Korea, India, Russia, and Saudi Arabia [5], [6]. Asano et al. made an analysis of the implementation of ICH E17 major points in 167 MRCTs approved in Japan, and found that insufficient consideration was given to scientific consultation meetings in Japan [7], [8]. Inadequate communication with regulatory authorities at the planning stage of MRCTs may significantly reduce the possibility of conducting MRCTs and the efficiency of late-stage simultaneous submission of study data of the new drug to regulatory authorities in multiple regions. Moreover, because regulatory authorities in multiple regions may impose very different or even conflicting requirements, failure to understand the requirements and communicate with regulatory authorities in advance is not conducive to forming a unified regulatory view and recognition of development programme in advance. In addition to the increased frequency of communication with regulatory authorities, it is also necessary to ensure the quality of communication [1]. It is recommended to actively promote the convening of scientific consultation meetings between sponsors and regulatory authorities to discuss the different aspects affecting the development of the trial protocol. For example, during the actual execution of MRCTs, some countries may not be able to initiate trials simultaneously as others, and have shortened enrollment window due to regulatory processes, clinical site setup, drug supply, etc. In addition, the competitive enrollment mechanism across regions may make it difficult to enroll the planned number of subjects in some regions. To address this issue, some sponsors propose the extended enrollment strategy (EES), i.e., to extend the enrollment in specific countries under the same MRCT protocol. However, the EES is a strategy under special circumstances rather than a preferred approach. In practice, sponsors should maintain active communication with regulatory authorities and prudently assess the applicability of the strategy based on specific circumstances. In the meantime, scientific discussions between regulatory authorities in different regions are also encouraged to harmonize study requirements at the global level.

Chapter II
Considerations at the Study Design Stage

Section I General Considerations

Although MRCTs generally become the preferred option for investigating a new drug with regulatory submission planned in multiple regions, the potential impact of regional differences on trial design and consistency evaluation must be carefully considered at the design stage of MRCTs. This means that at the design stage of the drug development plan and MRCTs, epidemiological and disease situations should be investigated in detail and relevant information should be summarized, so that intrinsic and extrinsic factors that have an important impact on drug development can be identified in advance. Potential impacts of these factors should be investigated at the exploratory stage while relevant information on these factors should continue to be collected at the confirmatory stage. This is essential for subsequent consistency evaluation. Based on known information on similarities between regions of MRCTs, the pooled regions or subpopulations could be pre-specified and this decision should be reflected in the trial design plan and the statistical analysis plan. This approach not only helps provide flexibility in sample size allocation, but also helps facilitate the assessment of consistency in treatment effects across regions, and supports regulatory decision-making. When MRCTs are used as a drug development strategy, the underlying assumption for MRCTs is that the treatment effect applies to the overall target population. Therefore, at the design stage, the assumption that the treatment effect applies to the overall target population should be assessed based on early recognition of the product and the target population, and efforts should be made to ensure that these considerations are reflected in the overall design of MRCTs. In doing so, sponsors are encouraged to communicate effectively with various regulatory authorities at the planning stage of MRCTs, with the goal of obtaining acceptance of a global approach to study design across the different regulatory regions. This communication helps to ensure that trial design complies with regulatory requirements in each region, thereby accelerating the global development, review and approval of new drugs. The following are more detailed considerations for design.

2.1.1 Epidemiological Characteristics of the Disease

The epidemiological characteristics of the disease should be an important factor to be considered during drug development, which is of guiding significance to the overall development strategy of MRCTs. Epidemiology investigates the incidence/prevalence, etiology and risk factors of the disease in specific populations, histological conditions (e.g., squamous cell carcinoma, adenocarcinoma), gene mutations (e.g., mutant or wild-type), and even prognosis situation. Because the epidemiological information can provide valuable reference for critical decision-making on inclusion and exclusion criteria, stratification factors, and sample size allocation, it is critical to the design of MRCTs. For example, knowledge of the incidence of the disease in different regions can help sponsors decide the regions for recruitment and the anticipated time and resources required for recruitment; knowledge of the etiology and risk factors can help identify patient populations that may respond better to treatment or be more likely to have particular adverse reactions.

To better understand and compare the global situation of epidemiology, text, figures and/or tables could be used to summarize and compare the disease situation across regions. This includes the overall incidence or

prevalence, incidence by age, gender and ethnicity, common etiologies, and distribution of various contributing factors. Through this comprehensive data collection and analysis, investigators can more accurately develop the design scheme of MRCTs to ensure that trial results are both scientifically reliable and globally applicable. This epidemiology-based approach not only improves the efficiency of MRCTs, but also helps to ensure that study results are of significance and value to various participating regions.

2.1.2 Status Quo of Disease Diagnosis and Treatment

An in-depth understanding of the status quo of diagnosis and treatment of a specific disease is critical to the design of MRCTs. This involves not only a general understanding of the disease, but also a comprehensive understanding of diagnostic methods, diagnostic criteria, standard of care processes (e.g., first- and second-line therapies), clinical practice, and treatment effects, as these elements directly affect key information such as the development of inclusion and exclusion criteria for the study, the selection of stratification factors, the setting of randomization methods, the selection of comparators, and the determination of endpoints.

Precise diagnosis in clinical studies is the basis for successful clinical trials. Diagnosis of different diseases may rely on a range of tests, including laboratory tests, imaging examination and possibly gene test. For example, in cancer clinical trials, precise diagnosis of tumor typing and staging is critical to the successful achievement of study results and their interpretation. Similarly, for cardiovascular diseases, it is just as important to know patients' cardiac function assessment methods such as electrocardiogram and echocardiogram. Especially when MRCTs are conducted in different countries, various international and regional diagnostic criteria may affect patient enrollment and have a greater impact on study results.

For any specific disease, knowledge of the current standard of care and its clinical use is key to developing an effective trial design. For example, when investigating a new oncology drug, knowledge of available chemotherapy, targeted therapy, or immunotherapy is essential for setting up appropriate comparators and assessing the treatment effects of the new therapy. Similarly, for chronic diseases like diabetes, knowledge of drug therapies and lifestyle interventions available globally is critical to designing a realistic and operable clinical trial.

2.1.3 Clinical Pharmacology Studies and Doses

Clinical pharmacology studies play a key role in MRCT design, and the role of such studies is not only reflected at the early exploratory stage of drug development, but also runs through the late confirmatory stage of the entire study. At the early stage, these studies not only include the assessment of pharmacokinetics (PK) and pharmacodynamics (PD) of the drug, but also involve studies of the dose-exposure-response relationship (including PD, safety, and efficacy), as well as key contents such as dose selection and confirmation. It is encouraged to collect genetic data (e.g., genotypes of drug metabolizing enzymes) from subjects enrolled, if necessary, to examine the effect of genetic factors on PK and PD, where subjects with specific genotypes may be considered a subpopulation.

It is recommended that clinical pharmacology and other early studies be conducted in China as early as possible to obtain evaluable data. In early studies, difference factors between regional populations and the overall population should be explored as early as possible. PK and PD data obtained from different populations in early exploratory clinical trials can be used to preliminarily analyze differences across ethnicities and regions, which could provide a basis for the design, conduct, and result analysis and interpretation of confirmatory MRCT studies. If necessary, considerations should be given to enhancing MRCT clinical risk management and control measures, or dose adjustment, or even conducting clinical pharmacology or other clinical trials independently in the regional (or Asian) population to ensure safety and adequate treatment benefit in regional subjects.

In a bid to ensure the risk of regional subjects under control, preclinical PK, PD and toxicology, as well as disease mechanism, clinical operation and treatment methods can be assessed. In simultaneous global drug development,

early participation of the region in early MRCT studies may be considered to facilitate efficient data collection and improve the assessment of differences across ethnicities and regions. In addition to data of the regional target population, existing early clinical data of overseas populations, especially those of oversea Chinese and Asian data, are also valid evidence for the assessment of differences across ethnicities and regions. Early clinical trials can obtain the dose-exposure-response relationship in the indicated population within a wide range of doses, investigate potential intrinsic and extrinsic factors as early as possible, provide sufficient support for conducting confirmatory MRCTs while protecting the safety of subjects, and finally support drug dosage and administration for registration and guide the rational post-marketing use of drugs in the overall regional target patient population and special populations.

In addition to the early studies, relevant investigations of clinical pharmacology will be continued in the later studies, such as population PK, population PD, PK in special populations. In principle, the same dosing regimens should generally be used for all participating ethnic populations in a confirmatory MRCT. If no important regional differences in PK and/or PD and dose-exposure-response relationships are expected (e.g., drug response may be insensitive to intrinsic and extrinsic factors), it is not necessary to obtain PK, PD, or dose-exposure-response data for subjects from all regions planned to be enrolled in the confirmatory MRCT, and the same dosing regimen could be used directly. If significant differences in dose-exposure-response are identified in the regional population, adjustments to the dosing regimen may be considered in the context of the totality of intrinsic and extrinsic factors and dose-response relationships. It is recommended to be careful in decision of using different dosing regimens.

2.1.4 Effect Modifiers

Effect modifiers are broadly defined as the intrinsic or extrinsic factors that may affect the treatment effects of drugs. Effect modifiers include, but are not limited to, age, gender, body weight, disease severity, ethnicity, region, medical practice, and environmental factors. In the early stage of clinical development, the true effect modifiers are often difficult to be identified because they are generally hidden among factors such as ethnicity and region. A geographic region or regulatory region is a composite or alternative factor for many ethnicity-related factors. Ethnicity-related factors are more specific than geographic region, while there are still many confounding factors involved. Various intrinsic and extrinsic factors are listed in ICH E5 guideline in detail, as shown in Figure 1. Intrinsic ethnic factors are factors that help to define and identify a sub-population and may influence the ability to extrapolate clinical data between regions, such as age, gender, race, and genetic factors; extrinsic ethnic factors are factors associated with the environment and culture in which a person resides, such as diet, medical practice, and air pollution.

Intrinsic		Extrinsic
Genetic	Physiological and pathological conditions	Environmental
Gender	Age (children - elderly)	Climate
		Sunlight
Height		Pollution
Body weight		
	Liver	Culture
	Kidney	Socioeconomic factors
	Cardiovascular functions	Educational status

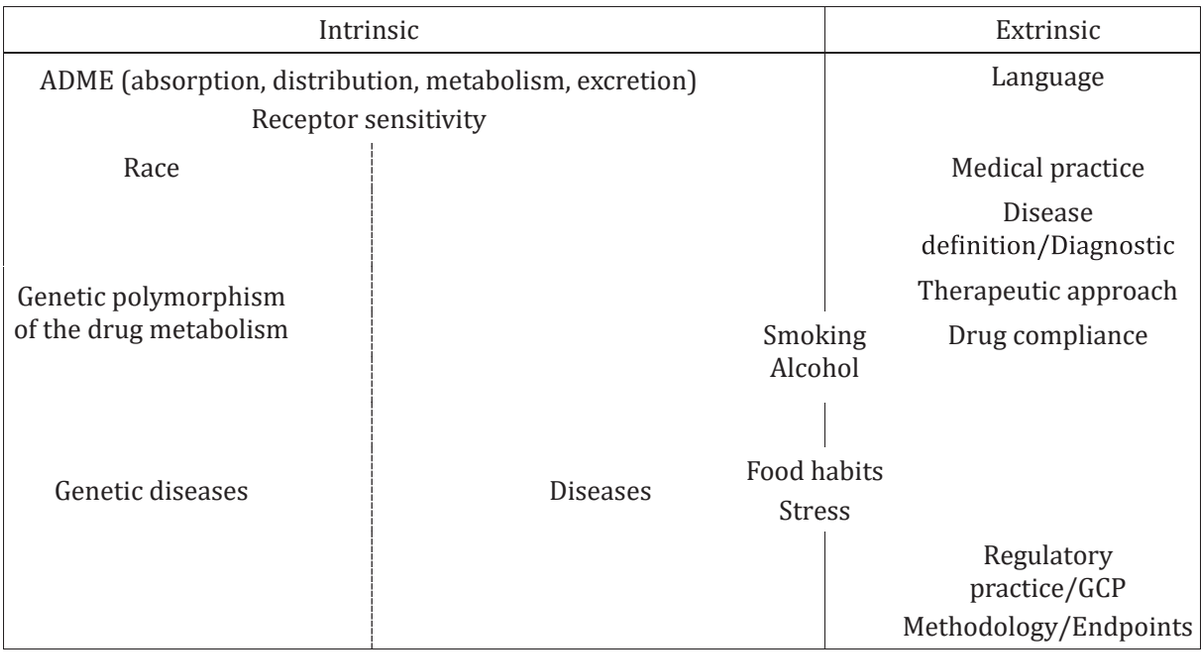

Figure 1 Classification of Intrinsic and Extrinsic Ethnic Factors

(1) Intrinsic factors

1) Gender

The levels of certain hormones in male and female organisms are greatly influenced by gender, while gender can also influence psychological and cultural behaviors due to ethnic, social, or religious backgrounds. In ICH E5, gender factors mainly involve the differences between women and men and their associated biological mechanisms, as well as the commonalities in human physiology, pathophysiology, and clinical outcomes between men and women. Since the guideline was issued, gender medicine has become an important horizontal dimension in medical practice and clinical trial design. Differences in symptoms, clinical evolution and treatment effects between men and women with the same disease should be considered in drug development. Therefore, understanding gender differences and including a representative gender population in MRCTs will help to better interpret the results.

2) Ethnicity

As pointed out in Appendix C of the ICH E5 guideline - Pharmacokinetic, Pharmacodynamic, and Dose Response Considerations, evaluation of the pharmacokinetics and pharmacodynamics, and their comparability, in the three major racial groups most relevant to the ICH regions (Asian, Black, and Caucasian) is critical to the registration of medicines in the ICH regions. However, how to define ethnicity is a major challenge. Collins et al. state that ethnicity alone is not a measure of clinically significant treatment effects and should be used with caution in drug development [10]. Because ethnicity is a self-reported social structure with a degree of subjectivity, it is even more important to identify predictive biomarkers with prognostic significance or with relevance to the etiology and treatment of the disease at the early stage of drug development. These prognostic or predictive biomarkers may or may not be related to ethnicity. Therefore, grouping populations based on self-reported race or ethnicity does not accurately predict patient genotypes or drug response, and other methods to define populations rather than ethnicity should be actively investigated to more effectively evaluate and facilitate drug development [11].

3) Genetic factors

Genetic factors mainly refer to all factors that can be passed on through genes so that offspring acquire parental characteristics, excluding epigenetic factors that are influenced by environment. To understand PK genetic polymorphisms, several studies and analyses have been conducted over the past decade to assess PK differences and similarities within East Asia (Japan, China, and South Korea) and between regions (e.g., East Asians vs. Caucasians) [8]. In these studies, the East Asian population has similar genetic background and can thus be treated as a population to be compared with the Caucasian population. These studies suggest directionally that such differences may be limited to drugs metabolized by CYP450 enzymes, and genetic polymorphisms in CYP450 have been widely investigated in different populations and regions [12]. In addition, when the ICH E5 guideline was issued, new drugs were mainly small molecules or new chemical entities that were more sensitive to within-population variation related to drug absorption, distribution, metabolism and excretion, which could lead to differences in drug metabolism and configuration across different populations (or regions) [13]. Currently, new therapies are mostly biological products, which are generally less affected thereby.

(2) Extrinsic factors

As the genetics and biology of diseases are investigated, the pharmaceutical industry and academia have considered the intrinsic factors proposed in ICH E5, but there is still a lag in considering external factors. By including different patient populations in MRCTs, extrinsic factors related to socioeconomic factors, educational status, and access to preventive health care can be addressed. However, identifying extrinsic factors is often quite subjective, so it is difficult to incorporate them into the study design, adding to the difficulty in conducting MRCTs. At the same time, there are often difficulties in identifying and collecting extrinsic factors. For example, it has been difficult to collect food effects or standardize health care.

(3) Identification of effect modifiers

The identification of effect modifiers is generally divided into two steps: 1) Data collection. In addition to the intrinsic and extrinsic factors listed in ICH E5, other factors such as biomarkers should also be included. Both the prognostic factors of disease and the predictive factors of treatment effects should be considered. Medical and scientific literature, guidelines and other publicly available information may be searched to collect disease information and genetic information. Various relevant medical databases may be retrieved, such as the WHO disease database and various registry studies, to collect epidemiological data and relevant historical data. Local healthcare professionals may be consulted to inform local clinical practices and special treatments. 2) Check and evaluation. Based on the information on drugs and the studies collected, check and evaluate the effects of intrinsic and extrinsic factors on the drugs. The evaluation is performed by conducting clinical trials or via models and extrapolation, such as early clinical trials, exposure-response analysis, PK/PD models and investigator-led trials.

The optimal condition is that the effect modifiers are identified prior to the initiation of the pivotal MRCT. Otherwise, some factors with a high potential to have an impact can be pre-defined and then verified step-by-step based on the data obtained during each phase of clinical development. In some cases, after completion of the pivotal clinical trial, the data results reveal potential confounding factors which however can only be further identified as effect modifiers through more analyses or even additional clinical trials

.

Section II Special Considerations

2.2.1 Pooling Strategy

The importance of pooling strategies in MRCTs is emphasized in the 4th basic principle of the ICH E17 guideline. As defined in the guideline, pre-specified pooling of regions or subpopulations, based on established knowledge about similarities among study subjects in certain intrinsic or extrinsic factors, may help provide flexibility in sample size allocation to regions, facilitate the assessment of consistency in treatment effects across regions, and support regulatory decision-making. In short, the pooling of regions or subpopulations based on similarity can ensure consistency in the assessment of efficacy and safety across regions and provide greater flexibility in sample size allocation to regions, thereby facilitating the successful conduct of MRCTs.

E17 covers two pooling strategies: pooled regions and pooled subpopulations. Pooled regions: Pooling some geographical regions, countries or regulatory regions at the planning stage, if subjects in those regions are thought to be similar enough with respect to intrinsic and/or extrinsic factors relevant to the disease and/or drug under study. Pooled subpopulations: Pooling a subset of the subjects from a particular region with similarly defined subsets from other regions whose members share one or more intrinsic or extrinsic factors important for the drug development programme at the planning stage. Pooled subpopulations are assumed as ethnicity-related subgroup particular important in the MRCT setting.

The pooling strategy should be justified based on the distribution of the intrinsic and extrinsic factors known to affect the treatment response, and the disease under investigation and similarity of those factors across regions. Region is often a surrogate for intrinsic and extrinsic factors that tend to differentiate regions or populations from each other. In MRCTs, pooling regions is also a common pooling strategy. For example, it is generally considered reasonable to pool Canada and the US into North America, and China, South Korea, and Japan into East Asia due to similar medical practices and similar concomitant medications. In some areas, where available information is not yet complete, common pooling strategies such as pooling East Asian, Asian and North American populations may be considered and information on potential key factors can be collected on an ongoing basis during the global development process to aid in subsequent consistency evaluation. Besides regions, subpopulations may be identified based on key intrinsic or extrinsic factors that influence treatment effects, and then these newly defined subpopulations may be included in the stratification, analysis, and consistency evaluation. Ethnicity can often cross regional boundaries and can be an important risk factor associated with disease or treatment effect. If applicable, the pooling strategy for stratification, sample size allocation, and consistency evaluation should be pre-defined in the protocol, statistical analysis plan, or other documentation.

The distribution of intrinsic or extrinsic factors known to affect the treatment effects of drug, disease characteristics and the similarity of these factors across regions should be used to find effect modifiers and justify the rationale for pooling strategy. The key to determining the pooling strategy is to find the true effect modifiers. If this cannot be clearly identified, a flexible pooling strategy can be specified based on scientific arguments and operability. A two-step approach is first adopted to explore and identify effect modifiers from relevant intrinsic and extrinsic factors. If true effect modifiers can be identified, specific subpopulations can be defined based on different levels (categories) of the modifiers. In other words, stratification could be performed based on the different levels of effect modifiers (e.g., Figure 2), with subpopulations pooled by the level of factors with impact on efficacy. For example, if it has been identified that the low-weight population benefits significantly from an investigational drug, then three subpopulations can be generated by body weight (low, medium, and high), with the low-weight population as subpopulation 1. Among the populations in various regions, the Region A population has lower weight. Subpopulation 1 provides a good representation of the overall Region A population and supports the assessment of treatment effects of the drug in the Region A population in the context of a larger sample size. As another example, if the populations for a certain biomarker have similar responses to a drug

and a regional population accounts for a higher proportion in a certain category of the biomarker, then similar populations outside the region can be pooled with the regional population to help support the estimation of treatment effects in the regional patient population.

	Stratum 1 (low weight)	Stratum 2 (medium weight)	Stratum 3 (high weight)	Total
Region A				
Region B				
Region C				
Region D				
Total	Pooled subpopulation 1	Pooled subpopulation 2	Pooled subpopulation 3	

Figure 2 Stratification and Subpopulation Pooling by Effect Modifier

It is extremely challenging to identify the true effect modifiers. If this cannot be clearly identified, a region pooling strategy for drug assessment may be used so that both the impact of potential confounding factors and operability can be taken into account. A region is defined as a geographical region, country or regulatory region. Often, commonalities are presented in a region in terms of both potential impact on efficacy and distribution of factor data for specific effect modifier(s). Thus, the pooling of either the entire Asian region, to which China belongs, or the East Asian region, could be considered. Priority should be given to the East Asian region due to the huge differences in various aspects across the West, Central, South and East Asian regions in the vast Asian region.

When the East Asian region pooling strategy is planned to be used, data should be obtained from epidemiological and early clinical studies whenever possible, especially ethnic sensitivity data, such as PK and PD, genetic data, biomarkers, etc., to explore whether there are differences across populations from countries and regions in East Asia. If no strong evidence of significant differences across populations is shown, the pooling of East Asian populations is feasible while attention is paid to the Chinese population. If there is evidence of efficacy differences across East Asian populations such as the Chinese, Japanese, and South Korean populations, the pooling strategy may not be appropriate, with the main focus on the Chinese population instead.

2.2.2 Sample Size Allocation to Regions

The sample size allocation should consider the planned analysis of variability of treatment effect among pooled subpopulations. There is no uniformly acceptable or optimal approach to sample size allocation in MRCTs. As shown in Figure 3, ICH E17 introduce 5 approaches to sample size allocation to regions. Each approach has its own pros and cons. In general, it is recommended that a balance be maintained between proportional allocation (#1) and equal allocation (#2) to ensure that recruitment is feasible, i.e., the trial can be completed in a timely manner to achieve a true MRCT, and that sufficient information can be provided to evaluate consistency across regions. The principles described above for allocation to regions may also apply to pooled subpopulations. Regarding sample size assessment, it is necessary to ensure that there is no dominance of the sample size of a single region or some regions, giving rise to the dominance of the trial results.

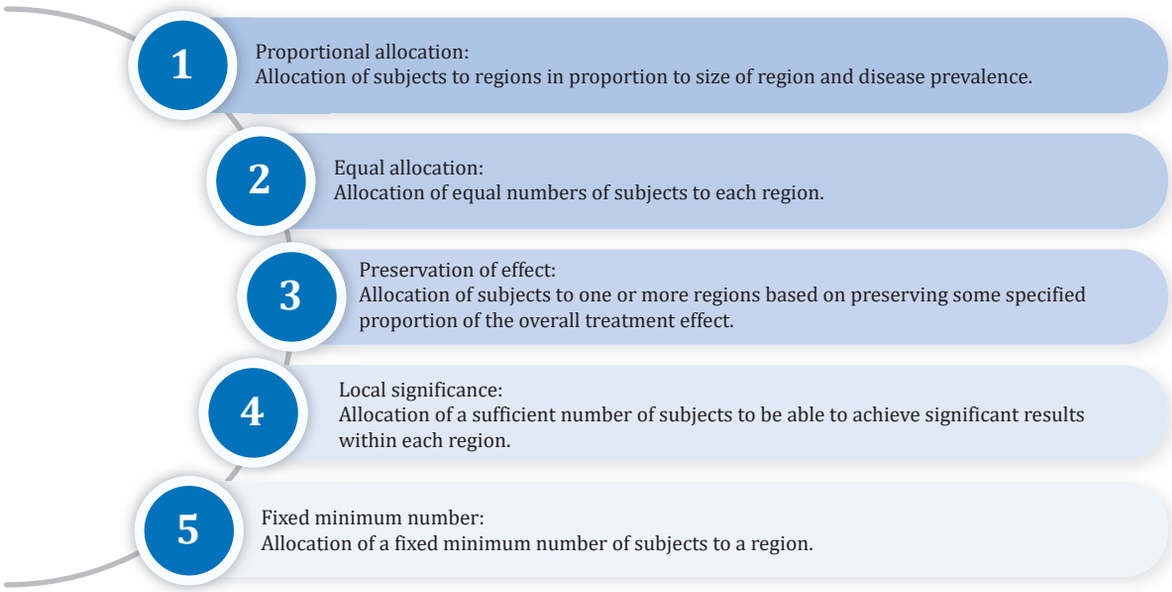

Figure 3 Five Approaches to Sample Size Allocation to Regions

One of the challenges in sample size allocation is how to select and meet criteria for consistency evaluation, which may require several rounds of discussions between different functions and between authorities and sponsors. According to the ICH E17 guideline, each consistency evaluation approach has its own advantage and limitation. Therefore, a holistic approach is recommended for consistency evaluation. Compared to sample size allocation using a fixed consistency evaluation criterion, it is recommended to select several candidate criteria or endpoints for consistency evaluation based on protocol design. Summaries of operational characteristics such as the probabilities of meeting these candidate consistency evaluation criteria over a range of sample sizes could be provided in text or tables. An approach with a balance maintained between proportional allocation (#1) and equal allocation (#2) can then be used to determine the sample size for countries and/or pooled regions based on the totality of evidence. Pre-specified pooling of regions or subpopulations, based on established knowledge about similarities, may help provide flexibility in sample size allocation to regions, facilitate the assessment of consistency in treatment effects across regions, and support regulatory decision-making.

Chapter III
Considerations at the Study Execution Stage

In an MRCT, high-quality MRCT requires high-quality execution. It is important to keep in mind execution and operation during the whole period of the MRCT. Operation should comply with GCP to ensure the quality and integrity of the clinical trial, and should be tracked and adjusted as needed. Every effort should be made to start the MRCT simultaneously across regions and reduce the gap between global enrollment and local enrollment. Enrollment speed and population characteristics should be monitored continuously during the enrollment period. At the same time, special attention should be paid to the consistency of MRCT operations across different regions. It is recommended to use consistent standard systems and processes whenever possible, including disease diagnosis, patient inclusion/exclusion criteria, randomization method, drug supply, data collection, data cleaning, medical monitoring, quality control, statistical analysis, and investigator training. Every effort should be made to ensure high quality and good consistency of the trial, to reduce variations and uninterpretable results caused by unnecessary reasons, facilitate more efficient drug development, and increase the possibility of submitting marketing authorization applications to multiple regulatory authorities in different regions simultaneously and the success of obtaining marketing approval.

When the pooling strategy is included in a clinical trial design, it is recommended to communicate with regulatory authorities on key issues as early as possible, including but not limited to the basis of pooling strategy, the statistical analysis method, and the determination of sample size. If a pooling strategy is used in a single MRCT, it should be clearly specified in the Clinical Study Protocol and detailed in the Statistical Analysis Plan. If a pooling strategy is used for multiple MRCTs, it should be documented in a separate file (e.g., Integrated Statistical Analysis Plan). During the whole period of clinical development, the pooling strategy may be adjusted based on the results from early clinical trials and the actual status of clinical studies. Adjustments can be made either at the design stage of a pivotal clinical trial or when the trial is ongoing. Sponsors should discuss with regulatory authorities in advance to clarify and document the final strategy, at least before the efficacy study database lock. If sponsors and regulatory authorities cannot agree on the specific implementation of a pooling strategy, other approaches, such as EES, may be explored to ensure that regional populations can participate in the pivotal global clinical trial.

Chapter IV
Study Result Interpretation

Section I Considerations for the Interpretation of Consistency Results

ICH E17 recommends a holistic approach to evaluating regional consistency from multiple perspectives with the totality of evidence rather than relying on results from one criterion on one endpoint. Evaluation should take into account biological plausibility, internal consistency, external consistency, the strength of evidence, clinical relevance, and statistical uncertainty for systematic analysis and interpretation.

4.1.1 Result Interpretation

A three-layer method is recommended for result interpretation. Layer 1: Evaluate overall efficacy and safety results. Layer 2: Evaluate the results across the pre-specified pooled regions or pooled subpopulations in the trial by utilizing a pooling strategy (if applicable). Note that the pooling strategy does not just include the pooling of regions/countries. The pooling of subpopulations may also be considered if applicable. Other analyses based on non-pre-specified pooled regions or pooled subpopulations may serve as supporting evidence. Layer 3: Evaluate the benefit/risk in specific countries or regions. At layer 2, utilizing a pooling strategy (if applicable) can facilitate consistency evaluation of treatment effects across regions or subpopulations and support regulatory decision-making. The pooling strategy should be justified based on the distribution of the intrinsic and extrinsic factors known to affect the treatment response, and the disease under investigation and similarity of those factors across regions.

The pooling strategy is one of the key principles of ICH E17. This strategy is important not only in the planning stage but also in the analysis and exploratory stage. It is generally recommended to pre-specify the pooling strategy. The pooling strategy should be justified based on the distribution of the intrinsic and extrinsic factors known to affect the treatment response, the disease under investigation, and the similarity of those factors across regions. During the clinical development of drugs, it is recommended to consider the pooling strategy in the planning stage, continuously collect and analyze relevant factors, and assess and adjust the pooling strategy as needed. Data should be obtained from epidemiological and early clinical studies whenever possible, especially ethnicity-related data, such as PK and PD, genetic data, and biomarkers, to identify an appropriate pooling strategy. If there is reasonable evidence supporting no significant differences between different populations, the pooling of corresponding regions and populations may be considered. It is generally not recommended to directly conduct the third layer of country-specific/regional-specific subgroup analyses and assess the corresponding benefit/risk profile without the evaluation of consistency across pooled regions and/or pooled subpopulations.

4.1.2 Clinical Pharmacology

Key considerations to assess the impact of drug exposure and ethnic factors usually include: ① PK profile: Exposure (local effects or not, no or minimal systematic exposure); dose-exposure relationship (linear or not);

metabolic pathways (single or multiple); genetic polymorphisms of relevant drug-metabolizing enzymes or transporters (CYP2C9, CYP2D6, OATP1B1); drug forms (e.g., antibodies, RNA). ② PD profile: Expression or genetic polymorphisms of target (e.g., VKORC, EGFR, HLA), dose-ranging exposure-response relationship profile (flat or steep); pathological mechanism subtype distribution of disease (HCV); clinical practice (accessibility of combination, concomitant medications). ③ Safety: Therapeutic window (wide or narrow); risk management measures or monitoring methods; concomitant medication (Chinese herbal medicine). ④ Others: Rare diseases, pediatric use. At the same time, at different stages of clinical development, make good use of model informed drug research and development, adopt modeling, simulation, and other methods and technologies to analyze the clinical pharmacology study data obtained locally and globally, distinguish the between-subpopulation variability and the between-subject variability to support the consistency evaluation and interpretation of drug dose, usage, efficacy and safety in MRCTs.

4.1.3 Efficacy/Pharmacodynamics

4.1.3.1 Internal Consistency

Internal consistency refers to whether consistent drug effects can be observed across multiple settings (e.g., endpoints, analyses, subgroups) of an MRCT. It can generally be assessed by evaluating the consistency of the primary and sensitivity analyses, the consistency of the primary and key secondary endpoints, and the consistency across subgroups. In an MRCT, a pooling strategy is usually adopted to evaluate the consistency of treatment effects across regions and/or across subpopulations. When evaluating consistency across regions, subpopulations, and/or subgroups, the effect of sample size on variability needs to be considered. Results should be interpreted with caution. The use of figures and/or tables to present results, if applicable, can be considered. For example, combine the results of important endpoints into one table; present the results into one or more tables/figures for important regions, subpopulations, and/or subgroups, as shown in Figure 4.

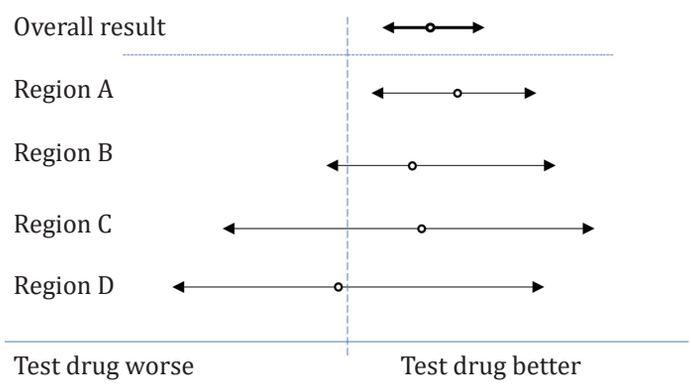

Figure 4 Graphical Examples of Regional Consistency (Source: ICH E17 Training Module 6)

4.1.3.2 External Consistency

External consistency refers to whether a consistent drug effect or a consistent impact on treatment effect by a particular factor is observed in multiple data sources. In some cases, multiple trials may be required to support drug submissions and regulatory decisions. When evaluating external consistency, it is common to consider consistency across similar phase III registration trials and/or phase II exploratory trials in the development program, as well as trials from outside the development program but with similar trial conditions. It may be possible to investigate whether the drug effect (between-group difference, treatment groups, or control groups) is consistent across trials. It may also be possible to investigate whether there is a consistent impact of specific factors, pooled regions, and pooled subpopulations on the treatment effect. In addition to presenting the results

of each trial separately, integrated efficacy analysis can also be used if needed. Integrated efficacy analysis refers to a systematic analysis of all clinical efficacy study data of the drug for the same indication to be registered, with the purpose of comparing the strengths and weaknesses of different study data to describe the overall efficacy profile. A pooling strategy may also be considered when performing the integrated efficacy analyses. The statistical analysis plan for the integrated efficacy analyses should be submitted to the regulatory authorities together with the integrated efficacy analysis report. Sufficient communication with regulatory authorities on the statistical analysis plan for the integrated efficacy analyses is encouraged.

4.1.4 Safety

When comparing safety in the region/subpopulation to that in the overall population, for regional/subpopulation analyses, it is not necessary to repeat all analyses on the overall population. Primary analyses may be determined based on the product safety profile, target population, and disease characteristics. Data analyses may include: overall summary of adverse events (AEs), severe AEs (grade 3 and above), serious AEs (SAEs), AEs leading to death, AEs leading to treatment discontinuation, AEs leading to drug interruption and/or dose reduction, significant AEs selected based on drug characteristics (e.g., AEs of special interests), and laboratory tests. In addition to investigating the incidence, severity, relationship to study drug, and category of the AEs, the time to onset, duration, outcome of certain AEs, or exposure-adjusted AE analyses may be further analyzed as necessary.

The analysis strategy includes assessing whether the safety profile of the investigational treatment group is consistent between the regional population and the overall population, whether the difference between the investigational treatment group and the control group is consistent between the regional population and the overall population, and whether the reported AE categories are similar. Results are generally analyzed using descriptive statistics and presented as tables and/or graphs. Unless pre-specified as study endpoints, post-hoc hypothesis testing comparing safety results between regions is not recommended. To assess whether the safety of the regional population is consistent with that of the overall population, the key consideration is whether there is a clinically meaningful difference. The numerical difference in the incidence of AEs between the subgroup and the overall population does not necessarily indicate a true difference. The reason is that the results may be affected by certain factors, such as the sample size of the overall population and that of the region. When the regional sample size is small, the numerical variation will be relatively large.

4.1.5 Benefits and Risks

Based on the previous evaluation, a comprehensive benefit-risk assessment could be conducted for the overall population. Subsequently, the consistency between regions and the overall population could be evaluated. Contents include: ① Treatment background: e.g. incidence/prevalence, etiology, risk factors, standard of care, prognosis; ② Clinical pharmacology: e.g., PK, PD, exposure-response relationship; ③ Benefits: e.g. beneficial endpoints, magnitude of benefits, duration of benefits; ④ Risks: such as the nature, incidence, and severity of AEs; ⑤ Benefit-risk assessment.

Section II Inconsistency exploration framework

Even with careful design and appropriate execution, unexpected regional differences may still be observed. If potential clinically relevant differences in treatment effects across regions are observed, a structured exploration should be utilized to investigate sources of differences. A holistic approach should be used to assess the magnitude and acceptability of differences. It is also important to note that data from a particular region/country should be evaluated using a descriptive framework, rather than hypothesis testing. In the exploration, we should systematically investigate the internal and external factors according to the principles of ICH E17 as shown in Figure 5.

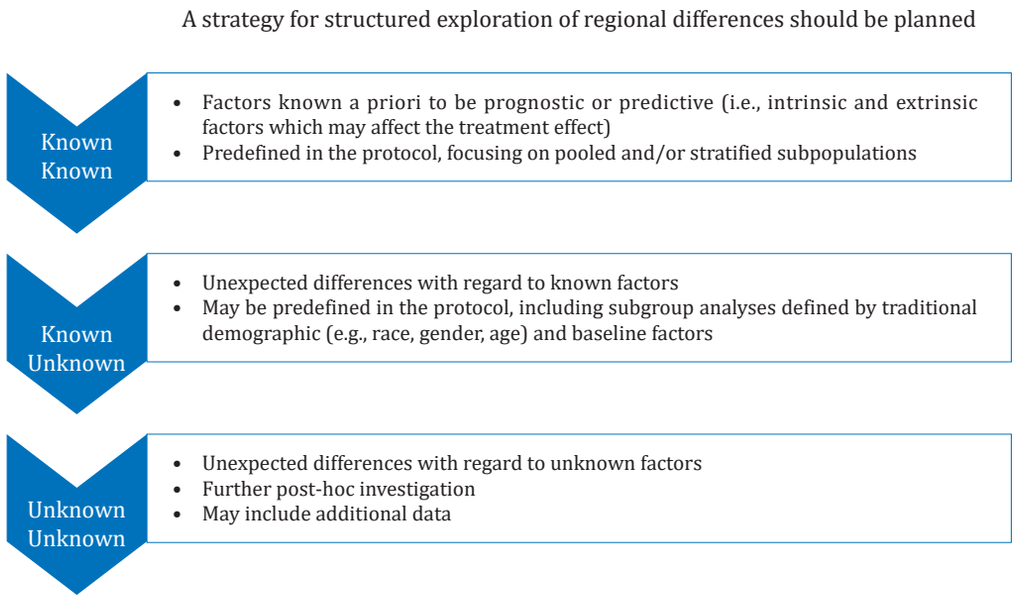

Figure 5 Structured Exploration of Regional Differences (Source: ICH E17 Training Module 6)

4.2.1 Clinical Relevance

In an MRCT context, the differences across regions or countries may not be clinically relevant. Points of consideration include the target population, burden of disease, study endpoints, and medical practice. Clinical relevance also depends on the profile of the treatment effect, including the metrics of treatment effect (e.g., hazard ratio or risk difference). If any difference between regions is observed, the first step should be to evaluate whether the regional differences are clinically relevant in size and significance. For example, when assessing clinical relevance, consider the regional context, and evaluate whether the benefits of the drug outweigh the risks in regions. This should be considered during the trial planning, and corresponding post-hoc in-depth analyses should be conducted.

4.2.2 Disease and Treatment

In addition to the summary of epidemics, diagnostics, and treatments in the design and planning stage, the corresponding contents should be re-examined using the most recent data at the interpretation stage. In some cases, important changes such as updates of trial results, drug approval, or guideline updates may occur in some regions/countries during the trial and result in impacts on trial execution (withdrawal from trial/treatment), subsequent treatment, and treatment effect.

4.2.3 Clinical Pharmacology

If clinically meaningful regional differences in drug exposure and dose-exposure-response (PD, safety, efficacy) are identified in clinical studies, *in vitro* and/or *in vivo* mechanistic studies should be conducted to explore the impact of PK and PD-related factors (as exemplified in Section 4.1.2). Population PK/PD, simulation and other methods may also be used to analyze the possible impact of each covariate.

4.2.4 Biological Plausibility

Biological plausibility here refers to the extent to which a particular effect (difference in treatment effect across regions or subgroups in this case) that might be predicted or expected, based on clinical, pharmacological, and mechanistic considerations associated with intrinsic and extrinsic factors. Plausibility is mainly clinical and pharmacological judgment, and is generally not directly quantifiable or measurable unless it has already been taken into account at the planning stage. In general, large regional differences are not expected for agents with local effects and targeted therapy targeting certain genetic mutations. Treatment effects of some drugs may be associated with baseline weight/BMI, baseline risk (high-risk, low-risk), histology (squamous cell carcinoma, adenocarcinoma), and biomarker expression levels (low-expression, high-expression), rather than directly with ethnicity or region. This biological plausibility can be used to investigate efficacy and safety, as well as differences between groups, or investigational treatment and control groups.

4.2.5 Enrollment and Sample Size

In the consistency evaluation exploration, detailed enrollment should be investigated, including overall and regional enrollment start time, enrollment end time/status, regional sample size allocation, proportion of target regions/countries/areas of interest, and sample sizes of each investigational treatment group. Enrollment has certain impacts on exposure time and follow-up time. The proportion of sample size and the balance of sample size between groups also affect statistical uncertainty.

4.2.6 Balance of Baseline Characteristics

Evaluate whether the regional population is consistent with the overall population and whether it is balanced between the investigational treatment group and the control group. In addition to pre-defined predictive and prognostic factors, special attentions should be paid to variables where baseline differences and treatment differences were observed. The multivariate model may be considered to adjust for important variables and assess whether differences are caused by baseline differences. When the regional sample size is small and there are many variables, the multivariate model may not be applicable. In that case, population resampling method may also be considered to assess whether the treatment effect is consistent across populations with similar characteristics, regardless of races.

4.2.7 Exposure, Follow-up, and Disposition

Exposure, follow-up, and disposition including reasons of trial discontinuation and reasons of treatment discontinuation should be summarized in detail. For the exposure of combination therapies, it may be considered to not only summarize the exposure of the combined treatment but also summarize the exposure of each component in the combination therapy. The similarity of exposure and study follow-up in the regional and overall population should be evaluated. If there are differences, it is necessary to check whether there are impacts on the treatment effect and whether exposure or follow-up adjusted analyses should be considered for evaluation.

4.2.8 Internal Consistency

Internal consistency refers to whether consistent treatment effects can be observed across multiple scenarios (e.g., endpoints, analyses, subgroups) of an MRCT. The strength to support internal consistency is reinforced if consistent result is shown with different analytical methods; different study endpoints (e.g., primary, secondary, and other supportive endpoints) correspond to consistent efficacy; the subgroup results are generally consistent and robust; trends are consistent over time. Some examples relevant to internal consistency evaluation are listed below:

(1) Results observed for specific regional endpoints differ from those observed for the overall population and this particular regional population differs from the overall population in one important factor. It turns out that in every subpopulation categorized by this factor, the regional population is consistent with the overall population (Figure 6 PLATO).

Region	ASA dose (mg)	Ticagrelor N	Ticagrelor E	Clopidogrel N	Clopidogrel E	HR(95% CI)
US						
	>=300	324	40	325	27	1.62(0.99, 2.64)
	>100 - <300	22	2	16	2	
	<=100	284	19	263	24	0.73(0.40, 1.33)
non-US						
	>=300	140	28	140	23	1.23(0.71, 2.14)
	>100 - <300	503	62	511	63	1.00(0.71,1.42)
	<=100	7449	546	7443	699	0.78(0.69, 0.87)

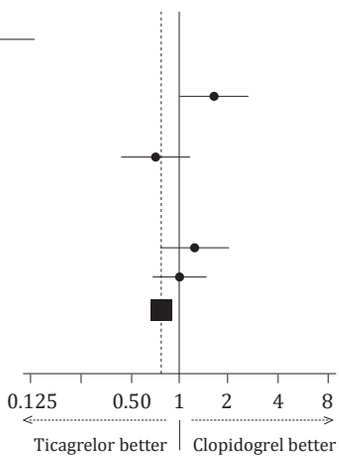

Figure 6 Estimates of Treatment Effect by Region and Aspirin Dose Level in PLATO Trial (source: ICH E17 Training Module 6)

(2) The results of the primary analysis of specific regional endpoints differ from those of the overall population using the same statistical model. Some complicated models are suitable for large sample size data. Using the same complicated statistical model on small sample size data may result in non-convergence or non-robustness of the model. Results became consistent after using an optimized model (AIC or BIC or other criterion selection). This observed difference may be caused by the instability of complicated models rather than true clinically relevant differences.

(3) The results of the analyses of the country-specific primary endpoints differ from those of the overall population but results in the pooled regions and pooled subpopulations defined by importance factors were generally consistent with the overall results and showed great robustness. In this case, the observed regional differences may be more due to data variability rather than true clinically relevant differences.

(4) Analyses of the region-specific primary endpoints differ from those of the overall population, but analyses of the important secondary endpoints showed generally consistent and clinically meaningful improvements from those of the overall results (12935 Pulmonary Hypertension Study). This observed difference is more likely due to instability of some measures (such as subjective measures) under small regional sample sizes rather than true clinical differences.

(5) Point estimates for specific regional endpoints differ from those for the overall population, but investigation of trends over time revealed temporal heterogeneity in the pattern of change (e.g., delayed treatment effects). Trends over time for regional populations are consistent with trends over time for the overall population (e.g., piecewise HR). This observed difference in point estimate summaries is more due to differences in follow-up time rather than true clinically relevant differences. Results tend to be consistent under similar follow-up time and tend to be consistent over time.

(6) Results for specific regional endpoints differ from those for the overall population at a particular time point, but trends over time show that results for the regional population are generally consistent with those for the overall population at both prior and subsequent time points. This observed difference is more likely due to the variability of the data rather than a true clinically relevant difference.

(7) Point estimates for specific regional endpoints differ from those for the overall population at the interim analysis. However, it is found that there is a trend that as the analysis time increases, the proportion of sample size in specific regions increases, the degree of variation decreases, and the degree of consistency increases. This observed difference may be due to variability caused by the limited sample size at the interim analysis.

4.2.9 External Consistency

External consistency refers to whether an effect of a particular factor on the treatment effect is observed across multiple data sources: In particular, whether inconsistency in a confirmatory clinical trial can also be found in external clinical trial data, such as another phase III trial, a phase II exploratory trial, or a trial from outside the development program but with similar trial conditions. External consistency can be assessed by examining consistency between similar studies, consistency with historical/external data, comprehensive analysis of efficacy, or meta-analysis, as appropriate. Before examining external consistency between different trial conditions, it is important to examine whether the trial conditions are similar in terms of treatment regimens, study population, endpoint measures and their summary metrics, as well as intercurrent events and multi-regional context. Points of consideration relevant to external consistency evaluation are listed below:

(1) Specific endpoint results of specific countries/regions observed in specific studies differ from those from the overall population, but no specific regional differences were observed in other similar studies. The integrated analysis of efficacy results in more robust results and the region-specific results are found to be consistent with the overall results.

(2) Specific endpoint results of specific countries/regions observed in specific studies differ from those from the overall population, but the treatment group effect is similar and the differences are mainly driven by the control group. However, no specific regional efficacy differences that could impact efficacy were observed in historical studies and in contemporary external data. The differences observed in this case may not be true clinically relevant differences.

4.2.10 Statistical Uncertainty

Statistical uncertainty arises from occasional or random fluctuations in a multi-regional trial. When multiple regions, countries, or subgroups are included in an MRCT, chance or random factors based on statistical uncertainty can result in numerical inconsistency of regional treatment estimates observed, particularly when regional sample sizes are small or there are too many regions. Pooling strategies or data-borrowing methods (Bayesian methods) could be considered to increase robustness. In addition, some graphical methods, such as Figure 7 funnel plot, can also be used to evaluate the expected variations (such as 95% CI) under different sample sizes to better understand the regional variability and thus perform further investigation.

4.2.11 Safety Analyses

When notable differences in safety results (incidence, severity, or category of AEs, etc.) are observed between the regional subgroup and the overall population, the reasons for such differences will need to be analyzed such as the relationship between PK exposure and safety, intrinsic and extrinsic factors (such as patient weight, baseline, regional medical practice and AE management, concomitant medication such as the use of traditional Chinese medicine). If needed, these analyses could also be combined with the AE-related auxiliary examination results, and/or previous historical data of this study drug in regional population. If there are differences in exposure time between the subgroups and the overall population, exposure-adjusted analyses can be used for safety consistency evaluation, particularly for certain events associated with exposure time, and it is more reasonable to combine exposure-adjusted analysis for result interpretation.

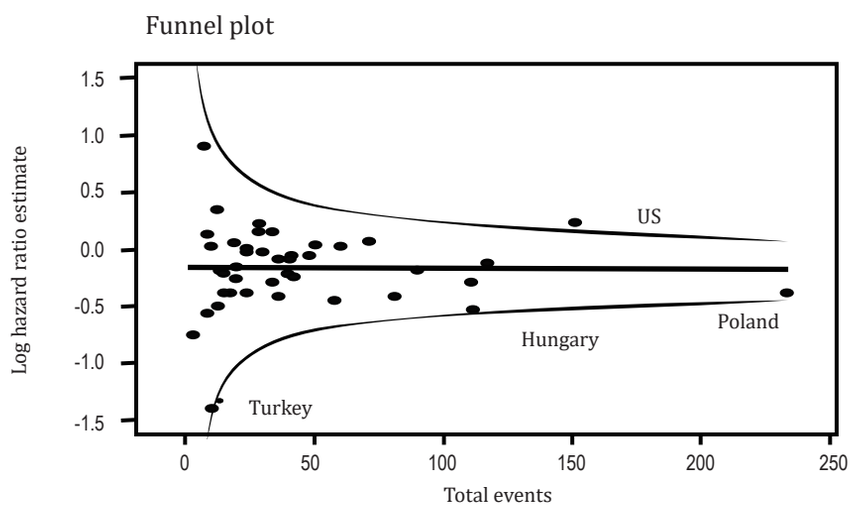

Figure 7 Funnel Plot of Treatment Effects by Sample Size (Source: ICH E17 Training Module 6)

Chapter V
Special Considerations

5.1 Type of Study Design

Superiority trials are generally the preferred design to demonstrate the efficacy of a drug. When a superiority trial is not feasible, a non-inferiority trial may be considered. Non-inferiority trials should use active control. The goal is to confirm the efficacy of the investigational drug, i.e., the new treatment is not inferior to a clinically unacceptable extent compared to the active control. As shown in Table 1, non-inferiority trials generally use a fixed margin approach and the non-inferiority margin M needs to be pre-defined in the protocol (use M > 0 for absolute metrics and M > 1 for relative metrics for illustration purpose). In non-inferiority studies, 'treatment effect' may be replaced with 'non-inferiority margin adjusted treatment effect'. After that, sample size assessment and consistency evaluation can be performed in a non-inferiority study using the same approach as that in a superiority study.

Table 1 Non-inferiority Margin-adjusted Treatment Effect

Absolute/relative indicator	Superior/inferior indicator	Non-inferiority margin-adjusted treatment effect
Absolute metrics	H0: T - C ≤-M vs H1: T - C >-M (M>0)	T-C+M
Absolute metrics	H0: T - C ≥M vs H1: T - C <M (M>0)	T-C-M
Relative metrics	H0: T / C ≤1/M vs H1: T / C >1/M (M>1)	T/C*M
Relative metrics	H0: T / C ≥M vs H1: T / C <M (M>1)	T/C/M

Absolute metrics include mean difference and rate difference, etc.; relative metrics include rate ratio, hazard ratio, odds ratio, etc.

Superior indicators are indicators whose higher values indicate better efficacy; inferior indicators are indicators whose lower values indicate better efficacy.

T represents the treatment group effect, C represents the active control group effect, and M represents the non-inferiority margin.

5.2 Multiple Primary Endpoints

An MRCT may have multiple primary endpoints. There are usually two scenarios. One is that the demonstration of treatment effects on all primary endpoints is recommended to establish clinical benefits (co-primary endpoints); the other is that the demonstration of treatment effects on at least one of several primary endpoints is sufficient (dual-primary endpoints etc.). Sample size evaluation should also be adjusted accordingly. For the co-primary endpoints, it is recommended to provide the probability of consistency for all primary endpoints. For the dual-primary endpoints, it is recommended to provide the probability of consistency for each primary endpoint. Consistency evaluation adopts a descriptive framework without hypothesis testing. Multiplicity adjustment for multiple endpoints is not required for consistency evaluation.

5.3 Interim Analyses

An MRCT may have one or more efficacy interim analyses. All efficacy interim analyses should be pre-defined. If one or more endpoints reach statistical significance at the time of interim analysis, the sponsor may choose to use interim analysis data for registration submission. Interim analyses and consistency assessments will be performed according to the protocol's pre-specified interim analysis plan. For consistency evaluation, the plan of hypothesis testing and the results of the global trial endpoints should be taken into consideration. In the interim analysis, it should be noted that the difference in enrollment time in different regions and the difference in sample size for interim analysis may have an impact on consistency evaluation. It is recommended to provide enrollment status and sample size to facilitate the assessment. Follow-up efficacy and safety follow-up data can be provided if needed and feasible. Blinding should be appropriately controlled if there are requirements of blinded subsequent long-term follow-up (primary endpoint not tested yet and/or regulatory requirement). There should be a dedicated team responsible for filing and a separate blinded team responsible for continuing operating the subsequent long-term follow-up of the trial to minimize the impact on the trial.

5.4 Delayed Treatment Effect

In some MRCTs, there may be delayed treatment effects due to disease or drug characteristics (e.g., HR = 1 in the first 6 months, followed by HR = 0.7). In this case, if there is delay in regional enrollment, such as 6 months later than global enrollment, it may cause challenges and biases in directly conducting the consistency evaluation. In this case, time-varying piecewise efficacy and consistency evaluation, longer follow-up data, trial simulation, or data imputation for supportive and sensitivity analysis purpose could be considered.

5.5 Adaptive Design

Currently, the application of adaptive trial design in MRCTs has become increasingly popular. Adaptive designs are generally defined as clinical trial designs that modify the trial according to a pre-specified plan and based on data accumulated during the trial at the time of interim analysis. Common adaptive design includes sample size re-estimation, group sequential design, removing treatment groups, and seamless design. The types of adaptations that can be performed include stopping the trial due to futility or efficacy; adjusting for sample size, patient population, treatment group, or patient allocation ratio. Compared with the traditional fixed design, the adaptive design can modify the trial process according to the pre-specified rules by using the cumulative data in the trial for interim analysis, so that the clinical trial is more flexible, efficient, and ethical.

Adaptive designs have become more and more common in MRCTs in recent years. Adaptive designs and MRCTs are complex in study design, planning, and execution. When used together in the implementation of an adaptive MRCT, special attention should be paid to ensure that the trial design and execution will serve its objectives. In practice, sites in some countries may be initiated several months later than in others due to differences in approval requirements of regulatory and institutional review committee. For an adaptive MRCT, regional sample size allocation should be considered not only for the entire trial but also for interim analysis. Because the plan for adaptive modification needs to be described in the trial protocol and statistical analysis plan before the start of the trial, if the MRCT is considered to support simultaneous global development, the consideration of the consistency of the final results should not only ensure that it is still applicable under the potential adaptive change of the protocol, but also ensure that it will not change based on the information causing the change of the protocol. Temporal heterogeneity also needs to be considered if it exists. For the adaptive MRCT, computer simulations may be used to determine the operational characteristics of the trial to aid trial design and sample size allocation.

5.6 Single Arm Study

Single arm trial (SAT) refers to a clinical trial designed to be open-label without parallel control group. Single-arm studies are commonly used in the early exploratory phase of new drugs. In some special cases, the development strategy of SATs may also be considered to support accelerated approval of drugs and significantly shorten the marketing time of new drugs, such as patients with advanced malignancies who had run out of treatment options, new drugs with very prominent treatment effects, or rare disease areas with very limited numbers of patients. Applicants are encouraged to communicate this strategy with regulatory authorities to discuss the rationale of this strategy. SATs typically use surrogate endpoints as primary endpoints, such as objective response rate, which is commonly used as primary endpoint in single-arm trials in oncology.

According to ICH E17 principles, a holistic approach should be applied to evaluate consistency using multiple candidate consistency evaluation criteria from different perspectives. There are generally small sample sizes for the overall population and limited enrollment windows in single-arm studies, and little time is left for enrollment in the region. Due to the inherent limitations in the design of SATs, it is difficult to control confounding factors, necessitating stricter adherence to the pre-defined protocol. It is also recommended to take into account the overall consistency evaluation of the clinical pharmacology evaluation in early studies and/or trials.

5.7 Rare Disease

Rare diseases refer to diseases with a very low incidence/prevalence. The challenges encountered in the drug development for rare diseases often surpass those encountered for more common diseases. This is due to the low incidence/prevalence of rare diseases, complexity of rare diseases, and limited understanding of rare diseases. Drug development for rare diseases is likely to have the following characteristics: limited data of epidemiology and the natural course of the diseases, insufficient medical information, possibly lack of established methods and endpoints to assess efficacy, limited clinical study opportunities for patients, limited drug development experiences in this area, multiple subtypes of the diseases, high patient heterogeneity, limited sample size, and narrow enrollment window.

Due to the limited sample size of rare diseases, MRCTs have historically been used as a quick way to recruit patients with rare diseases and enable global regulatory submissions. Regulatory authorities around the world may have different review criteria for drug development for "rare diseases". The actual practice may also vary in performing assessments for unmet needs for a particular disease and for consistency criteria to be met for a particular development program. Discussions with regulatory authorities in advance are necessary.

In general, for rare diseases, regional sample size allocation often requires greater flexibility given the limited number of patient population and strong unmet medical needs. Sample size may be more driven by practical feasibility considerations rather than statistical considerations. Global sample size allocation usually implements global competitive enrollment rather than fixed regional sample size allocation. Rare diseases typically have limited historical data and limited development experience. When historical data and clinical experience are insufficient, consider utilizing important factors and pooling strategies used in other relevant areas. It is generally difficult to develop drugs for rare diseases using traditional approaches. Global studies often employ a single MRCT study with an innovative complex design, which makes consistency evaluation more challenging. In this case, compared to traditional consistency evaluation, it may be more feasible to determine whether there is other evidence of plausible ethnic differences, and it is recommended to use a more flexible approach by taking both internal and external data for overall consistency evaluation.

References

[1] ICH, I.E., *General principles for planning and desingn of multi-regional clinical trials [EB/OL] (2017-11-16) [2023-09-26]*. https://database.ich.org/sites/default/files/E17EWG_Step4_2017_1116.pdf.

[2] ICH, I.E., *Ethnic factors in the acceptability of foreign clinical data [EB/OL] (1998-02-05) [2023-09-26*]. https://database.ich.org/sites/default/files/E5_R1__Guideline.pdf.

[3] Bunning, B.J., Hedlin, H., Chen, J.H., et al., *The evolving role of data & safety monitoring boards for real-world clinical trials*. J Clin Transl Sci, 2023. 7(1): p. e179.

[4] Wasser, J.S.,Greenblatt, D.J., *Applying real-world data from expanded-access ("compassionate use") patients to drug development*. J Clin Transl Sci, 2023. 7(1): p. e181.

[5] Associations, E.F.o.P.I., *ICH E17 Guideline on Multi-Regional Clinical Trials Catalyst for Simultaneous Global Development and Registration (2021 DIA Europe) [2023-09-26]*. https://globalforum.diaglobal.org/issue/june-2021/ich-e17-guideline-on-multi-regional-clinical-trials/#:~:text=A%20European%20Federation%20of%20Pharmaceutical%20Industry%20Associations%20%28EFPIA%29,and%20the%20existence%20of%20local%20legislation%20and%2For%20regulations.

[6] Associations, E.F.o.P.I., *Industry Survey on Implementation of ICH E17 Guideline General Principles on Planning and Designing Multi-Regional Clinical Trials (2021 DIA Europe) [2023-09-26]*. https://globalforum.diaglobal.org/issue/june-2021/industry-survey-on-implementation-of-ich-e17-guideline/.

[7] Asano, K., Aoi, Y., Kamada, S., et al., *Points to Consider for Implementation of the ICH E17 Guideline: Learning from Past Multiregional Clinical Trials in Japan*. Clin Pharmacol Ther, 2021. 109(6): p. 1555-1563.

[8] Asano, K., Uyama, Y.,Tohkin, M., *Factors Affecting Drug-Development Strategies in Asian Global Clinical Trials for Drug Approval in Japan*. Cts-Clinical and Translational Science, 2018. 11(2): p. 182-188.

[9] DIA 中国统计社区 , *ICH E9(R1)* 估计目标及敏感性分析蓝皮书 *(2021-11-01) [2023-09-26]*. http://www.diachampion.com/index.php?c=show&id=12.

[10] Collins, F.S., *What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era*. Nature Genetics, 2004. 36(11): p. S13-S15.

[11] Bonham, V.L., Callier, S.L.,Royal, C.D., *Will Precision Medicine Move Us beyond Race?* New England Journal of Medicine, 2016. 374(21): p. 2003-2005.

[12] Zanger, U.M.,Schwab, M., *Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation*. Pharmacology & Therapeutics, 2013. 138(1): p. 103-141.

[13] Wilson, J.F., Weale, M.E., Smith, A.C., et al., *Population genetic structure of variable drug response*. Nature Genetics, 2001. 29(3): p. 265-269.

[14] Pharmaceuticals and Medical Devices Agency, P.E.N.N., *Basic Principles on Global Clinical Trials [EB/OL] (2007-09-28) [2022-10-10]*. https://www.pmda.go.jp/files/000157727.pdf.

[15] Quan, H., Zhao, P.-L., Zhang, J., et al., *Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance*. Pharmaceutical Statistics, 2010. 9(2): p. 100-112.

[16] Hung, H.M.J., Wang, S.J.,O'Neill, R.T., *Consideration of regional difference in design and analysis of multi-regional trials*. Pharmaceutical Statistics, 2010. 9(3): p. 173-178.

[17] Lan, K.K.G.,Pinheiro, J., *Combined Estimation of Treatment Effects Under a Discrete Random Effects Model.* Statistics in Biosciences, 2012. 4(2): p. 235-244.

Appendix: Technical Details of Relevant Methods Covered in the Blue Paper

1. Statistical Analysis Model

A fixed-effect model (FEM), continuous random-effect model (CREM), and discrete random-effect model (DREM) can be considered for in-depth assessment of regional treatment effects and overall treatment effects as well as variation between regions [16], [17]. When using these models, attention should be paid to different model assumptions, FEM assumes that both the regional treatment effect and the regional sample size are fixed parameters, CREM assumes that the regional treatment effect is a random variable subject to a certain distribution while the regional sample size is a fixed parameter, and DREM assumes that the regional treatment effect is a fixed parameter while the regional sample size is a random variable subject to a polynomial distribution.

2. Relevant Technical Details of EES

(1) Basic idea of the EES

As described in 1.3.8, during the actual execution of MRCTs, some countries may not be able to initiate trials simultaneously as others, and have shortened enrollment window due to regulatory processes, clinical site setup, drug supply, etc. In addition, the competitive enrollment mechanism across regions may make it difficult to enroll the planned number of subjects in some regions. When other known approaches (e.g., pooling strategies) are not applicable, scientifically justifiable and operationally feasible remedies that all parties can accept can be actively explored. Some sponsors proposed to extend the enrollment in specific countries under the same MRCT protocol, i.e., EES.

If an EES is used due to the infeasibility of an MRCT without extended enrollment (e.g., failure to simultaneously meet the registration requirements of regulatory authorities in different countries), subjects will continue to be enrolled in the extended cohort in specific countries after the completion of enrollment in the main global cohort in a seamless way. The extended cohort will be identical to the main cohort with respect to enrollment criteria, study procedures, and important study endpoints. The data collected for the subjects in the extended cohort will follow the same data collection, data cleaning, and medical monitoring procedures conducted by the same study operation team following the same GCP quality control processes. The data collected for the extended cohort will have the same quality standard as that in the main cohort and be evaluated for the consistency of regional efficacy and safety under the same protocol. The analysis of the main cohort and country-specific analyses should be designed with caution to avoid biases.

The general framework of the EES can be described as follows: 1) For recruitment and enrollment in the global cohort, the trial period for the “main cohort” is minimized by starting recruitment for the MRCT as soon as the first participating country joins, ensuring fully competitive enrollment; 2) For extended enrollment in a country with insufficient sample size, the number of enrolled patients planned for a country can potentially be met in the "extended cohort" of the corresponding country. This can be illustrated with a hypothetical chart in the following figure.

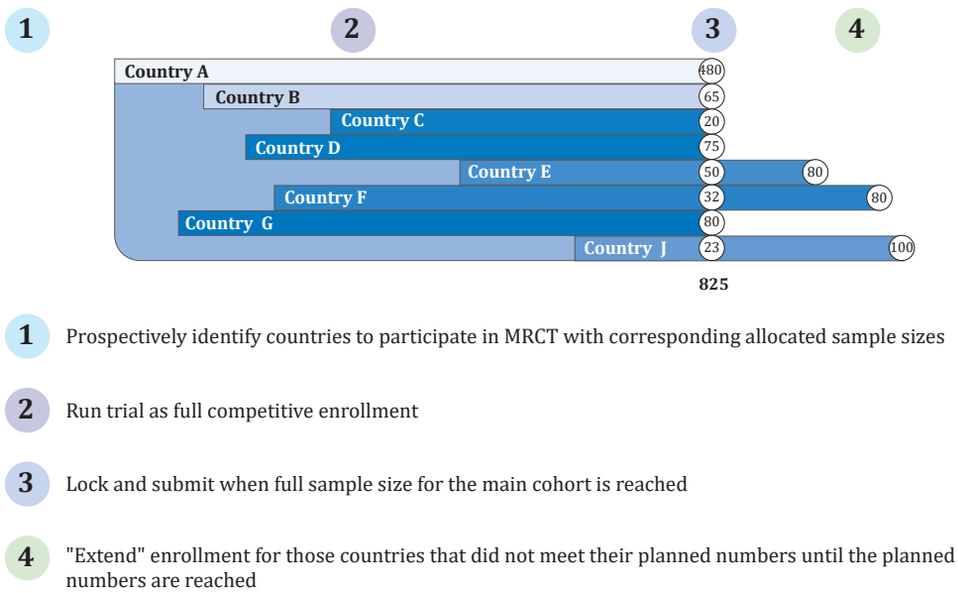

Figure 8 A Hypothetical Chart as an Illustration of EES

During the design stage, we determine the sample size of the main cohort for the MRCT based on the primary objective(s). In the example (Figure 8), the calculated size is 825. Regulatory authorities in Countries A, B, C, and D may not have specific requirements regarding allocated sample sizes in line with the competitive enrollment approach, while Countries E, F, G and J may have individual regulatory-driven requirements for sample size from either efficacy or safety perspectives. We recruit patients from all participating countries for the MRCT in a fully competitive enrollment manner. Recruitment for the "main cohort" stops when the sample size required for scientific evaluation of the primary study objective is reached according to the protocol. For countries with planned sample sizes, recruitment continues in an "extended cohort" for the corresponding country until the planned number is met (if not already fulfilled in the main cohort).

Compared to standalone bridging studies guided by ICH E5, MRCTs with EES are more aligned with the ICH E17 principle of conducting one study with one protocol and one operation team, allowing for consistent enrollment criteria, study procedures, data collection, data cleaning, medical monitoring, quality control processes, statistical analyses, trial planning, and investigator training. This can promote efficient operations in each region, and consistency evaluations between regions and the whole and across regions. Standalone studies may require larger sample sizes, longer enrollment times, longer study durations with more time heterogeneity and delays compared to MRCTs. Standalone studies may also operate under a different study framework, using separate protocols, teams, operation processes, and training, introducing more variability, reducing consistency, and posing challenges to consistency evaluation.

(2) Role of different cohorts in EES

There are a few analytical considerations after the EES has been established. Table 1 summarizes key considerations on how main cohort and extended cohort should be used for global submission purposes. Among them, main cohort (C1) will be used to support global registrations, while extended cohort (C2) will only be used for registration in a certain region. C1 is fully powered to meet the primary objective outlined in the protocol. The risk/benefit analysis derived from this cohort will be used to support global registrations; In contrast, C2 is designed to meet the sample size as required by the local regulatory authority and will include patients from that region only. C1 is considered the primary database, and overall efficacy and safety conclusions will be based

on data from this cohort. There will be no proactive trial-level pooling of C1 and C2 for global registration. The region-level analysis will be conducted as a subpopulation analysis. For the latter, regional patient data from both cohorts (C1+C2) will be pooled to evaluate the consistency with C1 in terms of efficacy, safety, and benefit/risk profile, and will be summarized in corresponding documents. C2 does not have a direct impact on conclusions from C1 because there will be no pooling of C1 and C2 in global submissions.

Table 2 Submission Purposes of Main and Extended Cohorts in an MRCT

Main cohort (C1) (for global regulatory registration)	Extended cohort (C2) (for regional regulatory registration)
• Pre-specified in protocol to address primary objective; • Representative of global study population; • Fully powered for the primary analysis and risk/benefit conclusions; • Safety/Efficacy analysis reported in CSR; • Includes regional subjects in the main cohort;	• Pre-specified in protocol (or a protocol addendum); • Not part of the global study population; • Efficacy and safety of patient data will be reported in a document describing regional data; • Begins after the global main cohort's last patient entering trial (LPET) (only in the region; regional patient data from C1+C2 will be used to evaluate consistency with global C1 data); • If C1 ends early for either benefit or futility (for example, based on an interim analysis), C2 plan will be reevaluated.

Note: CSR is clinical study report, hereinafter referred to as CSR.

(3) Protocol description

The MRCT with EES should be covered by a single protocol for both main and extended cohorts, and therefore trial procedures should be the same for both cohorts. The protocol should describe the EES, either in the main body of the protocol or through a protocol addendum. The protocol may use flexible languages regarding the EES as appropriate to ensure the optional nature of EES, as there may be adjustments in the speed of regional enrollment at the trial execution stage. In addition, having a single protocol could minimize the operational biases from having two separate studies, and could practically save time/resources needed for initiating a second (bridging) study.

(4) Impact of EES on global submission for MRCTs

In general, the use of EES for a region will not have any impact on submission review timelines from other regulatory authorities (e.g., FDA/EMA). The submission of the global data package, which is based on the main cohort, will follow the normal timelines, regardless of the use of EES in the region. The global data package can be submitted to, reviewed, and approved by other regulatory authorities (e.g., FDA/EMA) based on regulatory supplementary submissions while the extended cohort in the region is being completed. The following figure demonstrates the relationship between the main and extended cohorts in terms of their respective key milestones.

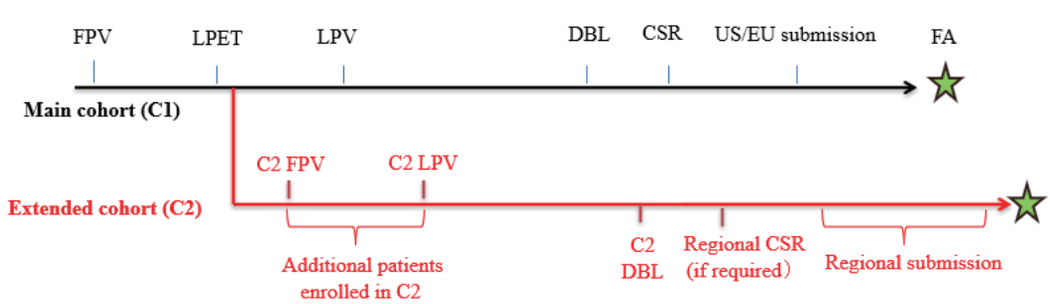

Figure 9 Relationship Diagram of Key Milestones between Main Cohort and Extended Cohort

Notes: FPV: first patient visit; LPET: last patient entering trial; LPV: last patient visit; DBL: database lock; CSR: clinical study report; FA: final approvall

(5) Sample size calculation of an MRCT with EES

The sample size in the EES involves samples for the main cohort (C1) and samples for the extended cohort (C2). Note that C1 can include some regional subjects where the EES may be needed to meet the requirements of regional drug regulatory authorities. The sample size for C1 is determined based on the primary objective of the MRCT and has no reference to C2. The sample size for C2 is often determined based on the statistical criteria for consistency between the data from regional subjects in C1 and C2 after pooling and global data in C1. Currently, there are multiple sample size calculation methods available for ideal MRCTs without any extended cohort (i.e., only C1 and no C2), which have been developed for consistency evaluation. Among them, Method 1 proposed in Japan's Ministry of Health, Labour and Welfare (MHLW) guideline on "Basic Principles on Global Clinical Trials" [14] is often mentioned in practice. This method stipulates that the Japanese (or any other country/region) sample size should be determined such that

$$P(\hat{D}_J / \hat{D}_A > \pi) \geq 1 - \beta' \qquad (1)$$

where the observed treatment effect in Japanese patients and all MRCT subjects are denoted by $\hat{D}_J$ and $\hat{D}_A$ respectively, π is some threshold for so-called consistency criteria, and $1 - \beta'$ is the desired level of probability showing such consistency, which by the guidance is at least 80%. Quan [15] states that Japanese patients need to be at least 22.4% of the overall MRCT sample size if π=0.5 in (1) and an overall MRCT power of 90%. Figure 10 illustrates the two scenarios for establishing consistency in R1 (for example, China) of interest with a regulatory-mandated sample size.

With some minor modifications, the above consistency criteria can be applied to an MRCT with the EES for a region. At the design stage, assuming a certain percentage of regional patients would be enrolled in the main cohort, then the number of regional patients for the extended cohort (and the overall number of regional patients needed) to satisfy the consistency criteria can be determined through simulations. Refer to (7) for more statistical details.

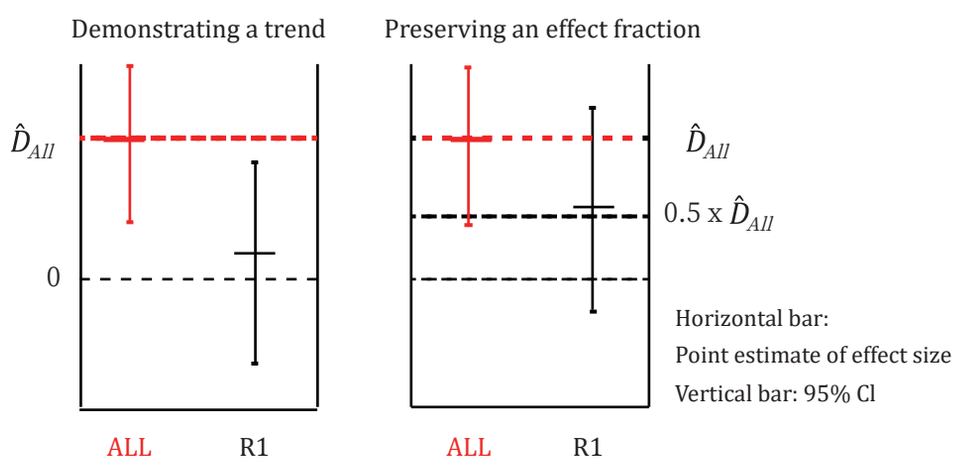

Figure 10 Illustration of Two Scenarios for Establishing Efficacy in R1

(6) Considerations for EES in practice

In the case of using EES in an MRCT, a protocol addendum (hereinafter referred to ESS addendum) may be necessary to address the need for additional patients in a region. However, submitting an addendum does not commit the implementation of EES. The content of the addendum may be flexible to allow for adjustments depending on the enrollment of the main cohort and regional needs. If the main protocol is already finalized, there is no need to amend it to include the EES addendum. However, if the main protocol and EES addendum are developed together, it may be helpful to briefly reference the addendum. The EES addendum can be written

and implemented after the trial has started, but agreement should be reached with the regulatory authorities and approval should be obtained prior to the enrollment of regional patients. It is not necessary to specify the number of patients in each cohort in the addendum. The statistical content should be concise, with further details described in the Statistical Analysis Plan (SAP). It is recommended that all regulatory authorities, especially the FDA and EMA, are informed about the potential EES for transparency.

In addition, a blinding and unblinding plan should be developed for the EES. Prior to the database lock (DBL) for the main cohort, all data should remain blinded. After the unblinding of the main cohort, the extended cohort remains blinded and it will not be unblinded until the extended cohort DBL. In EES, the regional cohort (C2) data is usually retained separately from the main cohort (C1) data and is labeled with date and country. The data from C2 will be removed from the primary database at the time of C1 DBL, and will be extracted at the time of C2 DBL. Maintaining blinding helps to prevent biases and accurately evaluate the drug's safety and efficacy in regional patients.

The degree of biases depends on the plan and implementation details of the EES. The possibility of biases may be reduced if the enrollment and treatment for the extended cohort have been completed before data from the main cohort are disclosed. However, if the enrollment and treatment of the extended cohort are still ongoing and there are significant efficacy demonstrated, study integrity and data interpretation may be impacted. Furthermore, biases can be introduced by including additional sites, so it is necessary to maintain operational consistency.

When an EES is used, the extended cohort should be identical to the main cohort with respect to criteria and procedures. The data collected for the subjects in the extended cohort will be collected and assessed by the same team following the same quality control criteria to ensure the consistency of efficacy and safety. This approach is also aligned with the quality principles of ICH E17. In contrast to the design with a second study (e.g., bridging study), when the number of regional patients included in the main cohort increases in an MRCT based on an EES design, the corresponding number of regional patients to be enrolled in the extended cohort decreases and hence the overall required number of regional patients decreases, using the consistency criteria illustrated in MHLW method 1 (e.g., Table 3). However, more thoughts regarding this issue should be taken into consideration in practical operations. If only a small proportion of regional patients will be included in the extended cohort (referred to as a small "tail"), and a significant number of regional patients are already enrolled in the main cohort during the competitive enrollment period, the necessity of the small tail is questionable, especially if an opportunity of simultaneous global submission may exist. In this scenario, the sponsor should evaluate whether the small tail is genuinely needed, taking into account all available evidence, such as regional patient data in early-phase global programs. In addition, the sponsor may consider the following two options:

① Cut the ‘tail’: The sponsor may assess whether it is feasible to submit NDA with other countries, using available regional data in the main cohort, after consulting with the regulatory authority. It may be possible to cut the small ‘tail’ or utilize other feasible approaches, such as an appropriate pooling strategy.

② Absorb the ‘tail’: The sponsor may consider including the small ‘tail’ in the main cohort for the same primary analysis of the MRCT.

As such, the global LPET is essentially postponed for the region to be able to absorb the tail in the main cohort. Conversely, if a large proportion of regional patients will be enrolled in the extended cohort (referred to as a large "tail"), caution should be exercised in planning the EES. Firstly, the EES, by definition, requires at least one patient enrolled/randomized in the main cohort to be eligible for an EES. Therefore, the way zero patient enrolled/randomized in the main cohort and all regional patients enrolled/randomized in the extended cohort is NOT an EES and should be regarded as a separate bridging study, with guidance on relevant requirements detailed in ICH E5. Secondly, if a large tail is likely to take a significant amount of time to enroll regional patients, and factors discussed earlier, such as blinding in the extended cohort and potential biases with known results from the main cohort, may affect the integrity and interpretation of the results from the extended cohort. We recommend that the EES might be considered for the scenario where a balanced tail size is expected. We further recommend that, in this scenario, the patient enrollment and treatment in the extended cohort should be completed before disclosing results from the main cohort.

Table 3 Required Number of Regional Patients in an MRCT with or without EES

	MRCT with EES (1)	MRCT with EES (2)	MRCT with a second study
Main cohort	800 (with 80 regional patients enrolled in main cohort)	800 (with 120 regional patients enrolled in main cohort)	800 (with 0 regional patients enrolled in main cohort)
Extended cohort (or second study)	132	66	242
Overall regional patients	132+80=212	120+66=186	0+242=242

In addition, the regional use of EES is to help with region specific enrollment needs, and the EES does not impact the main cohort. The main cohort is the basis of global submissions, and all risk/benefit conclusions are based on the results from patients in the main cohort. Regarding the impact of regional extension on regulatory submissions in other regions, there is no procedural impact, and the main cohort will follow the normal timelines regardless of the EES. However, it may be appropriate to discuss the EES with regulatory authorities, depending on their previous exposure to the concept. The usual route for notifying the FDA and EMA is during the end-of-phase 2 (EOP2) meeting, and the EES can be included in the briefing document. Data from the main cohort constitutes the common technical document (CTD) and is used for all global submissions. Regional patient data from the extended cohort will be provided to other regulatory authorities if requested. In certain circumstances, submissions to the regulatory authority may be made with the global submission package, even if the extended cohort is still ongoing. But this decision should take into account the percentage of regional patients that are already enrolled in the main cohort, additional resources needed to support the registration filing, benefit and regulatory risk of submitting early, etc. In summary, the use of the EES is not expected to have a significant impact on regulatory submissions in other regions or countries. However, it is important to consider the specific circumstances of each region or country and communicate with regulatory authorities as needed.

(7) Statistical technical details of EES

The sample size planning for an MRCT with EES involves two aspects: one for the main cohort and the other for the region with regulatory requirements. For convenience, we refer to the region of interest with a regulatory-mandated sample size as R1, while the rest of the regions grouped together are denoted as R2. The combination of R1 and R2 represents all the regions participating in the MRCT.

In this section, we will focus on the regional sample size since the sample size for the "main cohort" remains the same with or without EES. If R1 requires a fixed number of patients exposed to the study drug, this is usually addressed from a safety perspective and the issue is relatively straightforward. However, if R1 requires evidence of efficacy and safety in patients from the region, this can be challenging. The most straightforward approach is to generate such evidence through a standalone trial in R1 only, adequately powered. However, this approach can be criticized ethically for trial duplication if there is no evidence of intrinsic or extrinsic factors indicating the possibility of an unfavorable benefit-risk ratio in the R1 population. Within the MRCT framework, there have also been some practices where the trial in R1 subpopulation is statistically powered. Although there is nothing unethical about satisfying the regulatory need in R1, it inflates the MRCT sample size and defeats the purpose of running an MRCT, especially when more than two regions are involved.

It is common to use consistency as evidence of efficacy in R1 in drug development practices, with generally favorable opinions from regulatory authorities. Many methods have been developed by researchers in the past ten years for consistency evaluation. Among them, let us take Method 1 proposed in Japan's Ministry of Health, Labour and Welfare (MHLW) guideline on "Basic Principles on Global Clinical Trials" [14] as an example. This method stipulates that the Japanese sample size should be determined such that

$$P(\hat{D}_J / \hat{D}_A > \pi) \geq 1 - \beta', \quad (1)$$

where the observed treatment effect in Japanese patients and all MRCT subjects are denoted by $\hat{D}_J$ and $\hat{D}_A$ respectively, π is some threshold for so-called consistency criteria, and $1-\beta'$ is the desired level of probability showing such consistency, which by the guidance is at least 80%. Quan [15] states that Japanese patients need to be at least 22.4% of the overall MRCT sample size if π=0.5 in (1) and an overall MRCT power of 90%. Criterion in (1) is only interested in showing the treatment in R1. It does not require the examination of global consistency across all regions.

A few method variations have been developed. Without loss of generality, we focus on the criterion in (1) as the basis for the discussion of sample size determination in R1. Criterion in (1) can be re-written by emphasizing the dependence on the true treatment effects D_J and D_A :

$$P(\hat{D}_J / \hat{D}_A > \pi | D_J, D_A) \geq 1-\beta'$$

Provided that the idea in this section is not restricted by this assumption, we assume that $D_J = D_A$ for the subsequent discussion. In practice, π=0 and 0<π<1 may refer to two scenarios that the former describes a "trend" and the latter preserves a certain fraction of effect from R1. A "trend" may be appropriate for drugs addressing highly unmet medical needs with reasonable risk profiles while the latter may be more favored in a general setting.

Criterion (1) is only meaningful when the overall trial is a success. We can denote two events that A={$D_J / D_A > \pi$} and B={MRCT *is successful*}. Regulatory authorities are usually comfortable to approve the drug in R1 on the observation of event A∩B. For trial sponsors, the conditional event A|B is the focus interest. For the subsequent discussion, P(A|B) is denoted as the conditional probability of showing consistency and P(A) as the unconditional one.

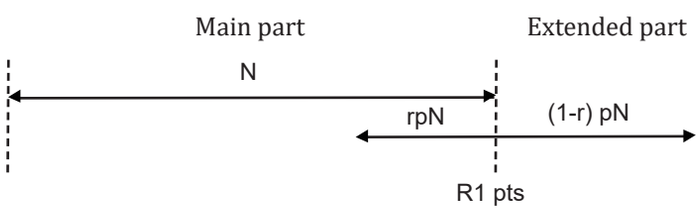

Figure 11 Sample Size Planning for an MRCT with EES

The sample size planning of an MRCT with EES is illustrated in Figure 11. The sample size for the main cohort will be determined irrespective of the EES. The following notations are straightforward:

- N = the sample size for the main cohort;
- p= R1/N, e.g. total sample size for R1 as a percentage of N;
- r= allocation to R1 in the main cohort, as a percentage of total sample size for R1; and
- pN = the sample size for R1, of which rpN is the allocation to R1 in the main cohort and (1-r)pN is the allocation to R1 in the extended cohort;
- [1+(1-r)p]N = the total sample size for the main and extended cohort combined, assuming R1 is the only region in the extended cohort, by discussion in the beginning of the section.

With competitive enrollment, r is normally assumed to be random. For a fixed r, it can be viewed as a special case.

(8) Demonstrating same-direction trend

Demonstration of a trend may be referred to π=0. The results below are straightforward.

For a fixed p, P(A) is independent of r.

Applying the normal approximation, to achieve P(A) of at least $1-\beta'$,

$$p = \left(\frac{Z_{1-\beta'}}{Z_{1-\frac{\alpha}{2}}+Z_{\beta}}\right)^2 \qquad (2)$$

which holds regardless of the true effect size and randomization ratio.

For a fixed p, P(A|B) is always greater than P(A). The difference increases as r increases and is maximized at r=1. The maximized difference depends on the power of the main cohort.

Provided by the nature of EES that r is random, Criterion (1) can be satisfied with π=0 such that

$$\int_0^1 P(\hat{D}_J / \hat{D}_A > \pi | p, r) f(r) dr \geq 1-\beta' \qquad (3)$$

f(r) is the distribution of r over a value space of [0,1] which is unknown due to the dynamic recruitment during the trial conduct. To calculate p, we may just use (2) knowing it is conservative and the true probability will always be no less than our anticipation of $1-\beta'$.

For illustration, assuming main cohort is powered at 90%. From (2), p=6.74% by setting $1-\beta'$=0.8. 10000 trials are simulated with an effect size of 0.3 which leads to N=470 (235 per arm with 1:1 randomization ratio) for a two-arm parallel design. P(A|B) is summarized in Figure 12. The difference between P(A | B) and P(A) is not more than 0.02.

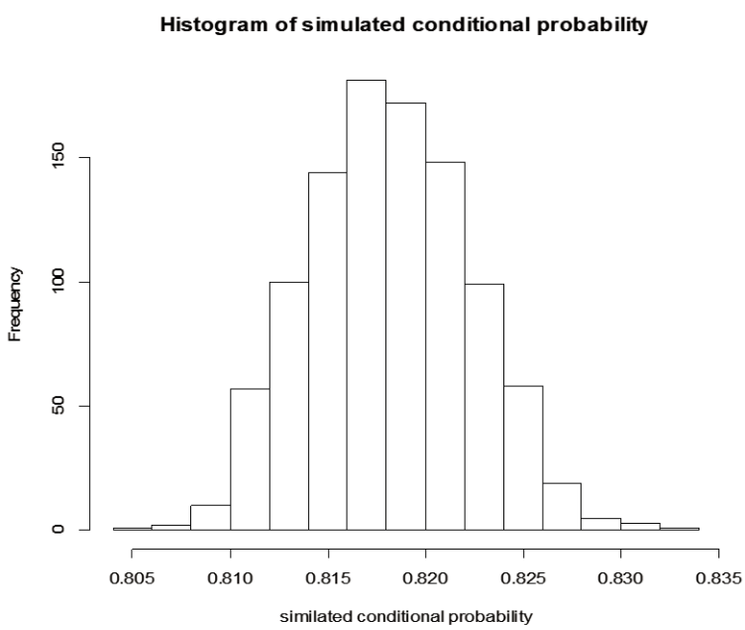

Figure 12 Status of Simulated Conditional Probability

(9) Preserving an effect fraction

When π>0, it is referred to a general case of consistency with MHLW Method 1 as a special case. The following results could be obvious.

For a fixed p, both P(A|B) and P(A) increase as r increases. For a given P(A|B) or P(A), p decreases as r increases;

P(A|B) is no longer always no less than P(A) over the value space of r in [0,1]. To understand heuristically, imagine that for r close to 1, event B indicates good treatment effects shown from R1, thus affecting the fraction in event A by adding some value to both the numerator and denominator, thus increasing the value of the fraction, and subsequently P(A). Same analogy reveals that the other direction may hold when r is close to 0.

p that satisfies (1) with π>0 should satisfy (3). The formula expression for P($\hat{D}_J / \hat{D}_A > \pi \mid p,r$) can be found in Luo(4). Similarly to the previous, it may be prudent to conservatively estimate p, by setting P($\hat{D}_J / \hat{D}_A > \pi \mid p,r^*$)= $1-\beta'$, where *r** is the most conservative estimate of r.

An explicit solution for p that satisfies (3) when r is 1 can be found in Quan [15].

The results from a simulated example is provided in Figure 13 in which x-axis represents rp, and y-axis represents p. With a power of 90% for the main cohort and =0.8, p=22.4% (the horizontal line in the plot) can be analytically calculated from Quan [15]. For P(A) (the blue line in the plot), when rp=p, it intersects with the horizontal line, verifying the results from Quan [15]. As r decreases, p also increases. Similar pattern can be shown for P(A|B). As r gets closer to 1, p corresponding P(A) is smaller than that for P(A|B). In Quan [15], it has been explained this is the case for r=1 (and they proceeded with unconditional probability noting the closed-form solution and the conservativeness). When r is close to 0, an opposite direction is anticipated.

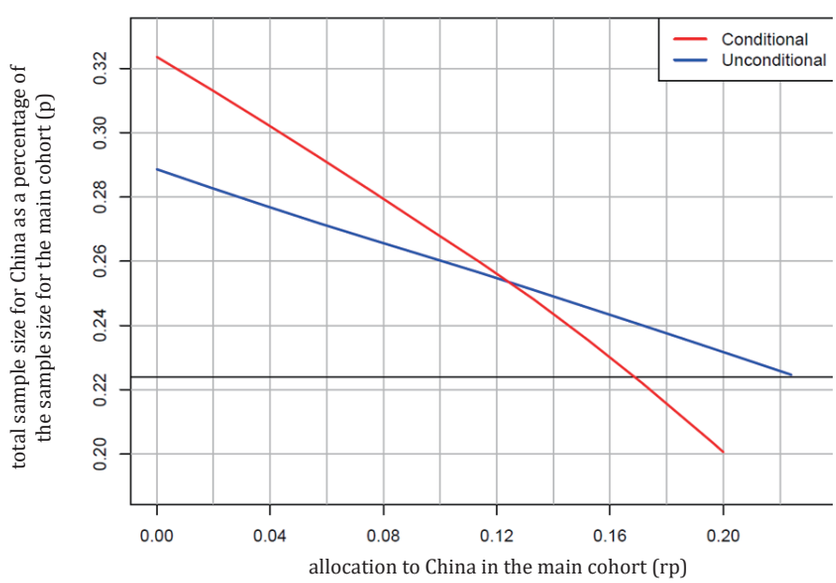

Figure 13 Relationship between Sample Size for R1 and Allocation to R1 in the Main Cohort